Horizon 2020

Call: H2020-FETHPC-2016-2017

(FET Proactive – High Performance Computing)

Topic: FETHPC-01-2016

Type of action: RIA

(Research and Innovation action)

Proposal number: 754313

Proposal acronym: ITHACA

Deadline Id: H2020-FETHPC-2016

Table of contents

How to fill in the forms

The administrative forms must be filled in for each proposal using the templates available in the submission system. Some data fields in the administrative forms are pre-filled based on the previous steps in the submission wizard.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA**

1 - General information

Topic	FETHPC-01-2016
Call Identifier	H2020-FETHPC-2016-2017
Type of Action	RIA
Deadline Id	H2020-FETHPC-2016

Acronym: ITHACA

Proposal title*: Innovative Technologies for HPC Applications and Computer Architectures

Note that for technical reasons, the following characters are not accepted in the Proposal Title and will be removed: < > " &

Duration in months: *36*

Fixed keyword 1: *Computer systems, parallel/distributed systems, grid, cloud proc* Add

Fixed keyword 2: *Software Design & Development* Add Remove

Fixed keyword 3: *High performance computing* Add Remove

Free keywords: *Data Management, Interconnect*

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA**

Abstract

The advent of Exascale HPC systems will enable significant advances in fundamental scientific fields (such as high energy physics, chemistry and material science…), important industrial sectors (including automotive, aeronautics, energy and complex system optimisation…) and complex societal challenges (in the areas of climatology, medicine, energy, large cities management…). However, building HPC systems for Extreme Scale Computing (ESC) poses a number of technological challenges. Exascale is characterized not just by the availability of exaflop computational capability, but also by the massive volumes of data required by simulations running on such systems.
ITHACA will benefit from the results of the ongoing H2020 SAGE project to innovate a new IO object storage framework, and also from the Bull eXascale Interconnect (BXI) technology. ITHACA will complement the H2020 Mont-Blanc 3 project which is focussed on providing a new efficient CPU for HPC systems.
Therefore, the ITHACA project will incorporate research findings innovation in hardware technologies and newly enabled software, with the objective of designing and developing a set of new components required for the efficient and scalable data management framework. Amongst the set of principal components, this project will implement an enriched generation of the interconnect, will update and evaluate the best programming models for such an architecture, and will develop advanced data containers objects, an efficient data manager service, a new global addressing mechanism, added-values tools and libraries, and will use these components via a set of key scientific and big data applications.
Thus, the ITHACA project aims to provide data management component developments, design specifications and associated simulation tools needed to deliver an Exascale compute-capable data management solution for horizon 2020.

Remaining characters 107

Has this proposal (or a very similar one) been submitted in the past 2 years in response to a call for proposals under the 7th Framework Programme, Horizon 2020 or any other EU programme(s)? ○ Yes ◉ No

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA**

Declarations

1) The coordinator declares to have the explicit consent of all applicants on their participation and on the content of this proposal.	☒
2) The information contained in this proposal is correct and complete.	☒
3) This proposal complies with ethical principles (including the highest standards of research integrity — as set out, for instance, in the European Code of Conduct for Research Integrity — and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct).	☒

4) The coordinator confirms:

- to have carried out the self-check of the financial capacity of the organisation on http://ec.europa.eu/research/participants/portal/desktop/en/organisations/lfv.html or to be covered by a financial viability check in an EU project for the last closed financial year. Where the result was "weak" or "insufficient", the coordinator confirms being aware of the measures that may be imposed in accordance with the H2020 Grants Manual (Chapter on Financial capacity check); or	◉
- is exempt from the financial capacity check being a public body including international organisations, higher or secondary education establishment or a legal entity, whose viability is guaranteed by a Member State or associated country, as defined in the H2020 Grants Manual (Chapter on Financial capacity check); or	○
- as sole participant in the proposal is exempt from the financial capacity check.	○

5) The coordinator hereby declares that each applicant has confirmed:

- they are fully eligible in accordance with the criteria set out in the specific call for proposals; and	☒
- they have the financial and operational capacity to carry out the proposed action.	☒

The coordinator is only responsible for the correctness of the information relating to his/her own organisation. Each applicant remains responsible for the correctness of the information related to him/her and declared above. Where the proposal to be retained for EU funding, the coordinator and each beneficiary applicant will be required to present a formal declaration in this respect.

According to Article 131 of the Financial Regulation of 25 October 2012 on the financial rules applicable to the general budget of the Union (Official Journal L 298 of 26.10.2012, p. 1) and Article 145 of its Rules of Application (Official Journal L 362, 31.12.2012, p.1) applicants found guilty of misrepresentation may be subject to administrative and financial penalties under certain conditions.

Personal data protection

The assessment of your grant application will involve the collection and processing of personal data (such as your name, address and CV), which will be performed pursuant to Regulation (EC) No 45/2001 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data. Unless indicated otherwise, your replies to the questions in this form and any personal data requested are required to assess your grant application in accordance with the specifications of the call for proposals and will be processed solely for that purpose. Details concerning the purposes and means of the processing of your personal data as well as information on how to exercise your rights are available in the privacy statement. Applicants may lodge a complaint about the processing of their personal data with the European Data Protection Supervisor at any time.

Your personal data may be registered in the Early Detection and Exclusion system of the European Commission (EDES), the new system established by the Commission to reinforce the protection of the Union's financial interests and to ensure sound financial management, in accordance with the provisions of articles 105a and 108 of the revised EU Financial Regulation (FR) (Regulation (EU, EURATOM) 2015/1929 of the European Parliament and of the Council of 28 October 2015 amending Regulation (EU, EURATOM) No 966/2012) and articles 143 - 144 of the corresponding Rules of Application (RAP) (COMMISSION DELEGATED REGULATION (EU) 2015/2462 of 30 October 2015 amending Delegated Regulation (EU) No 1268/2012) for more information see the Privacy statement for the EDES Database).

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA**

List of participants

#	Participant Legal Name	Country
1	BULL SAS	France
2	SEAGATE SYSTEMS UK LIMITED	United Kingdom
3	COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES	France
4	UNIVERSITAT POLITECNICA DE VALENCIA	Spain
5	UNIVERSIDAD DE CASTILLA - LA MANCHA	Spain
6	DEUTSCHES KLIMARECHENZENTRUM GMBH	Germany
7	FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V.	Germany
8	BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION	Spain
9	ALLINEA SOFTWARE LIMITED	United Kingdom
10	SURFSARA BV	Netherlands
11	INSTITUT NATIONAL DE RECHERCHE ENINFORMATIQUE ET AUTOMATIQUE	France
12	ARM LIMITED	United Kingdom
13	UNIVERSITAET ZU KOELN	Germany

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **BULL**

2 - Administrative data of participating organisations

PIC	***Legal name***
996058081	*BULL SAS*

Short name: BULL

Address of the organisation

Street RUE JEAN JAURES 68

Town LES CLAYES SOUS BOIS

Postcode 78340

Country France

Webpage www.bull.com

Legal Status of your organisation

Research and Innovation legal statuses

Public body ..... no
Non-profit ..... no
International organisation ..... no
International organisation of European interest ..... no
Secondary or Higher education establishment ..... no
Research organisation ..... no

Legal person ..... yes

Enterprise Data

SME self-declared status ..... unknown
SME self-assessment ..... unknown
SME validation sme ..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

NACE Code: 6202 - Computer consultancy and computer facilities management activities

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **BULL**

Department(s) carrying out the proposed work

Department 1

Department name: BDS R&D ☐ not applicable

☒ Same as organisation address

Street: RUE JEAN JAURES 68

Town: LES CLAYES SOUS BOIS

Postcode: 78340

Country: France

Dependencies with other proposal participants

Character of dependence	***Participant***	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **BULL**

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Mr. Sex: (•) Male ( ) Female

First name: **bruno** Last name: **FARCY**

E-Mail: **bruno.farcy@bull.net**

Position in org.: R&D Program Manager

Department: R&D ☐ Same as organisation

☒ Same as organisation address

Street: RUE JEAN JAURES 68

Town: LES CLAYES SOUS BOIS Post code: 78340

Country: France

Website: www.bull.com

Phone 1: +33 130807620 Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

Other contact persons

First Name	*Last Name*	*E-mail*	*Phone*
Medur	Sridharan	medur.sridharan@bull.net	+33130803024

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **SEAGATE SYSTEMS**

PIC	***Legal name***
999728076	*SEAGATE SYSTEMS UK LIMITED*

Short name: SEAGATE SYSTEMS

Address of the organisation

Street LANGSTONE ROAD

Town HAVANT

Postcode PO9 1SA

Country United Kingdom

Webpage

Legal Status of your organisation

Research and Innovation legal statuses

Public body ..... no
Non-profit ..... no
International organisation ..... no
International organisation of European interest ..... no
Secondary or Higher education establishment ..... no
Research organisation ..... no

Legal person ..... yes

Enterprise Data

SME self-declared status..... unknown
SME self-assessment ..... unknown
SME validation sme..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

NACE Code: 2620 - Manufacture of computers and peripheral equipment

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313**	*Acronym* **ITHACA**	*Short name* **SEAGATE SYSTEMS**

Department(s) carrying out the proposed work

Department 1

Department name: Seagate Systems Group (SSG) ☐ not applicable

☒ Same as organisation address

Street: LANGSTONE ROAD

Town: HAVANT

Postcode: PO9 1SA

Country: United Kingdom

Dependencies with other proposal participants

Character of dependence	***Participant***	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **SEAGATE SYSTEMS**

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Mr. Sex: (•) Male ( ) Female

First name: **Sai** Last name: **Narasimhamurthy**

E-Mail: **sai.narasimhamurthy@seagate.com**

Position in org.: Staff Engineer

Department: Seagate Systems Group ☐ Same as organisation

☒ Same as organisation address

Street: LANGSTONE ROAD

Town: HAVANT Post code: PO9 1SA

Country: United Kingdom

Website:

Phone 1: +44 2392496648 Phone 2: +44 7584080691 Fax: *+xxx xxxxxxxxx*

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **CEA**

PIC	***Legal name***
999992401	*COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES*

Short name: CEA

Address of the organisation

Street RUE LEBLANC 25

Town PARIS 15

Postcode 75015

Country France

Webpage www.cea.fr

Legal Status of your organisation

Research and Innovation legal statuses

Public body ..... yes

Non-profit ..... yes

International organisation ..... no

International organisation of European interest ..... no

Secondary or Higher education establishment ..... no

Research organisation ..... yes

Legal person ..... yes

Enterprise Data

SME self-declared status ..... 2007 - no

SME self-assessment ..... unknown

SME validation sme ..... 2007 - no

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

NACE Code: - - Not applicable

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **CEA**

Department(s) carrying out the proposed work

Department 1

Department name: DAM ☐ not applicable

☐ Same as organisation address

Street: CEA/DIF - Bruyères le Châtel

Town: ARPAJON

Postcode: 91297

Country: France

Dependencies with other proposal participants

Character of dependence	***Participant***	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **CEA**

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Mr. Sex: (•) Male ( ) Female

First name: **Jacques-Charles** Last name: **Lafoucrière**

E-Mail: **jacques-charles.lafoucriere@cea.fr**

Position in org.: Head of Service

Department: DAM ☐ Same as organisation

☐ Same as organisation address

Street: CEA/DIF - Bruyères le Châtel

Town: ARPAJON Post code: 91297

Country: France

Website:

Phone 1: +33 169266727 Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **UPV**

PIC	***Legal name***
999864846	*UNIVERSITAT POLITECNICA DE VALENCIA*

Short name: UPV

Address of the organisation

Street CAMINO DE VERA SN EDIFICIO 3A

Town VALENCIA

Postcode 46022

Country Spain

Webpage www.upv.es

Legal Status of your organisation

Research and Innovation legal statuses

Public body ..... yes

Non-profit ..... yes

International organisation ..... no

International organisation of European interest ..... no

Secondary or Higher education establishment ..... yes

Research organisation ..... yes

Legal person ..... yes

Enterprise Data

SME self-declared status ..... 2012 - no

SME self-assessment ..... unknown

SME validation sme ..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

NACE Code: 853 - Higher education

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **UPV**

Department(s) carrying out the proposed work

Department 1

Department name: DISCA ☐ not applicable

☒ Same as organisation address

Street: CAMINO DE VERA SN EDIFICIO 3A

Town: VALENCIA

Postcode: 46022

Country: Spain

Dependencies with other proposal participants

Character of dependence	***Participant***	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **UPV**

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Prof. Sex: (•) Male ( ) Female

First name: **Jose** Last name: **Duato**

E-Mail: **jduato@disca.upv.es**

Position in org.: Professor

Department: DISCA ☐ Same as organisation

☒ Same as organisation address

Street: CAMINO DE VERA SN EDIFICIO 3A

Town: VALENCIA Post code: 46022

Country: Spain

Website:

Phone 1: +34 963 879705 Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Maria Engracia	Gomez Requena	megomez@disca.upv.es	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **UCLM**

PIC	***Legal name***
999840208	*UNIVERSIDAD DE CASTILLA - LA MANCHA*

Short name: UCLM

Address of the organisation

Street CALLE ALTAGRACIA 50

Town CIUDAD REAL

Postcode 13071

Country Spain

Webpage www.uclm.es

Legal Status of your organisation

Research and Innovation legal statuses

Public body ..... yes

Non-profit ..... yes

International organisation ..... no

International organisation of European interest ..... no

Secondary or Higher education establishment ..... yes

Research organisation ..... yes

Legal person ..... yes

Enterprise Data

SME self-declared status ..... 1981 - no

SME self-assessment ..... unknown

SME validation sme ..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

NACE Code: 853 - Higher education

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **UCLM**

Department(s) carrying out the proposed work

Department 1

Department name: Computing Systems department ☐ not applicable

☐ Same as organisation address

Street: Edificio Infante D.Juan Manuel, Campus U

Town: Albacete

Postcode: 02071

Country: Spain

Dependencies with other proposal participants

Character of dependence	***Participant***	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **UCLM**

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Pedro Javier** Last name: **Garcia Garcia**

E-Mail: **pedrojavier.garcia@uclm.es**

Position in org.: Assistant professor

Department: Computing Systems ☐ Same as organisation

☐ Same as organisation address

Street: Edificio Infante D.Juan Manuel, Campus Universitario, s/n

Town: Albacete Post code: 02071

Country: Spain

Website:

Phone 1: +34 967599200 Phone 2: +34 657185891 Fax: +34 967599224

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **DKRZ**

PIC	***Legal name***
998692310	*DEUTSCHES KLIMARECHENZENTRUM GMBH*

Short name: DKRZ

Address of the organisation

Street BUNDESSTRASSE 45A

Town HAMBURG

Postcode 20146

Country Germany

Webpage http:/www.dkrz.de

Legal Status of your organisation

Research and Innovation legal statuses

Public body ..... yes
Non-profit ..... yes
International organisation ..... no
International organisation of European interest ..... no
Secondary or Higher education establishment ..... no
Research organisation ..... yes
Legal person ..... yes

Enterprise Data

SME self-declared status ..... 2007 - no
SME self-assessment ..... unknown
SME validation sme ..... 2007 - no

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

NACE Code: 721 - Research and experimental development on natural sciences and engineering

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Department(s) carrying out the proposed work

Department 1

Department name: Application ☐ not applicable

☒ Same as organisation address

Street: BUNDESSTRASSE 45A

Town: HAMBURG

Postcode: 20146

Country: Germany

Department 2

Department name: WR ☐ not applicable

☒ Same as organisation address

Street: BUNDESSTRASSE 45A

Town: HAMBURG

Postcode: 20146

Country: Germany

Dependencies with other proposal participants

Character of dependence	***Participant***	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **DKRZ**

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Joachim** Last name: **Biercamp**

E-Mail: **biercamp@dkrz.de**

Position in org.: head of department applications

Department: DEUTSCHES KLIMARECHENZENTRUM GMBH ☒ Same as organisation

☒ Same as organisation address

Street: BUNDESSTRASSE 45A

Town: HAMBURG Post code: 20146

Country: Germany

Website: www.dkrz.de

Phone 1: +49 40 460094314 Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Katja	Brendt	brendt@dkrz.de	+4940460094415
Tatjana	Grek	grek@dkrz.de	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **Fraunhofer**

PIC	***Legal name***
999984059	*FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V.*

Short name: Fraunhofer

Address of the organisation

Street HANSASTRASSE 27C

Town MUNCHEN

Postcode 80686

Country Germany

Webpage www.fraunhofer.de

Legal Status of your organisation

Research and Innovation legal statuses

Public body ..... no

Non-profit ..... yes

International organisation ..... no

International organisation of European interest ..... no

Secondary or Higher education establishment ..... no

Research organisation ..... yes

Legal person ..... yes

Enterprise Data

SME self-declared status ..... 2007 - no

SME self-assessment ..... unknown

SME validation sme ..... 2007 - no

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

NACE Code: 721 - Research and experimental development on natural sciences and engineering

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **Fraunhofer**

Department(s) carrying out the proposed work

Department 1

Department name: Fraunhofer-ITWM ☐ not applicable

☐ Same as organisation address

Street: Fraunhofer-Platz 1

Town: Kaiserslautern

Postcode: 67663

Country: Germany

Dependencies with other proposal participants

Character of dependence	***Participant***	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **Fraunhofer**

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Ms Sex: ○ Male ◉ Female

First name: **Valeria** Last name: **Bartsch**

E-Mail: **valeria.bartsch@itwm.fraunhofer.de**

Position in org.: Senior Scientist

Department: Fraunhofer-ITWM ☐ Same as organisation

☐ Same as organisation address

Street: Fraunhofer-Platz 1

Town: Kaiserslautern Post code: 67663

Country: Germany

Website: www.itwm.fraunhofer.de

Phone 1: +49 631316004741 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Franz-Josef	Pfreundt	franz-josef.pfreundt@itwm.fraunhofer.de	+49631316004459
Andrea	Zeumann	andrea.zeumann@zv.fraunhofer.de	+498912052723

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **BSC**

PIC	***Legal name***
999655520	*BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION*

Short name: BSC

Address of the organisation

Street Calle Jordi Girona 31

Town BARCELONA

Postcode 08034

Country Spain

Webpage www.bsc.es

Legal Status of your organisation

Research and Innovation legal statuses

Public body ..... yes

Non-profit ..... yes

International organisation ..... no

International organisation of European interest ..... no

Secondary or Higher education establishment ..... no

Research organisation ..... yes

Legal person ..... yes

Enterprise Data

SME self-declared status ..... 2011 - no

SME self-assessment ..... unknown

SME validation sme ..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

NACE Code: 72 - Scientific research and development

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **BSC**

Department(s) carrying out the proposed work

Department 1

Department name: Computer Science ☐ not applicable

☐ Same as organisation address

Street: Calle Jordi Girona, 29 (Nexus II)

Town: Barcelona

Postcode: 08034

Country: Spain

Dependencies with other proposal participants

Character of dependence	***Participant***	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **BSC**

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Prof. Sex: (•) Male ( ) Female

First name: **Toni** Last name: **Cortes**

E-Mail: **toni.cortes@bsc.es**

Position in org.: Computer Science- Storage Systems Group Manager

Department: Computer Science [ ] Same as organisation

[ ] Same as organisation address

Street: Calle Jordi Girona, 29 (Nexus II)

Town: Barcelona Post code: 08034

Country: Spain

Website: http://www.bsc.es/

Phone 1: +34 934137966 Phone 2: +34934137570 Fax: +34 93 413 77 21

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Isabel	Martinez	isabel.martinez@bsc.es	+34934137570
Rosa	Badia	rosa.m.badia@bsc.es	+34934134075

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **ALLINEA SOFTWARE LIMITED**

PIC	***Legal name***
969276478	*ALLINEA SOFTWARE LIMITED*

Short name: ALLINEA SOFTWARE LIMITED

Address of the organisation

Street WARWICK TECHNOLOGY PARK

Town WARWICK

Postcode CV34 6UW

Country United Kingdom

Webpage www.allinea.com

Legal Status of your organisation

Research and Innovation legal statuses

Public body ..... no

Non-profit ..... no

International organisation ..... no

International organisation of European interest ..... no

Secondary or Higher education establishment ..... no

Research organisation ..... no

Legal person ..... yes

Enterprise Data

SME self-declared status ..... 2010 - yes

SME self-assessment ..... 2014 - yes

SME validation sme ..... 2010 - yes

Based on the above details of the Beneficiary Registry the organisation is an SME (small- and medium-sized enterprise) for the call.

NACE Code: 72 - Scientific research and development

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313**	*Acronym* **ITHACA**	*Short name* **ALLINEA SOFTWARE LIMITED**

Department(s) carrying out the proposed work

Department 1

Department name: R&D ☐ not applicable

☒ Same as organisation address

Street: WARWICK TECHNOLOGY PARK

Town: WARWICK

Postcode: CV34 6UW

Country: United Kingdom

Dependencies with other proposal participants

Character of dependence	***Participant***	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **ALLINEA SOFTWARE LIMITED**

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **David** Last name: **Lecomber**

E-Mail: **david@allinea.com**

Position in org.: CEO

Department: HQ ☐ Same as organisation

☒ Same as organisation address

Street: WARWICK TECHNOLOGY PARK

Town: WARWICK Post code: CV34 6UW

Country: United Kingdom

Website:

Phone 1: +44 1926623231 Phone 2: *+xxx xxxxxxxxx* Fax: *+xxx xxxxxxxxx*

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **SURFSARA BV**

PIC	***Legal name***
952865630	*SURFSARA BV*

Short name: SURFSARA BV

Address of the organisation

Street SCIENCE PARK 140

Town AMSTERDAM

Postcode 1098XG

Country Netherlands

Webpage www.sara.nl

Legal Status of your organisation

Research and Innovation legal statuses

Public body ..... no

Non-profit ..... no

International organisation ..... no

International organisation of European interest ..... no

Secondary or Higher education establishment ..... no

Research organisation ..... no

Legal person ..... yes

Enterprise Data

SME self-declared status ..... 2011 - yes

SME self-assessment ..... unknown

SME validation sme ..... 2011 - no

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

NACE Code: 72 - Scientific research and development

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **SURFSARA BV**

Department(s) carrying out the proposed work

Department 1

Department name: Compute Services ☐ not applicable

☒ Same as organisation address

Street: SCIENCE PARK 140

Town: AMSTERDAM

Postcode: 1098XG

Country: Netherlands

Dependencies with other proposal participants

Character of dependence	***Participant***	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **SURFSARA BV**

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Mr. Sex: (•) Male ( ) Female

First name **John** Last name **Donners**

E-Mail **john.donners@surfsara.nl**

Position in org.: Senior Advisor

Department: Compute Services ☐ Same as organisation

☒ Same as organisation address

Street: SCIENCE PARK 140

Town: AMSTERDAM Post code: 1098XG

Country: Netherlands

Website: www.surfsara.nl

Phone 1: +31 208001300 Phone 2: +31 619039023 Fax: *+xxx xxxxxxxxx*

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Marcel	Van der Lann	marcel.vanderlaan@surfsara.nl	+31208001300

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **INRIA**

PIC	***Legal name***
999547074	*INSTITUT NATIONAL DE RECHERCHE ENINFORMATIQUE ET AUTOMATIQUE*

Short name: INRIA

Address of the organisation

Street DOMAINE DE VOLUCEAU ROCQUENCOURT

Town LE CHESNAY CEDEX

Postcode 78153

Country France

Webpage www.inria.fr

Legal Status of your organisation

Research and Innovation legal statuses

Public body ..... yes

Non-profit ..... yes

International organisation ..... no

International organisation of European interest ..... no

Secondary or Higher education establishment ..... no

Research organisation ..... yes

Legal person ..... yes

Enterprise Data

SME self-declared status ..... unknown

SME self-assessment ..... unknown

SME validation sme ..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

NACE Code: 72 - Scientific research and development

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **INRIA**

Department(s) carrying out the proposed work

Department 1

Department name: Inria research center Grenoble - DATAMOVE Team ☐ not applicable

☐ Same as organisation address

Street: 655 avenue de l'Europe - Inovallée

Town: Saint-Ismier Cedex

Postcode: 38334

Country: France

Dependencies with other proposal participants

Character of dependence	***Participant***	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **INRIA**

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Dr. Sex: (•) Male ( ) Female

First name: **Bruno** Last name: **Raffin**

E-Mail: **bruno.raffin@inria.fr**

Position in org.: Senior researcher

Department: Research Center Inria Grenoble Rhône-Alpes ☐ Same as organisation

☐ Same as organisation address

Street: 700 avenue Centrale - Batiment IMAG - Domaine Universitaire

Town: Saint-Martin-d'Hères Post code: 38401

Country: France

Website: https://team.inria.fr/datamove/

Phone 1: +33 457421549 Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Admin	INRIA	recettes-grenoble@inria.fr	+33476615231

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **ARM**

PIC	***Legal name***
999813824	*ARM LIMITED*

Short name: ARM

Address of the organisation

Street 110 FULBOURN ROAD

Town CAMBRIDGE

Postcode CB1 9NJ

Country United Kingdom

Webpage

Legal Status of your organisation

Research and Innovation legal statuses

Public body ..... no

Non-profit ..... no

International organisation ..... no

International organisation of European interest ..... no

Secondary or Higher education establishment ..... no

Research organisation ..... no

Legal person ..... yes

Enterprise Data

SME self-declared status ..... 2013 - no

SME self-assessment ..... unknown

SME validation sme ..... unknown

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

NACE Code: 72 - Scientific research and development

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **ARM**

Department(s) carrying out the proposed work

Department 1

Department name: Compute Services ☐ not applicable

☒ Same as organisation address

Street: 110 FULBOURN ROAD

Town: CAMBRIDGE

Postcode: CB1 9NJ

Country: United Kingdom

Dependencies with other proposal participants

Character of dependence	***Participant***	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **ARM**

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Mr. Sex: (•) Male ( ) Female

First name **Travis** Last name **Walton**

E-Mail **travis.walton@arm.com**

Position in org.: Staff Software Engineer

Department: Development Solutions Group ☐ Same as organisation

☐ Same as organisation address

Street: ARM Limited, York House, York Street

Town: Manchester Post code: M2 3BB

Country: United Kingdom

Website: www.arm.com

Phone 1: +44 1612349445 Phone 2: +xxx xxxxxxxxx Fax: +xxx xxxxxxxxx

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Geraint	North	geraint.north@arm.com	
Bruno	Jansen	bruno.jansen@arm.com	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **UoC**

PIC	***Legal name***
999852915	*UNIVERSITAET ZU KOELN*

Short name: UoC

Address of the organisation

Street ALBERTUS MAGNUS PLATZ

Town KOELN

Postcode 50923

Country Germany

Webpage www.uni-koeln.de

Legal Status of your organisation

Research and Innovation legal statuses

Public body ..... yes

Non-profit ..... yes

International organisation ..... no

International organisation of European interest ..... no

Secondary or Higher education establishment ..... yes

Research organisation ..... unknown

Legal person ..... yes

Enterprise Data

SME self-declared status ..... 2012 - no

SME self-assessment ..... unknown

SME validation sme ..... 2012 - no

Based on the above details of the Beneficiary Registry the organisation is not an SME (small- and medium-sized enterprise) for the call.

NACE Code: 853 - Higher education

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **UoC**

Department(s) carrying out the proposed work

Department 1

Department name: Regionales Rechenzentrum & Institut für Informatik ☐ not applicable

☒ Same as organisation address

Street: ALBERTUS MAGNUS PLATZ

Town: KOELN

Postcode: 50923

Country: Germany

Dependencies with other proposal participants

Character of dependence	***Participant***	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA** *Short name* **UoC**

Person in charge of the proposal

The name and e-mail of contact persons are read-only in the administrative form, only additional details can be edited here. To give access rights and basic contact details of contact persons, please go back to Step 4 of the submission wizard and save the changes.

Title: Prof. Sex: (•) Male ( ) Female

First name: **Ulrich** Last name: **Lang**

E-Mail: **lang@uni-koeln.de**

Position in org.: Professor

Department: Regionales Rechenzentrum & Institut für Informatik ☐ Same as organisation

☐ Same as organisation address

Street: Weyertal 121

Town: Köln Post code: D - 50931

Country: Germany

Website:

Phone 1: +49 221 470-89700 Phone 2: +xxx xxxxxxxxx Fax: 49 221 470-89669

Other contact persons

First Name	***Last Name***	***E-mail***	***Phone***
Ingo	Trempek	i.trempeck@verw.uni-koeln.de	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA**

3 - Budget for the proposal

No	Participant	Country	(A) Direct personnel costs/€	(B) Other direct costs/€	(C) Direct costs of sub-contracting/€	(D) Direct costs of providing financial support to third parties/€	(E) Costs of inkind contributions not used on the beneficiary's premises/€	(F) Indirect Costs / € (=0.25(A+B-E))	(G) Special unit costs covering direct & indirect costs / €	(H) Total estimated eligible costs / € (=A+B+C+D+F+G)	(I) Reimburse-ment rate (%)	(J) Max.EU Contribution / € (=H*I)	(K) Requested EU Contribution/ €
1	Bull	FR	5460000	950000	0	0	0	1602500,00	0	8012500,00	100	8012500,00	8012500,00
2	Seagate Systems	UK	1708533	205000	186563	0	0	478383,25	0	2578479,25	100	2578479,25	2578479,25
3	Cea	FR	1004365	10000	0	0	0	253591,25	0	1267956,25	100	1267956,25	1267956,25
4	Upv	ES	676400	40000	0	0	0	179100,00	0	895500,00	100	895500,00	895500,00
5	Uclm	ES	651924	72000	8000	0	0	180981,00	0	912905,00	100	912905,00	912905,00
6	Dkrz	DE	820800	75000	0	0	0	223950,00	0	1119750,00	100	1119750,00	1119750,00
7	Fraunhofer	DE	873924	7500	0	0	0	220356,00	0	1101780,00	100	1101780,00	1101780,00
8	Bsc	ES	865800	40000	0	0	0	226450,00	0	1132250,00	100	1132250,00	1132250,00
9	Allinea Software Limited	UK	440100	21000	23625	0	0	115275,00	0	600000,00	100	600000,00	600000,00
10	Surfsara Bv	NL	262600	10000	0	0	0	68150,00	0	340750,00	100	340750,00	340750,00

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA**

11	Inria	FR	243763	35000	0	0	0	69690,75	0	348453,75	100	348453,75	348453,75
12	Arm	UK	285876	13000	0	0	0	74719,00	0	373595,00	100	373595,00	373595,00
13	Uoc	DE	468000	10000	0	0	0	119500,00	0	597500,00	100	597500,00	597500,00
	Total		13762085	1488500	218188	0	0	3812646,25	0	19281419,25		19281419,25	19281419,25

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA**

4 - Ethics issues table

1. HUMAN EMBRYOS/FOETUSES		Page
Does your research involve Human Embryonic Stem Cells (hESCs)?	○ Yes ◉ No	
Does your research involve the use of human embryos?	○ Yes ◉ No	
Does your research involve the use of human foetal tissues / cells?	○ Yes ◉ No	
2. HUMANS		**Page**
Does your research involve human participants?	○ Yes ◉ No	
Does your research involve physical interventions on the study participants?	○ Yes ◉ No	
3. HUMAN CELLS / TISSUES		**Page**
Does your research involve human cells or tissues (other than from Human Embryos/ Foetuses, i.e. section 1)?	○ Yes ◉ No	
4. PERSONAL DATA		**Page**
Does your research involve personal data collection and/or processing?	○ Yes ◉ No	
Does your research involve further processing of previously collected personal data (secondary use)?	○ Yes ◉ No	
5. ANIMALS		**Page**
Does your research involve animals?	○ Yes ◉ No	
6. THIRD COUNTRIES		**Page**
In case non-EU countries are involved, do the research related activities undertaken in these countries raise potential ethics issues?	○ Yes ◉ No	
Do you plan to use local resources (e.g. animal and/or human tissue samples, genetic material, live animals, human remains, materials of historical value, endangered fauna or flora samples, etc.)?	○ Yes ◉ No	
Do you plan to import any material - including personal data - from non-EU countries into the EU?	○ Yes ◉ No	
Do you plan to export any material - including personal data - from the EU to non-EU countries?	○ Yes ◉ No	
In case your research involves low and/or lower middle income countries, are any benefits-sharing actions planned?	○ Yes ◉ No	
Could the situation in the country put the individuals taking part in the research at risk?	○ Yes ◉ No	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA**

7. ENVIRONMENT & HEALTH and SAFETY		Page
Does your research involve the use of elements that may cause harm to the environment, to animals or plants?	○ Yes ◉ No	
Does your research deal with endangered fauna and/or flora and/or protected areas?	○ Yes ◉ No	
Does your research involve the use of elements that may cause harm to humans, including research staff?	○ Yes ◉ No	
8. DUAL USE		**Page**
Does your research involve dual-use items in the sense of Regulation 428/2009, or other items for which an authorisation is required?	○ Yes ◉ No	
9. EXCLUSIVE FOCUS ON CIVIL APPLICATIONS		**Page**
Could your research raise concerns regarding the exclusive focus on civil applications?	○ Yes ◉ No	
10. MISUSE		**Page**
Does your research have the potential for misuse of research results?	○ Yes ◉ No	
11. OTHER ETHICS ISSUES		**Page**
Are there any other ethics issues that should be taken into consideration? Please specify	○ Yes ◉ No	

I confirm that I have taken into account all ethics issues described above and that, if any ethics issues apply, I will complete the ethics self-assessment and attach the required documents. ☒

How to Complete your Ethics Self-Assessment

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Proposal ID **754313** *Acronym* **ITHACA**

5 - Call specific questions

Data management activities

A new focus within Horizon 2020 is data management, for example through the use of Data Management Plan (DMP).

DMPs detail what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved.

The use of a DMP is required for projects participating in the Open Research Data Pilot in the form of a deliverable in the first 6 months of the project (possible updates during the project).

Other projects are invited to submit a DMP if relevant for their planned research.

Are data management activities relevant for your proposed project?	○ Yes	◉ No

Open Research Data Pilot in Horizon 2020

If selected, all applicants will participate in the Pilot on Open Research Data in Horizon 2020 , which aims to improve and maximise access to and re-use of research data generated by actions.

Participants in the Pilot will be invited to formulate a Data Management Plan (DMP). DMPs detail what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved.

Participating in the Pilot is flexible in the sense that it does not mean that all research data needs to be open. Rather, projects can define certain datasets to remain closed via a Data Management Plan (DMP).

Applicants also have the possibility to opt out of this Pilot. In this case, applicants must indicate a reason for this choice (see options below).

Please note that participation in this Pilot does not constitute part of the evaluation process. Proposals will not be penalised for opting out.

We wish to opt out of the Pilot on Open Research Data in Horizon 2020.	○ Yes	◉ No

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Innovative Technologies for HPC Applications and Computer Architectures

Participating Organisations

Participant No *	Participant Organisation Name	Participant short name	Country
1 (Coordinator)	Bull SAS	Bull	France
2	Seagate System UK Ltd	Seagate	United Kingdom
3	Commissariat à l'énergie atomique et aux énergies alternatives	CEA	France
4	Universidad Politecnica de Valencia	UPV	Spain
5	Universidad de Castilla – La Mancha	UCLM	Spain
6	Deutsches KlimaRechenZentrum GMBH	DKRZ	Germany
7	ITWM Fraunhofer	Fraunhofer	Germany
8	Barcelona Supercomputing Center	BSC	Spain
9	Allinea Software Limited	Allinea	United Kingdom
10	SURFsara BV	SURFsara	Netherlands
11	Institut National de Recherche en Informatique et Automatique	INRIA	France
12	ARM Limited	ARM	United Kingdom
13	Universitaet zu Koeln	UKOELN	Germany

* We use the same participant numbering as indicated in the administrative proposal forms.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Contents

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

1. Excellence

1.1. Objectives

1. ITHACA's roots

The advent of Exascale HPC systems will enable significant advances in fundamental scientific fields (such as high energy physics, chemistry and material science…), important industrial sectors (including automotive, aeronautics, energy and complex system optimisation…) and complex societal challenges (related to climatology, medicine, energy, large cities management…). However, building HPC systems for Extreme Scale Computing (ESC) poses a number of technological challenges. Exascale is characterized not just by the availability of exaflop computational capability, but also by the massive volumes of data required by simulations running on such systems.

The objective of the **I**nnovative **T**echnologies for **H**PC **A**pplications and **C**omputer **A**rchitectures (**ITHACA**) project is to develop key technologies needed for Exascale-level HPC architectures. HPC systems at Exascale will provide a 1000x increase in application performance, compared to what can be achieved on petaflop-scale systems. To deliver ESC capability, the specifications of Exascale solutions will differ from those of Petascale systems, as shown in Table 1:

	1 PFlops Reference system "Curie" Tier 0 PRACE system (2011)	**Exascale system expected by HPC and BigData users**
Compute performance	1 PFlops	1000 PFlops
Total Memory size	0.23 PB	50-80 PB
Total Memory BandWith	0.36 PB/s	300 PB/s
Bytes/Flops Ratio	0.36	> 0.3
Processor performance	140 GFlops	> 10 TFlops
Nb of nodes & processors	3500 nodes bi-sockets / 9k procs	50k-100k nodes-processors
Interconnect BW by node	40 Gb/s (QDR Infiniband)	200-400 Gb/s
Total Storage size	2 PB	500 PB
Total Storage BandWith	5GB/s	5TB/s
Energy consumption	2,2 MW	20-40 MW

Table 1: Petascale vs Exascale System specification

To attain the 1000x performance factor while increasing energy consumption by only a factor of 10 to 20, significant advances are required in the basic components (processor, interconnect and storage) used in HPC systems, as well as in the solution architecture. The main driver for architectural improvements to implement an Exascale-level infrastructure is to switch from a compute-centric view to a data-centric view—enabling architectures that leverage new technologies, such as an Exascale interconnect, fast memory, non-volatile memory, I/O forwarder and compute-capable storage. The schema shown in Figure 1 indicates the new components being implemented for Exascale (in blue), which are the subject of the ITHACA project:

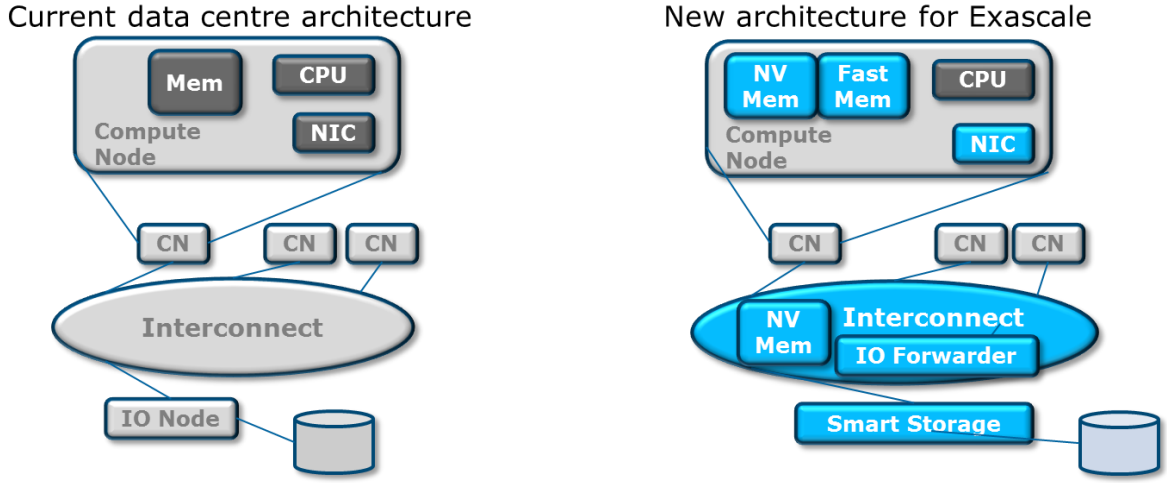


Figure 1: Data Centre Architecture (Current vs Exascale)

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

For Exascale, ITHACA proposes a novel approach to data access, data addressing and the data management paradigm and the entire supporting ecosystem, consisting of the following:

- New data management framework
- Novel persistent data storage methods.
- New programming models
- New interconnect
- New enriched software tools associated with these technologies.

To deliver both expected compute performance and energy efficiency, these architectural changes requires a new framework. ITHACA proposes new solutions for two key hardware components - interconnect and storage - and for a new Data Management middleware. Given the large project scope, ITHACA cannot undertake research in the domain of CPU design suitable for Exascale. Nevertheless ITHACA will take into account CPU design innovations and influence evolution in this domain (use of fast memory, vector registers and vector instructions, and tight integration of interconnect). This collaboration will ensure that the project's proposed solutions can use the Exascale CPU. The involment of several of ITHACA's key partners in the Mont-Blanc project (ARM, BSC and Bull) ensures that ITHACA will be aligned with advances in CPU design. Thus, pairing ITHACA with a complementary project on CPU evolution can lead to the actual emergence of an Exascale HPC system based on European technologies.

ITHACA will benefit from the results of the ongoing H2020 SAGE project to integrate a new I/O object storage framework, and also from the Bull eXascale Interconnect (BXI) technology. ITHACA will complement the H2020 Mont-Blanc 3 project which is focussed on providing a new efficient CPU for HPC systems. Moreover, a strong connection with the different Centres of Excellence will be established.

Therefore, the ITHACA project will incorporate research findings and innovation in hardware technologies and newly enabled software, with the objective of designing and developing a set of new components required for the efficient and scalable data management framework.
Amongst the set of principal components, this project will implement an enriched generation of the interconnect, update and evaluate the best programming models for such an architecture, develop advanced data containers objects, an efficient data manager service, a new global addressing mechanism, added-values tools and libraries, and enable the use and integration of these components by a set of key applications from the scientific and big data domains. Indeed, as data management raises issues for HPC as well as for big data applications and data centres, ITHACA's outcomes will have a significant impact and an extensive reach.

Thus, the ITHACA project aims to provide component developments, design specifications and associated simulation tools needed to deliver an Exascale compute-capable data management solution for horizon 2020.
For Seagate and Bull, ITHACA is a key milestone in the development of Exascale solutions that positions these European companies at the forefront of HPC advances worldwide. The research proposed in this project will enable technological developments that support future products to meet the needs and requirements of Exascale users.

2. ITHACA's main lines

Over the last twenty years, large HPC systems have evolved from a "compute centric" into a "data centric" architecture. Petascale systems are now characterized by extremely large volumes of data and high refresh rates. Exascale systems will exacerbate this requirements with much larger data volume and system size. Much bigger and complex interconnect will be needed to move data between compute nodes and storage at an affordable cost.
ITHACA aims to offer new mechanisms for moving compute resources to data, i.e. working on data locally. The intent is to execute a job or a portion of a job, close to the required data, but also to organize data movements transparently, so that the entire compute, network and storage infrastructure is permanently optimised to achieve the best performance level. Figure 2 illustrates the evolution from compute centric to data centric systems:

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

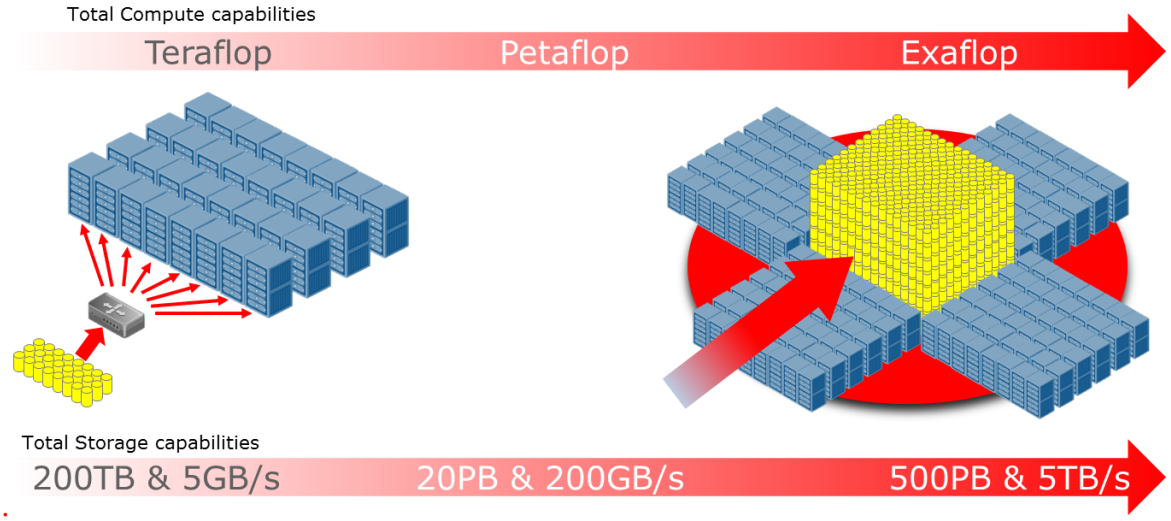


Figure 2: Compute Centric to Data Centric evolution

The I/O software stacks and APIs are important domains to improve. They have their roots in the very first UNIX implementations, when computers were monolithic and networks ran at very low speed. Over time, as network speeds improved, distributed computers (clusters) were built and, parallel file systems were introduced. For reasons of backward compatibility, these file systems used the same POSIX APIs and semantics for coherency.
Today's data storage architectures used on the largest supercomputers are extensions of traditional distributed parallel file systems - all compute nodes are connected to a parallel file system such as Lustre or GPFS and rely on the POSIX semantics to manage coherency. This model does not scale effectively, i.e. scalability of the Distributed Large Memories, prohibitive recovery time, etc. Solutions for HPC environments are emerging along two different axes:

- Keep the same architectural model (with POSIX I/Os issued by applications) and address its limitations. This category includes I/O proxies and burst buffers.
- Propose a new I/O paradigm for applications, redesign the I/O architecture and APIs to address scalability challenges and leverage emerging technologies (such as non-volatile, byte addressable memory) much more efficiently. This category includes object storage environments and new/enhanced APIs to expose new I/O concepts to applications.

Both approaches are meaningful—the first enables existing applications to run at higher scale without modification, while the second, a more disruptive approach, presents promising methods to manage data at large scale with increased performance.

ITHACA proposes to implement a new data access middleware that supports both modes (selected at runtime), a very innovative I/O stacks, an interconnect network, as well as programming environments that enable new technologies, such as NVMe. New I/O usage will be demonstrated by the project's applications. Simultaneous support for both approaches on the same infrastructure is also essential to facilitate migration from one model to the other one.

The primary goals and objectives of the ITHACA project are organised along six main lines:

a) Exascale Interconnect:

The interconnect network is a constituent component of a data centre, and considered one of its most important elements. Existing interconnect technologies are limited in terms of raw performances (bandwidth, latency and message rate) and scalability. Better performance shortens data transfer times, while scalability is mandatory to build Exascale systems. Thus, one of the co-design challenges of the ITHACA project will be to design a next generation interconnect (Exascale Interconnect) that scales to hundreds of thousands of nodes and that enables efficient data transfers between compute nodes and storage or service nodes.
To achieve this goal, it will be necessary to capitalize on existing technologies. Hence, testing and studying the behaviour of existing interconnects, such as the BXI interconnect, will be very valuable.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

b) Data Management Layer:

A new data access framework is needed to reach performance and scalability targets. In particular, this framework must address the new memory hierarchy models envisioned for Exascale systems.
Along with scalability, performance improvements are required in terms of data manipulation, to support the increased computing power. These improvements relie on a combination of hardware and software such as data prefetching or function shipping.
Thus, a complete Data Management Layer, capable of encompassing all of these aspects and able to cope with new hardware designs and optimisations, must be crafted.

c) Data Access Programming models & runtime level:

A principal objective of the ITHACA project is to present a data management system that maintains compatibility with existing applications by using interfaces that are compliant with POSIX, MPI or PGAS/GPI. Additionnal programming models can be created to leverage the new software and hardware architectures. Along with the implementation of an original Data Management Layer, an relevant level of abstraction must be provided to enable developers to avoid overly complicated application creation and tuning. More specifically, new concepts and innovative programming models (task based, as represented by COMPs and dataClay) must be employable and must benefit from the proposed ITHACA architecture.Reducing data traffic and minimizing the storage footprint remain key goals of the project, both to save energy and money, but also to support system scalability. A specific and complementary study about in-situ processing must be conducted. Moreover, a true co-study of the applications data access schemes and process scheduling must be performed. By placing an application's processes in the correct location, rather than transferring data to the processes, a huge gain in transferred volume can be observed. Another significant aspect of the ITHACA project is its co-design aspect - the project is considering use cases from several scientific and big data application domains, to enable the development of a hardware/software platform that reflects application-specific requirements. These requirements will be integrated in our programming models studies.

d) Enrich the Data Management EcoSystem:

Exascale-level applications will grow in both size and complexity, creating new challenges regarding how to analyse their behaviour. The growth in storage size is trivial and a direct consequence of system upscaling; the more space available for applications to execute, the more space they will use. The larger the application, the more difficult the analysis and profiling. These challenges are the consequence of larger data sets, a higher degree of parallelism and a much greater number of interactions between the application's constituent elements (processes, threads, execution atomics, etc.). This issue is directly connected to the second point, which relates to the behaviour of parallel applications. As stated earlier, new programming models will be needed to enable efficient application development on an Exascale system. Existing models will have difficulty scaling and benefiting from new hardware and low-level software innovations. Even if the proper level of abstraction is presented to developers, profiling an application requires a deeper dive and to consider interactions at a low-level. Without appropriate tools, this analysis will be almost impossible. These tools will be able to consider the performance impacts of a high number of elements (topology, application scheduling and interactions inside and between applications) and to rebuild a suitable level of abstraction for a human operator.

e) Ambitious Use Cases:

This project will target use cases from several scientific and big data application domains that are very data intensive and, as such, future candidate for Exascale platforms. These use cases will cover the following areas:

- **Earth System:** The ICOsahedral Non-hydrostatic (ICON) dynamical core is a next-generation earth system model designed to simulate multiple scales of atmosphere processes, enabling both climate simulations and numerical weather predictions. It provides the option to run locally nested, highly refined resolutions, allowing very fine scale simulations [1] ICON's system of equations is solved in grid point space on a geodesic icosahedral grid, which enables a quasi-isotropic horizontal resolution on the sphere as well as the restriction to regional domains. The primary cells of the grid are triangles, resulting from a Delaunay triangulation, which enable C-grid type discretisation and straightforward local refinement in selected areas.[2]

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

- **Earth Climate**: Currently, Earth System Models (ESMs), such as EC-Earth, are the only method of providing reliable information on future climatic conditions. EC-Earth generates valid in-house predictions and projections of global climate changes, which are a prerequisite to support the development of mitigation strategies and adaptive responses. EC-Earth has successfully contributed to international climate change projections such as CMIP5. Ongoing development by the consortium will ensure that increasingly more reliable projections can be offered to decision and policy-makers at regional, national and international levels.EC-Earth is a coupled model that uses the Integrated Forecast System (IFS) as the atmosphere model, the Nucleus for European Modelling of the Ocean (NEMO) as the ocean model and OASIS-MCT as the coupler between the components. A new version of EC-Earth is under development and the consortium plans to participate in CMIP6.
- **Seismic Imaging Method**: Seismic GRT migration computes images of the Earth's subsurface and is acknowledged to be superior to comparable methods, due to its ability to deliver true amplitudes. GRT requires permanent global access to all (several TB-scale) input data that are distributed across compute nodes. Double buffering strategies enable a communication-computation overlap, while local caching increases the reuse of non-local data between different cores (grouped together in multiple thread pools). However, depending on actual problem sizes, optimal locality and overlap cannot always be achieved. Thus, it is crucial that the interconnect offer large bandwidth and short latency and is highly scalable regarding the number of connections so as not to delay computations.
- **Deep Learning**: Distributed training of Deep Learning models. The recent success of Deep Learning methods have led to the development of ever larger models, focussed on amounts of data growing at faster rates. The evolutionary growth of deep neuronal networks (DNNs) has triggered demand for distributed optimisation algorithms to parallelise the very compute intensive training of DNNs. Current approaches [3] have failed to scale efficiently beyond 32 compute nodes. One reason for this limited scalability is the high communications load of the distributed optimisation algorithms which, when combined with linear growth of the data input stream, exceeds the interconnect bandwidth of the compute nodes. While a reduction in the communications load of the optimisation algorithms received a lot of recent attention [4], the data input stream, the most significant factor, has been neglected in recent literature.
- **Genomics:** Genetics pipelines have significant potential for clinical use, i.e. a high volume market with a large societal impact. Code structure and characteristics will show good performance improvements based on the following:
    - Single runs do not scale into the high petaflops performance range; high performance requirements result from multi patient use.
    - Emergent scientific field with significant potential (whole genome sequencing).

  Next-Generation Sequencing (NGS) is an increasingly cost efficient and reliable method to provide whole genomes or exomes in a relatively short time. As costs have fallen, it became feasible to widen the scope of sequencing research and applications. In the last decade, personalized medicine has been partially realized. It is expected that with the increasingly widespread availability of personalized medicine in the future the number of whole-genome sequencing will also increase dramatically, resulting in considerably more whole-genome computations. Apart from clinical applications they will also rise in scientific research [6][7][8]. Massive amounts of resulting data pose various challenges that need to be addressed to enable their exploration, analysis and effective dissemination. In particular, genetic data runs through a lifecycle: the generated input is organized and stored, depending on its type and origin, later it is retrieved, processed and analysed by high-throughput machines in an HPC cluster and, finally, once final results have been computed, they are made available for review and further comparison.

  The typical pipeline to sequence and analyse exomes and genomes includes several steps [5]. The next-generation sequencer generates short snippets of genetic sequences or “reads” in FastQ format, which are converted into the binary BAM format and stored in the institute’s local storage. In the next step, the pipeline aligns raw whole genome or exome sequencing data with the reference genome using the BWA-MEM algorithm. After alignment, the data is pre-processed to allow for mutation detection. Alignments are sorted and indexed, and potential

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

PCR-duplications are masked. In the next step, the difference compared to the reference genome is determined, together with quality control parameters of the sequencing run (e.g. mean coverage, insert size, etc.). Furthermore, genotype information, as well as local read depths, are extracted from the data. These derived data sets are the basis for mutation detection, in which single nucleotides substitutions, small insertions and deletions, copy number changes and genomic rearrangements (the latter only in case of whole genomes) are determined. Integration into a HPC system provides the computing power necessary to perform a high throughput data analysis framework. To facilitate a streamlined and robust workflow, pipeline automation as well as error handling procedures have been introduced to react to both internal (content based) as well as external (system based) errors.

- **Molecular Dynamics Simulations**: Molecular Dynamics (MD) simulations are amongst the traditional parallel applications that commonly running on supercomputers. GROMACS is one of the most popular codes for running MD simulations. Developed in Europe, GROMACS is a sophisticated code that efficiently supports hierarchical parallelism (MPI, OpenMP, GPU and SSE). Because of its high quality and widespread use, GROMACS is regularly utilised as a benchmark in HPC publications.

  As in many parallel simulations, I/Os are a growing performance bottleneck. Larger machines enable larger simulations that produce more data, while I/O and storage capabilities have progressed at a slower pace. Writing simulation results to disk can significantly slow down an application's performance. Then, the results need to be read back from disk for analysis. This traditional approach, relying on disk storage between the simulation and the analysis phases is time-consuming and energy-consuming. To circumvent this issue, a simple and commonly used solution is to save simulation results less frequently. However, in-situ analytics proposes a more sustainable approach by starting data analysis (compression, indexation, and computation of any kind of descriptor) as early as possible and as close to the data source as practical, i.e. as soons as data is available in the memory of the nodes running the simulation. In-situ analytics is seen by many as a paradigm shift. Because it radically changes data management and analysis, it has the potential to impact several aspects of HPC environments: resource allocation strategies, analysis algorithms, data storage and access.

  INRIA developed an in-situ analytics infrastructure for GROMACS, based on their open source FlowVR library. INRIA demonstrated that GROMAC's performance can be significantly improved by bypassing the disk while shortening the analysis time. For instance, they demonstrated that saving GROMACS results every 100 iterations, running with 2048 cores (an in-situ based approach that consists in aggregating results first on each node with a dedicated core, and then saving the results to disk) impacts GROMACS performance by about 3%, while the impact reaches 70% with the native GROMACS strategy that relies on classical MPI based data aggregation.
- **Dynamic Flow**: AFiD is a CFD model with a numerical scheme geared for high performance computation of wall-bounded turbulent flows. It utilizes the favourable scaling of the CFL time-step constraint as compared to the diffusive time-step constraint. As the CFL condition is more restrictive at high driving, implicit time integration of the viscous terms in the wall-parallel directions is no longer required. This avoids communication of non-local information to a process for computation of implicit derivatives in these directions. The number of all-to-all communications is decreased to only six instances by using a two-dimensional (pencil) domain decomposition. The code is shown to have very good strong and weak scaling to at least 64 K cores.

f) Research results dissemination:

To create a global stimulation of the related fields (both industrial and academic), dissemination of the ITHACA project's innovations will occur, to reach the following audiences:

- Key influencers of HPC and big data technology through organisations, such as the relevant ETPs: ETP4HPC and infrastructure technology with PRACE
- Scientific communities as they rely (increasingly) on data created through simulation or aggregation and homologation from other sources

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

- Wider markets, by disseminating into and exploiting markets other than those of extreme scale, increasing opportunities across a swath of technical, social and business analytics areas.

We next discuss the relationship of the ITHACA project to the H2020 programme.

1.2 Relation to the Work Programme

By definition and by nature, the proposed work items inside the ITHACA project are directly related to the FETHPC-1-2016 topics.
The new Data Management Layer that is integral to the ITHACA project is a core technology for future Exascale HPC systems. Its design will be driven by scientific and big data applications needs, by state-of-the-art technology studies, by energy efficiency and resilience aspects, and by I/O domain and scalability challenges. New programming models and tools to adapt and test applications on this groundbreaking Data Management framework will need to be designed and specified.
The project is aligned with the expected impacts of this call:

- *Contribution to the realization of the ETP4HPC SRA:*

  The project will design a Data Management layer and interconnect network to deliver performance efficiency in line with the roadmap specified in the Strategy Research Agenda (SRA) document. The schema shown in Figure 3 illustrates that the ITHACA project covers many of the strategic domains defined by ETP4HPC:

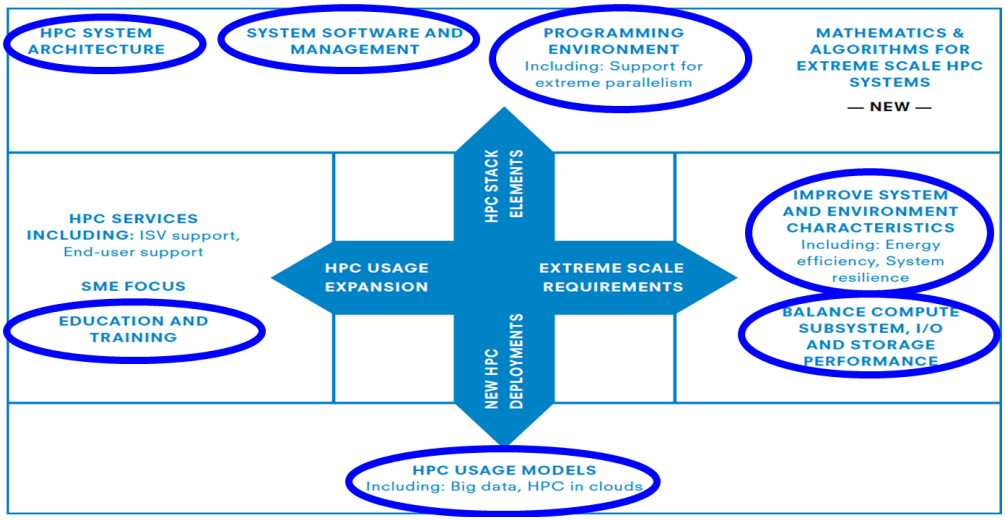


Figure 3: SRA Domain covered by ITHACA (circled in Dark Blue)

- *Covering important segments of the broader and/or emerging HPC markets*:
  A selection of scalable applications (+100 PFLOPS) covering a wide range of codes will ensure that the designed data management solution will be of interest to both large existing markets and emerging ones.
- *Proof-of concept through integrated pre-Exascale prototype*:
  Proposed miniapps and real applications will be tested on a tests vehicle, provided during the project and accessible to all partners. The prototype will support new advanced technologies and software developed during the project.
- *Impact on standard bodies and other relevant research programmes*:
  Several key open sources HPC organisations will benefit directly from the works undertaken in this project (Portals, openMPI, GPI, Lustre, openSHMEM, PyCOMPSs, dataClay, …). Virtually all project partners are also members of major international research programmes or cooperative consortium in the HPC domain (ETP4HPC, H2020 SAGE, H2020 Mont-Blanc 3, ESiWACE, PRACE, TERATEC, HiPINEB and H2020 NextGenIO).
  More details regarding this project's impacts are discussed in Section 2.1

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Globally the project will deliver the Data Management, data access and interconnect parts of an HPC solution that could, foreseeably, enter the market in the 2020 timeframe and, upon its introduction, provide leading edge performance and energy efficiency. ITHACA will provide technologies that enable Extreme-Scale Demonstrators to be built.
The ITHACA project is a cornerstone to enable Exascale solutions from European entities that are the goal of this FET HPC call.

1.3 Concept and Methodology

1.3.1 Concepts and Approach

The concepts and approach adopted by ITHACA have been influenced by and built upon previous EU cooperative projects. Indeed, using the results from H2020 SAGE project at the I/O storage side and from the Bull BXI technology at the interconnect level, the ITHACA project will create new concepts and technologies to provide a complete advanced Data Management architecture.

a) Global Data Management architecture to provide Distributed Data Container

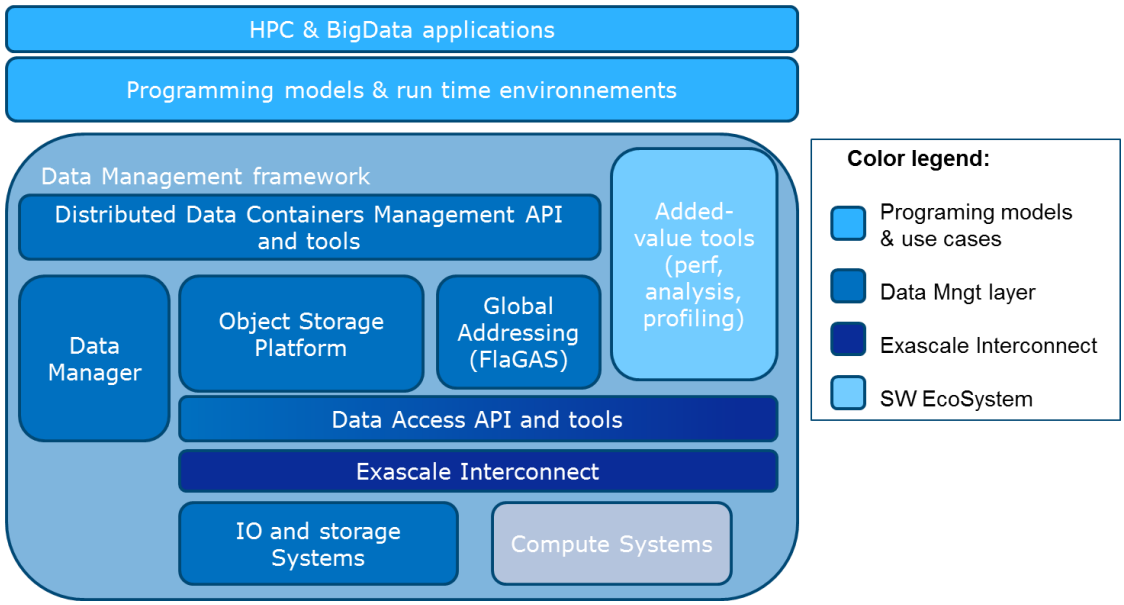


Figure 4: ITHACA Hardware and Software components

The schema shown in Figure 4 represents a complete data management framework based on new (conceptual) hardware and software components in ITHACA, described below:

- **Distributed Data Container (DDC)**: ITHACA's data access middleware will be able to dynamically build a **namespace** in which compute nodes find the data needed to run one or more applications using a specific protocol.
- **Flat Global Address Space (FlaGAS)**: At the lowest level, a software service called **FlaGAS** will be able to cross-map the address spaces of application processes, allowing transparent access to remote memory.
- **Object Storage Platform:** The Object Storage Platform is a major component of the new I/O stack that brings required scalability and data management features.
- **Data Manager**: The Data Manager sets up the run time environment (DDC) for applications.
- **Run Time Environnments for Applications**: ITHACA will leverage FlaGAS and the Object Storage Platform to study and develop different run time environments which, in turn, will be instantiated within DDCs.
- **Exascale Interconnect:** In addition to providing performance and scalability improvements, the interconnect will be enriched with specific data access and management features that correlate with other data management components (the offload of FlaGAS features for example).
- **Value-Added Tools for Data Management:** New simulators, performance and reporting tools, and analysers will be developed, in a co-design manner, with the ITHACA component designers

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

and developers. These tools will collect and assemble the most useful information for designers and later on for HPC and big data users.

b) *The Data Management layer*

ITHACA's data management framework will be able to dynamically build a namespace in which compute nodes find data needed to run one or more applications using a specific protocol. Within the project, this concept is called the DDC.
A DDC will be built as part of the application launch process. When the Batch and Resource Manager launches an application, it will require a DDC to be created for (or attached to) the application. The application must describe the data resources it needs to run (files, file systems and/or objects) and the protocols it supports.
The lifetime of a DDC is usually limited to a single application run. Users can also create a DDC to span a set of related runs and/or application workflows. When several applications work on the same data sets during a period of time, the DDC will improve "data locality".
In its simplest form, a DDC is composed of mount points to the distributed file systems available on the supercomputer. Currently, on most supercomputers, DDCs are not dynamic.
Within the ITHACA project, the first implementation will dynamically mount the required file systems on the allocated compute nodes when an application is launched. On sites with many different production file systems, it will enable the number of file systems handled by client nodes to be lowered and improve recovery times.
ITHACA will generalize this concept for all applications launched on a supercomputer. For each application, a DDC will be created, depending on its requirements:

- Accessable existing data
- Volume of data to be created, types (checkpoints, final results)
- Protocol (POSIX, other)

The Data Manager component controls the DDC. The Data Manager creates and populates the DDC upon request of the Batch and Resource Manager. Optionally, a DDC can use additional hardware resources such as I/O routers, I/O proxies, burst buffers and persistent storage resources.
In concrete terms, DDC creation will involve one or more of the following:

- Allocation and configuration of additional hardware resources (in addition to compute nodes)
- Configuration of the Data Store to provide access to files and objects
- Data movement to populate the allocated storage resources. For example, pre-fetching to use large blocks of a data file in a compute node's persistent memory, enabling improved performance and energy efficiency when this file is accessed in small random reads by the application—resulting in substantially less network overhead and latency
- System configuration on the compute nodes to establish access to remote data

Figures 5 shows possible instantiations of the DDC for regular compute nodes without locally attached persistent storage:

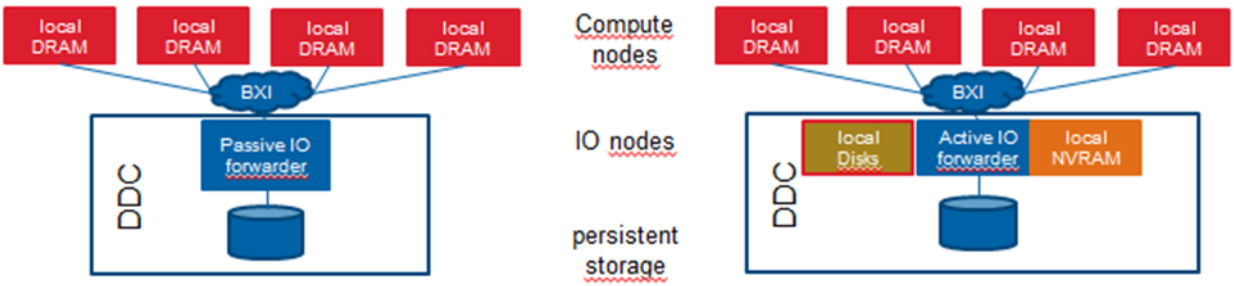


Figure 5: POSIX DDC

A Lustre router is an example of a passive I/O forwarder. The objective of setting up this type of DDC is to control which passive I/O forwarder is used by the compute nodes, thereby minimising traffic over the interconnect. These DDC instantiations address POSIX-compliant applications. In Figure 6, the DDC on the left is composed of file system mount points that are configured to go through the passive I/O forwarder. In this example, the I/O forwarder is "passive" because it does not provide

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

additional features, such as data caching; it simply receives I/O requests from compute nodes and forwards them to the backend file system. The DDC on the right is also composed of file system mount points, but the I/O forwarder is capable of providing advanced features such as burst buffering, local caching, prefetching, etc. In these cases, the DDC is considered to be a Software Defined Storage solution that contributes to performance, scalability and energy efficiency.

- New Global Address Service: FlaGAS

At the lowest level, the Flat Global Address Space (FlaGAS) software service will be able to cross-map the address spaces of application processes, allowing transparent access to remote memory (either RAM and/or byte addressable persistent storage, possibly including disks). FlaGAS can be located in compute nodes and/or I/O nodes, as illustrated in Figure 6:

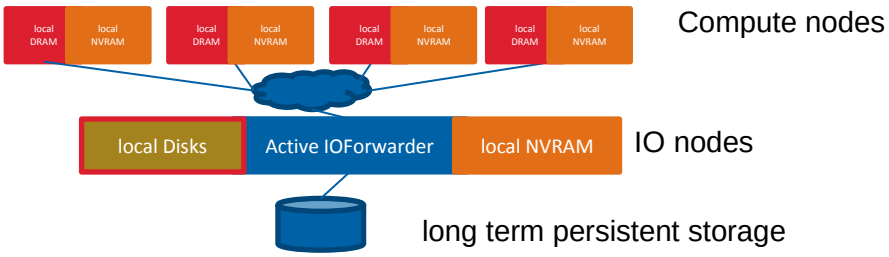


Figure 6: FlaGAS : Address Spaces Unification

Thanks to the FlaGAS service, any compute or I/O node can transparently access the DRAM and NVRAM of other nodes. **This transparent access to remote NVRAM is the foundation on which higher-level data management features will be built.**
The FlaGAS service is not intended to be used directly by applications. Instead, it is used by higher-level middleware services and programming tools (PGAS, HDF5, structured objects, dataClay) close to the object storage platform. Setting up the FlaGAS service for users will be handled by the Data Manager when it creates the DDC.

Once a process is attached to a FlaGAS instance, it can use pointers to access remote memory locations, just as if it was local memory. More specifically, as proposed, the FlaGAS implementation will function exactly as virtual memory. If a virtual address is currently mapped to a physical memory location, it can be accessed directly. Otherwise, an exception is raised to retrieve the required memory page from a remote node's addressable memory, and then the requested location(s) are accessed.
The FlaGAS service will be designed over BXI/portals to enable efficient access to remote memory.

- Object Storage Platform for Exascale

The Exascale Object Storage Platform will, through its API, provide critical new scalability features for the Exascale regime.
In term of semantics, each POSIX I/O is a transaction that, itself, must be executed in order. This is the root cause of complexity in distributed file systems and it ultimately limits their scalability. New IO APIs must provide:

- *Non blocking I/O Access*
  The I/O access interface must enable applications and/or middleware to create, write, read and free objects in a non-blocking manner. In addition to data buffers and offsets (as input arguments), these functions also take object attributes, allowing DDCs to manipulate available properties of the corresponding objects.
- *Application Driven I/O Transactions*
  I/O and metadata operations must be, ultimately, organized into transactions, which are atomic with respect to failures. In other words, either all or none of the updates corresponding to a transaction are visible to other users. A Distributed Transaction Manager (DTM) guarantees efficient management of system state consistency in an environment in which dependant data and metadata are scattered over multiple nodes to provide fault tolerance and scalability. Semantics for transactions can be specified through the API. Applications and data access middleware (MPI-IO, HDF, etc) can leverage transactional semantics.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Today data placement management in the memory hierarchy (persistent or not) is not exposed to applications and data management middleware. These features are necessary to optimise use of local persistent storage resources and ultimately data locality:

- *Layouts*
  A layout is a mapping of different parts or regions of an object to distributed storage locations (in NVRAM, Flash, Disk, etc). Each object has a layout attribute that defines how the object is stored in the cluster.
- *Containers*
  Containers provide the infrastructure to group objects as needed by the DDCs. The containers are where the "user" of the infrastructure, the DDCs in this case, specify the rules used to group objects. Mechanisms to "create", "add" and "remove" objects from containers will be provided. It is possible to specify a "bulk operation" on the containers as a whole, for migration, replication etc.
- *Reliability/High Availability*
  Layouts can be used by DDCs to specify various types of data distributions for reliability such as “parity declustered” RAID layouts with various combinations of data and parity blocks. It is possible to specify even simplistic RAID layouts such as those based on mirroring.
- *Built-in Function Shipping Infrastructure*
  A function-shipping infrastructure is used to distribute functions to data locations.
- *Telemetry*
  Monitoring and analysing data usage in real time on an Exascale machine will be needed to build a knowledge base capable of executing the proper decisions when setting up new DDCs.
  The data management interface will be used to obtain telemetry information from the object store which will be used for system analytics and diagnostics. Telemetry information will be very rich in structure and very amenable for analytics, unlike unstructured logs that system administrators resort to in today's parallel file system environments.

- Data Manager

When an application is launched, the Batch and Resource Manager provides a description of I/O resources to the Data Manager, and then one or more of the following tasks are performed:

1. Set up the system environment on the compute nodes (select an Operating System, set up a diskless environment, set up virtual machines or containers,etc.)
2. Set up the system environment on the I/O service nodes (I/O Proxies, routers, object storage configuration…)
3. Set up a FlaGAS instance
4. Prefetch data into persistent storage areas using user-provided guidelines and instrumentation from previous similar runs
5. Finalize the I/O environment setup on the compute nodes (mount remote resources, connect nodes together, etc.)

The Data Manager’s main role is to unify different I/O run time environments under the same global management tool and manage data locality, allocating I/O resources and initiating data movements.
The Data Manager is not visible from applications at run time (no API for applications). It is visible to supercomputer users at application launch time when they create or attach to a DDC.
A user who needs to run a complex workflow with many applications that data through persistent storage will first create a DDC with the required initial attributes (data objects, protocols, free space, localization, etc.) and then launch the applications attached to this DDC. During workflow execution, the user can alter the DDC (extend free space, for example) and manage data objects (move in or out, remove, etc.), and when the workflow is complete, delete the DDC.

Additionally, the Data Manager can alter object properties within a DDC during its lifetime to adapt to the supercomputer's current state. For instance, the Data Manager can move data from high speed local storage to lower speed external storage if such data is no longer being accessed by applications.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

- In-Situ Processing

  In-situ processing proposes to move away from the standard approach of saving raw data to disk and then performing results analysis post-mortem. In-situ aims to reduce data movement as well as speed up results analysis by processing the results of parallel simulations as soon as they become available in compute process memory, as shown in Figure 7.

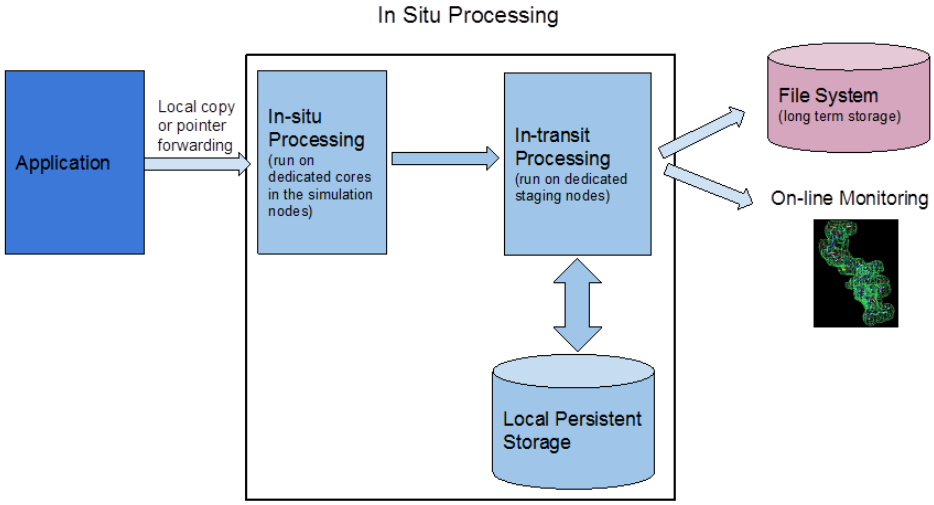


Figure 7: In-Situ Processing Pipeline

A regular in-situ pipeline takes control of generated data on each simulation node. It aggregates simulation results and performs initial processing (compression, indexation and computation of some descriptors) on one or several dedicated cores (known as helper cores). The data is then redistributed to staging nodes where it is further processed for later direct visualisation or stored to the file system for further analysis. The memory of the staging nodes may not always be sufficient to store intermediate results, in particular, when they aggregate the results of several simulation runs. ITHACA will investigate how to leverage the developed data management architecture for in-situ processing that relies on the FlowVR in situ framework.

c) *Exascale Interconnect*

The future Exascale Interconnect will increase performance, scalability, efficiency, reliability and Quality of Service (QoS) for extreme workloads. This interconnect architecture will be able to scale up to several hundred thousands nodes.

As in the current BXI, the core feature of the Exascale Interconnect will be full hardware-offloaded communication management that enables CPUs to be dedicated to computational tasks while communications are independently managed by the interconnect. Exascale Interconnect hardware primitives will map directly with communication libraries, such as Message Passing Interface (MPI) and Partitioned Global Address Space (PGAS). This hardware acceleration will enable Exascale Interconnect to deliver the highest level of large-scale communication performance, characterised by high bandwidth, low latency and a high message rate for HPC applications.

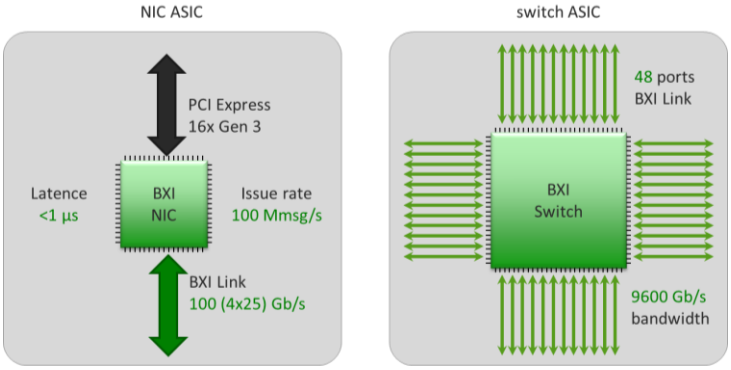


Figure 8: BXI v1 NIC and Switch ASIC Characteristics

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

As shown in Figure 8, the current BXI interconnect is based on a switch which offers a global bandwidth of 9600 Gb/s via 48 ports. The NIC contains a standard 16x Gen3 PCI Express link and provides a BXI port with a bandwidth of 100 Gb/s in each direction. A BXI NIC is connected to a BXI switch via an optical cable or via the back plane of a specific rack. Current BXI architecture is based on the Portals 4 communication library. This enables all MPI communication types to be fully optimised, including the latest MPI-2 and MPI-3 extensions and PGAS. The Portals 4 non-connected protocol guarantees a minimum constant memory footprint independent of the system size. The Exascale Interconnect architecture will be based on the next generation of Portals and will be enriched with the data management features specified during the ITHACA project.
The interconnect QoS will enable the definition of several virtual networks and will ensure, for example, that hefty I/O messages do not interfere with small data message flow. In addition, its adaptive routing capabilities will dynamically avoid communication bottlenecks. And finally, end-to-end error checking and link level retry have been implemented to enhance communication reliability and resilience without jeopardising communication performance. The Exascale Interconnect architecture will pick up all BXI features. In addition it will implement the next generation of Portals and will be enriched with the data management features specified during the ITHACA project.

d) Programming Models & Run Time Environments

All of the approaches and components described above will configure a new hardware and software stack that needs to be "exposed" to applications. We plan to provide this exposure though the programming models we will investigate. The idea is to first port and then extend at least three programming models: MPI, PGAS (represented by GPI), and task-based (represented by PyCOMPSs). These three programming models cover the most standard model used in HPC and the new proposals to make applications usable in Exascale systems.
Cooperatively to the programming model, it is also important to investigate how to leverage transparent access to remote memory (also covered by PGAS tools) and to remote persistent memory (in addition to the POSIX file standard), which is the main innovation to build upon:

- **MPI/MPI-IO**
  MPI itself has no exposed semantics to the programmer related to data persistency. It runs efficiently on top of RDMA-enabled interconnects, and as such, has no real interest in using FlaGAS. However, MPI will, at least, directly benefit from the improvements to the Exascale Interconnect. On the other hand, MPI-IO provides a means to collectively write data to files. MPI-IO could be extended to use distributed, shared persistent memory for some files (e.g. checkpoints).
- **PGAS GPI/GPI-SPACES**
  Within ITHACA, persistent memory segments will be added to GPI capabilities. This will offer GPI applications a means to store raw data to persistent storage at the lowest level (i.e. without any data management service such as naming, access rights management,etc.) and these aspects will have to be managed by the application itself, or using another middleware. The DDC concept and the FlaGAS service will help manage security (processes allowed to attach to the FlaGAS instance) and to report information such as localization of persistent memory segments (FlaGAS API).
- **dataClay**
  One further step will be supported by dataClay, that proposes an object-oriented programming style to manage persistent data within an application. A dataClay object contains data and methods to manage it, in the same way that object-oriented programming languages handle objects. Among these methods, some will be related to persistence management (make_persistent(), make_resilient()…) and others will be related to function shipping (standard object methods): dataClay will forward the method to data storage to be executed on the nodes where the data is stored.
- **HDF5**
  HDF5 extensions expose function shipping and use data persistence and object storage. Given the knowledge of the data stored in an HDF5 and the mechanisms already mentioned, many optimisations can be implemented.
- **Structured Objects**
  Following the idea of HDF5 and generalising it, we will propose a structured objects API to expose function shipping and use data persistence and object storage. The data will remain in the file, but the system will have information about its contents, and again be able to perform optimisations.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Of course, during the project, we will be ready to add other runtime environments if they become relevant for any partner.

e) Simulator, Performance and Analytic Tools

Several complementary tools are needed for the ITHACA project to succeed. Most of these tools already exist, but they will need further development to match the cutting-edge expectations of the ITHACA project. In fact, ITHACA will help create an ecosystem of technology suppliers, infrastructure providers and scientific users who are best positioned to both enrich the project and benefit from its innovations affecting almost all areas of computing.
On the innovations side, technical and academic contributions will occur regarding:

- The design, development and use of software simulators for significant components such as the interconnect layer.
- The development of the necessary software ecosystem (API, tools, added-values services, etc.) to enrich the proposed Data Management framework.
- Performance analysis and profiling tools - new specialised tools will be needed to work with the new hardware and software architectures, as well as new abstraction levels.

1.3.2 Project Positioning and Technology Readiness levels (TRL)

Despite the research nature of the project, ITHACA will deliver high Technology Readiness Level (TRL) outcomes, which will be fully integrated in a joint Data Management offering from world-leading industrials like Bull, Seagate and other partners. The Data Manager is expected to be implemented in pre-commercial deployments during the project execution phase (at least at the use case providers' locations) and, hopefully, at several other companies.
Hence, the results of the project will feature TRL levels of six (6) and above. This TRL will be reached based on: (A) the exploitation of background products and projects of the partners, which are described in the following paragraph and will provide a sound basis for bootstrapping the project's developments and (B) the iterative development methodology of the project, which will progressively produce releases at higher technology readiness levels. Additional exploitation details are described in Chapter 2.2.a).

1.3.3 Methodology

a) Driven by Co-Design and Component Implementations

The ITHACA project will use both classical and complementary R&D approaches in order to develop innovative solutions. One set of studies is driven by a bottom-up approach: new models, concepts and technologies are studied to determine how they can be used and integrated. In parallel, other tasks are driven by a top-down approach, based on new functional need studies and test results.
The experience of the project partners and previous studies that have resulted in the identification of technical or functional key domains which require research and improvement.
To concretise the co-design aspect of the project, each technical WP will contain at least one specific task that defines a detailed specification or architecture based on co-design methodology.

b) Simulations Tools

Several ITHACA studies will be based on simulations. Thus, simulation tools will be essential for the development of these tasks, and they must include support for all mechanisms and strategies that need to be evaluated. Finally, the evaluation results will indicate how to implement certain features in the technology ecosystem. For example, for communication protocol implementation, which may eventually use adaptive routing, the effect of communication protocol design decisions on the probability of activating power saving mechanisms, the interaction between load balancing and congestion management, and the interaction between fault-tolerant mechanisms and communication protocols, is significant.

c) Performance and Analytics Tools

Development and use of performance, monitoring and analysis tools will be driven by co-design considerations at each layer of the ITHACA architecture. Some low-level components will be enriched with useful and relevant data information, which will be provided to collector tools. High-level tools will

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

be created or improved with new semantics and presentation models to exploit and enrich newly collected data. In the project, outputs will be used to design and/or improve the next generation of ITHACA components.

d) Tests Vehicles

In the middle of the project's first year a complete platform that includes all items comprising a HPC cluster will be provided to the partners. The HPC platform will be a double rack cabinet that houses more than 100 compute nodes. The BXI interconnect will be configured as a two-level full fat tree topology. This platform will enable validation and performance evaluation of the first generation interconnect, and also be used to integrate Data Management components and test the project partners' applications. The test results and application benchmarks will be used to develop recommendations for the next-gen interconnect and identify improvements to the Data Management components and integration. Moreover, this platform will be open to any external partner invited during our research, for example, the European Centres of Excellence.

e) Applicable Use-Cases

Initially, we will investigate and document the requirements of relevant HPC and big data applications, and their use cases, as part of the co-design methodolgy. Then, we will demonstrate the full system for a relevant set of applications and benchmarks. The ITHACA prototype will be integrated into an HPC data cluster, on which its performance, scalability and data management efficiency can be validated. In Section 3 of this proposal, a table in the WP7 description (dedicated to applicable use cases), describes how each application will benefit from the ITHACA architecture:

- **Earth System:** ICOsahedral Nonhydrostatic general circulation model (ICON)

Within the ITHACA project, three types of data movement can be studied for the ICON climate/weather model. In order to minimize in-memory data movement, the YAXT communication library was developed at DKRZ. In this project, DKRZ plans to first explore and analyse various data access and communication patterns in ICON. Secondly, identified patterns will be mapped, as efficiently as possible, to the underlying interconnect. This phase will require feedback from the designers of the interconnect in a co-design manner. The ultimate goal will be to implement the scientific findings in the communication library YAXT and the parallel I/O library CDI-PIO and make them ready for extreme scale parallel computing.

- **Earth Climate**: In the ITHACA project, BSC will contribute by modifying their code to use a novel approach to calculating online diagnostics during computations. With the increased use of horizontal resolution in models over the next few years, an effort to manage the model's outputs will be performed to overcome the current strategy of saving the outputs to disk and then computing diagnostics over these saved outputs. BSC proposes to use innovative techniques (described in WP3) to compute diagnostics during the execution of the simulation, thereby dramatically reducing the occupied space and the time required to serve results to users. These modifications will be made at the workflow level, but also at the model code level, to enable the model to work with persistent data file techniques. More concretely, we plan to parallelise simulation and analysis codes using PyCOMPSs, to enable the simulation code to write dataClay objects in a way that is accessible for in-situ computation by the analysis codes. This change will enable the workflow to reduce data written to disk when it is not needed, and avoid mismatch impedance between the data formats of the simulator, the file and the analyser.
- **Seismic Imaging Method:** GRT is an important productive code that is working the on INFINIBAND cluster at Fraunhofer and Fraunhofer's clients' sites; however, we foresee problems with the application to continuously grow seismic data sets. GRT offers the possibility to handle the memory management itself with its GPI implementation. The two approaches can be benchmarked to each other. Within the ITHACA project, Fraunhofer will benchmark the new GASPI/GPI implementation, with extensions to the BXI, and then the Exascale Interconnect versus the current GRT-INFINIBAND version. Adaptations to the GRT code will address dimensioning of the local caches and thread-pools—the actual ratio between latency and bandwidth, and the ratio between latency plus bandwidth and CPU performance. The GRT would be a good place to test the checkpointing extensions.
- **Deep Learning:** Within the ITHACA project, Fraunhofer will follow two approaches to reduce the load of the input data stream on the network in order to free bandwidth capacities for the distributed

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

optimisation algorithm. In both cases, Fraunhofer will investigate novel data-layers for deep neuronal networks which will leverage technologies developed throughout the ITHACA project.

- **Genomics:** Using current methods, a main bottle neck of the described genomics work flows is the staging time. The ITHACA project with it's FlaGAS approach and the use of the BXI interconnect technology aims to reduce staging times. In addition, using of GPI-Space will improve the situation in two ways - the computation will be scheduled close to the stored data and the necessary communication steps will be started as early as possible, so they overlap with computation times. Together with the University of Cologne, Fraunhofer will analyse the current work flow and build a workflow generator with an customized API. GPI has provided an error state based on timeout mechanisms. Starting from genomics application requirements, the fault tolerance of GPI-Space will be collected, discussed and taken into account when designing and implementing a resilience technique for GPI-Space. Improvements of the FlaGAS hardware approach on its own and the FlaGAS plus GPI-Space approach will be compared.
- **Molecular Dynamics Simulations:** In this project, INRIA will provide GROMACS-based benchmarks with various data sets for testing ITHACA developments. INRIA will develop different output data management strategies. The first strategy is to directly store GROMACS results to file, based on ITHACA's MPI-related developments (GROMACS is MPI based) as well as with FlaGAS. The second strategy is to couple GROMACS with an advanced in-situ analytics pipeline, leveraging ITHACA's capabilities for staging intermediate results on different memory/storage levels.
- **Dynamic Flow:** For SURFsara**,** a very active research field is the interaction of fluids and particles. After water, granular materials are the second most processed materials in industry and their transport involves severe challenges, such as clustering and segregation. As a result, the simulation of Lagrangian particles is difficult to implement in a balanced and efficient way. Particles are implemented at different levels of complexity: as point particles, one-way coupled (i.e. no feedback from the particle to the fluid), as perfect balls, etc. The PGAS paradigm is better suited to the freely-moving nature of particles by giving access to the full simulation domain without the artificial boundaries of the domain decomposition. The GPI-2 API will be used to implement Lagrangian particles with all their complexities in the AFiD model.

f) *Links to other innovative research programmes:*

The Mont-Blanc project is designing the computing part of an Exascale system based on the ARM architecture, while SAGE is designing a performant and scalable IO storage framework. Together, the results of the Mont-Blanc, SAGE and ITHACA projects results will be used to propose an Exascale HPC solution comprising a very powerful data-centric cluster with optimised connections to the computing, network and data management engines. The results of these individual projects may be used separately, however their integration will enable development of a very powerful HPC solution, which will be tested in the ITHACA project.

GASPI is a project funded by the German Ministry of Education and Research (BmBF). Its main goal is to establish a standard for PGAS-APIs, namely GPI, and provide a reliable basis for future developments.

A full list of scientific and technical academic references for each project partner are included in Section 4 of this proposal.

1.4 Ambition

1.4.1 Progress from the State of the Art

a) *Advances in Interconnect*

One ambition of this project is to design an interconnect network that can be used to build an Exascale system. The global objective for the Exascale Interconnect is to quadruple the current BXI bandwidth per link and to support extreme scale topologies by connecting a huge number of nodes without increasing network diameter or latency.

This goal can only be achieved by technological advances in various domains:

- Scalability of the network and Latency of the communication
- Performance/energy ratio for the interconnect
- Readiness for data-centric topology
- Network topology and resilience

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

- Routing and congestion management
- Simulation and value-added software tools.

Besides validating the design, principles and performance of the first-generation interconnect (BXI), we plan to develop complete high-level hardware specification items (NIC and switch) for the next-generation interconnect (Exascale Interconnect). This specification will enable us to develop and industrialise a scalable, high performance interconnect, compatible with new CPU and accelerator families. It will include improvements and innovative features that will enable the Exascale step to be reached in the 2020 horizon.

b) Advances in Data Movement and Management

As presented earlier, the POSIX IO interface cannot scale anymore and a major evolution is needed in the programming models. ITHACA will address the following data movement and management points:

• **Data locality** : Moving data from persistent storage to RAM on compute nodes for processing is becoming the most energy consumptive activity on supercomputers. In many cases, it would be more efficient to move processing operations to the data, a method known as function shipping. For instance, an application that needs to locate the 3D cell with the highest temperature in an HDF5 file will send a request to the storage service, such as find_record(<file>, <temperature>, MAX). The search is run on the data node storing the file and only one cell is moved from the data node to the compute node. Another aspect where data locality can dramatically improve execution time and IO bandwidth is in complex applications worflows, in which many applications work together in pipeline mode and share data through persistent storage. Locating persistent storage close to the compute nodes is critical for scalability and energy efficiency.

• **IO transaction level**: The POSIX semantic considers that each individual IO operation is a transaction unto itself, which is the root cause of the scalability issues. This property is not particularly useful for scientific applications, for which IO transaction management at a higher level would be ideal, e.g all data related to a time step. In such a model, the application has the means to start a transaction, perform many IO operations and commit the transaction at a later time. If the commit operation is successful, the IO stack ensures that all IO operations included in this transaction are written to persistent and resilient storage. If the commit operation is not successful, then no IO belonging to the transaction is written to persistent and resilient storage.

• **Management of data persistency**: In today's architectures, data persistency is achieved through POSIX IOs—when an application wants to save data to a persistent location, it writes it to a file. With new technologies such as NVRAM, new methods for managing persistency will become possible, allowing data storage at much higher speeds and with much finer granularity than with POSIX IOs. For instance, applications will be able to use object storage with "methods" (like C++ methods) such as make_persistent(), and make_resilient() for management at a much finer grain data persistency. Writing that data to NVDimms will be done at a much higher speed than with current disk or flash memory, enabling new approaches to application check-pointing, in-situ processing, etc.

• **In-Situ Processing**:

ITHACA will enable new methods of data management by facilitating access to local persistent storage. Staging nodes can commit data to local storage, significantly extending their storage capabilities without having to rely on a traditional file system, thus saving on data movements. Fine grain data addressing will enable advanced distributed data structures to be maintained in persistent storage, allowing more efficient insertions and searches.

c) Advances in Programming Models

Current state of the art programming models distinguish between data in memory and data in persistent storage, which leads to non-optimal data access solutions. The arrival of non-volatile memories will change the infrastructure paradigm as there should be new ways to holistically address data. ITHACA will progress the state of the art model with a flat global address service which treats data, wherever it is located, in a unified way, resulting to superior time to solution and cleaner use case implementations. Further this global addressing service can be exploited by existing programming models. ITHACA also aims to connect highly optimised hardware with real-life applications. By providing optimised software implementations, the project will permit current HPC and big data applications to fully benefit from ITHACA's Data Management framework. It will also provide implementations for new Exascale-oriented programming models to enable development of next generation of HPC applications.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

The results of studies on programming models will free applications from management functions to move and locate data, as the ITHACA framework will propose the best optimisations for data movement and processing.
ITHACA's partners will make project study results available to the HPC and big data communities. At the same time, several open source initiatives are interested in the results of the ITHACA project. Consequently, the partners will maintain project-related contact and exchanges with a broad group of HPC and big data developers and entities.

d) *Advances in Simulators and Performance Analysis tools*

Bearing in mind the expected size of Exascale-level systems, developing tools that can accurately simulate these systems is a challenge, as it is even more difficult to achieve a balance between simulator accuracy and high scalability than is the case for current simulators modelling HPC (non-Exascale) systems. To the best of our knowledge, no simulation tool is currently available to accurately (and scalably) model the behavior of an HPC systems that consist of hundreds thousands of nodes. However, by applying the project partners' HPC system simulator development experience, we believe it is possible to develop an Exascale-level simulator through the use of modern simulation frameworks and advanced modelling strategies. This tool would be used by the ITHACA project, and would also be a significant contribution to the field of supercomputing simulation.
In parallel, ITHACA's performance and analysis tools will support Exascale scalability and will collect and provide all data management information required by high-level services. This data will be used dynamically, for example, by the network fabric to manage congestion, or by the resource manager or job scheduler. These tools will enable improved energy efficiency within global data centers.

e) *Advances in Application Areas*

The use cases put forward by HPC users and project partners are rich and ambitious. They include a large number of constraints that have become more and more problematic as scalability increases. These bottlenecks will be resolved by intrinsic performance improvements (latency, bandwidth, etc.), better programing models, and innovative ways of organizing and managing the data. The large scope of the proposed use cases will enable us to be as exhaustive as possible in our recommendations and conclusions. Also, and most importantly, submission of these use cases enables us to fully anticipate the future needs of HPC applications. At the conclusion of the ITHACA project, these applications will be ready to run on an Exascale system.

1.4.2 Innovation Potential

ITHACA's strategic objective to be exploitable as a primary European data-centric computing platform that can be used for both commercial big data and big science domains, in the various ways described in the following section. For big science applications, ITHACA plans to set the trend for building advanced and performant data management infrastructures and make them available to European Centres of Excellence, like ESiWACE or to a HPC data centre like PRACE.
ITHACA will integrate its interconnect and data management improvements in several open source HPC programming models and standards. This integration will allow the HPC community and big data applications and communities to directly benefit from the scalability and performance of ITHACA's components.
ITHACA will blur the lines between traditional memory and storage, between the data processing and the data movement mechanisms in future HPC or big data infrastructures leading to tangible benefits for many applications and users moving toward Exascale who today, are severely memory bound.

2. Impact

2.1 Expected Impacts

The ITHACA project is aligned with the objectives set by the FETHPC-1 call to develop core technologies and architectures. The project objective is to design a Data Management framework associated to an interconnect network that will be able to address Exascale computing. The project will target technological research issues and will deal with cross-cutting issues like energy efficiency and resilience. It will use a cooperative design approach, looking at various applications to specify and design

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

a data management solution to support Exascale applications. The project will contribute significantly to the development of HPC and big data technologies in Europe.

2.1.1 : Impacts:

a) Community impact

The effects of today's convergence of big data with compute is becoming increasingly obvious to the HPC community, especially in the areas of social data analysis. The results of the scientific work are, perhaps, less obvious to the public but of equal importance, as there are many examples of its effects on our daily lives.
Several academic members of the ITHACA project are involved in international and strategic scientific or technical programs:

- CEA – energy domain.
- BSC – climatic and chemical domains.
- DKRZ – climatic domain
- SURFsara – mechanical domain
- Fraunhofer – seismic domain
- UKOELN – life-science and genomic domains
- INRIA – life-science domain
- UPV & UCLM – network architecture domain

The project will help increase HPC's value to the scientific community and enable technologies that support economic prosperity and the wellbeing of the public.

b) Market Impact

The principal expected market impact will be ITHACA providing the right technologies at the right moment versus the market expectations listed in the following schema. Indeed, ITHACA will present an authentic European solution to building Exascale data-centric solutions.
The project covers important segments of broader HPC markets. Because of the applications targeted by the project, the results will positively impact the domains of life sciences and genetics, seismic and climatic applications, and complex flow simulations.
As described above, the result of this project will be of strategic interest to the big data market via data analysis applications. The Commercial Technology delivery partners in ITHACA include world-leading industrials and a European SME:

- Seagate – The world's largest data storage technology provider,
- Bull (ATOS company) – Europes premier HPC systems technology provider and a subsidiary of ATOS Europe's largest digital services deliverer
- ARM – A world leader in the processor architecture industry.
- Allinea – A fast growing SME with an extreme HPC toolkit

The direct exploitation of these market opportunities by the ITHACA consortium, during and after the project's conclusion, will be discussed in section 2.2.a)

c) Impact in Relation to the Work Programme and the Strategic European Agenda

The project meets the objectives set forth in the roadmap in the Strategic Research Agenda (SRA) issued by ETP4HPC, regarding the following areas:

- HPC system architecture and components: The project will design a data management architecture and an interconnect that will improve scalability and performance in compliance with the SRA's objectives. The project will also address the resilience and energy efficiency issues underlined in the SRA.
- Programming environment: ITHACA's programming model contributes to the objectives outlined in this section of the SRA.
- Balance compute subsystem, I/O and storage performance: The project will address the optimisation of the cluster architecture for communication between nodes and storage.
- Simulator and Performance analysis tools: The project will provide advanced tools to enable improved knowledge and analysis of data intensive applications.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

- Extreme Scale Demonstrator (ESD): The project will provide the technologies and infrastructure for 2018, the timescale to integrate into a pre-Exascale data centre, as planned by SRA.

2.1.2: Barriers and Obstacles

The project will base the realization of its impacts on technological innovations in both hardware design and middleware architectures. However, there are also external factors (threats) that could adversely affect realisation of the project's impacts:

- **Emergence of competing platforms:** Other Data Management systems may improve on the operational side and in areas that differentiate the ITHACA solution. This risk will be mitigated by continuous innovation within ITHACA that will render current technologies obsolete even as they are improved by competitors.
- **Limited market acceptance:** Market penetration is always a big challenge and new solutions, such as proposed by ITHACA, are often perceived as risky by decision makers. This risk will be overcome by identifying market segments in which the requirement for a real-time analytics solutions is so acute that the perception of risk is overcome by urgent need. Additionally, the partners will make a significant effort to identify market segments comfortable with innovation and early adoption. Acceptance will be more likely in early stage commercialisation.

2.2 Measures to Maximise Impact

a) Dissemination and Exploitation of Results

As the consortium represents several countries and includes partners participating in a number of markets, exploitation of the project's results is likely to rapidly benefit major European HPC users, in research as well as in industry, across many domains. Thus, the HPC innovations realized in the ITHACA project will contribute to the competitiveness and independence of Europe in the face of the vast research and industrial capability wielded today by North America and Asia.

About Dissemination:
Most of the project's results are expected to be published in highly-rated journals and at major conferences. Open access to these publications will be granted with a budget allocation specifically for this purpose. There is a possibility of patents ensuing from specific research results.. All these areas of dissemination will be strongly encouraged and supported during the lifetime of the project.
Project kick-off and results will be highlighted in online HPC journals that provides technical information, news and strategic business insights for executives.
Dissemination of results will be via multiple paths to ensure the broadest impact for the project, including:

- Conferences, forums, and industry events
- Open-access publication of partner papers and articles
- Other communications activities addressed in the following section

Researchers participating in the consortium have some experience publishing in top venues and creating a research impact based on citations (as shown in Table 2, information from Google Scholar):

Project Researcher	Number of citations (source Google Scholar)	Index h (source Google Scholar)
Jose Duato (UPV)	12506	51
Rosa Badia (BSC)	4827	35
Toni Cortes (BSC)	1653	22
Bruno Raffin (INRIA)	1465	20
Francisco J. Quiles (UCLM)	1176	17
Walter Lioen (SURFsara)	530	10
Julian Kunkel (DKRZ)	447	9
Pedro Garcia (UCLM)	350	10
Ulrich Lang (UKOELN)	312	11
John Donners (SURFsara)	181	6

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Marc Perache (CEA)	169	7
Pierre Vigneras (BULL)	94	6

Table 2: Project Researcher Citations (Google Scholar)

By associating open-source initiatives with our research, the development results will automatically be disseminated widely within the HPC development community.
Dissemination activities are described with more detail in the WP2.

About Exploitation:
Exploitation of the project's results falls mainly into three categories:

- Direct commercial and industrial investment returns: creating a new HPC system (hardware and/or software) integrates new, innovative technologies and also offers new expertise and services associated with these new technologies.
- Recognized expertise and active participation in this cooperative project by several major standards and open sources initiatives within the HPC domain.
- Direct use of ITHACA's results by scientific research labs, e.g. European Centres of Excellence, such as ESiWACE.

The partners will directly benefit from this project as described belows:

- **Bull** will industrialise the next generation of interconnect hardware products (NIC and switch) that are studied and specified during this project and integrate them into its future Exascale supercomputing offering (around 2020). In parallel, Bull will integrate improvements and new Data Management features for the software tools and APIs that are selected, studied and developed by the ITHACA project. This software will be integrated in the open source HPC ecosystem thereby providing it to HPC development communities.
  Bull plans to combine the results of the ITHACA project with those of the Mont-Blanc and SAGE projects to produce a new HPC solution planned for the 2020 timeframe. The Data Management middleware, the new interconnect features and the small size and reduced energy consumption of the nodes will allow this solution to scale to several hundred thousand nodes, an important step in "Bull's Exascale program".
  The ITHACA results may also be exploited by processor architectures other than the one developed for the Mont-Blanc project. ITHACA's Data Management framework, containing the BXI interconnect, will be a central element for an Exascale system with very promising exploitation possibilities.
- The key market driver for **Seagate**'s exploitation in ITHACA is the High Performance Data Analysis (HPDA) market as defined by Gartner which includes data intensive modeling and simulation, and new higher performance data analytics. IDC anticipates that the HPDA storage sector will grow very robustly at 26.5% CAGR between 2013-2018, with a market value of $1.5B in 2018. IDC also expects the landscape for storage to change significantly for HPDA in reaction to emerging and evolving requirements.
  Features of the next generation object storage technology that will be developed in the project can be integrated into near-future product offerings. One objective is to make next-generation object storage and its API be the defacto storage software technology of choice for future extreme scale and data-centric computing platforms. Indeed, there is an urgent need within the community for storage solutions that specifically address data-centric extreme computing. ITHACA will develop many of the key features of next-generation object storage technology, to be integrated into future offerings; the base architecture of which is already provided in the SAGE project. We also anticipate the creation of new IP around the software technology coming out of the project that can be exploited in future products.
  ITHACA presents a wide variety of use cases, introducing new networking technologies such as BXI, data management solutions such as Phobos and programming models such as FlaGAS. The project exposes object storage technology to an entirely new ecosystem of tools methods and techniques - providing strong inroads into a potential future customer base.
  Seagate also anticipates increasing its exposure to HPC and cloud storage hardware platforms, including disk drives, SSDs, Flash technology and enclosures within the community through the dissemination of ITHACA'results. Existing products in the market based on related technology are exemplified by the ClusterStor series and A200 product. The co-design of storage system

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

components will be used to provide inputs to company technology roadmaps (e.g. optimising NVRAM utilisation focussing on data-centric computing aspects).
When planning or implementing storage solutions today, there are no tools to assist or guide the implementer to provide an optimised solution that matches user needs; over provisioning is the only effective (but expensive) way to ensure that performance and other system goals will be met in production systems. The simulation and modelling capabilities created in the project can be leveraged to both provide such capabilities; it is anticipated to build a community effort to continue tool development for such optimisation.
Seagate and Bull's joint work on the ITHACA project is expected to further strengthen their existing product collaborations and provide the European community with industry-leading HPC product offerings in the area of extreme scale data-centric computing, including the provision and use of BXI, and then Exascale Interconnect and Mero technologies.

- In ITHACA, **CEA** will develop multiple open source projects. The project results will be included in the following opensource projects MPC, MALP, Phobos and Pcocc. Work around MPC and MALP, will help provide an alternative European MPI library suitable for the Exascale era with MPC and a profiling tool dedicated to large systems with an optimized IO support. Work on Phobos will provide new features in management of object storage resources to HPC community. Pcocc will bring major improvement in virtual cluster support on large HPC systems. CEA will work with Seagate and Bull to improve security for HPC cluster networking.
- **UPV** is a technical university that develops mostly applied research. The research group participating in ITHACA, known as GAP, is widely recognized at the international level as one of the leading groups on interconnect architectures. The ITHACA project constitutes a unique oportunity to transfer GAP's research results to industry. GAP will also benefit from direct collaboration with industrial partners on the specification and design of state-of-the-art interconnect technology. Additionally, the simulation tools for scalable interconnects developed in ITHACA will be used by GAP for other studies on interconnection networks. These tools will be released to the public domain, thus enabling other groups to develop this kind of study and increase GAP's international visibility.
- For **UCLM**, the different results obtained from ITHACA will significantly impact several aspects of research activity developed by RAAP, the group involved in the project. First, the overall background of the group, as well as its future research lines, will benefit from first-hand knowledge of the state-of-the-art advances in the interconnects which are expected from ITHACA. The different simulators which are planned for development in ITHACA will significantly improve and expand the capabilities and performance of current interconnect simulators, thereby allowing RAAP to exploit the novel tools that strongly support future research lines. Moreover, these simulation frameworks will also benefit from new traces derived from the novel architectures, applications and programming models developed in ITHACA. Finally, collaborations with industry, especially joint publications and patents, are highly appreciated in the academic world, so the members of RAAP will benefit from this aspect of the expected research results.
- **Fraunhofer** is a centre that undertakes applied research to drive economic development, and it is experienced in understanding the marketplace. Fraunhofer foresees several exploitation channels for the work of ITHACA. The project's results will be included in the licensed software stack (GPI and BeeGFS) and the applications (GRT and Deep Learning) will be prepared for the exascale era and will be commercially exploited.
- **BSC** will exploit the results in three ways: COMPSs, dataClay, and climate model. The COMPSs distributed computing platform that is been used in production at BSC in its infrastructures. Applications that have benefited from COMPSs are from the areas of bioinformatics (cancer research, drug research), astrophysics (Gaia project), or green buildings design, among others. COMPSs will be extended in the project in order to be able to deal with future architectures. BSC plans to continue exploiting COMPSs in its infrastructures as well as promoting it for its use in other funded projects and in contracts with companies. dataClay is a key technology designed to be transferred for exploitation to a Spin-off that should commercialize it (mid-term), through research by the spin-off. IP will stay at BSC, however. In addition, dataClay is planned to be used internally at BSC for science applications. dataClay will be extended in Ithaca to be ready for the new architectures based on NVRAM that we expect will become mainstream in HPC furst, but will later

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

become of much wider use. This will enable the exploitation of dataClay both in the academic arenas as well as by the industry (trhogh the spin-off). Finally, BSC will improve the weather model and make a new version that can take advantage of the new kind of archtiectures.

- **Allinea** will commercialize and exploit the results of the ITHACA project by industrializing and integrating the most promising components into its existing product offerings for software development and analysis. As the leading cross-platform vendor of software tools for high performance computing it is in a unique position to be able to bring such work to market and to enable the results to be used within its existing and new customer base. This customer base is global and currently counts over 50% of the Top 500 HPC systems, and 7 of the top 10 systems.
  Specific extensions for improved profiling on this and other platforms of the future where I/O is increasingly significant and adding support for additional standards enables better support and applicability of technology worldwide. By working with project partners that provide their own tools or components, such as CEA and ARM, we can help widen the route to the HPC market for outputs of the project.
  Furthermore, by working with the application partners during the project, the benefits of applying prototype tools and expertise in performance optimization will be of immediate value to the partners enabling them to target their work more efficiently and improve the results.
- **DKRZ** is a centre that undertakes applied research to improve the HPC infrastructure for climate modelling. In order to enhance the quality of climate projections in Europe, experiments with very high resolution are necessary. ICON is a next generation earth system model, which is used by the German weather service and which has the ability to cary out simulations at very fine resolution, up to 100 meter on regional scale. Within the framework of ESiWACE (one of the European Centers of Excellence in HPC applications funded through H2020), which is being coordinated by DKRZ, ICON has been chosen as an Exascale demonstrator by performing experiments with a resolution of 1 km globally. Such a resolution on global scale is required to resolve clouds and small ocean eddies and thus improve representation of high-impact extreme events. The hard- and software components required so as to run such computational and data intensive simulations within a reasonable time frame in operational mode are not yet available. For this reason, the results of ITHACA will be of great value to ESiWACE and the community of climate and weather modellers so as to enable breakthrough in climate and weather research.
- **SURFsara** supports research in the Netherlands by developing and offering advanced and sustainable ICT infrastructure, services and expertise. ITHACA will keep SURFsara at the forefront of HPC technology, to ensure that it can provide relevant information about the ITHACA infrastructure and toolset to its users in its daily support tasks. SURFsara expects to create new business models by collaborating with and assisting some SMEs that are targeting the market for large scale industrial HPC applications.
  The ITHACA developments will be exploited in the AFiD code, an open-source computational fluid dynamics code. This code is representative of many codes in fluid dynamics, which is highly relevant for energy research, life-sciences, the food processing industry and turbulence research in general. It is expected that the general design and optimization strategies developed in this project will find their way to other applications, possibly running on the ITHACA architecture
- **INRIA** is a public computer science and applied math research institute. ITHACA co-design strategy brings vertical integration from hardware to applications that will be highly beneficial to INRIA research effort. ITHACA will gain access to low-level networking and storage technology, advanced monitoring capabilities, as well as a wide range of mini-apps that are of high value for INRIA's research. INRIA will develop and validate state-of-the-art algorithms and multi-criteria optimisation strategies for reducing data movements. Results will be accessible first to ITHACA's partners and then published in topranked journals and presented at conferences. ITHACA will enable INRIA to strengthen FlowVR, its framework for in-situ processing, in a domain for which a de facto standard has not yet been defined. INRIA will also develop new data management strategies to analyse molecular dynamics data, enabling to enforce its effort to develop multidisciplinary research with life science groups. INRIA will involve young scientists (Postdoc, PhD sand Master students) in these efforts, providing them with a high quality training that they, in turn will transfert to European companies or research labs where they often are hired.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

- **ARM** today provides commercially-supported compilers, libraries and performance analysis tools for HPC applications running on the ARM architecture. These performance tools, through analysis of a user's application runs and interaction with the other components, present the user with actionable advice to improve the performance of their application. ARM aims to add the results of its research done under ITHACA to these existing HPC tools as a practical and available application of the research. A core goal of ARM's HPC performance analysis tools is that, after collecting complex low level data, this data is transformed into a series of actionable advice items. Actionable advice is high-level code restructuring advice aimed at allowing those without a deep understanding of the low level hardware and systems to make full use of those systems. ARM's project participation will broaden the range of end users able to take advantage of the innovative HPC systems developed under ITHACA. The ARM HPC User Group meets at each annual U.S, SuperComputing conference. ARM can use this event to promote and showcase progress on the ITHACA project. ARM also runs an annual invitation-only NDA event, the ARM Partner Meeting (APM) that provides an opportunity to demo and discuss novel and emerging applications of the ARM architecture. 2016 has seen HPC-specific booths at this event for the first time and ARM can promote ITHACA and the tools developed under the project to its partners at the APM.
- **University of Cologne (UKOELN)** is a public university. Contributions to ITHACA will come from the Regional Computing Centre at the University of Cologne (RRZK). RRZK has a large expertise in the field of genetic pipelines on HPC-Systems and will contribute this knowledge to ITHACA to help develop middleware to enable modern supercomputers to efficiently drive such use cases. Together with Fraunhofer, the University of Cologne will analyze important genetic pipelines and build a workflow generator with a customized API under ITHACA. By enabling the applications in the pipeline to use a common memory space, IO driven by staging will be reduced to a minimum, effectively reducing the required bandwidth and enabling much higher quantities of genomes being processed in an HPC-Environment under ITHACA**.**

b) Communication activities

Besides being used for dissemination, a project website and social network accounts (publicly available) will be created to publish information about the project and to enable information exchange. An internal project website will be created to serve as the central platform for information exchange (internal documents generated as part of the project) between project partners and as a software repository (benchmarks, applications, simulators, ecosystem software, etc.).

The external website and social network accounts will not only disseminate project results, but also be a collection point for research received from the world's top universities and research groups in the high performance interconnects domain, thus helping to promote the next breakthrough in design technology for Exascale networks and data management middleware. The website will become a reference point for highly scalable data application developers worldwide and the place where the most recent and interesting solutions for the industry can be found.

We will propose to show and explain our Data Management architecture to the European Centres of Excellence, which will benefit from our results and contribute to the co-design effort. We will offer them the use of our vehicles of tests to realise proof of concept with their specific use cases.

We will published two videos—the first one at the beginning of the project to introduce ITHACA and explains its objectives, and a second one at the end of the project, to show its results and benefits realised.

We will also offer tutorials and workshops about our Data Management solution and highly-scalable Data Management framework, at several conference, including: HotInterconnect (US), HIPINEB (US), ISC (Germany) and/or SuperComputing (US). These workshops will be prepared by the consortium in order to present the most significant project results with presentations and some demos. The most suitable dates for these workshops will be towards the end of the project, so ISC in June 2020 appears to be a promising (although tentative) date.

Finally, the consortium will also organize a summer school (at the end of the project) targeted at students and professionals interested in the data management and interconnection networks fields. This school will include introductory talks about our architecture and components, and discuss the principal project outcomes (technology, algorithms, programming models and results).

See WP2 description for more details about our dissemination workplan.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

3. Implementation

3.1. Work plan — Work packages, deliverables

1. List of work packages (WP)

In accordance with the objectives, concepts and approach outlined in the preceding paragraphs, we have structured our work plan so that it consists of four groups of complementary work packages :

- Management and Dissemination:

The two first work packages are dedicated to the administration of this cooperative project: WP1 for Operational and Technical Management and WP2 for Communication and Dissemination actions, which will be the results of other work packages.

- Data Management concepts and programming models:

The WP3 aims at studying current and future models of Data Management mechanisms and API in order to propose new improved programming models that evaluate the strong and weak point of current programming models for the new architecture and data management mechanisms.

- Data Management framework & Interconnect:

The two next technical work packages will interact to propose a new Data Management framework based on the new concepts and models from WP3, and will be implemented via a set of hardware and software components. The WP4 will specify and implement the different needed components (hardware and software) to efficiently provide a uniform access to the data (independently of where they are stored) for an application, to put the data near processing and reciprocally. The WP5 will contribute to this efficiency by proposing a performant and enriched interconnect to move the data and requests in a scalable cluster. This work package will provide at first a new advanced interconnect, and then a future generation will be specified via a co-design way with the other WPs. These studies integrate cooperative works with other consortium and OpenSource communities. The WP6 will complement this group of technical work packages by developing added-value services, high-level API and tools based on low layers and API of WP3, WP4 and WP5. These items will help to define the I/O profiling of an application, to improve the resource management and the power-efficiency of a cluster, to enhance the monitoring and reporting. These developments will enrich the ecosystem around ITHACA's Data Management architecture and technologies.

- Use cases and benefits:

The Seventh work package will interact with all previous work packages to integrate the newly defined programming models, to use the enriched monitoring tools and manage the applications' data more efficiently. The applications will be run and the results will be analysed to show the benefits of ITHACA architecture and to give feedback and new requests to the other work packages for future improvements. The following table summarizes the ITHACA work packages information:

WP No	Work Package Title	Lead Participant No	Lead Participant Short Name	Total Person-Months	Start Month	End month
1	Management	1	Bull	79	m1	m36
2	Dissemination	5	UCLM	80	m1	m36
3	Data Management concepts & Programming models	8	BSC	328	m1	m36
4	Co-Design & Data Access framework	2	Seagate	610	m1	m36
5	Co-Design & Data Movement interconnect	1	Bull	578	m1	m36
6	Data Management Ecosystem	3	CEA	268	m1	m36
7	Data Applications use cases	6	DKRZ	287	m1	m36
			Total person-months	**2230**		

Table 3: List of Work Packages

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Each work package is composed by a set of tasks which are organized during all the time period of the project as shown in the following planning:

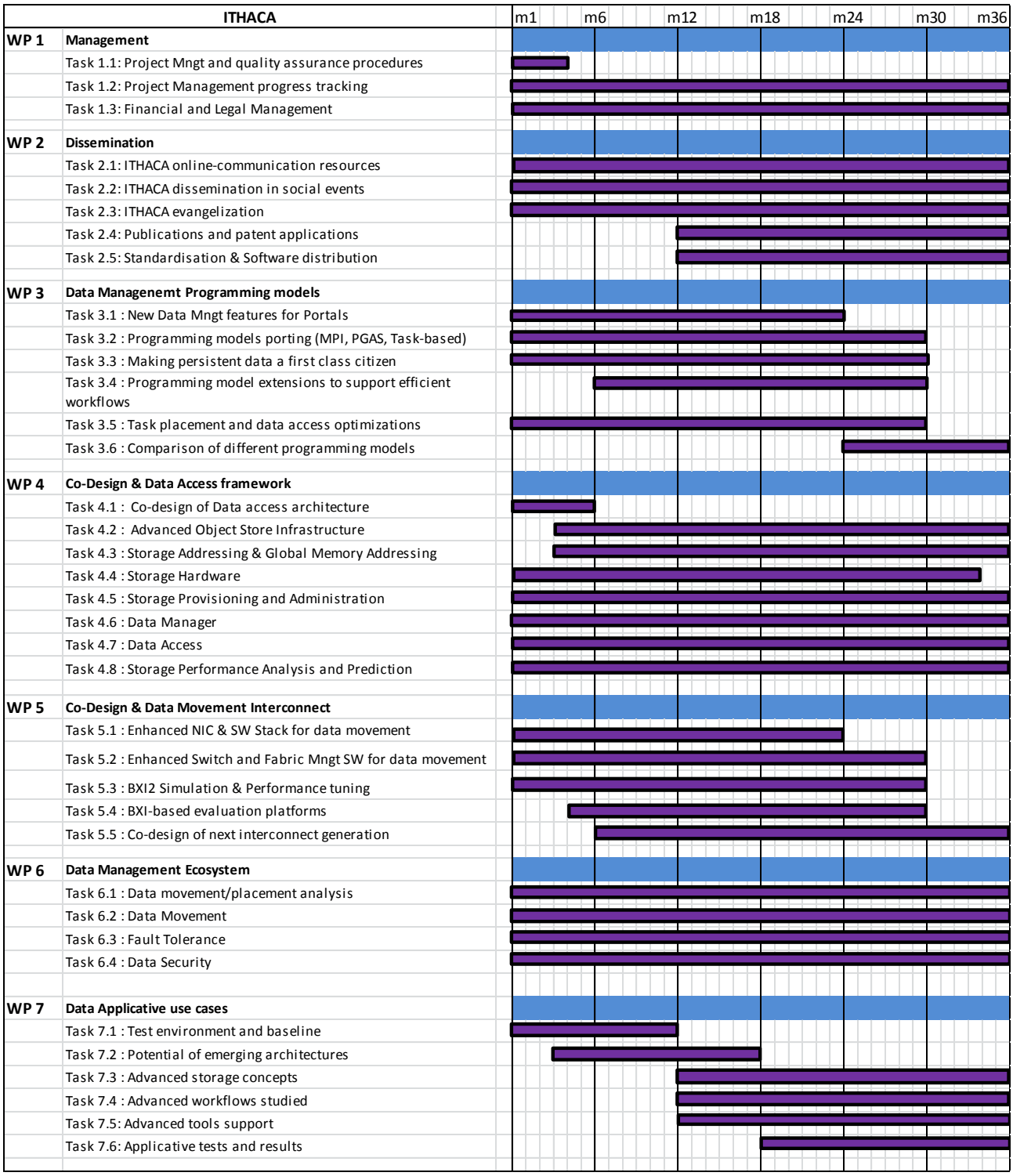


Table 4: Duration of tasks

The following chapters describe in detail each work package and their associated tasks.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

2. *Work package 1: Management*

Work package number	**1**		**Start Date or Starting Event**						**m1**				
Work package title	**Management**												
Participant number	**1**	2	3	4	5	6	7	8	9	10	11	12	13
Short name of participant	**Bull**	Seagate	CEA	UPV	UCLM	DKRZ	Fraunhofer	BSC	Allinea	SURFsara	INRIA	ARM	UKOELN
Person/months per partner	**18**	9	6	8	8	6	2	12	4	2	1	2	1

Objectives :

This work package is devoted to project management and administrative tasks.
The main objective is to facilitate technical work, encourage synergy among the partners as well as consistency among the various technical tasks, and ensure that expected results are achieved in time, within the allocated budget, and are of good quality.

Description of work: **Lead partner: Bull**

All partners contribute to this work package.
The following is a list of the tasks required to achieve the objectives of this work package. The high-level Management Structure as well as the individual roles and responsibilities within this structure are explained in § 3.2 of this document. The chapter 3.2 also includes a brief overview of the most important procedures of the project which will be further defined in the early months of the project as described in T1.2.

T1.1. Project management and quality assurance procedures, Collaborative tools (m1:m4) [Bull & all partners**].**

In this task, we will define and implement the appropriate administrative project management processes (and tools if applicable) that ensure accurate documentation, reporting and justification of the work being carried out. We will develop a process to ensure that the deliverables have been reviewed by a broad spectrum of individuals against a well-defined set of criteria. Moreover, we will determine the minimum level of quality required for presentation to the European Commission as official outcomes of the project. The high-level principles guiding these procedures will be agreed to at the start of the project at the Kick-off Meeting. Appropriate strategy will be defined to ensure clear communication channels and tools between all partners in order to facilitate the exchange of critical project documentation and news and to encourage participation in the decision-making process. The task will define adaptations and maintain the internal collaborative tools from the ITHACA project for sharing documentation and communicating work status. These administrative project management and quality assurance processes including collaborative tools will be documented in the Project Handbook.

T1.2. Project and Technical management progress tracking (m1:m36) [Bull & all partners**].**

In this task, we will ensure that technical project progress is in line with the plan of record as described in this Description of Work. The task will primarily consist of gathering and assessing technical progress (and its relationship to the effort tracked in T 1.3) on a regular basis in the form of a monthly General Assembly and Technical Board teleconferences. The status of technical progress will be provided on a regular basis to the European Commission in the form of Annual Progress Reports. It will also consist of organizing the face-to-face General Assembly (GA) and Technical Board (TB) meetings every six months that will focus on detailed project planning. This task will also handle the grant agreement including amendments as well as the Consortium Agreement with T1.3; the organisation and preparation for three reviews; the quality control of all deliverables and external documents in accordance with outputs of task 1.1.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

T1.3 Financial and Legal Management
(m1:m36) [Bull & all partners**]**

This task will manage the financial and legal aspects of the project collaborating with T1.2. In particular the following activities will be undertaken:

- The control and monitoring of the timely and accurate production of cost claims;
- Liaison with the beneficiaries with regard to the financial and contractual aspects of the project;
- Liaison with the European Commission with regard to the financial and contractual aspects of project.

Deliverables :

D1.1) ITHACA Project Management Handbook: **(m4): [Bull** & all partners**]**
This document will provide an overview of the management and administrative procedures of the project in order to ensure efficient project execution as well as high quality project results. The deliverable will be accompanied by a document that will describe the collaborative tool and/or website for sharing project technical and administrative status and documentation with all project participants

D1.2) First annual Progress Report: **(m12): [Bull** & all partners**]**
This report will be based on the Guidelines provided by the European Commission as in the article 20.3 of the H2020 annotated Model Grant agreement.

D1.3) Second annual Progress Report: (m24): [Bull & all partners**]**
This report will be based on the Guidelines provided by the European Commission as in the article 20.3 of the H2020 annotated Model Grant agreement.

D1.4) Third annual Progress Report: (m36): [Bull & all partners**]**
This report will be based on the Guidelines provided by the European Commission as in the article 20.3 of the H2020 annotated Model Grant agreement. The report will include any required Certificate on the Financial Statement (CFS).

D1.5) Project Final Report: (m36): [Bull & all partners**]**
This report will be based on the Guidelines provided by the European Commission (article 20.4 of the H2020 annotated Model Grant agreement) and will include the following: a final publishable summary of the work completed to date covering results, the conclusions and socio-economic impact of the project, a chapter on awareness and wider societal implications as well as a report on the distribution of the Community financial contribution. It will be presented in conjunction with the final report of Dissemination (deliverable D2.10).

3. Work package 2: Dissemination

Work package number	**2**		**Start Date or Starting Event**						**m1**				
Work package title	**Dissemination**												
Participant number	1	2	3	4	**5**	6	7	8	9	10	11	12	13
Short name of participant	Bull	Seagate	CEA	UPV	**UCLM**	DKRZ	Fraunhofer	BSC	Allinea	SURFsara	INRIA	ARM	UKOELN
Person/months per partner	12	9	3	10	**26**	6	2	3	2	2	2	2	1

Objectives :

To consolidate the relationships between the project and specific target audiences (including scientists, scientific bodies, industry, industrial advisors and general public), ensure their understanding of its progress and achievements, and disseminate its results, products or services created by the project.
Specific objectives:
- To stimulate the demand of the hardware and software solutions developed in the ITHACA framework

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

- To continue to build awareness of the project through the communication campaign
- To reach additional key stakeholders in new sectors using appropriate communication channels, as well as to maintain contacts realized during the Mont-Blanc3 & SAGE projects
- To continue liaising with other Exascale initiatives in Europe and worldwide with the aim to position Europe as innovation leader in the Exascale race
- To ensure the collaboration with related industry from HPC and other related markets (server, IT, etc.)

Description of work : **Lead partner: UCLM**

All partners contribute to this work package.
The following is a list of the tasks required to achieve the objectives of this work package.

T2.1: ITHACA online-communication resources
(m1-m36) [UCLM & All partners]

This task will provide a public website and social-network profiles intended to promote and make visible ITHACA, as well as an online platform intended to ease the exchange of information and software inside the project community.

sT2.1.1: Public project website
(m1-m36) [UCLM & All partners]

A public project website will be created to include information of all the partners, as well as of the achievements, publications or events organized in the project framework. Besides, this website will include a section of relevant news not derived from the project but related to the project topics in order to make the project much more visible. Indeed, this website is intended to become a reference point for anyone interested in any aspect of HPC and BigData, where the most recent and industry-appealing solutions coming from the world's top universities and research groups in the high performance interconnects domain can be found. This subtask will contribute to deliverables D2.3, D2.5, D2.6 and D2.10.

sT2.1.2: Project-community online platform
(m1-m36) [Bull & All partners]

A project-community online platform will be created to serve as a point for technical information exchange as well as a software repository. Note "community" does not mean necessarily "project partner", as we consider inviting external developers to contribute (with the logical access restrictions). This platform is intended to be a fundamental management resource as well as a place to keep essential communication among partners. For instance, this platform will support calls for meetings, task-progress monitoring, software updates, deliverables in progress and submitted, administrative information, etc... This subtask will contribute to deliverables D2.1.

sT2.1.3: Social networks
(m1-m36) [UCLM & All partners]

During this project the most common current social networks will be used for dissemination purposes. Specifically, Twitter and Facebook profiles of the project will be created, plus a LinkedIn group, intended to spread project achievements, as well as to promote events organized from the project framework. Specific strategies and policies will be agreed to use these profiles as efficiently and co-ordinately as possible, in order to maximize their impact and visibility. This subtask will contribute to deliverables D2.2, D2.5, D2.6 and D2.10.

T2.2: ITHACA dissemination in social events
(m1-m36) [UCLM & All partners]

This task includes activities to spread the main achievements and results of ITHACA through direct contact with different types of audiences

sT2.2.1: Organization of dissemination activities in conferences and industry events
(m1-m36) [UCLM & All partners]

Project partners will organize booths, stands, tutorials and/or workshops in conferences and industry events (or any other similar forum) to show the main goals and the most outstanding results of the project through posters, presentations, videos, and (if possible) some demos. It will be especially encouraged and supported the organization of these activities in those conferences and events which are reputed to have a high number of attendants (for instance, HiPEAC and ISC conferences), in order to maximize the potential audience. All these activities will be announced, promoted, and later summarized through the project website, social

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

networks and press releases. This subtask will contribute to deliverables D2.5, D2.6 and D2.10.

sT2.2.2: ITHACA Summer School

(m30-m36) [UCLM & All partners]

The consortium will organize a summer school targeted to students and professionals interested in HPC and BigData systems. This school will include some introductory keynotes and talks about HPC and BigData generalities given by reputed experts in these fields, but also other talks to explain the main outcomes of the project (technology, algorithms and results). In order to offer the widest vision of ITHACA, this summer school is tentatively scheduled for the summer of the last year of the project. The summer school will be announced, promoted, and later summarized through the project website, social networks and press releases. This subtask will contribute to deliverable D2.9 .

T2.3: ITHACA evangelization

(m1-m36) [UCLM & All partners]

This task includes activities to emphasize the importance of the ITHACA solutions, targeting not only at the general public but especially at the European Centers of Excellence

sT2.3.1: European Centers of Excellence

(m1-m36) [BULL & All partners]

Special care will be taken to let the European Centers of Excellence know the relevance of the ITHACA solutions through messages, summaries and technical reports specially addressed to these Centers. In addition, demos will be organized in those Centers interested in receiving the corresponding ITHACA experts to be directly aware of the ITHACA solutions. This subtask will contribute to deliverables D2.5, D2.6 and D2.10 .

sT2.3.2: Press releases

(m12-m36) [UCLM & All partners]

Press releases will be elaborated to announce the most outstanding project achievements. They will be sent not only to mass communication media but also to any communication channel accessible to ITHACA partners, such as newsletters or forums belonging to associations, research networks, companies or universities that ITHACA partners belong to and/or collaborate with. Similar to social networks, a common policy on press releases will be agreed among ITHACA partners. This subtask will contribute to deliverables D2.6 and D2.10.

sT2.3.3: ITHACA Videos

(m1-m36) [UCLM & All partners]

Two dissemination videos will be elaborated by professional video producers. The first one will be released at the beginning of the project, stating the project objectives and expected benefits. The second one will be released at the end of the project, summarizing the project results and achievements. These videos will be released as contents of the project website, as well as part of the activities in conferences and events indicated in Task 2.6. This subtask will contribute to deliverables D2.4 and D2.7.

T2.4: Publications and patent applications

(m12-m36) [UCLM & All partners]

Technical papers and articles reflecting project developments and proposals should be published in highly-ranked journals and major conferences. Besides, some other results from the project will possibly be patented, provided the corresponding agreements between all the involved project partners. This type of dissemination activities will be strongly encouraged and supported during the duration of the project. All the publications and granted patents will be spread through the project website, social networks and press releases. This task will contribute to deliverables D2.5, D2.6 and D2.10 .

T2.5: Standardisation & Software Distribution

(m12-m36) [UCLM & All partners]

Special care will be taken to promote the ITHACA new concepts and improvements among the HPC and BigData developers and users communities. We will aim at using or at creating standard API and tools for these communities. Ideally, this would contribute not only to make visible the ITHACA solutions but also to get valuable feedback from these potential users of these solutions. In that sense, as part of this task, some of the software developed in ITHACA will be distributed through OpenSource. Note this activity only covers those applications or tools whose distribution would not be in conflict with confidentiality or property issues inside the project. Moreover, the specific applications or tools to be distributed through

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

OpenSource will be agreed by all the ITHACA partners involved in their respective developments.
This task will contribute to deliverable D2.8

Deliverables: (Ordered by delivery date; note task and deliverable numbers are unrelated)

D2.1) Release of the project-community platform: (m1): [Bull & All partners**]**
The ITHACA-community platform addressed in subtask 2.1.2 will be active from this point until the end of the project, with constant updates.

D2.2: Release of the social-network profiles of the project: (m4): [**UCLM** & All partners]
The ITHACA Twitter, Facebook and LinkedIn profiles addressed in Task 2.1.3 will be active from this point until the end of the project.

D2.3: Release of the public project website: (m6): [**UCLM** & All partners]
The public ITHACA website addressed in Task 2.1.1 will be active from this point until the end of the project, with constant updates.

D2.4: Release of the initial video: (m6): [**UCLM** & All partners]
The first dissemination video addressed in subtask 2.3.3 will be released at this point.

D2.5: First Intermediate report on dissemination activities: (m12): [**UCLM** & All partners]
This report will summarize all the dissemination activities developed until the end of the first year of the project, including website and social-networks activity, press releases, publications, patent applications, presentations, tutorials, workshops, etc.

D2.6: Second Intermediate report on dissemination activities: (m24): [**UCLM** & All partners]
This report will summarize all the dissemination activities developed during the second year of the project, including website and social-networks activity, press releases, publications, patent applications, presentations, tutorials, workshops, etc.

D2.7: Release of the final video: (m34): [**UCLM** & All partners]
The second dissemination video addressed in subtask 2.3.3 will be released at this point.

D2.8: OpenSource software repository report: (m34): [**UCLM** & All partners]
This report will summarize the software included in the OpenSource repository addressed in T2.5, indicating also the contributors and the main benefits obtained (for instance, how, where and by who the software were used).

D2.9: Summer school report : (m36): [**UCLM** & All partners]
Once finished the summer school addressed in subtask sT2.2.2, we will elaborate a report including all the relevant information about this event (i.e., the number of attendants, the name and a short CV of the speakers, the presentations used in their talks and keynotes, etc.).

D2.10: Final report on dissemination activities: (m36): [**UCLM** & All partners]
This report will summarize all the dissemination activities developed all along the project's duration, including website and social-networks activity, press releases, publications, patent applications, presentations, tutorials, workshops, etc.

4. Work package 3: Data Management concepts & programming models

Work package number	**3**		**Start Date or Starting Event**					**m1**				
Work package title	**Data Management concepts & Programming models**											
Participant number	1	2	3	4	5	6	7	**8**	9	10	11	12	13
Short name of participant	Bull	Seagate	CEA	UPV	UCLM	DKRZ	Fraunhofer	**BSC**	Allinea	SURFsara	INRIA	ARM	UKOELN
Person/months per partner	84	12	12	3	18	30	34	**105**			27		3

Objectives:
The main objective of WP3 is to do research on the best programming model to be supported by the

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

proposed architecture. For this reason, we will evaluate three different alternatives, either currently well positioned or upcoming in the HPC world: MPI, PGAS, and task-based.
The three proposed programming models are important, because they cover all possibilities with respect to popularity in HPC and their closeness to a FlaGAS model. MPI is clearly the most used paradigm, but it fully relies on distributed address spaces. PGAS (represented by GPI and GPI-Space) is an upcoming programming paradigm that, though not 100% FlaGAS, is closer to a shared address space. Finally, PyCOMPSs (task based) offers a unified address space to the application, which is key in ITHACA, and its popularity is starting to grow in the HPC community. Offering these three models will enable the project to evaluate the benefits obtained by the proposed architecture with legacy applications, as well as the benefits of the advances in the programming models for applications willing to adapt to them.
Given that, the three programming models proposed are based on communicating components, we have decided to implement Portals as a communication mechanism. This implementation will be based on BXI and then Exascale Interconnect, that are implemented and/or designed in WP5.
In addition to the more traditional programming models specification, in this WP, we plan to extend ideas such as offering abstractions that integrate persistent data as first-class citizens (which seems to be a needed abstraction, especially for an architecture based on NVRAM). We will improve the programming models to effectively support workflow optimization such as data placement in the hierarchy or "in-situ computation", and we will implement polices to decide how to collocate data and task execution.
Another important objective in this WP is to really understand the FlaGAS model and needs to support, and what characteristics are key, to be offered to the proposed programming models.
Finally, we will compare the different programming models to understand the advantages and weaknesses of each of the programming models in the proposed architecture.

Description of work : **Lead partner: BSC**

BSC, UCLM, Bull, UPV, CEA, Seagate, DKRZ, Fraunhofer, INRIA and UKOELN contribute to this work package.

T3.1: New Data Management features for Portals

(m1-m24) [**UCLM,** Bull, UPV, CEA, Fraunhofer**]**

In order to implement the proposed programming models, we first need to extend the current implementation of Portals4 with new features in order to support the new programming models used in this work package. Currently, there exists a hardware implementation of Portals4 in BXI, according to published datasheets. The work in this task will be mainly on adapting a Portals4.1 implementation to the new architecture.
Specifically, we will enrich Portals to be able to implement the FlaGAS layer defined in this project. First we need to obtain the knowledge of hardware capabilities exposed by BXI hardware to drivers and libraries, in cooperation with Bull. In parallel, we will define a specification of new functionalities to be included in Portals in order to implement the FlaGAS layer. This task will be done in cooperation with UPV and BULL in order to obtain inputs from the FlaGAS layer specification. Then we will implement the new defined features into the Portals library. As a first approach, this implementation will be done in software by means of a prototype, which can be tested and validated in WP5.
This task will contribute to the deliverables D3.1 and D3.3.

T3.2: Programming model porting

(m1-m30) [BSC, CEA, Bull, UCLM, Fraunhofer]

sT3.2.1: New Data Management features for MPI

(m1-m30) [CEA, Bull, UCLM]

In this task, we will extend the Portals support in the MPC framework, already available for current version of Portals4, with the new Portals features from task 3.1. We will provide a standard (process based) MPI implementation. This MPI implementation is required to be able to run legacy applications on the top of Portals4. We will also provide a thread based implementation (one unique OS process per compute node and many MPI ranks in this OS process). This implementation may help users to incrementally update their code from the standard MPI programing model to PGAS or task-based programming models via a mix of MP , PGAS and/or task. This subtask will contribute to the deliverables D3.1, D3.3 and D3.4.

sT3.2.2: : New Data Management features for PGAS (GPI)

(m1-m30) [Fraunhofer]

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

The second programming model is GPI that belongs to the PGAS family. The performance and the features of the interconnect device are important for GPI. This may result in new requirements for the interconnect.
As a first step, a new implementation for the Portals specification has to be developed, evaluated and optimized within GPI. The initial implementation has to be optimized for performance on real BXI hardware with the help of the GPI tests, benchmarks and mini-apps. The results possibly will raise new software and hardware requirements for a next generation of interconnect. For instance, GPI provides a lightweight mechanism for synchronization called notifications. As the name implies, the target is notified about a data movement (or a data dependency) and can proceed with computation on that data as soon a notification has arrived. A fully hardware-supported notification could be provided by the BXI. Other improvements, if not yet provided, could include the offload of some computations (e.g. reductions) to the NIC or hardware-based active messages. This subtask will contribute to the deliverables D3.1, D3.3 and D3.4.

sT3.2.3: Task-based implementation (PyCOMPSs)
(m1-m30) [BSC]

The third programming model is PyCOMPSs that will leverage the idea of offering a single address space from the programming model itself, which will enable the implementation of workflows, whose tasks can be implemented in other programming models (i.e., MPI). In this task we will implement an adapted version of the programming model where the communication between components is based on the Portals porting done in Task 3.1. While the persistent data will be delivered to PyCOMPSs through dataClay, interaction with Portals will still be necessary for the non-persistent data communication and for synchronization between the different processes of the PyCOMPSs applications. The communication/synchronization layer of the PyCOMPSs runtime is based on the library NIO, which runs on top of TCP. The first implementation of PyCOMPSs in Ithaca will be based on this protocol, but BSC will perform a more specific implementation on top of Portals that will benefit of more efficiency in subsequent versions. This subtask will contribute to the deliverables D3.1, D3.3 and D3.4.

T3.3: Making persistent data a first class citizen
(m1-m30) [**BSC,** DKRZ, Fraunhofer, Seagate**]**

Extend the programming models in the proposal so that persistent data can be accessed as a first class citizen and be shared between applications. The idea of data becoming a first class citizen is to enable mechanisms such as iterators to be able to iterate over large datasets. This will not prevent applications from being able to use files in the standard way (open/close/read/write/…) but will open a new way. This mechanism should be more flexible than a standard mmap where the size of the data set is limited by address space, and sharing data between applications (such as in "in-situ analysis" is not trivial), especially if the simulation is implemented in Fortran and the analysis in Python. Summarizing, making persistent data a first class citizen into the programming model.
Two main models will be studied and implemented:

- Object Oriented Data Access: These extensions will be implemented as part of GPI-Spaces and PyCOMPSs integrated with dataClay.
- Semantic Aware Data Access: For this model, an intermediate abstraction layer is developed that describes the data types of scientific and big data use cases. Storage systems such as object storage and file systems implementing this interface or middleware exporting it, abstract from the traditional array of bytes file system interfaces. Instead the interface provides interfaces to create and query data by its semantics – it allows to define data types and (scientific) meaning similarly to HDF5 but on the abstraction level of FlaGAS.
  This semantic data will enable the system optimize data placement but also for co-scheduling of code, i.e., offloading of computation over the storage path would become doable. For example, instead of retrieving a subset of data from each object storage, computing the minimum/maximum could be performed directly on the storage. It can be considered an alternative to dataClay but also could allow data clay to access the self-describing data. Its interface will allow to define the data types and the semantics in memory and make this information queryable and available to the storage infrastructure.

This task will contribute to the deliverables D3.2 and D3.4.

T3.4: Programming model extensions to support efficient workflows
(m6-m30) [**INRIA,** DKRZ**,** BSC, Fraunhofer**]**

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Workflows are essential to define and orchestrate the different steps associated with scientific applications (data preparation, simulations, data analysis,…) and, as such need, to deal with data movements and I/Os. In this project we aim to develope concepts and create an API to enable workflows and in order to benefit from ITHACA developments.
The API allows users to provide information (hints) about the future I/O demands of advanced workflows. The middleware (in WP4) exploits this information by orchestrating data placement across the storage tiers to reduce the demand for (fast) online storage (e.g. NVRAM) in terms of performance, energy-efficiency and cost-efficiency. Additionally, relying on ITHACA monitoring capabilities (WP6) enables to automatically capturing information on existing workflows This information can then be feed into the model/simulation component to determine good hint sets to the system. Thus, this approach allows optimizing re-running of workflows without requiring users to set the hints explicitly.
We will also focus on enabling efficient in situ processing capabilities for data intensive workflows. In this task we will propose ways to express, from the programming point of view, mechanism for in situ processing based on the idea that persistent data is a first class citizen. The idea is that applications do not need to serialize the data to store into a persistent format, but be able to share objects between applications. This sharing is enabled based on the concepts of dataClay and the semantic aware data access interface. i.e. one application writes certain variables with scientific data out and another searches for it and visualizes the connected data on the fly. Such mechanism should ease programming in situ data processing applications as well as and make them more efficient given that unnecessary transformations are avoided.
Finally, when defining workflows, it is important to be able to define a good naming scheme for output data. Defining a hierarchical directory structure for the result data of experiments is time consuming for scientists, should the top-level directory be the model name, the data or relevant scenario? This setting results in deep directory structures and does not cope with changing requirements in the accesses, for instance, for climate simulations, when scientists compare results of different models. In this task, a data model is developed to allow scientists to use scientific metadata that use the semantic data to organize the namespace.
These extensions will be implemented as part of GPI-Spaces, PyCOMPSs and FlowVR, and a FUSE module using the semantic data will be created. This task will contribute to the deliverables D3.2 and D3.4.

T3.5: Task placement and data access optimizations
(m1-m30) [DKRZ , BSC, Fraunhofer, INRIA, Seagate**]**
The objective of this task is to optimize the runtimes and data access methods of the different programming models to be able to take decisions, based on the programming model extensions, on where data needs to be moved, how they can be efficiently accessed (through appropriate APIs), as well as, where computation should be performed. These decisions will be based on performance metrics as well as on energy consumption. These decisions will be implemented at the runtime layer or will be forwarded to the WP4 runtimes, depending on the nature of the operation. The implementation will utilize the knowledge about the workflows (Task T3.4) and the semantic data (task T3.3). This task will contribute to the deliverables D3.2 and D3.4.

T3.6: Comparison of different programming models
(m24-m36) [DKRZ and All partners]
The idea of this WP is to implement and extend different programming models to understand their potential and limitations in the proposed architecture. This task will evaluate, by comparing the productivity of porting a few selected benchmarks or kernels as well as the performance and scalability obtained by them using the different programming models. This task will contribute to the deliverable D3.5.

Deliverables: (Ordered by delivery date; note task and deliverable numbers are unrelated)

D3.1) First specifications Report (m12): [**BSC** and all WP partners]
In this deliverable we will present the specification of the extensions for Portals, the first definition of the API to access data from the programming models.
D3.2) Programming models extensions specification: (m18): [**INRIA** and all WP partners]
In this deliverable we will present the extensions planned for the programing models in order to support persistent data as a first class citizens as well as efficient workflows.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

D3.3) First prototype of the programming models: **(m20)**: [**BSC** and all WP partners]
In this deliverable we will deliver the first prototype of Portals extensions and the first implementations of the 3 programming models supported in the project: MPC, GPI, and PyCOMPSs

D3.4) Prototype of the programming models with the proposed extension:(m30):[**BSC** and WP partners]
In this deliverable we will deliver the prototype of the programming models with the extensions defined in deliverable D3.2

D3.5) Performance evaluation of programming models: **(m32)**: [**DKRZ** and all WP partners]
In this deliverable we will deliver the final results of the evaluation of the different programming models performed in Task 3.6, based on the performance and analysis tools developed in WP 6.

5. Work package 4: Co-Design & Data Access framework

Work package number	**4**		**Start Date or Starting Event**						**m1**				
Work package title	**Co-Design & Data Access framework**												
Participant number	1	**2**	3	4	5	6	7	8	9	10	11	12	13
Short name of participant	Bull	**Seagate**	CEA	UPV	UCLM	DKRZ	Fraunhofer	BSC	Allinea	SURFsara	INRIA	ARM	UKOELN
Person/months per partner	270	**159**	36	37	54	18		36					

Objectives :

The key goal of WP4 is to provide the data access middleware driving the storage I/O for ITHACA use cases to bring about application efficiency in turn contributing to reduced time to solution for the applications, working closely with all the other components of ITHACA in WP3, 5 and 6. This work package implements the necessary infrastructure for Distributed Data Containers (DDC)s which are namespaces unique for each applications' and workflows' data and I/O requirements.

The work package will define the data access middleware architecture and components needed for DDCs including design, validation and integration of the individual components of the architecture.

The components of the DDC-based Data Access middleware architecture will consist of an Object Storage Software Framework termed "Mero" that abstracts data in the distributed storage resources incl. memory, NVRAM, Flash storage and disk storage - and provide the necessary infrastructure for applications to exploit compute resources close to these storage tiers (for compute shipping). The DDCs along with Mero will eventually meet the I/O requirements of all programming models in ITHACA and expose a flat data addressing structure (using FlaGAS software support as needed) for application processes. The architecture will include a data manager that drives the data movement across these different resources to bring about data locality and reduced data movement for applications, eventually contributing to application efficiency. The data manager and the storage provisioning tools virtualize the storage resources to provide the required logical isolation between the various data elements for bringing about the "Containerization" aspect of the DDCs. The data in the containers can easily be synchronised independently to centralised storage. A service registry framework will be provided on top of Mero that can be exploited by layers above. For example, a data format abstraction layer will provide mappings for various types of data models, for ex: scientific data models that can be independently managed/processed - through appropriate metadata services. Mero will also map to next generation Object Oriented programming paradigms that can exploit compute shipping capabilities and encapsulate data and functions that operate on the data.

The architecture will address an ecosystem of Performance analysis tools, provisioning tools and simulation tools providing insights into "what if" scenarios. The work package will also provide the storage platform elements for ITHACA (Disk pools & NVRAM pools).

The work package is divided into tasks providing different modules of the data access middleware, namely:

1. Advanced Object Storage Infrastructure (Mero)

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

2. Storage addressing
3. Storage hardware
4. Storage Provisioning and Administration
5. Data Management
6. Data Access
7. Storage Performance Analysis and Prediction

The co-design inputs for these modules are derived by a ITHACA middleware top level architecture which will be tasked at the beginning of the WP.

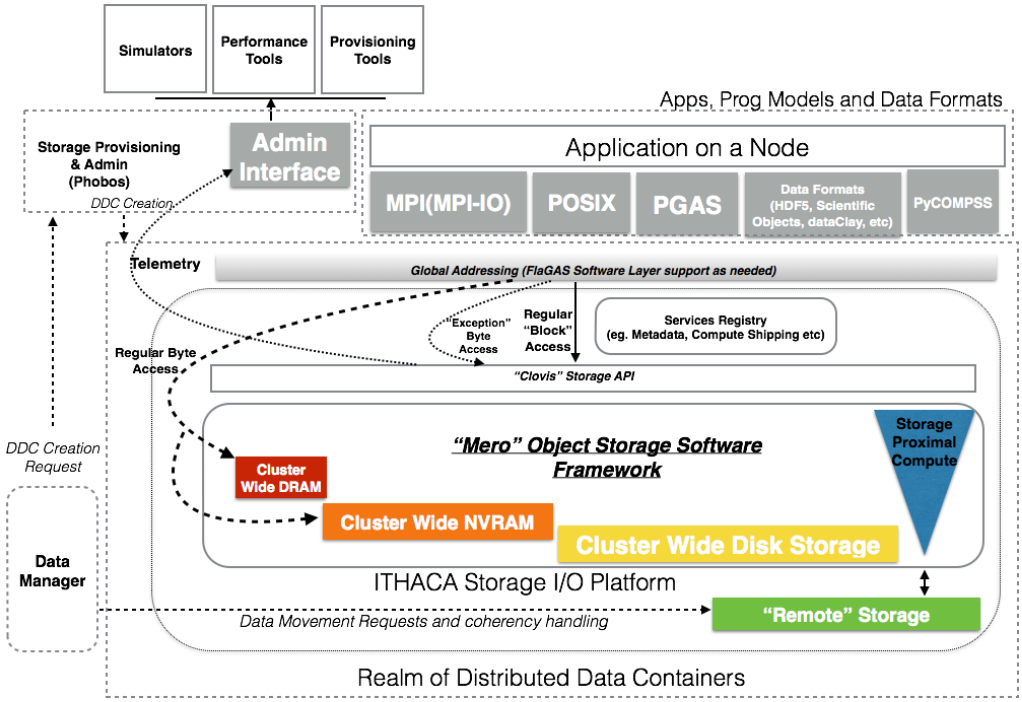


Figure 9: DDC Based Storage in ITHACA: Top Level Architecture

Description of work : **Lead partner: Seagate**

Seagate, BSC, DKRZ, CEA, Bull, UCLM, UPV and Allinea contribute to this work package.
The following is a list of the tasks and subtasks required to achieve the objectives of this work package.

T4.1: Co-design of Data Access architecture
(m1-m6) [**Bull**, Seagate**,** CEA, BSC, UPV, DKRZ**]**

This task co-designs the top level architecture of the data access middleware & associated APIs based on inputs from the use cases and programming models and establishes key considerations for Distributed Data Containers and all the subcomponents described in other tasks in this WP.
This task will contribute to the deliverables D4.1.

T4.2: Advanced Object Store Infrastructure
(m3-m36) [**Seagate,** CEA, Bull, BSC**]**

This task will provide the Object storage infrastructure, Mero that drives the DDCs.

sT4.2.1: New Object Storage platform: Mero Components
(m3:m34) [Seagate, Bull**,** CEA**]**

This subtask will co-design and develop various components of Mero along with ITHACA use cases and the needs of global addressing of data. We start from a base design and then study how those need to be adapted to ITHACA. The various components will be for addressing:

- Scalability – Concepts such as containerization of data(lower level of containerization within Mero, and different from DDCs), advanced caching mechanisms and resource management across the I/O hierarchy. This also includes implementation aspects to implement Mero services in compute nodes to exploit node local storage resources.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

- Infrastructure Resiliency – High Availability infrastructure obtaining inputs from telemetry data
- Application Resiliency – Transaction and Epoch management infrastructure
- Flexible data layouts – Providing various techniques to spread data across the I/O hierarchy as desired by applications and the infrastructure provider
- Infrastructure to support ITHACA Data Sets and scientific data Formats (as defined in "Data Access" task).
- Multiple Views of the same data, which implies the provision to infrastructure to have HDF5 views, Posix views, etc of the same raw stored data without copies.
- Support for HSM tools in ITHACA

This subtask will contribute to the deliverables D4.4, D4.6, D4.7 and D4.11.

sT4.2.2: Service Registry Framework
(m3:m28) [**Seagate**]

This subtask will provide a service registry framework to add additional services as "plugins" on top of the storage platform. The additional plugin framework will be provided on top of the object storage interface and will exploit the logging information available from the storage infrastructure to drive the functionality of the plugin. This subtask will contribute to the deliverables D4.4, D4.6, D4.7 and D4.11.

sT4.2.3: In-Storage Compute
(m3:m34) [**Seagate**]

This task will research and implement the leveraging of storage resources through storage services, close to compute wherever they are present (Storage system and the compute nodes). In-storage compute capability will be part of the Mero object storage infrastructure which will send compute processes to data and also handle in-storage compute to object locations in the case of storage failures.
This subtask will contribute to the deliverables D4.4, D4.6, D4.7 and D4.11.

sT4.2.4: Garbage Collection
(m12:m30) [BSC]

Using NVRAM as standard memory (not file system) implies that we may have memory leak problems. In volatile memory, these leaks are either solved by closing the application or running some kind of garbage collection by the runtime. In the case of NVRAM, closing the application will not solve the problem because we may want the use memory to be available for other applications in the future, thus we cannot remove it. Standard Garbage Collection mechanisms are not ready to scale to the size of non-volatile memory.
In this subtask we plan to understand how and when persistent data becomes useless and implement mechanisms to recover the space used by this obsolete data.
This subtask will contribute to the deliverables D4.7 and D4.11.

T4.3: **Storage Addressing & Global Memory Addressing (FlaGAS)**
(m3:m36) [**UPV**, Bull]

This task will provide the FlaGAS service. The following are the sub-tasks.

sT4.3.1: FlaGAS Low Level Software
(m3:m24) [**UPV**, Bull]

This subtask provides a software implementation of FlaGAS low level mechanisms needed for DDCs. The goal is to implement a memory access extension to enable the software to address cluster-wide DRAM and NVRAM, as well as byte-addressed storage devices, using a flat global address space (FlaGAS).

This infrastructure will provide support for accessing any memory region, including non-volatile RAM, from any core in the system. It will also provide support for memory region address definition. In order to achieve high scalability, cache coherence support will not be implemented.

This software implementation is based on leveraging the virtual memory mechanisms for translating from virtual addresses to physical addresses. Such a mechanism will be extended by rewriting the exception handlers so that virtual addresses could be mapped to remote DRAM and NVRAM as well as remote block devices when accessed through byte addresses. In those cases, requests will be sent through the interconnection network so that remote blocks or pages are brought to local DRAM or NVRAM. Those transfers will use the Portals network support developed in task T3.1. In particular, they will use the support for active messages and RDMA write operations to minimize access time to remote data.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

This implementation requires solving several problems, including potential protocol deadlocks, race conditions, coherency between memory maps, booting, as well as developing a very efficient implementation so that it does not become a bottleneck. This subtask will contribute to the deliverables D4.9.

sT4.3.2: FlaGAS Management Environment

(m3:m36) [**Bull**, UPV]

This subtask provides the FlaGAS management functions. We will define and implement the software services needed to manage the FlaGAS feature at run time, from the creation of the shared address space to the attachment/detachment by the compute and IO service nodes and finally the deletion when the jobs using it are terminated. Security is an important point in this design, as sharing address spaces over the interconnect is opening a door on the application data.

APIs will be defined to allow creation/deletion by the Provisioning Tools and attachment/detachment by the compute and IO service nodes. This subtask will contribute to the deliverable D4.9.

T4.4: **Storage Hardware**

(m1:m34) [**Seagate**]

This task will define and develop the storage hardware platform (Storage enclosures, etc) as needed for ITHACA use cases and the software infrastructure components.

This task will contribute to the deliverables D4.2, D4.3, D4.5 and D4.12.

T4.5: **Storage Provisioning and Administration**

(m1:m36) [**CEA,** Bull]

The goal of this task is to implement a "storage on-demand" framework within ITHACA suitable for Exascale, which will drive provisioning of storage and storage administration aspects. This framework is called "Phobos".

This infrastructure will provide unified access and administration interface for DDCs that will consist of a wide variety and the large numbers of storage resources that will be available in Extreme scale computing centers, including Non-Volatile Memories, Solid State Drives, spinning disks and magnetic tapes.

Such a framework will implement services of cluster-wide management of boot devices for diskless servers, on-demand storage resources for compute nodes (for checkpoint/restart files, temporary data and a common front-end to various object stores. It will better manage DDCs by allocating storage resources according to the application needs in term of persistence, capacity, performance, and data safety.

To implement these services, the framework will take advantage of the advanced technologies developed in this project including high-speed interconnect, and scalable object storage.

Phobos aims to manage a wide variety of storage resources through a common administration interface. It will implement a distributed management of storage resources available in a compute center, and makes it possible to set up and access remote or local storage, depending on available resources and their locality.

It will support several backend adapters to manage and access various storage resources: IO proxies, filesystems, object stores, block devices, cloud storage protocols. These resources can be made available for remote hosts using various protocols, depending on the kind of device.

Phobos provides the following front-end services:

The goal of this task is to implement a "storage on-demand" framework within ITHACA suitable for Exascale, which will drive provisioning of storage and storage administration aspects. This framework is called "Phobos".

This infrastructure will provide unified access and administration interface for DDCs that will consist of a wide variety and the large numbers of storage resources that will be available in Extreme scale computing centers, including Non-Volatile Memories, Solid State Drives, spinning disks and magnetic tapes.

Such a framework will implement services of cluster-wide management of boot devices for diskless servers, on-demand storage resources for compute nodes (for checkpoint/restart files, temporary data and a common front-end to various object stores. It will better manage DDCs by allocating storage resources according to the application needs in term of persistence, capacity, performance, and data safety.

To implement these services, the framework will take advantage of the advanced technologies developed in this project including high-speed interconnect, and scalable object storage.

Phobos aims to manage a wide variety of storage resources through a common administration interface. It will implement a distributed management of storage resources available in a compute center, thus allowing setting up and accessing remote or local storage, depending on available resources and their locality.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

It can support several backend interfaces to access storage resources: IO proxies, filesystems, object stores, block devices, cloud storage protocols. These resources can be made available for remote hosts using various protocols, depending on the kind of device.
Phobos provides the following front-end services:

- Access to block devices that can be used as disk image for diskless servers, or temporary storage. Phobos will manage allocation (provisioning), remote access mechanisms, and access control rights for these resources. This requires a protocol to access a remote block device, like iSCSI or its RDMA extension ISER. The support of such RDMA-based protocol over BXI will be a key requirement to enable maximum access performance of this solution.
- It instantiates DDCs on-demand, and make them available to applications.
- Object storage makes it possible to store data addressed by keys, in a distributed environment. Phobos provides a common API to access objects, whatever the underlying storage resources: block devices, filesystems (Lustre, BeeGFS), and objects stores (Mero), object-based devices (eg: "Kinetic" drives).

This task will contribute to the deliverables D4.7 and D4.11.

T4.6: **Data Manager**
(m1:m36) [**Bull,** Seagate, BSC, CEA]
This task will provide the Data Manager component and the DDC management API.
Within the ITHACA proposed architecture, the Data Manager is the component in charge of the overall coordination to create a Distributed Data Containers (DDC) for applications. A Distributed Data Container is an abstraction that is dynamically set up to give compute nodes access to storage objects such as regular Posix files and Mero objects. A distributed Data Container can be seen as a namespace shared by the nodes objects for a period of time to run one or more jobs. So it is not a persistent namespace as in the traditional HPC approach, but more a temporary namespace created specifically for a (set of) job(s). The namespace can be created using many different technical solutions, such as by copying files and objects to compute nodes' persistent storage devices (NVMe or local disks), or using an IO Proxy to give access to storage objects hosted elsewhere. The choice between the different solutions is done by the Data Manager using metadata provided by the users (which storage objects need to be in the DDC?) and using its own findings elaborated from previous runs observation and current telemetry data.
The Data Manager creates and populates the DDC upon request of the Provisioning tools (when a job is to be launched) or through the command line interface for interactive use of a set of compute nodes.
It interacts with Phobos and Mero to set up access to files and objects and eventually move/replicate them closer to compute resources. Conversely, it allocates the closest storage resources such as IO Proxies or, IO Cache appliances to the allocated compute nodes.
It interacts with "Data Movement services" such as IO Proxies and HSM data movers.
It interacts with Performance Tools to get Telemetry data and learn from current IO activity.
The Data Manager will be developed and demonstrated on 3 use cases in this task:

- Traditional Posix DDC using an IO Proxy to export files from a Lustre File system (with CEA)
- POSIX export from Mero Object Store (with Seagate)
- Innovative Object based export leveraging dataClay and making use of FlaGAS and NVM DIMMs in compute nodes and IO Cache appliances. Long term persistent back end will be Mero

This subtask will contribute to the deliverables D4.7 and D4.11.

T4.7: Data Access
(m1:m36) [DKRZ, BSC, Seagate]
This task studies new methods for data access as needed by ITHACA use cases with focus on an abstraction layer for new data formats and novel object oriented paradigms.
sT4.7.1: Object Oriented Data Access
(m1:m36) [BSC, Seagate, Fraunhofer**]**
In this subtask we propose to port dataClay that is an Object Oriented data access method, to be able to use all storage systems in the project such as NVRAM (direct access), BeeGFS, and object-store by Seagate (Clovis API). In addition, we will also extend the current implementation to support all features proposed in WP3. This subtask will contribute to the deliverables D4.1, D4.4, D4.6, D4.7, and D4.11.
sT4.7.2: Abstraction Layer for Data Access
(m1:m36) [DKRZ]

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

File systems and object storage rely on a very low-level data abstraction; a file is just an array of bytes. Optimizing storage for this simple data structure is barely possible as access granularity and access patterns are hard to predict. In this task, an intermediate abstraction layer is developed that describes the data types of scientific and big data use cases.
Storage systems such as object storage and file systems implementing this interface or middleware exporting it, abstract even further from the storage technology. For example, users can now provide hints based on scientific elements and not for files and directories. Also implicit programming models benefit from this abstracting since data mappings of complex data objects are easier to support reducing the burden of runtime environments to persist data. This abstraction allows the storage backend to make intelligent decisions by, e.g., storing scientific metadata in a suitable key-value store and relevant scientific variables on fast storage tiers. Additionally, offloading of computation over the storage path would become doable. For example, instead of retrieving a subset of data from each object storage; computing the minimum/maximum could be performed directly on the storage. The implementation will be done on top of existing interfaces for object storage and (shared) file systems. We will integrate the abstraction level as backend into HDF5 allowing it to talk to this scientific object storage directly.
This subtask will contribute to the deliverables D4.12

T4.8: Distributed Storage Simulator
(m1:m36) [**UCLM**, UPV,Seagate]
When defining a new distributed storage infrastructure combining RAM, NVRAM and massive storage, it is desirable to model certain critical aspects of that architecture within a simulation model before making design decisions, which could lead to bottlenecks or mistakes in the storage subsystem. Simulation models aid the storage system designers to avoid those bottlenecks and maximize the IOPS of the storage architecture. Apart from the hardware components of the storage architecture, it is desirable to model other software aspects of the I/O architecture, like data containers, I/O API or FlaGAS. Therefore, this contribution to the WP4 proposes to develop (in cooperation with interested partners) a distributed storage simulator (DSS) including those critical aspects. The idea is simple: based on the feedback of partners specialized in storage and data observability, we should obtain some information about the storage-based applications. This knowledge could be applied to model workloads that feed the DSS simulator. In addition, the DSS tool can be also include some functionality to interact with the network simulation model in order to analyze how the storage-based workloads have influence in the network performance, if they contribute to generate bottlenecks, testing QoS policies, etc.
The following are the studies of this task:

- Define (in cooperation with interested partners) the key aspects of the storage architecture to include in the DSS tool.
- Based on these requisites the partners want to include in the model, develop the DSS tool, following a stream programing methodology. We plan periodic meetings to adapt this model to involved partners demands, testing planning and validate specific functionality, etc.
- Cooperation with the network simulator team to develop an integration framework between network simulator and DSS tool, so that with some abstractions that reduce granularity, it could be possible for these two simulators to interact.
- Based on the use cases of the storage architecture, it is required to model workloads to feed the DSS tool
- Seagate will study and develop models for the various storage and I/O software and hardware elements in the ITHACA infrastructure that will feed into a DSS.

This task will contribute to the deliverables D4.8 and D4.10.

Deliverables: (Ordered by delivery date; note task and deliverable numbers are unrelated)

D4.1) Overall architecture of the Data Access Middleware **(m4) [Bull]**
D4.2) Hardware Platform Framework Report **(m6) [Seagate]**
D4.3) Hardware Platform Detailed Architecture Report **(m9) [Seagate]**
D4.4) Data Access Middleware Modules: Concept & Architecture Report (First Version**) (m12) [Seagate]**
D4.5) Hardware Platform Availability Report **(m18) [Seagate]**
D4.6) Data Access Middleware Modules: Initial Proof of Concepts **(m20) [Seagate]**

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

D4.7) Data Access Middleware Modules: Concept & Architecture Report (Second Version) **(m24) [Seagate]**
D4.8) Provide storage and I/O device models for feeding into DSS **(m24) [Seagate]**
D4.9) FlaGAS memory protocol extension and implementation description report **(m24) [UPV]**
D4.10) Demonstration infrastructure of the Distributed Storage Simulator **(m30) [UCLM]**
D4.11) Data Access Middleware Modules: Final Prototypes **(m35) [Seagate]**
D4.12) Hardware Platform Optimization and Use case Access Report **(m35) [Seagate]**

6. Work package 5: Co-Design & Data Movement Interconnect

Work package number	**5**		**Start Date or Starting Event**					**m1**					
Work package title	**Co-Design & Data Movement interconnect**												
Participant number	**1**	2	3	4	5	6	7	8	9	10	11	12	13
Short name of participant	**Bull**	Seagate	CEA	UPV	UCLM	DKRZ	Fraunhofer	BSC	Allinea	SURFsara	INRIA	ARM	UKOELN
Person/months per partner	**360**	9	12	117	74		6						

Objectives :

The goal of WP5 is to provide a scalable, reliable, and cost-effective high-performance network interconnect optimized for data management across the entire system. This interconnect is a critical component for ITHACA, as it is the hardware support for fast data movement between compute nodes, and between them and storage devices. It must offer a high efficiency for the programming models studied in WP3 and the Data Access middleware developed in WP4. This work package capitalizes on the BXI technology developed by Bull, which first generation will become available in 2017. For convenience, this generation of interconnect will be named BXI1 in WP5.

WP5 covers 3 main objectives :

- The design of a second generation of BXI that significantly improves data movement (tasks 5.1 to 5.3). For convenience, this generation will be named BXI2 in WP5.
- The provision of BXI-based evaluation platforms for partners to evaluate this technology and propose enhancements for data movement (task 5.4).
- The specification of a next generation of interconnect taking into account the outcome of the study performed at the other work packages to provide an optimal hardware and software implementation for the selected applications (task 5.5). For convenience, this generation will be named Exascale Interconnect in WP5.

During the project, it is important to develop some hardware prototype with associated low level software to validate the ITHACA architecture. As a consequence, BXI2 must be developed in parallel with the studies performed at the other work packages, and will take into account the general requirements available during the first months of the project.

One of the main ITHACA concepts is that a compute node may access data anywhere in the system. The efficiency of such model requires very high bandwidth and low latency. In order to increase bandwidth, BXI2 could double all raw peak throughput and message rate figures compared to first generation. The latency would be reduced mainly by working on the congestion management to sustain good performance under heavy load and under targeted benchmarks.

Depending on the way the data are organized, ITHACA architecture may require more endpoints (at least 128,000) and a larger data address space (at least 64 bits of addressing) than the ones supported by BXI1. BXI2 must support ARM V8 architecture since efficient compute nodes based on ARM will be available in 2020, and BXI2 might also be closely coupled with an ARM CPU.

BXI2 would also increase the level of monitoring, profiling and security, in accordance with the data management ecosystem needed in WP6.

Even if the hardware is defined early in the project, ITHACA will contribute to tune the many parameters

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

and algorithms to optimize BXI2 for the applications targeted in the project. This will be achieved first by developing software network simulators and running appropriate simulations, and later, at a smaller scale, by using the BXI2 prototype.
In parallel with those simulations, the partners will evaluate the ITHACA architecture concepts on BXI technology with platforms based on the first generation.
Finally, using the results of simulations and evaluations, and using also the feedback and results from the others technical work packages, the partners will co-define the specifications of Exascale Interconnect. It will have to provide a feasible implementation of the characteristics required by the ITHACA architecture.

Description of work:

WP5 must provide interconnect optimized for the programming models studied in WP3 and the Data Access framework developed in WP4. It must also provide efficient basic accesses required by the data management ecosystem developed in WP6 and applicative use cases of WP7 which will integrate the components and results from WP3 to WP6.
In this work package, we will first develop and tune a new version of interconnect named BXI2.
In order to achieve this goal before the end of this project, BXI2 will be based on the structure of BXI1, adding new functionalities. It consists in two hardware components, the Network Interface Controller (NIC) and the switch, and two main associated software components, the low level software stack and the fabric management suite. In parallel we will evaluate the BXI technology for the ITHACA architecture, aiming to define a new generation fully optimized for data movement.
The following is a list of the tasks and subtasks required to achieve the objectives of this work package.

T5.1. Enhanced NIC & software Stack for data movement
(m1:m24) [Bull, UPV, UCLM, Fraunhofer, CEA**].**
In this task, we focus on the components at the host side: the NIC and software stack.
sT5.1.1. Specification of the NIC side enhancements in BXI2
(m1:m6) [Bull, UPV, UCLM, Fraunhofer, CEA**].**
In this subtask, we analyse the evolution of BXI that can be developed on host side during this project for data movement optimization and write the specification of the associated hardware and software components of BXI2. The new hardware and software features will be defined by doing a cooperative study between hardware and software teams. The candidates are for example: full support of ARM architecture, lower latency with double bandwidth, offload optimization for IO traffic, Support of new API and programming models (Portals 4.x, GPI, NVMe, …), Enriched data and metrics for this WP5 and the others, especially for WP6 ecosystem: monitoring, application profiling, power management, resource management, batch scheduler … This study combines a bottom-up approach, based on what the hardware technology could bring (7nm, new interfaces) and a top-down approach, based on the needs for data movement optimization and a prioritization on the potential requirements. This subtask will contribute to one deliverable D5.1.
sT5.1.2. Design hardware of the NIC
(m5:m24) [Bull].
In this subtask, we describe the NIC ASIC in Verilog language, to implement the hardware specification defined in subtask sT5.1.1. In parallel, a validation environment will be built to verify this implementation via software simulation of Verilog code. The simulation will be run according to a testplan derived from the functional specifications, until all features are covered. ITHACA covers only the development of the NIC code, not the ASIC fabrication which cost exceeds the budget of the project.
This subtask will contribute to one deliverable: D5.7
sT5.1.3. Development of the associated software components
(m7:m24) [Bull, Fraunhofer, CEA**].**
In this subtask, starting from the BXI1 software stack solutions, we will:

- Define the architecture of the BXI2 software stack.
- Define the functional and improvements specifications of components (e.g. Kernel driver, API, …).
- Develop a beta version of this software stack
- Analyse and optimise the performance of this software stack
- Develop an enhanced version of this software stack.

This subtask will contribute to two deliverables: D5.3 and D5.7.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

T5.2. Enhanced Switch and Fabric Management software for data movement
(m1:m30) [Bull, UPV, UCLM**].**

In this task, we focus on the network interconnecting the NICs : the switch and the Fabric Management.

sT5.2.1. Specification of the Switch side enhancements in BXI2
(m1:m6) [Bull, UPV, UCLM**].**

In this subtask, we analyse the evolution of BXI that can be developed on switch side during this project for data movement optimization and write the specification of the associated hardware and software components of BXI2. The new hardware and software features will be defined by doing a cooperative study between hardware and software teams. The candidates are for example: Double peak bandwidth, High effective bandwidth, by implementing efficient congestion management, based on mixed hardware and software mechanisms, New topologies with same or higher radix switches, Increase of interconnect power efficiency, Enriched data and metrics for WP5 and WP6 ecosystem : monitoring, application profiling, power management, resource management, etc. Again, this study combines a bottom-up approach, based on what the technology could bring (7nm, 56Gb/s Serdes) and a top-down approach, based on the needs for data movement optimization. This subtask will contribute to one deliverable: D5.1

sT5.2.2. Design of the Switch hardware
(m5:m24) [Bull].

In this subtask, we describe the Switch ASIC in Verilog language, to implement the hardware specification defined in subtask T5.1.2. In parallel, a validation environment will be built to verify this implementation via software simulation of Verilog code. The simulation will be run according to a testplan derived from the functional specifications, until all features are covered.

The final simulation integrates both NIC and switch models, to ensure that they communicate properly. ITHACA covers only the development of the Switch code, not the ASIC fabrication, which cost exceeds the budget of the project. This subtask will contribute to one deliverable: D5.7.

sT5.2.3. : Development of the associated Fabric Management software
(m7:m24) [Bull].

In this subtask, starting from the BXI1 Fabric Management solutions we will:

- Define the architecture of the BXI2 Fabric Management software.
- Define the functional and improvements specifications of each subcomponent (e.g. Routing, Topology checking, Supervision,…).
- Develop a beta version of this Fabric Management software
- Analyse and optimise the performance and the scalability of this software.
- Develop an enhanced version of this Fabric Management software

This subtask will contribute to three deliverables: D5.3, D5.7 and D5.11.

T5.3. BXI2 Simulation & Performance tuning
(m1:m30) [UCLM, UPV, Bull**].**

In this task, we develop simulators and use them to validate and evaluate the BXI2 architecture.

sT5.3.1. SW-based simulator modeling BXI2
(m1:m18) [UCLM, Bull**,** UPV**].**

In this subtask, we develop two types of simulators to allow the characterization and specification of next generations of the interconnect :

- A detailed one very close to the hardware, for ASIC level simulation, useful to validate and evaluate the behaviour of software components. This requires constant contact with (and feedback from) the partners involved in tasks 5.1 and 5.2.
- A scalable one that is able to simulate the behaviour of the entire interconnection network, even for the largest size considered, but only models the most relevant mechanisms and events from each component as well as the interactions among them. It is useful for network level simulation and to evaluate network-wide algorithms and events. Similarly to the detailed-level simulator, developing the scalable simulator requires interaction with all the partners involved in tasks 5.1 and 5.2.

At detailed level, we will develop a time-approximate SW-based simulator modelling the NIC and Switch components, and basic traffic generators to ensure efficient software validation and evaluation. The simulator will be optimized throughout the duration of the task. This part will contribute to one deliverable: D5.4

At scalable level, we will develop a SW-based simulator modelling the same components, but at a higher degree of abstraction in order to be able to simulate a cluster consisting of hundreds of thousands of nodes.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

This simulator focuses mainly on topology and routing optimization, and congestion management.

Given the huge size of the targeted systems, it will become mandatory to optimize the modelling of network components to minimize the memory required at run-time and to prevent the simulation time from skyrocketing. Hence, we will analyse several simulation frameworks that reputedly allow an efficient use of memory, such as Omnet++ and SimGrid to determine the optimal framework to build this simulator from. If despite these optimizations the simulation time was not acceptable, we may consider developing some techniques to parallelize the simulator. The validation of this simulator could be performed against the BXI2 platforms and prototypes provided through task 5.4, when they become available. In addition, we may consider modelling in the simulator other technologies, strategies, and system configurations different from those addressed in ITHACA, in order to perform simulation-based comparisons between systems modelled according to existing solutions against systems based on the ITHACA proposals. We will also develop tools (a toolchain) to ease the use of the simulator, especially the generation of different topologies, routing tables, and synthetic traffic patterns to run the simulations, as well as tools to easily visualize the results. Besides, this toolchain will handle the configuration of workload traces as an input of the simulator, especially those traces extracted from WP7 applications. In a second step, this model will be enriched with power management add-ons, in order to take into account the power efficiency in the selection of routing algorithms, congestion management mechanisms and recovery on error. This modelization will help defining an efficient power strategy. In particular, the improvements on power efficiency addressed in task 5.2.1 will be modelled. This part will contribute to two deliverables: D5.4 and D5.5.

sT5.3.2. Performance tuning based on BXI2 simulator

(m13:m30) [UCLM, UPV**,** Bull**].**

This subtask uses the simulators developed in sT5.3.1 to validate the BXI2 architecture and provide optimal tuning for the ITHACA use cases. One objective is to evaluate the different network design components, using specific synthetic traffic patterns for each purpose. For example, we need different scenarios to evaluate the behaviour of routing algorithms and power management, where load has to vary over time. In the process, we will fine tune multiple network parameters, such as topologies and associated routing algorithms, Virtual Channel buffer configuration, Virtual Channel mapping strategies, Quality of Service, congestion management, etc. A second objective is to allow the evaluation using traces from workloads defined in WP7, doing some final tuning if necessary. It is not possible to scale an execution-driven simulator to hundreds of thousands of nodes. However, it is possible to scale using traces and very optimized network simulators. This subtask will contribute to two deliverables, phased to take into account the stepping of the different simulations :

- Via a performance and parameter tuning based on specific synthetic traffic patterns: deliverable D5.5.
- Via an evaluation of power management, Quality of Service management and congestion management strategies and an evaluation of overall performance and parameters tuning on workloads defined in WP7: deliverable D5.12.

T5.4. BXI-based evaluation platforms

(m4:m34) [Bull].

In this task, we will provide two types of BXI evaluation platforms: one standard system based on BXI1 ASICs, and one FPGA prototype to validate early BXI2 evolutions. The BXI1 ASICs one will be partially updated with some BXI2 prototypes and ARM CPUs as soon as they become available.

sT5.4.1. Provision of the cooperative BXI-based Platform

(m4:m30) [Bull].

In this subtask, we will provide a platform for partners to evaluate the ITHACA architecture on BXI technology. The chosen platform is the Bull Sequana platform, which consists in a very dense cell including up to 288 compute nodes and two levels of full fat-tree. The platform will be available by the end of 2017 (m6), and is based on the BXI1 and optimized for a full fat-tree using the 48 ports of the BXI1 switch (24 compute nodes per first-level switch). The compute nodes available at this time frame will be two-socket Xeon nodes and mono-socket XeonPhi nodes. The NIC in the compute nodes will be implemented on mezzanine, allowing a smooth upgrade of the interconnect when a new one becomes available. The Sequana platform will be hosted by Bull and access will be provided to partners for their own tests.

In the last year of the project, Bull will update this platform with the components available at that time:

- New compute node with ARM CPU: this node is being developed in the Mont-Blanc3 H2020 project

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

and is built with one of the first HPC ARM CPUs. It will help validating the ITHACA architecture in the ARM ecosystem.

- BXI2 hardware and software: NIC and Switch, software stack and Fabric management suite defined and developed in tasks T5.1 and T5.2 will be available. It will help evaluating the contribution of BXI2 to ITHACA architecture efficiency.

Bull will ensure the maintenance and support for this platform throughout the project
This subtask will contribute to two deliverables: D5.2 and D5.10.

sT5.4.2. Development and provision of a FPGA prototype for BXI2
(m13:m28) [Bull].

In this subtask, we will develop a FPGA prototype for BXI2. The main objective is to provide an accelerated environment to validate the software stack on the BXI2 hardware. It is a critical point as the many offloads in the NIC create a tight dependency between hardware and software, and the software simulation is too slow to fully validate this integration.
The FPGA platform will emulate the BXI2 NIC and Switch. It will require specific boards equipped with several interconnected FPGAs which can be configured as a NIC or as a switch. Several boards will be interconnected to build a small system with several NICs and switches. The maximal configuration is limited due to the cost of large FPGAs, but it will be sufficient to fully test the software-NIC interface, and the basic Fabric Management features.
This prototype will be developed during the project and will be operational when the Verilog development is completed. It will be delivered for the third year of ITHACA and hosted by Bull. It will be first used for hardware and software validation. When the design is stabilized, an access will be provided to selected partners. As debugging is complex with such platform, it is difficult to support too many users, and priorities will be discussed among partners inside the project, to evaluate which applications and when require access prior to the BXI2-based platform upgrade. This subtask will contribute to two deliverables: D5.7 and D5.9.

T5.5. Co-design of next interconnect generation
(m7:m36) [Bull, CEA, UPV, UCLM, Fraunhofer, Seagate, ARM**].**

Due to the duration of ASIC design, BXI2 development must start early in the project, and cannot take advantage of the outputs from the different ITHACA work packages. As a consequence, the next interconnect generation will be defined in this task, leveraging the different studies realized by several partners of this consortium in the ITHACA work packages.
This task includes the evaluation of BXI1 and BXI2 for data movement, the definition of Exascale Interconnect potential optimizations, the evaluation of these optimizations by means of simulation and feasibility analysis, and eventually the specification of Exascale Interconnect.

sT5.5.1. Evaluation of BXI1 and BXI2 interconnect technologies for data movement
(m7:m34) [CEA, Bull**,** Fraunhofer**].**

This subtask focuses on the low layers of BXI (NIC and software stack) and it is complementary to the studies done in WP7.
We will run tests and benchmarks to evaluate the performance and the specific new features of BXI1 and BXI2, and will collect requirements for Exascale Interconnect. At first, this study will start with BXI specification information, and quickly it will continue with real test vehicles provided by sT5.4.1.
For IO performance, we will collect trace information on the current NIC with the help of the BeeGFS file system, which is well instrumented to collect them. The instrumentation will be extended to analyse multi-tier architectures. We will analyse the file system communications on the current interconnect and, based on the tests cases with real world I/O-intensive applications, will give recommendations for Exascale Interconnect. This subtask will contribute to two deliverables: D5.6 and D5.15.

sT5.5.2. Definition of interconnect (Exascale Interconnect) optimized for data movement
(m13:m30) [Bull, CEA, UPV, UCLM, Fraunhofer, Seagate, ARM**].**

The final definition will integrate specific results of interconnect studies and the results and feedbacks from WP3, WP4, WP6 and WP7, that is, the collection of requirements from partners.
For example, the performance and the features of the interconnect devices are important for GPI and might result in new requirements for BXI arising from GPI, what will be analysed in this task. New requirements may also arise for GPI from BXI features, but they are not part of this work package.
We will also analyse the new functionality in the Portals5 library to leverage new capabilities of BXI regarding offload capabilities, network efficiency, congestion control, etc.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Such requirements may induce hardware modifications, such as new memory interfaces, collective communication operation offloading, hardware support of FlaGAS in the NIC, specific offloading in the switch, new topologies optimized for data movement with ITHACA architecture, etc.
They will also induce software modifications, such as impact of new programing models, support for bigger configurations, virtual memory improvement, congestion management enhancements, internal power efficiency mechanism, etc.
Gradually, as the requirements are collected, we will analyse the associated features in terms of feasibility, cost, and priority. This analysis will also take into account the evaluation done in sT5.5.3.
This subtask will produce two deliverables: D5.8 and D5.14.

sT5.5.3. Evaluation by simulation of Exascale Interconnect technology
(m19:m36) [UCLM, Bull, UPV**].**

In this subtask, we will evaluate the complexity of the mechanisms needed to achieve the requirements collected in sT5.5.2, and the performance improvement that is achievable.

Similarly to the BXI2-based simulator, the Exascale-Interconnect-based simulator is intended to scale up to hundreds of thousands of nodes. We will reuse the toolchain developed as part of subtask sT5.3.1, adapting and updating it (if necessary) according to the characteristics of the new components to be simulated. Once developed, the Exascale-Interconnect-based simulator will be used to evaluate the novel proposals introduced in the Exascale-Interconnect architecture. This will be done, as usual, through an iterative process of simulation runs until we achieve a nearly optimal configuration. In these simulation experiments, both significant synthetic-traffic patterns and traces from real applications will be used as the simulator workload, so that the Exascale Interconnect architecture will be tested under a wide range of scenarios. This task will produce two deliverables: D5.13 and D5.16.

sT5.5.4. Specification of the next interconnect generation (Exascale Interconnect)
(m25:m36) [Bull].

In this subtask, we will write the specification of Exascale Interconnect according to the previous analysis.

We will define both the architecture and the features list of the NIC and the switch, and the needed software components to provide with them. This task will produce one deliverable: D5.17

Deliverables : The deliverables of these WP5 are sorted by the time schedule.

D5.1) Specification of the BXI2 **(m6)** [**Bull**, UPV, UCLM, Fraunhofer, CEA]
D5.2) BXI-based platform delivery with BXI1 and Intel Xeon/XeonPhi CPU **(m6)** [**Bull]**
D5.3) First versions of BXI2 SW part **(m12)** [**Bull,** Fraunhofer, CEA]
D5.4) Report on BXI2 simulators developments **(m12)** [**UCLM**, UPV, Bull]
D5.5) First Report on simulation & new add-ons for simulators **(m18)** [**UCLM**, UPV, Bull]
D5.6) Report on BXI1 technology, including all partners contributions **(m18)** [**CEA**, Fraunhofer, Bull]
D5.7) Report on BXI2 HW & SW implementations (including the FPGA based prototype) **(m24)** [**Bull,** Fraunhofer, CEA]
D5.8) Report on the requirements collection for Exascale Interconnect **(m24)** [**Bull**, UPV, UCLM, Fraunhofer, Seagate, CEA, ARM]
D5.9) Report on FPGA simulations of enhanced features for BXI2 **(m28)** [**Bull**, UCLM,UPV]
D5.10) BXI-based platform upgraded with ARM CPU and BXI2 **(m28)** [**Bull]**
D5.11) SW version of a profiling analyser **(m30)** [**Bull**]
D5.12) Report on SW simulations of enhanced features for BXI2 **(m30)** [**UCLM**, Bull, UPV]
D5.13) Simulators delivery for Exascale Interconnect architecture **(m30)** [**UCLM**, Bull, UPV]
D5.14) Report on feasibility analysis for Exascale Interconnect **(m30)** [**Bull**]
D5.15) Report on BXI2 technology, including all partners contributions **(m34)** [**CEA**, Bull, Fraunhofer]
D5.16) Report on simulation for Exascale Interconnect architecture evaluation **(m36)** [**UCLM**, Bull]
D5.17) Specification of Exascale Interconnect **(m36)** [**Bull**]

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

7. *Work package 6: Data Management Ecosystem*

Work package number	**6**		**Start Date or Starting Event**						**m1**				
Work package title	**Data Management Ecosystem**												
Participant number	1	2	**3**	4	5	6	7	8	9	10	11	12	13
Short name of participant	Bull	Seagate	**CEA**	UPV	UCLM	DKRZ	Fraunhofer	BSC	Allinea	SURFsara	INRIA	ARM	UKOELN
Person/months per partner:	36	51	**53**	3		9	36	12	30		3	32	3

Objectives :

The ITHACA architecture is designed to provide new technologies for data management. In WP6, we will study added value services and libraries which enrich the ecosystem around the technologies provided by the WP3, WP4 and WP5. We will focus on performance measurement, access efficiency through data placement, data security and data resiliency. ITHACA platform will provide to applications an interconnect low layer access method, a legacy access method (Posix) and an innovative access method (Object). Posix based will use BeeGFS from Fraunhofer and Object based will use the object store middleware from Seagate. The WP6 studies will cover these data access methods.

Description of work:

The following is a list of the tasks and subtasks required to achieve the objectives of this work package.
The subtasks are grouped by functional domains: analysis tool and monitoring, data movement, fault tolerance and data security.

T6.1. Data movement/placement analysis
(m1-m36) [Allinea, Seagate, ARM, Fraunhofer, DKRZ, Bull, CEA, UPV, INRIA**]**

sT6.1.1: Memory Access Analysis Tool
(m1-m18) [ARM]

ARM currently develops tools that offer HPC developers relevant and actionable code restructuring advice for improved code performance. This is done through automated static analysis of source codes and runtime analysis of the generated binary. ARM will conduct research into the expansion of this tool to provide on-node data energy efficiency advice. The existing performance tools will be expanded, enabling them to first gather cache/memory performance data from software being profiled on a particular node and then process this data along with the results of static analysis of the source code by our compiler. The performance tool will then be able to provide the developer with code restructuring advice to allow them to alter their codes for better data-energy efficiency. At the end of this research and development effort, an end user should be able to profile the execution of a binary using our performance tools and receive precise and relevant code restructuring advice on how to improve their C, C++ or Fortran codes for greater energy efficiency.
This subtask will produce the deliverable D6.3.

sT6.1.2: Performance analysis and other development tools
(m1-m30) [**ARM,** Allinea, Bull, UPV, CEA, INRIA**]**

In this task we will expand ARM's (single-node code profiling and restructuring) and CEA and Allinea's (whole application profiling) analysis tools to include performance metrics available to them from new systems developed during the course of ITHACA and from whole system power metrics available to the tools. This will be a practical and necessary implementation of the results of other ITHACA work made available in commercial software products. The exact nature of this work will depend on the content of this and other work packages. We will investigate expanding these HPC tools to include support for new APIs and programming models developed under ITHACA including support for IO and metrics needed by the programming model's runtimes. UPV will discuss and interact with other partners to define the interface between different components. In particular to define the FlaGAS services and API and how higher levels will use them. We also plan to extend existing tools and runtimes in order to be able to efficiently cache all

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

the required information. For instance, in MPI, the new MPI-T interface to get some runtime internals information (buffer size, thresholds …) may help to understand the exact behavior of applications – for both debugging and performance profiling. We will further aim to extend scalability of tools to address specific needs for the extreme scale. We aim to add functionality to the existing software tools that will allow users of the system to optimize their codes for the new hardware and programming models without specialist knowledge of the underlying system.We will develop methods for converting the large quantities of complex technical information collected from the system into a form that can be presented in such a way as to be useable by those without specialist knowledge of the underlying platform ensuring the tools can be used by widest possible audience. This will be done by combining runtime system data with various static analysis techniques. We will also investigate adding a data sharing API to these HPC performance tools allowing 3rd parties to export collected data in a form suitable to their own analysis. This task will contribute to the deliverables D6.1 and D6.12.

sT6.1.3: Telemetry Infrastructure

(m1-m27) [Seagate, Allinea, DKRZ**]**

This task will research and implement a telemetry infrastructure of the I/O subsystem that gathers fine-grain performance and behaviour data. This task will also co-design the telemetry/measurement interfaces and implementation of metric receiver for Allinea MAP and Performance Reports. Allinea will work with Seagate and DKRZ on key storage hierarchy performance measurement requirements. We will first implement a generic telemetry framework and after we will identify in collaboration with applications the useful metrics. The subtask will also provide inputs into performance management tools (sT6.1.2) and to policy engine (sT6.2.1). This subtask will contribute to the deliverable D6.9.

sT6.1.4: Debugger and MAP for GPI

(m1-m36) [Fraunhofer, Allinea**]**

The GPI framework has not at this point the support of a debugging or profiling tool. Both are essential to enabling developers of applications to adopt GPI and mark a step in maturity of a model.

In order to provide this support, the GPI framework can adopt standards used by other parallel models such as OpenSHMEM and MPI and thereby lever the existing work of tools vendors such as Allinea and of other tools projects. To provide specific GPI data on application performance, a custom metric (a user extensible performance API) will be implemented by Fraunhofer.

Allinea's tools currently support parallel frameworks including MPI, UPC, SHMEM and Coarray Fortran.

The sampling based parallel tools – Allinea MAP and Allinea Performance Reports – will be enhanced to enable profiling of GASPI applications and integrate key metrics from the runtime. Allinea MAP is a tool aimed at developers of HPC codes and is part of Allinea's single unified environment of tools – giving line-by-line performance information on parallel applications. By contrast, the perspective offered by Allinea Performance Reports is a single page summary of application performance – enabling a high level overview to be shared with system owners and application users. Allinea MAP currently supports MPI, OpenMP.

Debugger: Identify missing features in the GASPI specification needed to support debugging tool; Define appropriate APIs and make a proposal to the GASPI standardisation forum; Enable the GASPI applications to provide support for debuggers according to the specifications identified; Support the GPI API within debugger by implementing modifications to support process acquisition and consequent debugging of GASPI applications.

Performance Analysis Tool: Extend tool to support the PGAS programming model; Implement new modules or support mechanisms similar to the existing for MPI and OpenMP; Identify missing features in the GASPI specification needed to support the performance analysis tools, define appropriate APIs; Proposal to the GASPI standardisation forum if necessary.

This subtask will contribute to the deliverable D6.5

T6.2. Data Movement

(m1-m36) [Fraunhofer, Seagate, CEA, Bull, BSC, DKRZ]

sT6.2.1: Policy Engine

(m1-m30) [Seagate, Bull, CEA, BSC, DKRZ**]**

This task will research and implement a policy engine that studies I/O workloads and optimizes the various elements of the I/O infrastructure for the various ITHACA use cases helping towards end-end I/O performance optimization. The policy engine can be driven by inputs from guided I/O or can be automated based on machine learning, etc. The policy Engine works closely with the data manager (WP4) and with the telemetry infrastructure (subtask 6.1.3). The Policy engine will also research and provide rudimentary I/O

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

QoS capability. This subtask will contribute to the deliverable D6.6 and D6.10

sT6.2.2 BeeGFS Cache API

(m1-m36) [**Fraunhofer**]:

BeeGFS implements a cache on the client. However, often the application developer is not willing to handle data movement explicitly. Therefore further development must include automation of prefetch and flush operations. The goal is to handle data movement between the caches and the central storage transparently for the user. For instance, this could be based on policies and heuristics implemented in the cache daemon. This way, it is possible to simply use the familiar POSIX interface to access data, while still benefiting from the caching mechanism. To fully utilize the BXI network technology: Portals (of WP5), needs to be integrated into BeeGFS. This subtask will contribute to the deliverable D6.13

T6.3. Fault Tolerance

(m1-m34) [Fraunhofer, Seagate, BSC**]**

sT6.3.1 Erasure coding support in BeeGFS

(m1-m34) [Fraunhofer]

One aspect that needs to be considered is the fact that the BeeGFS cache domains can be built by using BeeOND. This is a powerful and flexible way of dynamically creating the caches, as we could use storage devices that are directly attached to the compute nodes. Therefore, there is no need for an additional, dedicated storage environment for the caches (of type NVRAM for instance). However, given the fact that compute nodes are likely to fail during jobs, additional resiliency features need to be implemented. For this use case the most sensible solution is erasure coding which can be implemented with different, configurable code rates to achieve a good trade-off between safety, capacity and performance.

This subtask will contribute to the deliverable D6.4 and D6.14

sT6.3.2 Data Resiliency in Mero

(m1-m30) [Seagate, Fraunhofer**]**

This subtask will provide resilient data layouts as needed by the ITHACA Use cases. The types of layouts will be based on the resiliency requirements of the use cases. This task will leverage the concept of layouts developed in WP4. This subtask will contribute to the deliverable D6.7 and D6.11

sT6.3.3 Making dataClay resilient

(m1-m34) [BSC, Seagate]

In this subtask we plan to take advantage of the resiliency semantics offered by Mero, an Advanced Object Store, in order to make dataClay objects resilient. This integration does not only consist on a simple porting, but we will investigate the real resilience needs of objects from the programming model (implemented by dataClay) and what functionality needs to be offered by the object store to enable the right functionality and flexibility.This subtask will contribute to the deliverable D6.15

T6.4 Data Security

(m1-m36) [CEA, Bull, Seagate]

sT6.4.1 Data Movement Security

(m1-m36) [CEA, Bull]

Processing various levels of sensitive information and sharing distributed resources among users require the development of a multi-level security framework that can guarantee authentication and isolation of users' data. The goal of this topic is to work on points of progress that aim to enhance data movement security over the infrastructure for Exascale. Global resources (compute and storage) of the computing center are controlled by the resource allocator but interconnect is still out of its scope. Provisioning on-demand network resources requires including the fabric manager that controls interconnect. We plan to specify the interfaces between the resource manager, network fabric and security infrastructure.

Last point will be to test and evaluate the BXI v2 interconnect under the perspective of security, identification and classification of flows to deliver the proper security specification for the future Exascale Interconnect, compatible with a multi-level security framework. This subtask will contribute to the deliverable D6.2 and D6.8

sT6.4.2 Data Security

(m1-m30) [Seagate, CEA]

This subtask will look at providing various levels of security to data, leveraging the concept of layouts developed in WP4. In a co-design approach we will define a security model for data access with WP7 and implement it. This subtask will contribute to the deliverable D6.2 and D6.8

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Deliverables: The deliverables of these WP6 are sorted by the time schedule.

D6.1) First report for analysis and performance tools **(m12) [Allinea**]
D6.2) First Data security report **(m12)** [**CEA**, Seagate, Bull]
D6.3) A Memory Access Analysis Tool software version **(m18) [ARM]**
D6.4) Integration of NVRAMs in the in-memory checkpointing library for GPI **(m18) [Fraunhofer**]
D6.5) a GPI analysis and debugger Tools software version **(m24) [Fraunhofer**]
D6.6) Prototype of policy engine to improved data moving (**m24) [Seagate**]
D6.7) Integration of the resilience mechanism in GPI-Space **(m24) [Fraunhofer**]
D6.8) Data security Specification report **(m24**) [**CEA**]
D6.9) Telemetry Infrastructure: Validation & inputs to Tools report **(m27) [Seagate**]
D6.10) Validation of IO policy engine report (**m27) [Seagate**]
D6.11) Mero Data Resiliency: Validation report (**m27) [Seagate**]
D6.12) an enriched version of performance tools (**m32) [Allinea**]
D6.13) BeeGFS Benchmarks with Portals and final report (**m32) [Fraunhofer**]
D6.14) Integration of Erasure Coding Algorithms in BeeGFS (**m34) [Fraunhofer**]
D6.15) A reliable dataClay implementation (**m34) [BSC**]

8. Work package 7: Applicative use cases

Work package number	**7**		**Start Date or Starting Event**					**m1**				
Work package title	**Data Applications use cases**											
Participant number	1	2	3	4	5	**6**	7	8	9	10	11	12	13
Short name of participant	Bull	Seagate	CEA	UPV	UCLM	**DKRZ**	Fraunhofer	BSC	Allinea	SURFsara	INRIA	ARM	UKOELN
Person/months per participant		12				**45**	34	54	18	36	24		64

Objectives :

This WP7 deals with use cases from several scientific and big data application domains and fosters co-design between the development of the hardware and software platform and the application-specific requirements. Progress of this WP is aligned with the development of the other WPs.

Description of work:

Lead partner: DKRZ

Allinea, BSC, DKRZ, Fraunhofer, INRIA, SURFsara, Seagate and UKOELN partners contribute to this WP. The WP7 follows a strategy to which all application use-cases align: firstly, in T7.1 a test environment is prepared and a thorough performance analysis on large scale systems is conducted, then an initial set of requirements for future systems is prepared together with predictions of application execution key characteristics (e.g., performance, run-time, power consumption) on the created platform based on the availability of hardware features (T7.2). Over the course of the RD&E of ITHACA, application codes and workflows are modified to harness the advanced communication (T7.3) and data movement capabilities (T7.4). This process includes some application-specific evaluations and code-refinements that would be beneficial for the community as a whole. They will be described individually in the task lists. During this process, we follow the idea of test-driven development as a feedback channel to the technological WPs:
The complexity of the modified code will be gradually increased; initially, only small code portions are refined as tests, then into benchmarks and mini-apps that can be used of the development teams in the other WPs to judge the quality of the hardware platform and to refine the estimates for the full system (T7.2). This iterative and feedback-driven process will drive the co-design and allows us to assess the potential of the new architecture without porting complete workflows.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Ultimately, we aim to port some workflows to the new system architecture and directly demonstrate the benefit (T7.5). However, which applications and use-cases will finally be completely ported depends on the estimated coding effort and results obtained from intermediate mini-apps.
The applications are selected based on the following criteria:

- 1) The applications are representative for its domain
- 2) One partner has long experience with this particular application and is involved in its development
- 3) There is a demand for performance with 100+ PFLOP
- 4) The applications have already demonstrated high scalability
- 5) Strong and weak scaling are relevant. Consequently all selected applications are candidates to be executed on the Exascale demonstrators developed within H2020.

The applications involved in this WP are listed in the following table to show the enriched data features and aspects that will be used by these applications. Only the more impacting WPs are listed in the following table:

Applicative Domain	*Partner* **/ application**	**Progr. paradigm**	Candidate to use or take benefits from other WPs features in respect to		
			I/O	Workflows	Tool support
Deep Learning	*Fraunhofer* **CaffeGPI**	PGAS			6.1. Telemetry for I/O Subsystem +++
					6.1. Allinea Debugger/MAP for GPI +++
					6.2. BeeGFS Cache API +++
					6.3.Erase coding in BeeGFS ++
Seismics imaging	*Fraunhofer* **GRT**	PGAS	4.3 Storage Hardware ++		6.1. Telemetry for I/O Subsystem +++
					6.1. Allinea Debugger / MAP for GPI +++
Climate models	*DKRZ* **ICON**	MPI	3.3. Structured objects ++	3.4. Workflow APIs ++	5. BXI features (for YAXT) ++
			4.6 HDF5 ++	3.4 In-situ ++	6.1. Memory analysis +
			4.6. POSIX +		
	BSC **EC-Earth**	PyCOMPSs	3.3. dataClay/PyCOMPs +++	3.4. In-situ ++	
Molecular dynamics	*Inria* **Gromacs**	MPI	3.3. Structured objects ++	3.4 In-situ +++	6.1. Memory analysis +
			4.2. Storage adressing with FlaGAS ++		
			4.6. HDF5 +		
			4.6. MPI-IO ++		
Fluid Dynamics	*SURFsara* **AFiD**	MPI	4.5 HDF5 +	3.4 Workflow API ++	6.1. Allinea Debugger/Map for GPI +
		PGAS		3.4 In-situ ++	
Genetics Pipeline	*UKOELN* **Genomics**	PGAS	3.3. dataClay/GPI-Spaces +++	3.4. Workflow APIs ++	6.2. BeeGFS Cache API ++
			3.3. Structured objects +		
			4.6. POSIX ++		

The color scheme indicates the current status of the development:

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

- Green means it is ready to be used within ITHACA and, thus, with only minor adjustments the feature will be evaluated in this project.
- Blue means we aim to achieve **full support** within ITHACA: the scientific application will be ported to use the feature on relevant scientific experiments.
- Orange means we aim to achieve **partial support**: some code regions will be migrated to assess the benefit or a used library is adjusted.
- Velvet means mini-app: a **mini-app** will be written to test the benefit on key code regions.

For example, all applications with exception to the genetics pipeline are already parallelized using the listed programming paradigm. They will all benefit from improvements in the communication infrastructure.

After the task and shortened task name, the number of pluses indicate the expected benefit:

- +: Some benefit for performance/energy efficiency and/or software development.
- ++: Significant changes
- +++: Game changer

All other tasks provide implicit benefit to all applications but are not listed in the table, for example:

- 3.5. Task placement will be implicitly used when Task 3.4 is applicable
- 3.6. Performance analysis tools and development tools are applied to all applications
- 4. Object store infrastructure, all applications will be evaluated on the new storage hardware and take benefit of DDCs
- 4.7. Simulation, all applications should be evaluated on the simulator
- 5. BXI features are indirectly used for all programming model
- 6.1. Telemetry for I/O subsystem, we will collect the data and analyse it for all application
- 6.2. Policy engine, all applications will evaluate the impact on security.
- 6.3. Data resiliency in Mero, all applications will evaluate the benefit of using the object storage
- 6.4. Data movement security, all applications will apply their typical security level in production

Note that the table is subject for adjustment during the project; depending on the projected benefit according to Task 7.2., the priorities may change.

The following is a list of the tasks required to achieve the objectives of this work package.

T7.1. Test environment and baseline
(m1:m12) [Allinea with WP's partners].

In this task, firstly, the current application workflows are prepared on several HPC systems. Then, a thorough performance analysis on (medium) scale runs is conducted using state-of-the-art performance tools from Allinea. The obtained data of the different applications is used as reference values for performance and power consumption. Based on these values the application characteristics for future systems (100 PFLOP/250 PFLOP) are estimated using light-weight extrapolation models.
This task will produce the deliverable D7.1

T7.2. Potential of emerging architectures
(m3:m18) [DKRZ with WP's partners].

The main goal of this task is to understand the benefit of alternative hardware/software features and feedback this information to the development of the APIs and hardware platform.
We will extend the simple model and incorporate the availability and characteristics of advanced hardware and software platforms as parameters. This allows us to roughly estimate the benefit these features will bring to the execution of the applications and direct the hardware/software co-design. This explicitly covers predictions for performance, energy consumption but also the impact of failure rates.
An initial analysis will lead to a white paper that documents these requirements and the initial predictions and kick-off the co-design cycle with the other work-packages. The model and relevant characteristics will be update during the project. This task will produce the deliverable D7.2

T7.3 Advanced storage concepts
(m12:m36) [BSC with WP's partners]

This task bundles the responsibility and expertise to modify application code (or create suitable mini-apps) to utilize the new I/O features (see the column "I/O" for individual tasks). To illustrate the use of the features and potentially measure performance once they are implemented by WP3, WP4 and WP6, we follow the idea

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

of test-driven development (as described above).
This task will produce the deliverable D7.3 and contribute to D7.5

T7.4 Advanced workflows studies
(m12:m36) [DKRZ with WP's partners]
The idea of this task is similar to the previous task, but deals with the workflow features provided by WP3 and complete them with some studies about in-situ mechanisms. For instance, we will harness the active storage concepts to outsource post-processing and in-situ analysis for the applications..
This task will produce the deliverable D7.4 and contribute to D7.5

T7.5 Advanced tools support
(m12:m36) [SURFsara with other WP's partners]
Similarly to task 7.3 and 7.4, this task bundles utilization and evaluation of tools developed/extended in this project that primarily stem from WP6. It also covers the application of the I/O simulator and insight gained from the telemetry data of the storage. The benefit of the tools is qualitatively assessed by the partners and the gained insight documented. Insight is documented in D7.5.

T7.6 Applicative Tests and Results
(m18:m36) [DKRZ with other WP's partners]
This task integrates all the software and hardware components from WP3/4/5 & 6, hence providing the full ITHACA platform for demonstration with the various use cases in this WP. Benchmarks, applications and complete workflows are executed on the ITHACA platform hosted by Bull in France. The results are assessed and compared to our baseline (T7.1) and predictions (T7.2).
Performance analysis tools will be used to identify and investigate bottlenecks. To benefit the scientific merit, scientific results of selected use cases are visualized and further analysed.
This task will contribute to the deliverable D7.5

Deliverables : (Ordered by delivery date; note task and deliverable numbers are unrelated)

D7.1) Applications state of the art: (m9): [Allinea & WP7 partners**]**
This first WP deliverable will contain the obtained data as reference values and characteristics for the set of applications.
D7.2) Application requirements and changes for emerging architectures: (m12) [DKRZ & WP7 partners**]**
This deliverable documents relevant requirements and predictions for the various use-cases and describes our base prediction models. It directly feeds into the co-design of WP3 (and indirectly WP4 and WP5).
D7.3) Utilizing new storage concepts (m24) [**BSC** & WP7 partners]
This deliverable documents our insight from Task7.3, It documents the challenges in the software development to adjust I/O of this WP7 applications and mini-apps to the new concepts, describes the code changes needed and provides code-snippets illustrating the changes. It contains the extended models predicting the impact of the various hardware/software features on application execution key characteristics and the initial measurements on the WP3 and WP4.
D7.4) Utilizing new workflow concepts (m24) [DKRZ & WP7 partners]
This deliverable documents our intermediate insight from Task7.4. It is similar to D7.3 but addresses the new workflow concepts and in-situ data processing.
D7.5) Challenges for Exascale demonstrators (m36) [DKRZ & WP7 partners]
This deliverable extends the deliverables D7.2, D7.3 and D7.4 that illustrate examples on small code snippets and tests, to complete application workflows. Additionally, the tool benefit as part of D7.4 is now described. It provides an overview of the results to demonstrate the new level of science that can be achieved with ITHACA, as well as the lessons learned and refined predictions for future systems, especially for 100+ PFLOP demonstrators.

9. Work packages inter-relationship

The following diagram illustrates these co-design relationships between work packages at a high level point of view:

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

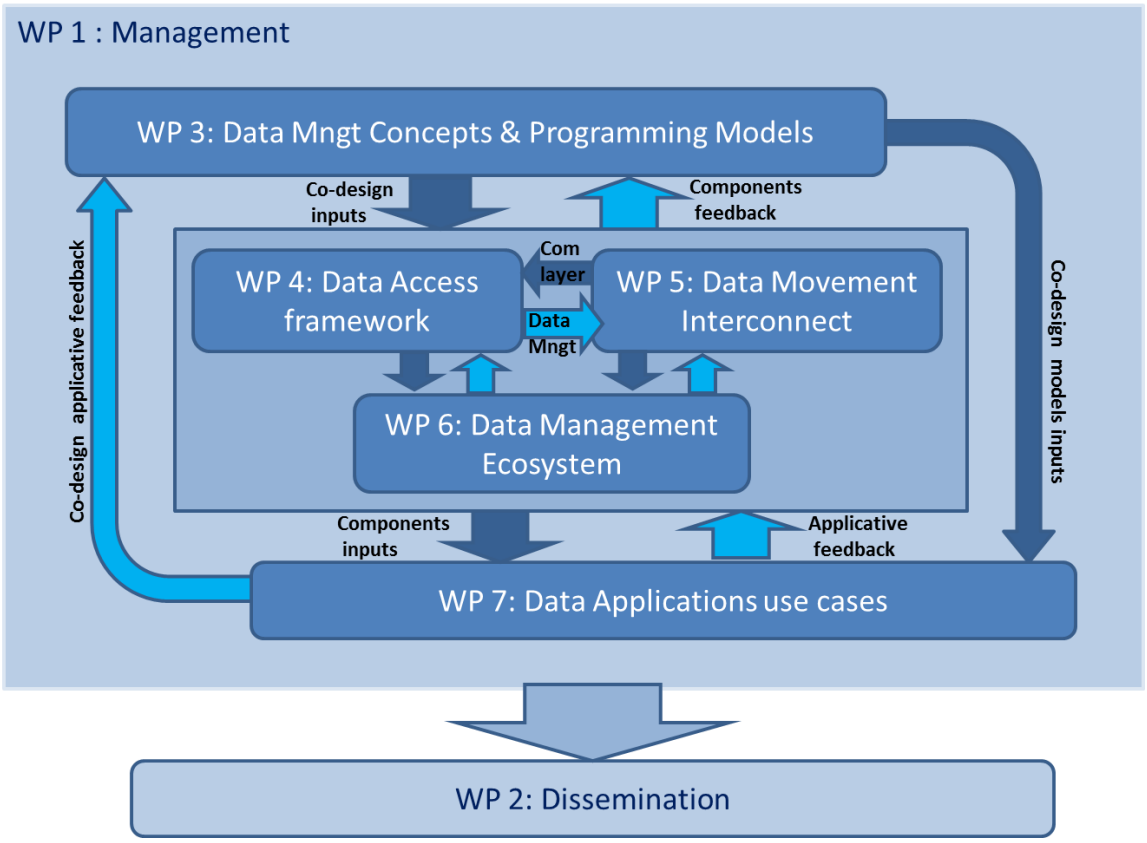


Figure 10: Work Packages Relationships

The first work package is dedicated to the project management, and the second one to the dissemination domain. In the first period of the project, the WP3 will define the concepts and will study several programming models to support the proposed Data Management architecture, which collapse RAM, NVRAM and IO storage. The first results will be used by the WP4 and WP5. WP4 will be focused on the developments of the Data Access middleware components: low level System services, Data Manager Engine, FlaGAS layer... While WP5 will provide the Hardware and Software component of a new Interconnect framework based on the Portals protocol to improve Data movement. In a second time, these three WP will interact to define new features and improvements. In parallel, the WP6 will inherit of the previous WP components and new features, to enrich or create new added-value tools, services or high-level API. The WP7 will be driven by applicative use cases and will integrate the components and models provided by the WP 3, 4, 5 and 6. It will aim at demonstrating the benefits of this new Data Management framework. During the third and last period of this project, all these studies will furnish data and results to all work packages, which will concern directly the next generation design. The WP3 and WP4 will concentrate the studies at Data Management middleware level, while the WP5 will work about the evolvement and new specification of the different components of the next generation of Interconnect.
You will find the complete table of effort and detailed information about resources in chapter 3.4: “Resources to be committed”.

10. List of deliverables

The following **Table 3.1c: List of Deliverables** includes the complete list of ITHACA project deliverables:

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Deliverable (number)	Deliverable Name	Work Package Number	Lead Participant (Short Name)	Type	Dissemination Level	Delivery Date
D1.1	ITHACA Project Management Handbook	1	Bull	R	CO	m4
D1.2	First annual Progress Report	1	Bull	R	CO	m12
D1.3	Second annual Progress Report	1	Bull	R	CO	m24
D1.4	Third annual Progress Report	1	Bull	R	CO	m36
D1.5	Project Final Report	1	Bull	R	CO	m36
D2.1	Release of the project-community platform	2	Bull	DEC	CO	m1
D2.2	Release of the social-network profiles of the project	2	UCLM	DEC	PU	m4
D2.3	Release of the public project website	2	UCLM	DEC	PU	m6
D2.4	Release of the initial video	2	UCLM	DEC	PU	m6
D2.5	First Intermediate report on dissemination activities	2	UCLM	R	PU	m12
D2.6	Second Intermediate report on dissemination activities	2	UCLM	R	PU	m24
D2.7	Release of the final video	2	UCLM	DEC	PU	m34
D2.8	OpenSource software repository report	2	UCLM	R	PU	m34
D2.9	Summer school report	2	UCLM	DEC	PU	m36
D2.10	Final report on dissemination activities	2	UCLM	R	PU	m36
D3.1	First specifications Report	3	BSC	R	CO	m12
D3.2	Programming models extensions specification	3	INRIA	R	CO	m18
D3.3	First prototype of the programming models	3	BSC	OTHER	CO	m20
D3.4	Prototype of the programming models with the proposed extension	3	BSC	OTHER	CO	m30
D3.5	Performance evaluation of programming models	3	DKRZ	DEM	CO	m32
D4.1	Overall architecture of the Data Access Middleware	4	Seagate	R	CO	m4
D4.2	Hardware Platform Framework Report	4	Seagate	R	CO	m6

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Deliverable (number)	Deliverable Name	Work Package Number	Lead Participant (Short Name)	Type	Dissemination Level	Delivery Date
D4.3	Hardware Platform Detailed Architecture Report	4	Seagate	R	CO	m9
D4.4	Data Access Middleware Modules: Concept & Architecture Report (First Version)	4	Seagate	R	CO	m12
D4.5	Hardware Platform Availability Report	4	Seagate	R	CO	m18
D4.6	Data Access Middleware Modules: Initial Proof of Concepts	4	Seagate	DEM		m20
D4.7	Data Access Middleware Modules: Second Concept & Architecture Report	4	Seagate	R	CO	m24
D4.8	storage and I/O device models for feeding into a Distributed Storage Simulator	4	Seagate	R	CO	m24
D4.9	FlaGAS memory protocol extension and implementation description report	4	UPV	OTHER	CO	m24
D4.10	Demonstration infrastructure of the Distributed Storage Simulator	4	UCLM	DEM	CO	m30
D4.11	Data Access Middleware Modules: Final Prototypes	4	Seagate	DEM	CO	m35
D4.12	Hardware Platform Optimization and Use case Access Report	4	Seagate	R	CO	m35
D5.1	Specification of the BXI2	5	Bull	R	CO	m6
D5.2	BXI-based platform delivery with BXI1	5	Bull	OTHER		m6
D5.3	First version of BXI2 SW part	5	Bull	OTHER		m12
D5.4	Report on BXI2 simulators developments	5	UCLM	R	CO	m12
D5.5	First Report on simulation & new add-ons for simulators	5	UCLM	OTHER		m18
D5.6	Report on BXI1 technology	5	CEA	R	CO	m18
D5.7	Report on BXI2 HW & SW implementation, including the FPGA prototype	5	Bull	R	CO	m24
D5.8	Report on the requirements collection for Exascale Interconnect	5	Bull	R	CO	m24

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Deliverable (number)	Deliverable Name	Work Package Number	Lead Participant (Short Name)	Type	Dissemination Level	Delivery Date
D5.9	Report on FPGA simulations of enhanced features of BXI2	5	Bull	R	CO	m28
D5.10	BXI-based platform upgraded with BXI2	5	Bull	DEM		m28
D5.11	SW version of a profiling analyser	5	Bull	OTHER		m30
D5.12	Report on SW simulations of enhanced features of BXI2	5	Bull	R	CO	m30
D5.13	Simulators delivery for Exascale Interconnect architecture	5	UCLM	R	CO	m30
D5.14	Report on feasibility analysis for Exascale Interconnect	5	Bull	R	CO	m30
D5.15	Report on BXI2 technology	5	CEA	R	CO	m34
D5.16	Report on simulation for Exascale Interconnect architecture evaluation	5	UCLM	R	CO	m36
D5.17	Specification of Exascale Interconnect	5	Bull	R	CO	m36
D6.1	First report for analysis and performance tools	6	Allinea	R	CO	m12
D6.2	First Report on Data Security	6	CEA	R	CO	m12
D6.3	SW version of a Memory Access analyser	6	ARM	OTHER	CO	m18
D6.4	NVRAM integration in the in-memory checkpointing with GPI	6	Fraunhofer	OTHER	CO	m18
D6.5	SW version of a GPI analyser & debugger	6	Fraunhofer	OTHER	CO	m24
D6.6	SW version of Policy Engine to improved data moving	6	Seagate	OTHER	CO	m24
D6.7	Resilience mechanism in GPI-space	6	Fraunhofer	OTHER	CO	m24
D6.8	Data Security specification	6	CEA	R	CO	m24
D6.9	Report on Telemetry infrastructure	6	Seagate	R	CO	m27
D6.10	Report on IO Policy engine	6	Seagate	R	CO	m27
D6.11	Report on Validation of Mero Data Resiliency	6	Seagate	R	CO	m27
D6.12	An enriched version of performance tools	6	Allinea	OTHER	CO	m32
D6.13	BeeGFS benchmark with Portals	6	Fraunhofer	R	CO	m32

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Deliverable (number)	Deliverable Name	Work Package Number	Lead Participant (Short Name)	Type	Dissemination level	Delivery Date
D6.14	Integration of Erasure Coding Algorithm in BeeGFS6	6	Fraunhofer	OTHER	CO	m34
D6.15	A reliable dataClay implementation	6	BSC	OTHER	CO	m34
D7.1	Applications state of the art	7	Allinea	R	PU	m9
D7.2	Application requirements and changes for emerging architectures	7	DKRZ	R	CO	m12
D7.3	Utilizing new communication concepts	7	BSC	R	CO	m24
D7.4	Utilizing new data management concepts	7	DKRZ	R	CO	m24
D7.5	Challenges for Exascale demonstrators	7	DKRZ	DEM + R	CO	m36

3.2 Management Structure, Milestones and Procedures

Project coordination will be achieved through following structures:

- The **Project Management Board (PMB)** is composed of one representative per partner, with each representative having a deputy. The PMB, chaired by the Project Coordinator, Bruno Farcy (Bull), will be the main decision body of the consortium: The board will make all formal decisions regarding direction of work, policies for promotion and exploitation of results, administrative arrangements and conflict resolution. PMB remains responsible for: contractual amendments; project budgetary issues; addressing recommendations made by other management bodies; addressing ethical, legal and gender issues relevant to the project.The PMB will host remote quarterly meetings, in addition to at least two face to face meetings per year.

- The **Technical Management Board (TMB)** consists of at least one representative per partner and the Work Packages leaders: Bruno Farcy (Bull) for WP1, Pedro Garcia (UCLM) for WP2, Toni Cortes (BSC) for WP3, Sai Narasimhamurthy (Seagate) for WP4, Sylvie Lesmanne (Bull) for WP5, Jacques-Charles Lafoucrière (CEA) for WP6 and Julian Kunkel (DKRZ) for WP7. The TMB, chaired by the Technical Manager, will
  - Be responsible for the overall technical consistency,
  - Monitor technical execution and performance of the project's activities,
  - Advise the PMB on any necessary plan adjustment
  - Ensure accomplishment of the project's technical and business objectives

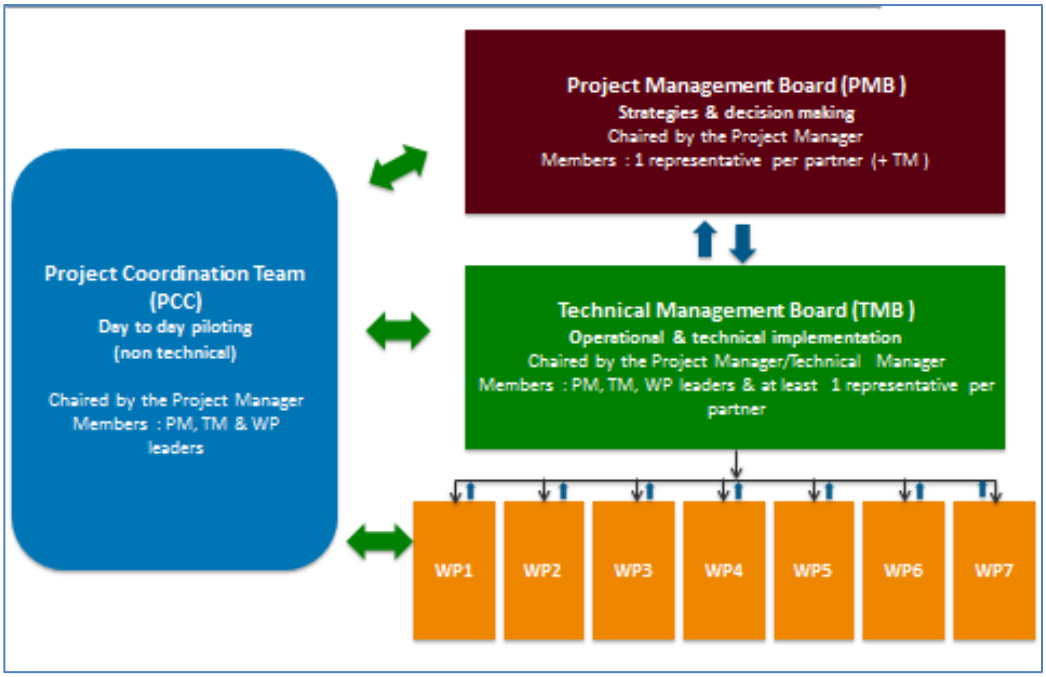
  The TMB will host monthly meetings on remote basis, and face to face meetings in conjunction with the PMB (with additional meetings held on an exceptions basis if needed).

- The **Project Coordination Committee (PCC)**, chaired by the Project Coordinator or a deputy includes the Work Packages leaders (or deputies). The PCC complements the TMB. Its mission is to execute the PMB's decisions as well as manage the day-to-day activities of the project. The PCC's responsabilities include: a) advising the Project Manager regarding management of the project when rapid decisions are required, b) together with the TMB, assuring effective coordination of project's assets to realise ITHACA objectives, such as preparation of project events such as project reviews and face-to-face meetings, c) dealing with IPR, Communication and Exploitation potential issues, and d) reporting to the PMB during any point in the project that requires a PMB decision.

The following schema illustrates the management structure of this cooperative project:

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

ITHACA Management Structure



Within this organisation, each **Work Package Team (WPT)** for WP2, WP3, WP4, WP5, WP6 and WP7, will be chaired by a corresponding Work Package Leader. These Work Package Leaders will be members of the TMB and of the PCC, to ensurie consistency between these two boards.
In addition, a **Project Advisory Committee (PAC),** chaired by the Project Manager or a deputy, includes one scientific advisor: Jose Duato (UPV) and one technical advisor: Jean-Pierre Panziera (Bull). The PAC's mission is to reinforce strategic management and complete scientific and technical expertise areas of the project.

Project management will organize **two general assemblies by year**. Each assembly will gather all of ITHACA's participants, in a face-to-face meeting, to facilitate project-wide consistency and understanding, and share information about technical issues that may extend beyond one work package. These general assemblies shall integrate PMB and TMP meetings, and report on the status of tasks and deliverables specified in the description of the WP1.
The project's management structure will enable technical aspects and management issues to be addressed quickly, at a work package level, and permit decisions to be made at a global level to avoid blocking progress of the project at a task level.
The Project Coordinator will be responsible for operation of the entire project and to maintain communication with the European Commission and external bodies.
The following table lists the principal milestones of the project which illustrate its cooperative nature:

Milestone Number	*Milestone Name*	Related Work Package(s)	Due Date (in month)	Means of Verification
M1	*Tests vehicle with BXI1*	5	m6	D5.2
M2	*Specification of BXI2*	4 & 5	m6	D5.1
M3	*Application requirements*	7	m12	D7.2
M4	*Specification of DataAccess Middleware architecture & Programming models*	3,4 & 7	m12	D3.1 & D4.4
M5	*First prototype of the programming models*	3, 4 & 7	m20	D3.3
M6	*First version of HW and SW interconnect, components and added-value tools*	4 & 6	m24	D4.7, D5.7, D6.5, D6.6 & D6.7

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

M7	*Integration of new programming modelsby applicative use cases*	3 & 7	m24	D7.3 & D7.4
M8	*Advanced Programming models with extensions*	4, 5, 6 & 7	m30	D3.4
M9	*Enhanced version of Data Acces middleware and Value-Added tools*	4, 5 & 6	m34	D4.11, D6.14 & D6.15
M10	*Spec Exascale Interconnect*	4, 5, 6 & 7	m36	D5.17
M11	*Applicative Results*	3, 4, 5, 6 & 7	m36	D7.5

Table 5: Main project milestones

Dependencies exist between some milestones. The Exascale Interconnect specifications (M5.5) will use the results from the M6.2 milestone. The milestones M3.1, M4.1 and M7.1 will result from a co-design effort and constitute an interdependent set. M6.1 relies on Data Access Middleware Architecture (M4.1). Finally, M7.3 depends on major milestones from the other packages.

The Project Manager is responsible, along with the PMB and TMB, to manage risks that arise in the project and be vigilant about the prospect of uncertain situations that have (or may) develop, in order to detect and contain them. The WP Leaders or a partner representative will report the identified risks and issues to the PMB and TMB (management boards) and PCC will maintain a Risk / Issue Log for the project's duration that lists assigned actions and contingency plans to enable the project's objectives and overall outcome to be met successfully.

The major risks initially identified in the project are listed in Table 6 below. The project's structures (described above) are charged with managing these risks and maintaining current status of all potential risks that may critically impact the project. Table 6 also lists recommended responses for the identified risks:

Description of Risk	WP(s) Involved	Severity	Probability	Proposed Risk-Mitigation Measures
R1: Losing a critical partner at a crucial point in the project This risk belongs to the category of inevitable ones in any partnership - namely changes to the direction and/or business of one or more partners leading to their withdrawal from the project.	All WPs	H	L	Risk effect will be dependent on the partner, the work affected, and the timing. Relatively to the core partners, already present in previous H2020 projects (ARM, BSC, Seagate, Bull, CEA,…), the likelihood of this risk is extremely low, since these partners already have a close collaboration history, sharing a common vision and a strong and long-term commitment. In the quite unlikely case of partner withdrawing, an adapted contingency plan would be quickly established and shared with the H2020 EU officers, permitting to pursue the global objectives and to limit the impacts on the scope of work.
R2: New features of Portals are not ready on time. Some technical complexities may delay the integration of new features in Portals API.	3	M	M	Partners can progress with the other new features offered by WP4 to be integrated in to the programming models
R3: Poorly defined needs for Data Access Middleware on time. Because of the difficulties to translate the applicative needs into new features.	4	M	L	Continuous and close interaction between WP4 and, WP3 and WP7 helps to hone the requirements. Indeed the Data Access Middleware high level architecture will be defined based on co-design.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Description of risk	WP(s) involved	Severity	Probability	Proposed risk-mitigation measures
R4: Data Management features from WP4 are not ready on time. FlaGAS, Data Manager,… implementation are delayed.	3 & 4	M	M	Partners will focus on adapting the features implemented on time and take full advantage of them.
R5: Simulators not ready on time with relevant features. Some important simulator features may not be detected and specified at the right moment during the project.	4 & 5	H	L	From the beginning of the project, an active co-design between providers and users should reduce this risk.
R6: Tests Vehicles not ready on time. BXI interconnect or some Data Access components may be delayed.	4, 6 & 7	M	L	Partners can use the simulators developed in the WP4 and WP5 to start their studies.
R7: New hardware technologies: ASIC 7nm vs 10nm , and new serdes (56Gb/s). The use at first time of new HW technologies.	5	M	M	The project is well aware of this risk. These new semi-conductor and serdes technologies will be studied before and at the beginning of ITHACA project, in order to select the suitable ones for this project.
R8: Concurrent IPR. External organisations patent elements of the developments in storage or interconnect software.	3, 4, 5 & 6	M	M	Establish IPR as soon as practical. Monitor IPR for technology trends and conflicting ownership.
R9: Use of NVRAM in distributed mode. Non-availability of NVRAM technologies in the time frame of the project.	4	H	L	Today, we lack information to know exactly how to use them and their availability. The first WP4 task will study this item in order to specify the exact use of them, or to define a backup plan. (NVMe emulation in RAM for instance).
R10: No agreement for features list between partners: The selection of some features for Data Movement optimisation, in the interconnect for example may be difficult because of the large number of partners.	4, 5 & 6	H	M	Detailed and open analysis between advantages vs complexity and cost of each feature between all concerned partners will be made frequently in order to align the expectations
R11: Disagreement among partners Again the likelihood of this risk is low, since most of the partners have collaborated successfully in the past. To reduce the risk, there should be strong leadership at the work package and project level.	All WPs	M	L	If disagreements arise, the project coordinator is responsible for solving conflict situations and propose adapted plan, helped by the project management structures.

Table 6: Critical Risks for Implementation

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

3.3 Consortium as a Whole

The ITHACA consortium is composed of a diverse range of partners with expertise covering three different fields: Academic, Industrial and HPC users. All project partners have extensive experience in the HPC domain. Academic institutions will be heavily involved in design and modelling tasks, while industrial partners will add their own technical and business constraints to the models before applying them. HPC users will contribute application use cases and their experience at the start of the project, and at the end they will test the newly defined models and provide initial conclusions regarding the simulation tools. Many of the project partners have several roles and will participate in multiple work packages.

Table 7 details the capabilities and/or roles of each partner:

Partner	Model & Concepts Provider	Technology (HW) Provider	Ecosystem (SW tools & APIs) Provider	HPC Use Case Provider
Bull	x	x	x	
Seagate	x	x	x	
CEA	x		x	
UPV	x		x	
UCLM	x		x	
DKRZ	x			x
Fraunhofer	x		x	x
BSC	x		x	x
Allinea			x	
SURFsara			x	x
INRIA	x		x	x
ARM	x		x	
UKOELN	x			x

Table 7: Project Partners Capabilities and Roles

3.4 Resources to be Committed

Resource commitment by partner and by work package: The following table shows the planned commitment of person/months to the project for each partner and for each work package:

	WP1	WP2	WP3	WP4	WP5	WP6	WP7	Total Person/ Months per Participant
1/ Bull	**18**	12	84	270	**360**	36		**780**
2/ Seagate	9	9	12	**159**	9	51	12	**261**
3/ CEA	6	3	12	36	12	**53**		**122**
4/ UPV	8	10	3	37	117	3		**178**
5/ UCLM	8	**26**	18	54	74			**180**
6/ DKRZ	6	6	30	18		9	**45**	**114**
7/ Fraunhofer	2	2	34		6	36	34	**114**
8/ BSC	12	3	**105**	36		12	54	**222**
9/ Allinea	4	2				30	18	**54**
10/ SURFsara	2	2					36	**40**
11/ INRIA	1	2	27			3	24	**57**
12/ ARM	2	2				32		**36**
13/ UKOELN	1	1	3			3	64	**72**
Total Person/Months	**79**	**80**	**328**	**610**	**578**	**268**	**287**	**2230**

Table 8: Commitment by Partner and by Work Package

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

All partners will contribute resources in the form of person/months to the management and dissemination of this cooperative project (WP1 and WP2).

For WP3: **BSC** will lead the work package and contribute by extending and evaluating PyCOMPSs, a task-based parallel programming model designed and exploited by BSC. Among the extensions we will include are the integration with dataClay, the data platform, the use of Portals4 and the proposed extensions on workflows, in-situ, and data as a first class citizen by providing our experience in managing data from a programming model. BSC will also contribute to the design of the afore mentioned extensions and final comparison of the programming models.
Bull, UPV, UCLM and CEA will work on improvements and new data management features in the communication layers (Portals) and in HPC programming models (MPI).
Seagate will work to extend the Mero object store API (Clovis) for the different programming models.
INRIA will work on integrating new data management capabilities for scientific workflows, in particular, data intensive ones requiring in-situ processing capabilities that rely on the FlowVR framework.
Together with Fraunhofer and Seagate, **DKRZ** will create an API for semantic-aware data access to allow complex data structures to be described and enable storage to exploit this information. The API will build programming extensions for workflows and optimise task and data placement using this information.
Fraunhofer will develop and evaluate a new implementation for the Portals specification within GPI. Fraunhofer will optimise the runtime and data access methods, GPI-Space and GPI, to be able to make decisions, based on the programming model extensions, on where data needs to be moved, how they can be efficiently accessed, as well as where computation should be performed.
UKOELN will supply requirements of the considered genetic use-cases and provide performance data as feedback to the ITHACA partners supporting new or emerging standards that extend the reach and capability of tools related to the programming models of WP3. All requirements of the genomics application will be collected. These include requirements regarding fault tolerance and parallelism, that ITHACA's have to deliver. This work product will form the basis for designing and implementing resilience techniques in WP6.
DKRZ will develop a data model that allows scientists to use scientific metadata and lead the task that concludes this WP with a comparison of the newly provided programming models.
Finally, all partners will work to provide benchmarks demonstrating the API's features.

For WP4: **Seagate** will lead this effort and provide functionalities of the advanced object store (Mero) as required by ITHACA, and provide the required storage system hardware components. Seagate will also participate in the simulation and modelling effort.
BSC will contribute to co-design the data management middleware as well as the porting of dataClay (a BSC data platform) into the ITHACA stack. We will also contribute our expertise in data management tools to the object store, garbage collection issues, and distributed data containers.
Bull will concentrate its effort on several components of the ITHACA data management architecture such as the Data Manager and the FlaGAS layer, and will help specify and implement the other components.
In WP4, **UPV** will design and implement a software version of FlaGAS.
UCLM will lead development of a distributed storage simulator for the ITHACA architecture, in cooperation with Seagate and UPV.
CEA, with the help of Bull, will work on data resource allocation.
DKRZ will integrate the abstraction layer for data structures allowing APIs to describe the semantics of the objects and the storage in order to exploit this information.

For WP5: **Bull**'s hardware ASIC teams will work on the BXI2 NIC ASIC and Switch ASIC design and verification. The prototyping engineers will provide the FPGA prototype to test the hardware and software parts of BXI2. Other hardware teams (mechanical, electronical cards, etc.) will be involved in development of new BXI cards (at NIC and Switch sides). The Bull architects, assisted by several expert engineers, will be involved in the study of next generation hardware items (NIC and Switch) and software components (communication stack and Fabric Management tools). Another Bull effort will focus on the installation and integration of a Sequana platform hardware and software. Several engineers in short sequential missions will be involved in the entire project.
UPV and UCLM will collaborate with Bull mainly to design and develop the software simulators necessary to study the characterisation of the current interconnect, and they will also help to enhance BXI2 NIC and Switch design, as well as to define and evaluate the optimizations for the next generation.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Seagate will help in the definition of the Exascale Interconnect by providing inputs related to storage I/O.
CEA will extend PCOCC's capability of virtualising high performance networks to the BXI network. CEA will improve PCOCC's effectiveness as a validation tool so that it can be more widely used to validate development of the Exascale Interconnect.
Fraunhofer will optimise the initial implementation of GPI for performance on real BXI hardware with the help of GPI tests, benchmarks and mini-apps, possibly raising new software and hardware requirements for a next-generation interconnect.

For WP6: **CEA** will lead the work package and contribute to security activities around network technologies and their use in HPC and data analysis environments. CEA will extend its profiling tools with MPI-IO oriented metrics.
Seagate will provide the telemetry infrastructure and associated policy engine to optimise access to persistent storage. Seagate will also provide data resiliency functionalities for persistent data.
This extension will be compatible with Allinea's profiling tools. Indeed, **Allinea** will address extensions for scaling tools on the ITHACA system, support Fraunhofer's GPI work, and aim to interoperate with CEA's MPI-IO and the data provided by ARM giving the work a wider applicability to other platforms.
Bull will, with the help of UPV, work on its hardware and software instrumentation to provide monitoring and performance information at the interconnect level and new security features for developed data management components.
Fraunhofer will optimize the cache on the BeeGFS client. Fraunhofer will extend the client to handle data movement between the caches and the central storage transparently for the user. Additional resiliency features for BeeGFS will be implemented. Fraunhofer will extend BeeGFS to add new security features.
BSC will work on leveraging the resilience capabilities of its layer to add resiliency to dataClay. **ARM** will conduct research into methods to allow developers to optimise local CPU cache usage by their code to enable optimal data energy efficiency without those users needing to understand the underlying cache structures. ARM will work to extend its HPC software development tools by adding support for new systems and programming models developed under ITHACA, helping to provide users of its tools with practical access to the results of this project.
UKOELN will develop resilience techniques, specified in WP3, that will be supported and tested in WP7.
INRIA will integrate the monitoring capabilities developed in the project with in-situ processing.
DKRZ will contribute to data placement analysis, telemetry infrastructure and the data movement policy engine as these infrastructures are mandatory to improve data placement algorithms and workload scheduling provided by WP3.
UPV will work with other partners to define the interface between different components.

For WP7: **DKRZ** will lead this WP. DKRZ**, Fraunhofer, BSC, SURFsara, INRIA and UKOELN** provide individual application use cases and port specific parts to the ITHACA platform using the newly proposed programming models and data access tools.
Allinea will support this activity with performance analysis through its skills and by applying its tools, and **Seagate** with its expertise on the storage infrastructure.
SURFsara will lead the evaluation of advanced tools developed mainly in WP6.
All teams working on performance analysis and modelling the benefit of the new features, will then perform the adjustments via ported applications or via ported mini-apps. Based on the expected benefit, the code will be iteratively modified. Finally, all partners will evaluate the benefit on the ITHACA platform**.**

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Direct costs breakdown

The following table shows detailed information about the direct costs for Bull, reaching 17% of its personnel costs mainly due to the provision of the two ITHACA prototype platforms.

1 / Bull	Cost (€)	Justification
Travel	20 000	Trips
Equipment	720 000	A Sequana platform of compute nodes that will integrate the first interconnect generations and associated software during the project.
	200 000	A FPGA Platform prototype to validate the BXI2
Other goods and services	10 000	- Management board logistics and organization. - project-community Web site creation and hosting - Accounting audit
Total	**950 000**	

The following table shows detailed information about the direct costs for Seagate, reaching 23% of its personnel costs mainly due to the subcontracting effort and the provision of equipment.

2 / Seagate	Cost (€)	Justification
Travel	31 000	This includes travel for general assembly meetings, conferences, workshops, events appropriate to the project and any one-one face2face meetings as demanded by the project.
Equipment	155 000	A tiered storage system will be made available as part of the development & test environment, and also to be part of the storage pools for ITHACA use cases.
Other goods and services	205 563	The budget for subcontracting (French SME: EURL Tweag) is 186 563 €. Seagate will use subcontracting resources based in France employed through an agency mainly for development work task 4.2 in WP4.This also includes €19,000 for financial audits, dissemination materials and IP costs.
Total	**391 563**	

No other partner's direct costs exceed 15% of their personnel costs.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Glossary:

- **ASIC:** An application-specific integrated circuit (ASIC), is an integrated circuit (IC) customized for a particular use, rather than intended for general-purpose use.
- **CFD:** Computational fluid dynamics (CFD) is a branch of fluid mechanics that uses numerical analysis and algorithms to solve and analyze problems that involve fluid flows.
- **COMPs:** COMP Superscalar (COMPSs) is a programming model which aims to ease the development of applications for distributed infrastructures, such as Clusters, Grids and Clouds. COMP superscalar also features a runtime system that exploits the inherent parallelism of applications at execution time.
- **CPU:** A central processing unit (CPU) is the electronic circuitry within a computer that carries out the instructions of a computer program by performing the basic arithmetic, logical, control and input/output (I/O) operations specified by the instructions.
- **dataClay:** a platform that manages Self-Contained Objects (data and code)
- **FPGA:** A field-programmable gate array is an integrated circuit designed to be configured by a customer or a designer after manufacturing.
- **GPI**: Global Address Space Programming Interface is an API for the development of scalable, asynchronous and fault tolerant parallel applications
- **HDF5:** HDF5 is a data model, library, and file format for storing and managing data. It is designed for flexible and efficient I/O and for high volume and complex data.
- **HSM:** Hierarchical storage management is a data storage technique, which automatically moves data between high-cost and low-cost storage media.
- **ISER**: The iSCSI Extensions for RDMA is a computer network protocol that extends the Internet Small Computer System Interface (iSCSI) protocol to use Remote Direct Memory Access (RDMA).
- **IOPS**: IOPS (Input/Output Operations Per Second, pronounced eye-ops) is a performance measurement used to characterize computer storage devices.
- **MPI**: Message Passing Interface (MPI) is a standardized and portable message-passing system designed by a group of researchers from academia and industry to function on a wide variety of parallel computers.
- **NIC**: A network interface controller is a computer hardware component that connects a computer to a computer network.
- **NVM:** Non-Volatile Memory**.**
- **OmpSs:** OmpSs is a shared-memory parallel programming model based on compiler directives for C/C++ and Fortran.
- **PGAS:** In computer science, a partitioned global address space (PGAS) is a parallel programming model. It assumes a global memory address space that is logically partitioned and a portion of it is local to each process or thread.
- **RDMA:** In computing, remote direct memory access (RDMA) is a direct memory access from the memory of one computer into that of another without involving either one's operating system.
- **SerDes:** A Serializer/Deserializer (SerDes) is a pair of functional blocks commonly used in high speed communications to compensate for limited input/output. These blocks convert data between serial data and parallel interfaces in each direction.
- **SHMEM**: (from Symmetric Hierarchical Memory access) - family of parallel programming libraries, initially providing remote memory access for big shared memory supercomputers using one-sided communications.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

References for section 1:

BXI:

- Ron Brightwell, "**Portals 4: Enabling Application/Architecture Co-Design for High-Performance Interconnects**"
- S. Derradji, T. Palfer-Sollier, J.-P. Panziera, A. Poudes and F. Wellen-reiter, "**The BXI interconnect architecture**", High-Performance Interconnects (HOTI) 2015 IEEE 23th Annual Symposium on.
- Vignéras and Jean-Noël, "**Fault-Tolerant Routing for Exascale Supercomputer: The BXI Routing Architecture**", Cluster Computing (CLUSTER) 2015 IEEE International Conference on, pp. 793-800, 2015.
- J. Duato, S. Yalamanchili and L. Ni, **Interconnection Networks: An Engineering Approach**, 2002, Morgan Kaufmann Publishers Inc.

HPC nd BigData Uses Cases:

ICON:

- [1] Gassmann, A., and H.-J. Herzog, Q. J. R. Meteorol. S., 2008. : Towards a consistent numerical compressible non-hydrostatic model using generalized Hamiltonian tools.
- [2] Gassmann, A., J. Comp. Phys., 2011 : Inspection of hexagonal and triangular C-grid discretizations of the shallow water equations.

EC-Earth:

- Hazeleger, W. and Coauthors, 2010: EC-Earth: A seamless Earth-system prediction approach in action. **Bull. Amer. Meteor. Soc.**, 91, 1357–1363, doi: 10.1175/2010BAMS2877.1

Deep Learning:

- [3] Keuper, Janis; Pfreundt, Franz-Josef, Asynchronous Parallel Stochastic Gradient Descent: A Numeric Core for Scalable Distributed Machine Learning Algorithms,Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments; doi.acm.org/10.1145/2834892.2834893, ISBN 978-1-4503-4006-9. Publisher: ACM.
- [4] Iandola, Forrest N., et al. "FireCaffe: near-linear acceleration of deep neural network training on compute clusters." arXiv preprint arXiv:1511.00175 (2015).

Genomic:

- [5] Nieroda L., Peifer M., Achter V., Velder J., Lang U.: Application of iRODS metadata management for cancer genome analysis workflow. iRODS UGM 2016 June 8-9, 2016, Chapel Hill, NC.
- [6] Kawalia, A.; Motameny, S.; Wonczak, S.; Thiele, H.; Nieroda, L.; Jabbari, K.; Borowski, S.; Sinha, V.; Gunia, W.; Lang, U.; Achter, V.; Nürnberg, P.: Leveraging the Power of High Performance Computing for Next Generation Sequencing Data Analysis: Tricks and Twists from a High Throughput Exome Workflow. In: PLoS ONE 10 (2015), 5. http://dx.doi.org/10.1371/journal.pone.0126321. – DOI 10.1371/journal .pone.0126321
- [7] Peifer M., Hertwig F., Roels F., Dreidax D., Gartlgruber M., Menon R., Krämer A., …, Achter V., Lang U., Peifer M., et al.: Telomerase activation by genomic rearrangements in high-risk neuroblastoma. Nature 526:700-704, 2015
- [8] George J., Lim J.S., Jang S.J., Cun Y., Ozretic L., Kong G., Leenders F., …, Peifer M, Achter V., Lang U., et al.: Comprehensive genomic profiles of small cell lung cancer. Nature 524:47-53, 2015.

Molecular Dynamics:

- Matthieu Dreher and Bruno Raffin. A Flexible Framework for Asynchronous In Situ and In Transit Ana- lytics for Scientific Simulations. In 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Chicago, United States, May 2014. IEEE Computer Science Press.
- Matthieu Dorier, Matthieu Dreher, Tom Peterka, Gabriel Antoniu, Bruno Raffin, and Justin M. Wozniak. Lessons Learned from Building In Situ Coupling Frameworks. In First Workshop on In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization, Austin, United States, November 2015.

AFiDs:

- van der Poel et al (2015), http://dx.doi.org/10.1016/j.compfluid.2015.04.007

URL: https://www.afid.eu/

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Section 4: Members of the consortium

4.1. Participants (applicants)

1. Bull SAS:

Short Name: **Bull**

Description of the Organisation:

Bull (Bull SAS) is the newest member of the Atos family. Atos SE is a leader in digital services with pro forma annual revenue of € 12 billion and 100,000 employees in 72 countries. Serving a global client base, the Group provides Consulting & Systems Integration services, Managed Services & BPO, Cloud operations, Big Data & Cyber-security solutions, as well as transactional services. With its deep technology expertise and industry knowledge, the Group works with clients across different business sectors: Defense, Financial Services, Health, Manufacturing, Media, Utilities, Public sector, Retail, Telecommunications, and Transportation. With 80+ years of technology innovation expertise, the new « Big Data & Cyber-security » Service Line (ATOS BDS) gathers the expertise in Big Data, Security and Critical Systems brought by Bull acquisition and the ones already developed by Atos in this domain.

The Service Line is structured into 3 complementary activities: Big Data, Cyber-security and Critical Systems.

- Big Data: the expertise of extreme performance to unleash the value of data (detailed below)
- Cyber-security: the expertise of extreme security for business trust
- Critical Systems: the expertise of extreme safety for mission-critical activities.

In recent years, the Bull R&D labs have developed many major products that are recognized for their originality and quality. These include the Sequana supercomputer which concretizes the first results of the “Bull Exascale Program” announced during SuperComputing 2014, bullion servers for the private Clouds and Big Data, the Shadow intelligent jamming system designed to counter RCIEDs, the libertp tool for modernization of legacy applications and hoox, the first European smartphone featuring native security. To explore new areas and develop tomorrow's solutions, today, Bull R&D is investing heavily in customers – with whom it has forged many successful technological partnerships – as well as in institutional collaborative programs (such as competitiveness clusters and European projects) and in partnerships with industry (Open Source, consortiums). Bull is involved with the strategy toward HPC in Europe with the active leadership of ETP4HPC and contribution to the Strategic Research Agenda.

Role and responsibilities:

On the long-term, the objective of Bull is to create, with the H2020 ITHACA project, market differentiators. The main areas are the efficient support of an innovative and performant Data Management framework based on its own interconnect technology, well adapted to HPC market needs. With these differentiators, Bull aims at developing its market share in the HPC market worldwide and especially in Europe. In the data management and data infrastructure segments, HPC solution is a key element; hence the H2020 ITHACA project is essential for Bull.
Bull initiated and leads this H2020 project.
Bull will participate to the Data Management infrastructure specification and development. Moreover, Bull will provide to all partners an access to the first generation of interconnect (BXI) in hosting a complete HPC platform containing this innovative technology, as proposed in WP5. In parallel, Bull will focus its studies on several domains. The first one concerns a set of API and services which will contribute to the Data Management framework via the WP3 and WP4. The second will start with the hardware and software implementations of the next generation of the interconnect inside the WP5, in order to continue to integrate this new interconnect technology in several HPC open sources initiatives

(Portals, MPI, Lustre, …) as already initiated with the previous generation. And for the last one but not the least step, Bull will integrate the co-design results and feedback of the studies not only of the WP5 but also of the WP3, WP4, WP6 and WP7, into a specification of a future generation of Bull eXascale Interconnect which will be able to offload also specific Data Management features.

Bull leads the WP1 and WP5.

Key Personnel:

Bruno Farcy (male): R&D Project Director at Bull in charge with the new interconnect technology program: BXI, since 2014. He joined Bull in 1998. He worked at first as network low layers software developer. In 2002, he joined the Server Design & Development entity as technical leader and architect of system management and reporting tools for HPC and big data systems. For 10 years, he has managed several cooperative french HPC or big data projects (FUI, PIA,…) and lead work packages for European ones.

Sylvie Lesmanne, (female) : Senior architect at Bull, Sylvie Lesmanne has been working first on processor design and then on several types of interconnection ASICs, some interconnecting processors and focusing on cache consistency and others interconnecting HPC nodes and focusing on latency, bandwidth and resiliency. She worked both in architecture definition and in ASIC design and verification. Since 2009, she is at the head of the hardware architecture team at Bull, which specified recently the open exascale supercomputer, code-named SEQUANA, and studies processor options to optimize performance, cost and power efficiency.

Matthieu Pérotin, (male): is a HPC architect at Bull. He has been working for 10 years in the HPC field, starting with a Ph.D. about process scheduling (received in 2008) and has joined Bull in 2009 as a software developer. Since joining Bull, he has worked on cluster management tools for petaflopic scale clusters and more specifically on on-line event management and processing. He has joined the BXI project in 2011, to define the architecture of the fabric management software suite, which includes fabric monitoring and routing. He also provides expertise on HPC fabric topologies.

Pascale Rosse-Laurent, (female): Senior Architect in HPC Competence Centre is in charge of advanced technology and software analysis. Until 2010, she was in charge of Software architecture of Bull HPC offer and of the Petaflop CEA-BULL research program during 2008-2010. She was also technical leader of Bull in the joint BULL-IBM software program on RS6000 SP2 platform. She obtained her PhD in Geology and a Master of Artificial Intelligence (CERICS) before joining BULL.

Anne-Marie Fourel (female): a Bull staff member since 1984, Anne-Marie Fourel has been working in various technical areas, mainly communications stacks and storage as developer then as manager of development teams. Previously product manager for backup software and tape libraries, she joined the HPC R&D department in 2011 where she manages the software part of the BXI project with a team of architects and engineers in charge of the integration of the BXI interconnect in the HPC middleware (MPI, PGAS, File systems) and of the development of the fabric management solution for BXI. She is also involved as project manager in a R&D collaboration program between Bull and the CEA.

Ben Bratu, (male): HPC Technical Manager at Bull, Ben Bratu has been working on software solutions for infrastructure and platform management for 12 years. After a PhD and first experiences in data mining and data management solutions for telecom platforms, he has joined Bull in 2011. Since joining Bull, he has managed the BXI software team responsible for the design and implementation of the fabric management solution for BXI and from 2015 he is the Interconnect Technical Manager in charge of the network stack and fabric management software suits for BXI and IB interconnects.

Philippe Couvée (male) : Senior software architect at Bull, Philippe Couvée has been working first as an AIX and LINUX kernel and driver developer and then, as a Storage architect. He joined the High Performance Computing R&D in 2004 where he was in charge of HPC middleware development

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

(MPI, distributed file systems), then as chief architect of the SuperComputer software products. Since early 2016, he is focusing on Data Management for exascale.

François Wellenreiter (male): is a HPC architect at Bull. He has been working for 12 years in the HPC field. Starting his career in wide area of software development (industrial and embedded systems, linux kernel, application development and network stacks). He has joined Bull in 2004 as a software developer. Since then he has worked on linux kernel drivers and network stacks, performance tools, the OpenMPI library, Lustre network drivers for tera- and peta-flopic scale clusters. He has joined the BXI project in 2011, to develop libraries and network stacks needed by the BXI software. Since the beginning of 2015, he is the software architect of the BXI dedicated software suite, which includes the BXI driver, Portals4, the IP stack, the Lustre and Plan9 network drivers, Sandia OpenShmem, UPC and various PGAS related implementations.

Publications and products:

As an industrial, Bull rarely submits publication in academic journals or to conferences. However, in association with academics organization Bull participated to some publications, and Bull is present publicly with innovative technologies presentations at HPC major events such as ISC or SuperComputing.

As an expert in delivering ultra-high performance, Bull is one of the world leaders in Extreme Computing. As an IT manufacturer, Bull has a strong presence in the list of the world's top supercomputers. With more HPC specialists than any other player in Europe, Bull is recognized for the technological excellence of its bullx range, its HPC applications expertise and its ability to manage large-scale projects. Bull has developed a complete vertical HPC offer containing Hardware and Software technologies dedicated to scalable HPC systems:

- **Bull Sequana**: The open exascale-class supercomputer.
  With the new Bull sequana range of supercomputers, Atos confirms its strategic commitment to the development of innovative high performance computing systems, the systems needed to meet the major challenges of the 21st century. Designed by the Bull R&D in close cooperation with major customers, the Bull sequana X1000 supercomputer leverages the latest technological advances, so as to guarantee maximum performance for a minimized operation cost.

- **bullx S6000 series**: This scalable server has been specially developed to meet today's challenges of mass data processing, high availability, virtualization and eco-efficiency. The bullx S6130, the first available model of the bullx 6000 series, is fully scalable up to 16 CPUs/240 cores/24 TB. It supports full memory protection and features hot-swappable memory and I/O capabilities. The bullx S6130 is the HPC version of the bullion S16 enterprise server, currently recognized as the fastest x86 server on the market (SPECint_rate2006 for the 16 socket configuration).

- **bullx supercomputer suite (scs)** is a comprehensive, powerful and robust software solution that meets the requirements of even the most challenging high performance computing needs. It is the result of Bull's long experience in deploying HPC software to the strictest specifications and major investments and continued efforts in R&D

Relevant Projects or Activities:

Bull is involved with the strategy toward HPC in Europe with the active leadership of **ETP4HPC** and contribution to the Strategic Research Agenda.
Bull participated to the following cooperative projects connected to the subject of this proposal:

- **Mont-Blanc 3 :**
  The main target of the H2020 Mont-Blanc 3 project is the creation of a new high-end HPC platform that is able to deliver a new level of performance / energy ratio whilst executing real applications. The project builds upon the previous Mont-Blanc & Mont-Blanc 2 FP7 projects, with ARM, BSC & Bull, it adopts a co-design approach to ensure that hardware and system

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

innovations are readily translated into benefits for HPC applications. This encompasses the three following objectives:

- To design a well-balanced architecture and to deliver the design for an ARM based SoC (System-on-a-Chip) or SoP (System-on-Package) capable of providing pre-exascale performance when implemented in the 2019-2020 time frame.
- To maximise the benefit for HPC applications with new high-performance ARM processors and throughput-oriented compute accelerators designed to work together.
- To develop the necessary software ecosystem for the future SoC. This additional objective is key to ensure that the project successfully translates into an industrial offer for the HPC market.

- **SAGE:** The SAGE project is working on building an Extreme scale data centric computing capable storage system. The project is building a storage system with in-storage compute capability, consisting of multiple tiers of storage devices. The use cases consist of extreme scale data generators (eg: scientific experiments) as well as extreme scale I/O HPC use cases.
- **PerfCloud :** It was a French FSN project. The objective of this Perfcloud activity is to develop advanced technologies for the building of a new generation of large HPC systems that can be used in cloud infrastructures and to demonstrate the usefulness of these technologies on some applications. The main goals are to increase computation performance (decrease the time to solution) and to get a more energy-efficient system (less energy for the same amount of computation and less energy to operate the system). Besides testing with classical benchmarks, the prototypes will be validated with two large applications: atmospheric dispersion and weather simulation and image retrieval in large data bases of images and videos.
- **ADAMme :** It is a French ”Projet Investissement d’Avenir” project. The aim of the ADAMme project is to develop a new server hosting 8 processors and up to 24 Terabytes of memory. This server aims at applying "in-memory computing". The project will rely on this quantitative advance to drive a qualitative breakthrough on Business Intelligence and Service Management applications, in the Industrial and scientific areas.
- **Mont-Blanc 2 :** Mont-Blanc 2 is a FP7 project, which contributes to the development of extreme scale energy-efficient platforms, with potential for exascale computing, addressing the challenges of massive parallelism, heterogeneous computing, and resiliency. It will enable further development of the OmpSs parallel programming model to automatically exploit multiple cluster nodes, transparent application check pointing for fault tolerance, support for ARMv8 64-bit processors, and the initial design of the Mont-Blanc exascale architecture.

Bull is also involved in other cooperative projects described by other partners (PRACE, DATASCALE, …)

Infrastructure description:

Platform NOVA: Bull makes available for collaborative R&D projects, a cluster architecture hosted in Bull France facilities providing a small datacenter for Cloud/Big Data as well as HPC/Simulation projects. This heterogeneous platform includes a mix of standard x86 servers, large In-Memory Servers (Bullion 4/8TB up to 16 sockets), HPC servers (bullx B500, 505, 510, 515) with or without accelerators (NVIDIA GPU or INTEL XeonPhi) as well as different types of storage systems. This platform that can be tuned, is providing cloud environments (various OpenStack, VMWare, …) or an HPC environment running an enhanced operating system, development tools and middleware software for executing compute-intensive applications. Within this platform partners can bring their data and then install, test and validate their developments throughout the project whether they are technology providers or applications developers. The platform is located in a DMZ zone so as to allow all projects partners to have access to the platform resources and to their own private dedicated network zone and space. In addition and specifically to ITHACA project, Bull will propose to R&D Cooperative projects a double rack Sequana packaging containing n Compute nodes (n/3 blades of 3 compute nodes), that means the use of 2n CPU sockets (Intel Broadwell family) and the Interconnect is composed of 2 level switch) for application tests and experiments. Then, this configuration will be enriched with new generation of Interconnect and new ARM CPU.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

2. Seagate Systems UK Ltd:

SEAGATE

Short Name: **Seagate**

Description of the Organisation:

Seagate is the world's leading provider of Data Storage devices, equipment and services.
The organisation is a worldwide multi-national registered in Ireland (Seagate Technology plc) with more than 50,000 employees.
Seagate operates two primary divisions within its corporate operations, Seagate Technology develops and produces data storage devices including disk drives, solid state drives and solid state storage for integration within servers, a large facility is located in Northern Ireland. Seagates' Cloud Systems and Silicon Group provides enterprise data storage solutions and core silicon technologies. Key within its portfolio are storage systems targeted at the High Performance Computing marketplace. The division of the organisation responsible for this project is Seagate Systems UK Ltd which is part of the CSSG organisation.
The HPC products are part of the ClusterStor product range. These are fully engineered data storage systems with all hardware, file systems software and system management provided. Systems are provided through our OEM or business partnerships including with ATOS, which features a ~60PB installation at DKRZ, Germany, and major installations at CEA in France. ClusterStor systems in general support some of the worlds most powerful supercomputers including NCSA Bluewaters where the storage system achieved the worlds first 1TByte / second performance.
Skills in all technology disciplines are needed to create these diverse range of products from molecular scientists to systems software and Seagate Systems will draw upon the knowledge and skills of the whole Seagate organisation to ensure the success of the project bringing knowledge on fundamental storage devices and techniques alongside methods to harness the huge capacity of systems for the HPC as well as accelerating their performance to future needs.
Within Seagate Systems (UK) the Emerging Technology Group manages collaborative research activities within Europe and will work in concert with development engineering groups based in UK. The group has experience with H2020, FP7 and UK National collaborative projects.
Current projects include:

- SAGE - Percipient storage for Extreme scale era of which Seagate is project coordinator and technical lead
- EsiWace - Climate and Weather centre of excellence where Seagate is working to optimise access to the vast databases of these communities
- BigStorage - a European Training program (ETN) of convergence of HPC storage and Data Science.
- EXDCI - European Extreme Data and Computing Initiative

The skills of the team and the wider Seagate Systems organisation have been harnessed to create a next generation object storage platform which will be used as the base storage software platform for ITHACA. The platform has been a 'ground up' development explicitly considering, from its outset, the needs of Extreme scale storage for the exascale era.

Role and responsibilities:

The base Object storage software platform that will be used in the ITHACA proposal is well developed and commercially available in seagate products (at the time of writing). It however lacks the advanced features that are required to meet the needs of Exascale systems and there is a lack of data and system management control and direction to effectively utilise multi-tiered storage systems incl.NVM. Developing further advanced capabilities of the object store for the new interconnects and global data addressing paradigms is the key role for Seagate, apart from providing the key elements of the storage hardware platform.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Seagate will provide resources across many technical activities and will provide elements of the software and system hardware required to effectively demonstrate advanced I/O and storage capabilities for ITHACA.
Seagate will be part of all the work packages with most the effort and the lead in **Data Access Middleware (WP4).**

Key Personnel:

Sai Narasimhamurthy PhD (male) is currently Staff Engineer, Seagate Research (formerly Lead Researcher, Emerging Tech, Xyratex) working on Research and Development for next generation storage systems (2010-). He has also actively led and contributed to many European led HPC and Cloud research initiatives on behalf of Xyratex (2010-) currently coordinating and providing technical leadership for SAGE. Previously(2005 - 2009) , Sai was CTO and Co-founder at 4Blox, inc, a venture capital backed storage infrastructure software company in California addressing IP SAN(Storage Area Network) performance issues as a software only solution. During the course of his doctoral dissertation at Arizona State University (2001-2005), Sai has worked on IP SAN protocol issues from the early days of iSCSI(2001). Sai also worked with Intel R&D and was a contributing participant in the first stages of the RDMA consortium (put together by IBM, Cisco and Intel) for IP Storage and 10GbE (2002). Earlier in his career, Sai worked as Systems Engineer with Nortel Networks through Wipro, India focussing on Broadband Networking solutions (2000-2001).

Malcolm Muggeridge (male) is Senior Director, Engineering; responsible for collaborative research at Seagate Systems UK. He joined Seagate through its acquisition of Xyratex in 2014 and was with Xyratex at its creation as a management buyout from IBM in 1994.
Malcolm has more than 37 years experience through his employment with IBM and Xyratex in the Technology, manufacturing, quality and reliability of Disk drives and Networked data storage systems and in recent years in HPC data storage, architecting and managing designs and new technologies across many products.
More recently he has been focused on Strategic Innovation and Business development,
Research and Technology. He is a steering board member of the ETP4HPC defining research objectives for future within Euorpe and is active in the Partnership board of the cPPP on HPC. He is a member of the UK eInfrastructure board with Special interest in HPC.
Malcolm has a B.Eng degree in Electronics from Liverpool University.

Nikita Danilov PhD (male) is a Consultant Software Architect at Seagate. His work on storage started in 2001, when he joined Namesys to develop the reiserfs file system for Linux. Since 2004 he worked on Lustre in ClusterFS, later acquired by Sun. Leaving in 2009 to join a start-up company, Clusterstor to design and implement an Exascale storage system, this technology was acquired by Xyratex and forms the basis of the core SAGE system. He received a PhD in mathematical cybernetics from Moscow Institute of Physics and Technology.

Publications and Products:

As a commercial organisation Seagate does not generally submit material for publication in academic journals or to conferences however they do present publicly on selected technical aspects of the systems and solutions with presentations at major events such as SuperComputing, ISC and other events such as Lustre developer conferences LAD and LUG.
Seagate is solution provider for HPC storage systems, thanks to its offer "**Enterprise Storage Systems**" (http://www.seagate.com/products/enterprise-servers-storage/enterprise-storage-systems/).

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Relevant Projects or Activities:

Seagate is involved in the H2020 through the **SAGE** project (http://www.sagestorage.eu) as co-ordinator and lead.
The SAGE project has introduced the base Object storage platform and the base storage API and a repertoire of ecosystem components (Hierarchical Storage Management, Programming models, etc.) that works on top of the API. The Object storage platform features developed was early stage. ITHACA will co-design, identify and further develop new features and capabilities on top of the base Object storage platform that was introduced in SAGE, for the new interconnect and the new programming models and use cases. The API will also be extended to suit the needs of the programming models and global addressing infrastructure. There is no direct dependency between SAGE and ITHACA.

Seagate is involved also in H2020 Big Storage (http://bigstorage-project.eu/) and EsiWACE(https://www.esiwace.eu/). Seagate is developing a next generation multi-tiered active storage system in SAGE, assessing the storage needs of SKA (https://www.skatelescope.org/sdp/) in Big Storage which is a MCITN (http://www.cipris.eu) and providing object storage interfacing solutions for the Climate and Weather community in EsiWACE which is a CoE(https://ec.europa.eu/programmes/horizon2020/en/news/eight-new-centres-excellence-computing-applications) . Seagate is also actively involved in the ETP4HPC activities that helps to define the direction of European HPC research and is on the board(http://www.etp4hpc.eu/) .

Seagate Systems has been involved in a number of FP7 projects including IRMOS (http://www.irmos.eu); creating Quality of Service capability for storage in 'real time' cloud systems and currently is a member of the DEEP-ER(http://www.deep-er.eu) project particular focused on improved IO guidance mechanisms.
The organisation also has involvement in a number of Research projects in the area of Optical Interconnects including PHOXTROT (http://phoxtrot.eu).

Seagate Systems has also involved in UK funded research initiatives (eg : http://www.avatar-m.org.uk/Avatarm/) as Xyratex.

Infrastructure description:

Seagate will provide its enclosures with Seagate Disk Drives/SSDs and seagate flash cards to help build the storage platform.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

3. *Commissariat à l'énergie atomique et aux énergies alternatives:*

Short Name: **CEA**

Description of the Organisation:

CEA as a whole is a major player throughout the value chain of HPC from R&D in the development of silicon technology, architecture of processors, system integration, software environments and tools through to use of numerical simulation in many different areas related to the missions of CEA. It has more than 16 000 staff members in the development of low carbon energies, technologies for health, information technology, defence and global security, and underlying fundamental research for all these objectives.

CEA also owns and operates two world-class computing infrastructures (TERA and TGCC), and deploys related HPC services, for the benefit of national and European research, defense, and industry, Industrial access to HPC has been deployed for 10 years at CEA CCRT, through an original partnership business model and a dedicated supercomputer of nearly 0.5 petaflop/s. TGCC hosts the PRACE French Tier-0 system CURIE funded by GENCI.

CEA DIF DSSI is the division in charge of HPC at CEA, located in the Paris Region. DSSI operates the aforementioned computing centres and leads R&D in hardware and software technologies for HPC systems. (See htpp://www-hpc.cea.fr)

Role and responsibilities:

CEA is involved in WP3, WP4, WP5 and WP6. Besides leading WP6, the main contribution of CEA will be through development of MPI-IO in MPC and development of Phobos. In WP3, CEA/DAM will focus on MPI-IO. In WP4, CEA/DAM will use its knowledge on object store to implement Phobos. In WP5, CEA/DAM will focus on virtualization, this WP will have a strong interaction with WP4 . Finally in WP6 CEA/DAM will leverage his 20 years' experience in security HPC datacenter to improve data movement security.

Key Personnel:

Elodie Ardoin, CEA/DIF (female): Mrs. Ardoin joined CEA Seismic research division in 2002 as electronic and signal processing engineer where she worked 7 years on embedded real-time data acquisition, transmission and analysis. She joined the HPC division in 2009 where she worked on high-performance networks. She participated on CCRT and TGCC storage network designs and operations. She is now focusing on cluster interconnect networks.

Jacques-Charles Lafoucrière (male) is a Department leader and International Expert at CEA. He received his Engineering degree Ecole Centrale Paris in 1989 and joined CEA teams. He initially worked as a storage administrator and a system developer. Since 2005 he is a Lustre developer. His area of expertise covers parallel file systems, storage technologies, supercomputer and datacentre architectures. He has an Engineering degree from Ecole Centrale Paris

Thomas Leibovici (male) is a storage expert and a system developer at CEA. He received his Engineering degree ENSIMAG in 2003. He is a system developer and lead the Robinhood and Phobos development community since it's creation. His areas of expertise are artificial intelligence, storage systems and distributed file systems. He is also a Lustre developer.

Marc Pérache (male) is a research engineer at CEA. He has been working in the HPC field since 2003, starting with a Ph.D. about runtime systems for cluster of NUMA nodes (received in 2006). During his Ph.D., he started the MPC (MultiProcessor Computing) framework. He joined CEA in

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

2006 as a research engineer to extend the MPC framework and provide building blocks for high performance multithreaded applications. Since 2013, he lead an R&D team dealing with runtime systems and tools for HPC systems (MPI, OpenMP, memory allocation, high speed networks, scalable profiling tools …). He received the "Habilitation à Diriger des Recherches" (French degree which accredits to supervise researches) from Versailles Saint Quentin-en-Yvelines University in 2015. Also, he has guided 8 doctoral theses and is co-author of more than 20 articles in conferences and journals.

Francois Diakhate (male): After a Ph.D, he joined CEA in 2011 and is in charge of designs and operations of TGCC HPC clusters. Besides this, he is involved since 2014 in HPC R&D and took the lead of the project "Virtualization in HPC" where he developed PCOCC tool (Private Cloud on a Compute Cluster).

Publications and products:

- Marc Pérache, Hervé Jourdren et Raymond Namyst : MPC : A unified parallel runtime for clusters of NUMA machines. In Emilio Luque, Tomàs Margalef et Domingo Benítez, Euro-Par 2008 – Parallel Processing, volume 5168 de Lecture Notes in Computer Science, pages 78–88. Springer Berlin Heidelberg, 2008.
- Marc Tchiboukdjian, Patrick Carribault et Marc Pérache : Hierarchical local storage : Exploiting flexible user-data sharing between MPI tasks. In Parallel Distributed Processing Symposium (IPDPS), 2012 IEEE 26th International, pages 366–377, May 2012.
- Marc Pérache, Patrick Carribault et Hervé Jourdren : MPC-MPI : An MPI implementation reducing the overall memory consumption. In Matti Ropo, Jan Westerholm et Jack Dongarra, Recent Advances in Parallel Virtual Machine and Message Passing Interface, volume 5759 de Lecture Notes in Computer Science, pages 94–103. Springer Berlin Heidelberg, 2009.
- Jérôme Clet-Ortega, Patrick Carribault et Marc Pérache : Evaluation of openmp task scheduling algorithms for large NUMA architectures. In Euro-Par 2014 – Parallel Processing. 2014.
- Aurèle Mahéo, Patrick Carribault, Marc Pérache et William Jalby : Optimizing collective operations in hybrid applications. In Proceedings of the 21st European MPI Users' Group Meeting, EuroMPI/ASIA '14, pages 121–122, New York, NY, USA, 2014. ACM.
- Jean-Baptiste Besnard, Allen D. Malony, Sameer Shende, Marc Pérache, Patrick Carribault, Julien Jaeger : An MPI Halo-Cell Implementation for Zero-Copy Abstraction. In proceedings of the 22nd European MPI Users' Group Meeting, EuroMPI '15, pages 1–9.
- J.-B. Besnard, Marc Pérache et William Jalby : Event streaming for online performance measurements reduction. In Parallel Processing (ICPP), 2013 42nd International Conference on, pages 985–994, Oct 2013.
- Patrick Carribault, Marc Pérache et Hervé Jourdren : Enabling low-overhead hybrid MPI/OpenMP parallelism with MPC. In Mitsuhisa Sato, Toshihiro Hanawa, MatthiasS. Müller, BarbaraM. Chapman et BronisR. Supinski, Beyond Loop Level Parallelism in OpenMP : Accelerators, Tasking and More, volume 6132 de Lecture Notes in Computer Science, pages 1–14. Springer Berlin Heidelberg, 2010.
- Sylvain Didelot, Patrick Carribault, Marc Pérache et William Jalby : Improving MPI communication overlap with collaborative polling. Computing, 96(4) : 263– 278, 2014. Editor: Springer-Verlag New York.

- **MPC** (MultiProcessor Computing) is a framework which provides a unified parallel runtime designed to improve the scalability and performances of applications running on clusters of (very) large multiprocessor/multicore NUMA nodes. It supports mixed-mode programming with POSIX Threads, Intel TBB, OpenMP 2.5 and MPI 1.3 standards. MPC is freely available under the CeCILL-C license (http://mpc.hpcframework.com).

- **PCOCC** (Private Cloud On a Compute Cluster, pronounced like "peacock") is an OpenSource tool which allows users of a HPC cluster to host their own clusters of VMs on compute nodes alongside regular jobs. Users can leverage this tool to fully customize their software

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

environments for development, testing or facilitating application deployment. Compute nodes remain managed by the batch scheduler as usual, since the clusters of VMs are seen as regular jobs. From the point of view of the batch scheduler, each VM is a task for which it allocates the requested CPUs and memory and the resource usage is billed to the user, just as for any other job. For each virtual cluster, pcocc instantiates private networks isolated from the host networks, creates temporary disk images from the selected templates (using Copy-on-Write) and instantiates the requested VMs.

Relevant Projects or Activities:

- **H4H/Perfcloud**: Perfcloud is an ITEA project, in which CEA is actively involved, aiming at demonstrating that HPC and cloud techniques can share a number of common technologies and benefit from each other. (http://www.h4h-itea2.org/)

- **Prace**: CEA is an active member of Prace since its implementation phase. CEA is actively supporting European scientists using the CURIE system and therefore has a strong knowledge of the requirements for a new generation of interconnect. (http://www.prace-ri.eu/)

- **Datascale**: Datascale is a French government funded project whose goal is to use HPC techniques for Big Data. CEA brings to this project its knowledge of running large facilities as well as its ability to process large quantities of seismic sensor data. In this project, the network is one of the essential assets that are well understood by CEA. (http://datascale.org/)

- **SAGE** : See above (Seagate)

Infrastructure description

Beside the aforementioned very large production systems, which can be accessed under certain conditions for large scale experimentation, CEA DIF has a dedicated experimental facility. This infrastructure is permanently updated with the latest emerging technologies for their assessment or collaborative research on HPC systems. This includes computing and storage systems of various sizes, including clusters of significant scale.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

4. University Polytechnic of Valencia:

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Short Name: **UPV**

Description of the Organisation:

Universitat Politècnica de València (www.upv.es) is a public university with three campus sites, over 35,000 students and 2,600 faculty members and research staff. It consists of 42 Departments, most of them in engineering areas, and is the top University in Spain regarding technology transfer and patent production.
Contributions to this project will come from the Parallel Architectures Group (GAP) (see http://www.gap.upv.es/). GAP has a 20-year research expertise in different aspects of system architecture, especially on interconnection networks. The general mission of GAP is to develop applied research considering the market trends, thus enabling technology transfer to industry. This proved effective with the development of fully adaptive routing algorithms for commercial supercomputers and microprocessors (IBM BlueGene/L, Cray T3E, Compaq Alpha 21364). GAP also developed very efficient routing algorithms for multistage networks, compact and versatile routers for networks on chip, and efficient techniques for congestion control. These techniques were incorporated into Sun Microsystems Magnum switch (3456 ports), used in the Ranger supercomputer, a NoC prototype at Intel Labs, and a joint patent with Xyratex on congestion control. In addition to solutions for interconnection networks, GAP also designed scalable shared-memory architectures, implementing a 64-node 1024-core cluster prototype with hardware support for a flat global address space. It was based on an extension of the HyperTransport protocol that was later standardized by the HyperTransport Consortium.

Role and responsibilities:

Currently, the group is formed by twenty-three researchers, eight of them being faculty members, and other fifteen members being students developing their PhD theses in our group. UPV will be mainly focused on WP5 and WP4. Our group has a large expertise on all the topics covered by WP5. The role of UPV is to provide the best combination of interconnect architectural solutions in order to achieve the highest stable performance and reliability with minimal power consumption. Our responsibility is to design, develop simulation models, evaluate, and fine tune such a combination of architectural solutions for the next-generation interconnection network. In WP4 UPV will collaborate to a software implementation of FlaGAS low level mechanisms to enable applications and system software to address cluster-wide DRAM and NVRAM, as well as byte-addressed storage devices, using a flat global address space (FlaGAS).

Key Personnel:

Dr. José Duato (male) is Professor in the Department of Computer Engineering (DISCA) at UPV. He published over 500 refereed papers, which received more than 12,500 citations (according to Google Scholar). His research results have been used in the design of the IBM BlueGene/L and Cray T3E supercomputers, the Alpha 21364 microprocessor, RECN (a scalable congestion management technique for Advanced Switching), and Sun Microsystem's 3456-port InfiniBand Magnum switch. Prof. Duato was the main contributor to the High Node Count HyperTransport Specification 1.0. He also led the development of rCUDA, which enables remote virtualized access to CUDA accelerators. Prof. Duato is the first author of the book "Interconnection Networks: An Engineering Approach". He served in the editorial boards of IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Computers, and IEEE Computer Architecture Letters. Prof. Duato was awarded the Julio Rey Pastor National Research Prize in 2009 and the “Rey Jaime I” Prize in 2006. He is a Correspondent Member of the Royal Academy of Sciences.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Dr. Pedro López (male) received the B.Eng. degree in electrical engineering and the M.S. and Ph.D. degrees in computer engineering from the Universitat Politècnica de València, València, Spain, in 1984, 1990, and 1995, respectively. Since 2002, he has been a Professor with the Department of Computer Engineering, Universitat Politècnica de València. His research interests include high performance interconnection networks for multiprocessor systems, clusters and networks on chip. He has published over 120 refereed conference and journal papers. He served in the editorial board of Parallel Computing journal for ten years.

Dr. Antonio Robles (male) received the MS degree in physics (electricity and electronics) from the Universitat de València, Spain, in 1984 and the PhD degree in computer engineering from the Universitat Politècnica de Valènncia in 1995. Since 2003, he is a full professor in the Department of Computer Engineering at the Universitat Politècnica de València. He has taught several courses on computer organization and architecture. His research interests include high-performance interconnection networks for multiprocessor systems, clusters, and networks on chip and scalable cache coherence protocols. He has published more than 80 refereed conference and journal papers. He has served on program committees for several major conferences.

Dr. María Engracia Gómez (female) received her PhD in 2000 in Computer Engineering and is an Associate Professor at UPV. Her current research activity is focused on interconnection networks, NoCs and cache coherence protocols. María Engracia Gómez has collaborated with different Institutions (SIMULA Labs, University of Ferrara, Universidad de Murcia) and companies (STM, Thales, ARM). She has acted as coordinator of the vIrtical project (7th Framework Programme) GA number 288574. Currently she is participating as researcher in the ExaNest project (H2020) GA number 671553.
She has co-invented different routing strategies, topologies and fault-tolerant strategies in the interconnection network and NoC fields. She is co-author of more than 60 articles in conferences and JCR journals, according to DBLP.

Publications and Products:

- J. Duato, S. Yalamanchili and L. M. Ni, **Interconnection Networks: An Engineering Approach**, Morgan Kaufmann Publishers, ISBN 1-55860-852-4, 2002.
- J. Duato, **A new theory of deadlock-free adaptive routing in wormhole networks**, IEEE Transactions on Parallel and Distributed Systems, Vol. 4, No. 12, pp. 1320-1331, 1993.
- J. Duato, I. Johnson, J. Flich, F. Naven, P. García, T. Nachiondo, **A new scalable and cost-effective congestion management strategy for lossless multistage interconnection networks**, 11th International Symposium on High-Performance Computer Architecture (HPCA-11), 2005.
- M. E. Gómez, N. A. Nordbotten, J. Flich, P. López, A. Robles, J. Duato, T. Skeie, O. Lysne, **A routing methodology for achieving fault tolerance in direct networks**, IEEE Transactions on Computers, Vol. 55, No. 4, pp. 400-415, 2006.
- M. Alonso, S. Coll, J.M. Martínez, V. Santonja, P. López, J. Duato: **Power saving in regular interconnection networks. Parallel Computing** 36(12): pp. 696-712, 2010.

Relevant Projects or Activities:

- **SARC**: Integrated project (IST-FET-ACA) from the 6th Framework Programme. Coordinator: Stamatis Vassiliadis. It is concerned with long term research in advanced computer architecture, and focuses on a systematic scalable approach to systems design ranging from small energy critical embedded systems right up to large scale networked data servers.

- **High-performance, Reliable Architectures for Data Centers and Internet Servers: Consolider – Ingenio 2010 project** (ref. CSD2006-00046). Coordinator: José Duato. This project aims at developing several techniques to improve the performance and reliability of current server architectures for data centers and Internet servers, for a given cost and power consumption budget.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

- **NaNoC**: STREP **project** from the 7th Framework Programme (ref. 248972). Coordinator: José Flich. It aims at developing a design platform for future Network-on-Chip (NoC) based multi-core systems. This design platform intends to master the design complexity of advanced microelectronic systems by enabling strict component oriented architectural design.

- **vIrtica**l: STREP project from the 7th Framework Programme (ref. 288574). Coordinator: Maria Engracia Gómez. It addresses embedded system virtualization with special emphasis in NoC design and support for virtualization.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

5. University of Castilla - La Mancha:

Short Name: **UCLM**

Description of the Organisation:

The University of Castilla-La Mancha (UCLM) is a regional university founded in October 1985. It has grown at high rate attracting qualified human resources and providing the people in the region with high educational skills. It has nowadays four Campuses with 37 buildings, integrating 36 academic departments offering 45 degree qualifications. More than two thousand researchers work at the UCLM. This number increases by achieving greater success at national and European levels from the point of view of scientific research, higher education, and technological development. The UCLM has two peculiar features: a) it is multi-disciplinary, undertaking work in practically all branches of knowledge; and b) its activities are very wide-ranging, embracing the spectrum from basic research to technological development.
Inside the UCLM, the researchers involved in this project belong to the Albacete Research Institute of Informatics (I3A). More specifically, they belong to the largest group among the six integrated in the I3A: the "High-Performance Architectures and Networks" group (RAAP). Some major current research effort of this group is devoted to the design and simulation of advanced strategies for high-performance interconnection networks and computer architectures, as well as other research lines such as the design of robust video communication systems over wireless links, parallel video compression systems, design and evaluation of high-capacity storage systems, and network planning. This expertise has been achieved through numerous national and international research projects, and through collaboration with important companies and research centres in Europe, such as Bull, Mellanox, Huawei, Telefónica, Thales, RENFE, INDRA, Eurocopter (EADS), MTP, ISIS, ACORDE, Telvent, Worldnet21, Boeing Research Center Madrid, Avaya Networks and OVERON Audiovisual Services. Through the development of these projects, we have also achieved a wide experience in the organization, co-ordination and dissemination of results of R&D projects both at national and international level, for instance through the HiPEAC European Network of Excellence, which the involved researchers belong to. The RAAP group counts with the required infrastructures for providing an excellent environment to conduct basic and applied research.

Role and responsibilities:

One of the main research targets of the RAAP group of UCLM has been always the improvement of the performance of the interconnection networks which are at the core of parallel systems. In that sense, during the last three decades this group has proposed strategies which overall have led to significant advances in the interconnects state-of-the-art. Based on this background, the group will contribute to the design of strategies to leverage and enhance the BXI components and the BXI-based systems targeted by ITHACA, focusing mainly on routing algorithms, queuing schemes, congestion-control techniques, quality-of-service provision strategies and power-management mechanisms. Besides, the RAAP group will provide its expertise as developer of interconnection-network simulators to lead the development of the simulation tools that will be used to evaluate the BXI-based networks designed in ITHACA. Moreover, UCLM will lead the development of a distributed storage simulator that will be used to aid in the design and evaluation of the Data Access Middleware defined in ITHACA. Also, UCLM expertise in interconnects control software will be used to enrich the BXI control software with new features in order to support the new programming models defined in ITHACA. Finally, the group will lead the dissemination activities of ITHACA based on the wide experience of its members as authors and reviewers of scientific publications, as well as organizers of conferences, workshops and summer schools.
UCLM leads the Work Package 2.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Key Personnel:

The five experts described below will participate actively to the ITHACA project:

Pedro J. García (male) is an Assistant Professor at UCLM. He has developed several strategies for high-performance interconnection networks, especially congestion management schemes and routing algorithms, which have led to tens of publications in top-ranked journals and conferences. He has served as Organizer Committee member in several international workshops such as WHTRA, HiPINEB and WOPSSS. He has been the coordinator of two research projects supported respectively by the Spanish Government and by the Government of Castilla-La Mancha, as well as the coordinator of two Research & Development Agreements. Besides, he has participated in other (more than 30) research projects, some of them supported by the European Commission.

Francisco J. Alfaro-Cortés (male) is since July 2007 a full-time associate professor at UCLM. He has published twenty papers in ranked technical journals and more than 50 papers in international peer-reviewed conferences. He has been a member of the program committees of 8 international conferences and workshops. He has reviewed papers for several ranked journals. He has advised 5 PhD students, and currently, he advises other 2 PhD students. He has participated in more than 30 research projects funded by the Regional Government, the Spanish Government or by the European Commission.

Francisco J. Quiles (male) is a Full Professor of Computer Architecture and Technology at UCLM. His research interests include: high-performance interconnection networks for multiprocessor systems and clusters, parallel algorithms for video compression and video transmission. He has served as Program Committee member in several conferences. He has published over 200 refereed papers, which received more than 1,000 citations (according to Google Scholar) and participated in 68 research projects supported by the NFS, European Commission, the Spanish Government and Research & Development Agreements with different companies. Also, he has guided 9 doctoral theses.

José L. Sánchez (male) is an Associate Professor at UCLM. He has focused his research on developing several techniques to improve some aspects of the high-speed interconnection networks. In particular, he has worked on switch architecture, network reconfiguration, quality of service and energy consumption. He has been the coordinator of 10 research projects, supported by national and regional Spanish Governments, besides participating in other research projects, some of them supported by the European Commission.

Jesus Escudero-Sahuquillo (male) is a PostDoc research assistant at UCLM. Along his career, he has worked for Oracle (Norway) as a Senior Software Engineer, and in the Technical University of Valencia, UPV, (Spain) as a Postdoc. He has participated in tens of research projects funded by the European Commission and the Spanish Government, and R&D agreements. He has served as reviewer in top-ranked journals, and he participates in the organization of international workshops (HiPINEB, WOPSSS). His research interests, which have led to more than 20 international publications, include high-performance computing and Big-Data, interconnection networks and all the strategies related to improve them, such as congestion management, routing algorithms, network topologies and power saving.

Publications and Products:

- **N-dimensional Twin Torus Topology,** Andújar, F.J., Villar, J.A., Sánchez, J.L., Alfaro, F.J., Duato, J., IEEE Transactions on Computers, 64(10), pp. 2847-2861. 2015
- **Efficient and Cost-Effective Hybrid Congestion Control for HPC Interconnection Networks**, Escudero, J., Gran, E.G., García, P.J., Flich, J., Skeie, T., Lysne, O., Quiles, F.J., Duato, J., IEEE Transactions on Parallel and Distributed Systems, 26(1), pp. 107-119. 2015
- **Building 3D Torus using Low-Profile Expansion Cards**, Andújar, F.J., Villar, J.A., Sánchez, J.L., Alfaro, F.J., Duato, J., IEEE Transactions on Computers. 63(11), pp. 2701-2715. 2014

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

- **A New Proposal to Deal with Congestion in InfiniBand-based Fat-Trees** Escudero, J., García, P.J., Quiles, F.J., Reinemo, S., Skeie, T., Lysne, O., Duato, J., Elsevier Journal of Parallel and Distributed Computing, 74(1), pp. 1802 – 1819. 2014
- **An Effective and Feasible Congestion Management Technique for High-Performance MINs with Tag-Based Distributed Routing**, Escudero, J., García, P.J., Quiles, F.J., Flich, J., Duato, J., IEEE Transactions on Parallel and Distributed Systems. 24 (10), pp. 1918 - 1929. 2013

Relevant Projects or Activities:

UCLM participated to cooperative projects connected to the subject of this proposal:

- **TecMASAS (Techniques to improve the architecture of servers, applications and services)** (2016-2018) is a research project supported by the Spanish Government and EU (under FEDER funds) which aims at developing several techniques to improve the performance and reliability of current high-performance servers, as well as reducing their cost and power consumption. Applications with a large impact on digital society and health services will be developed that will benefit from those architectural enhancements. Among other contributions, the RAAP group from UCLM participates by developing research on adaptive routing, congestion- and power-management techniques for off-chip interconnect architectures, on-chip optical networks, etc.

- **Scalable Hybrid Convergent Network Design and Simulation** (2015) was a Research & Development Agreement between the RAAP group of UCLM, GAP group of UPV and Huawei Technologies Co. Ltd. The main goal is twofold: to design and develop a convergence network simulator tool able to simulate networks interconnecting up to 100,000 end nodes, and to develop a NoC able to cope with 256- and 512-node systems implemented on an emulator-based prototype (FPGA-based), following current decisions for NoC design. The RAAP group from UCLM participated by developing the convergence network simulator.

- **MASSA (Enhancement of Server Architecture, Services and Applications)** (2013-2015) was a research project supported by the Spanish Government aiming at proposing architectural enhancements for cluster-based high-performance servers, as well as at improving the services offered by these servers. Among other contributions, the RAAP group from UCLM developed research on routing, congestion management, and QoS for interconnection networks, to increase their performance and reliability, while reducing power consumption.

- **Testing of Congestion Control techniques in a simulator of HPC** systems (2012) was a Research & Development Agreement between the RAAP group of UCLM and Simula Research Laboratory (Norway), focused on the modeling and simulation-based evaluation of congestion control techniques for HPC interconnects.

- **Enhancement of the Quality of Service Provided by the Internet Infrastructure** (2010-2013) was a research project supported by the Government of Castilla-La Mancha and developed entirely by the RAAP group of UCLM. The main objective was the improvement of the quality of service of the clusters which support high-performance Internet servers by using techniques that increase the performance of the interconnection network that is the core (and the bottleneck) of these systems.

- **COMCAS** (2009-2011) was a project from the 7th Framework Programme, EUREKA Programme: Catrene (CA501 label). The aim of the COMCAS project was to create a breakthrough in low power design solutions for new heterogeneous platforms, which include sophisticated on-chip communication infrastructures for higher efficiency and performances, crucial for fighting the complexity increase in embedded systems in the years to come. RAAP from UCLM participated in the tasks related to the specification, design and tools development of the on-chip network.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

6. *Deutsches KlimaRechenZentrum GMBH*

Short Name: **DKRZ**

Description of the Organisation:

DKRZ, the German Climate Computing Centre, is a national service provider which constitutes an outstanding research infrastructure for model-based simulations of global and regional climate and the investigation of the processes in the climate system. DKRZ's principal objectives are provision of adequate computer performance, data management, and service and support to use these tools efficiently. DKRZ operates one of the largest supercomputers in Germany and provides its more than 1000 scientific users with the technical infrastructure needed for the processing and analysis of huge amounts of data from climate simulations. This also includes training and support for related application software and data processing issues. DKRZ participates in many national and international projects aiming to improve the infrastructure for climate modeling. Through its research group on scientific computing DKRZ is linked to the Department of Informatics of the University of Hamburg.
DKRZ is a non-profit and non-commercial limited company with four shareholders. MPG (Partner 5) holds 55% of the shares of DKRZ (see http://www.dkrz.de/about-en/Organisation/gesellschafter for more references). The dependency relationship has been declared in the Part 2 –Administrative data of participating organisation of this application form

Role and responsibilities:

In ITHACA, DKRZ will lead the WP7 dedicated to the applicative use cases and contribute to all others WP except WP5.

Key Personnel:

Dr. Joachim Biercamp (male) holds a PhD in Physical Oceanography and has a long standing experience in supporting data intensive climate simulations. He is leading the Application department of DKRZ. His responsibilities include the organization of user support and the interaction with DKRZ's user group and scientific steering committee. He coordinated the procurement and benchmarking of the of several DKRZ super computers all ranked within the TOP 35 of the TOP500 list. Joachim is involved in several national and international projects dealing with infrastructure for climate modeling. In particular he is coordinator of the ESiWACE Center of Excellence and member of the steering committee of the German project HD(CP)2 aiming at development and operation of a cloud resolving version of the ICON model which is used for both, climate research and numerical weather prediction.

Dr. Julian Kunkel (male), he is Principal Investigator in the group Scientific Computing at the DKRZ. Julian gained interest in the topic of HPC storage during his studies of computer science in 2003. Since then, he researches methods to improve efficiency of storage systems in general. Besides his main goal to provide efficient and performance-portable I/O, his HPC-related interests are: data reduction techniques, performance analysis of parallel applications and parallel I/O, management of cluster systems, cost-efficiency considerations, and software engineering of scientific software. In 2013, he defended his thesis about monitoring and simulation of parallel programs on application and system level. Dr. Kunkel is member of many international program committees. He is currently coordinating the German projects AIMES and PeCoH, and he is participating in the ESiWACE project to advance I/O methods for climate applications. Previously he was coordinating the contributions to the ICOMEX and the SIOX projects. Julian will lead WP7.

Dr. Panagiotis Adamidis (male) holds a PhD in Mechanical Engineering. He is working in the area of High Performance Computing with emphasis on parallel numerical algorithms and parallel programming models since 1997. He joined DKRZ in 2006. The main focus of his work is the development of parallel algorithms for earth system models and optimization issues at application level. He participates in the project HDCP2 (High Definition Clouds and Precipitation for Climate

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Prediction) having as goal to enhance the scalability of the ICON model for future generation supercomputers.

Publications and Products:

These are an extract of relevant publications and products from DKRZ, connected to the subject of this proposal:

- **Utilizing In-Memory Storage for MPI-IO**. Julian Kunkel, Eugen Betke. 2016. Poster. The International Conference for High Performance Computing, Networking, Storage, and Analysis
- **The SIOX Architecture – Coupling Automatic Monitoring and Optimization of Parallel I/O**. Julian Kunkel, Michaela Zimmer, Nathanael Hübbe, Alvaro Aguilera, Holger Mickler, Xuan. 2014. Wang, Andriy Chut, Thomas Bönisch, Jakob Lüttgau, Roman Michel, Johann Weging. In Supercomputing, 29th International Supercomputing Conference, ISC 2014, Springer, ISBN: 978-3-319-07518-1, pp. 245-260
- **Simulating Parallel Programs on Application and System Level**. Julian Kunkel. 2013. In Computer Science - Research and Development (28, 2-3), pp. 167-174
- **SIOX, an open-source monitoring infrastructure**: https://github.com/JulianKunkel/siox
- M. Lautenschlager, P. Adamidis and M. Kuhn (2015): Big Data Research at DKRZ – **Climate Model Data Production Workflow**, pp 133 - 155 (DOI:10.3233/978-1-61499-583-8-133). In "Big Data and High Performance Computing", Eds.: L. Grandinetti, G. Joubert, M.Kunze, V. Pascucci. Vol. 26 of Advances in Parallel Computing. IOS Press.

Relevant Projects or Activities:

DKRZ currently participates to cooperative activities connected to the subject of this proposal:

- DKRZ is coordinator of **Centre of Excellence in Simulation of Weather and Climate in Europe** (**ESiWACE**), one of the nine Centers of Excellence in HPC applications funded under H2020. Amongst others, within ESiWACE we define a use case for the H2020 exascale demonstrator and we will feedback the knowledge gained within ITHACA to ESiWace. Additionally, we are developing a prototype for alternative data type mappings to optimize storage layout that is needed in ITHACA.
- DKRZ is coordinating the **Advanced Computation and I/O Methods for Earth-System Simulations (AIMES)**, an international project involving IPSL (France) and RIKEN (Japan). It aims to address the key issues of programmability, computational efficiency and I/O limitations that are common in next-generation icosahedral earth-system models.
- DKRZ was involved in the **Scalable I/O for Extreme Performance (SIOX),** which aims at developing interfaces and a tool for monitoring storage telemetry data of traditional storage environments.

A full list of projects led or participated by DKRZ is available here: http://www.dkrz.de/Klimaforschung-en/projects

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

7. Fraunhofer ITWM:

Short Name: **Fraunhofer**

Description of the Organisation:

The Fraunhofer-Gesellschaft (FhG) is Europe's largest organisation for application-oriented research. Founded in 1949, the organisation undertakes applied research that drives economic development and serves the wider benefit of society. Its services are solicited by customers and contractual partners in industry, the service sector and public administration. The majority of the more than 20,000 staff are qualified scientists and engineers.
The Fraunhofer Institute for Mathematics (ITWM) in Kaiserslautern, Germany, focuses on mathematical approaches to practical challenges like optimisation and visualisation. With computer simulations being an indispensable tool in the design and optimisation of products and production processes, real models are being replaced by virtual models and mathematics play a fundamental role in the creation of this virtual world. Core competences of the ITWM include processing of large data sets, drafting of mathematical models, problem-solving in numerical algorithms, summarisation of data sets, interactive optimisation of solutions, and visualisation of simulation and sensor data.

Role and responsibilities:

As part of Fraunhofer ITWM, the Competence Center for High Performance Computing (CC-HPC) develops innovative HPC solutions for the industry and participates in national and international research programs. Since its foundation in 2002, the CC-HPC focuses primarily on parallel application development and development of HPC tools. This includes the communication middleware GPI (Global Address Space Programming Interface), the GPI-Space programming environment for parallel and big data applications, and the BeeGFS File System formerly known as FhGFS.
Fraunhofer ITWM will lead the Ecosystem tools and API: Work Package 9.

Key Personnel:

The experts described below will participate actively in the ITHACA project:

Dr. Franz-Josef Pfreundt (male) studied Mathematics, Physics and Computer Science, receiving a Diploma in Mathematics and a PhD in Mathematical Physics (1986). From 1986 to 1995, he was Head of the Research Group for Industrial Mathematics at the University of Kaiserslautern. In 1995, he became Department Head of the Fraunhofer Institute for Industrial Mathematics (ITWM). His research topics are: Fluid dynamics, porous media, image analysis and parallel computing. At the ITWM, he founded the departments "Flow in complex structures" and "Models and algorithms in image analysis". Since 1999, he has been Division Director at Fraunhofer ITWM and Head of the "Competence Center for HPC and Visualisation".

Dr. Mirko Rahn (male) studied Computer Science and Logics resulting in a Diploma and received a Ph.D. in theoretical Computer Science (2008) from the University of Karlsruhe. In 2009 he was member of a team that set a new record in sorting huge amounts of data. In 2009 he joined the Fraunhofer ITWM in Kaiserslautern, Germany. He worked on new concepts and tools to manage and process very large data sets, especially data sets from the seismic domain. His key interests are high performance parallel processing, compiler and language technology and workflow management.

Norman Ettrich (male) studied Geophysics resulting in a Diploma and received a Ph. D. in Geophysics (1996) from the University of Hamburg. From 1996 to 1998 he had an assistant position at the University of Hamburg. Between 1998 and 2002 he filled a position as research geophysicist in Statoil's research center (Trondheim, Norway). Since 2002 he has been working for Fraunhofer ITWM (Kaiserslautern, Germany), first, in the field of fluid dynamics. Since 2005, he helped building-up the group doing research in the field of applied seismology. He has been in charge of several (big) projects with the

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

oil&gas industry. His key interests are seismic imaging, seismic data processing, and, in particular, developing and implementing production-ready methods and software for both aforementioned fields.

Sven Breuner (male) decided early to focus on design and development of parallel and distributed applications. He joined the Fraunhofer Competence Center for High Performance Computing (CC-HPC) in 2005, after receiving his Bachelor degree in computer science with a thesis on process management on heterogeneous clusters. At Fraunhofer, he developed the initial design of the Fraunhofer Parallel File System (FhGFS) which has been recently renamed BeeGFS and is currently leading the file system development team within the CC-HPC. In 2008, he received a Master degree in computer sciencewith a thesis on efficient distributed metadata management in parallel file systems.

Publications and products:

These are an extract of relevant publications and products from Fraunhofer, connected to the subject of this proposal:

- **GPI API**: open-source licences for researchers, commercial license for commercial users: http://www.gpi-site.com/gpi2/
- **GASPI** – A Partitioned Global Address Space Programming Interface, Alrutz, et. al., DOI: 10.1007/978-3-642-35893-7_18 In book: Facing the Multicore-Challenge III, Publisher: Springer Berlin Heidelberg, Editors: Keller, Rainer and Kramer, David and Weiss, Jan-Philipp, pp.135-136
- Unbalanced tree search on a manycore system using the GPI programming model, Machado et. al., Computer Science - Research and Development 01/2011; 26:229-236. DOI: 10.1007/s00450-011-0163-3
- Foss, S.k., Merten, D., Ettrich, N., Stangeland-Karlsen, E., and Mispel, J. [2014] Amplitude-friendly angle migration with stabilized Q: EAGE, Expanded Abstracts.
- **GRT-software** product web-site: http://www.seismic-grt.com/

Relevant Projects or Activities:

Fraunhofer participated to cooperative projects connected to the subject of this proposal:

- **EXA2CT** is an EC-funded FP7 project on exascale computing. The project brings together experts at the cutting edge of the development of solvers, related algorithmic techniques, and HPC software architects for programming models and communication. Fraunhofer ITWM provides the GPI API to the project and consults the applications on their communication models.
- **EPiGRAM** is an EC-funded FP7 project on exascale computing. The aim of the EPiGRAM project is to prepare Message Passing and PGAS programming models for exascale systems by fundamentally addressing their main current limitations. Fraunhofer ITWM is work package leader for exascale PGAS.
- **GASPI** is a project funded by the German ministry of education and research (BmBF). Its main goal is to establish a standard for PGAS-APIs, namely GPI and provide a reliable basis for future developments.
- Approx. 10 industry-funded research projects with partners from oil&gas area
- BMBF project: **CO2DEPTH**, Optimisation of 3D depth models by fast reflection tomography; special program: Geotechnologien, 2008-2001

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

8. *Barcelona Supercomputing Center*

Short name: **BSC**

Description of the Organisation:

The Barcelona Supercomputing Center (BSC) was established in 2005 and is the Spanish national supercomputing facility and a hosting member of the PRACE distributed supercomputing infrastructure. The Center houses MareNostrum, one of the most powerful supercomputers in Europe. The mission of BSC is to research, develop and manage information technologies in order to facilitate scientific progress.

BSC was a pioneer in combining HPC service provision, and R&D into both computer and computational science (life, earth and engineering sciences) under one roof. The centre fosters multidisciplinary scientific collaboration and innovation and currently has over 400 staff from 41 countries. In 2011, BSC was one of only eight Spanish research centres recognized by the national government as a "Severo Ochoa Centre of Excellence".

BSC has collaborated with industry since its creation, and has participated in projects with companies such as ARM, Bull and Airbus as well as numerous SMEs. BSC also participates in various bilateral joint research centers with companies such as IBM, Microsoft, Intel, NVIDIA and Spanish oil company Repsol. The centre has been extremely active in the EC Framework Programmes and has participated in over one hundred projects funded by it. BSC is a founding member of HiPEAC, the ETP4HPC and participates in the most relevant international roadmapping and discussion forums and has strong links to Latin America.

Education and Training is a priority for the centre and many of BSCs researchers are also university lecturers. BSC offers courses as a PRACE Advanced Training Centre, and through the Spanish national supercomputing network among others.

Two BSC departments will participate in the ITHACA project:

- Computer Sciences: The BSC-CNS Computer Sciences Department focuses on building upon currently available hardware and software technologies and adapting these technologies to make efficient use of supercomputing infrastructures. The department proposes novel architectures for processors and memory hierarchy and develops programming models and innovative implementation approaches for these models as well as tools for performance analysis and prediction.

- Earth Sciences: The Earth Sciences Department was established with the objective of conducting research in Earth system modelling. The research focuses on atmospheric emissions, air quality, mineral dust transport, global and regional climate modelling and prediction, climate services for private users and computational Earth Sciences.

Role and responsibilities:

BSC will contribute to the Ithaca project in three different aspects. First, BSC will provide its experience in parallel programming models and PyCOMPSs (an in-house task-based parallel programming model) in order to evaluate the effectiveness of task based programming in the Ithaca architecture. Second, BSC will contribute with its experience in data management, and especially by providing dataClay, a next-generation object store that perfectly fits the philosophy of Ithaca. Finally, BSC will also provide an application in the area of earth sciences (EC-Earth 3.2) to validate the benefits of the proposed architecture and software stack.

BSC will lead WP3 (programming models) and will actively participate in WP 4 (data management) and WP7 (applications).

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Key Personnel:

Prof. Toni Cortes (male) is the manager of the storage-system group at the BSC (since 2006) and is also an associate professor at Universitat Politècnica de Catalunya (since 1998). He received his Ph.D. in computer science in 1997 (at Universitat Politècnica de Catalunya). Since 1992, Toni has been teaching operating system and computer architecture courses at the Barcelona school of informatics (UPC) and from 2000 to 2004 he also served as Vice Dean for international affairs at the same school. His research concentrates in storage systems, programming models for scalable distributed systems, and operating systems. He has published 26 journal papers, 76 papers in international conferences and workshops. In addition, he has also advised 10 PhD theses since 1997. Dr. Cortes has been involved in several EU projects (Paros, Nanos, POP, XtreemOS, SCALUS, IOlanes, PRACE, MontBlanc, IOStack, BigStorage and NextGenIO) and has also participated in cooperation with IBM (TJW research lab) on scalability issues both for MPI and UPC.

Dr. Rosa M Badia (female) holds a PhD on Computer Science (1994) from the Technical University of Catalonia (UPC). She is the manager of the Workflows and Disttributed Computing group at the BSC since 2005 and coordinator of the Big Data activities at BSC. Since is also a Scientific Researcher from the Consejo Superior de Investigaciones Científicas (CSIC). She was Associated Professor at UPC from 1997 to 2008. From 1999 to 2005, she was involved in research and development activities at CEPBA. Her current research interests cover the programming models for distributed computing platforms and its integration with novel storage technologies for Big Data. She has participated in several European projects. Dr. Badia has been IP of project SIENA, and of project EU-Brazil OpenBIO. She is also a member of the HiPEAC2 NoE. She is currently participating in EU funded projects: HBP, EUBra BIGSEA, ASCETIC, TANGO, EUROSERVER, NEXTGENIO, MUG, BioExcel.

Msc. Kim Serradell Maronda, (male). Is Bachelor (2005) in Computer Sciences for the Facultat d'Informàtica de Barcelona (FIB-UPC) and for the Grande école publique d'ingénieurs en informatique, mathématiques appliquées et télécommunications de Grenoble (ENSIMAG). Since 2014 is also Master on High Performance Computing from the Facultat d'Informàtica de Barcelona (FIB-UPC). Currently, he is the manager of the Computational Earth Science (CES) group at the Earth Sciences department in the Barcelona Supercomputing Center (BSC). The CES group is a multidisciplinary team of 15 members with different IT profiles that interacts closely with all the other groups of the Earth Sciences Dept. In the last years, he has been in charge for the system administration of all the computational resources of the department and he was also responsible of supervising the operational runs of the NMMB/BSC-Dust model and CALIOPE Air Quality System in the HPC infrastructures of the BSC. He has been involved in European projects like IS-ENES (1 & 2), ESiWACE, SDS-WAS, BDFC or CONSOLIDER.

Publications and Products:

PyCOMPSs/COMPSs is a task based programming model that aims to ease the development of parallel applications/workflows that runs on distributed infrastructures. The COMPSs syntax allows developers to compose parallel applications being totally agnostic of the infrastructure and the parallelism details. COMPSs is in the process of being integrated with dataClay.
PyCOMPSs/COMPSs is distributed through the BSC website as open source and packaged in Linux packages: compss.bsc.es
PyCOMPSs/COMPSs will be used in Ithaca to validate the possibility of developing applications from a task-based programming model.

dataClay is a next-generation object store designed and implemented by BSC that is able to store objects as in an OO programming models by keeping both the data and its related methods. The functionality of storing objects at byte-level granularity as well as the possibility to execute methods where the data is located are a perfect fit for the philosophy of Ithaca.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

PyCOMPSs: Parallel computational workflows in Python Enric Tejedor, Yolanda Becerra, Guillem Alomar, Anna Queralt, Rosa M Badia, Jordi Torres, Toni Cortes, and Jesús Labarta International Journal of High Performance Computing Applications first published on August 19, 2015 as doi:10.1177/1094342015594678

EC-Earth is a coupled climate model developed as part of a Europe-wide consortium where BSC is an active partner. EC-Earth made successful contributions to international climate change projections such as CMIP5 (and next CMPI6). Ongoing development by the consortium will ensure that increasingly more reliable projections can be offered to decision and policy makers at regional, national and international levels.

Relevant Projects or Activities:

- **MontBlanc 1 and 2:** Mont-Blanc is a FP7 project, which contributes to the development of extreme scale energy-efficient platforms, with potential for exascale computing, addressing the challenges of massive parallelism, heterogeneous computing, and resiliency. It will enable further development of the OmpSs parallel programming model to automatically exploit multiple cluster nodes, transparent application check pointing for fault tolerance, support for ARMv8 64-bit processors, and the initial design of the Mont-Blanc exascale architecture.

- **NextGenIO:** NEXTGenIO targets to solve the IO problem in HPC architectures by bridging the gap between memory and storage. This will use Intel's revolutionary new 3D XPoint non-volatile memory, which will sit between conventional memory and disk storage. NEXTGenIO will design the hardware and software to exploit the new memory technology. The goal is to build a system with 100x faster I/O than current HPC systems, a significant step towards Exascale computation.

- **ESiWACE:** ESiWACE stands for Centre of Excellence in Simulation of Weather and Climate in Europe. The goal is to substantially improve efficiency and productivity of numerical weather and climate simulation on high-performance computing platforms by supporting the end-to-end workflow of global Earth system modelling in HPC environment. The project has been funded by Horizon 2020, call H2020-EINFRA-2015-1 “Centres of Excellence for computing applications” of the DG Connect.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

9. *Allinea*

Short Name: **Allinea**

Description of the Organisation:

Allinea Software is a successful European software SME within HPC. The company creates and sells tools for HPC software developers, scientists and applications consultants that enable applications and their creators to achieve more from the available resources of hardware and development time.

It has two core product lines – Allinea Forge, the scalable HPC development tool suite including the debugger Allinea DDT and profiler Allinea MAP, and Allinea Performance Reports which benchmarks and analyses key performance characteristics of applications.

Its tools are deployed worldwide – across all the core HPC segments in government, academia and industry.

Allinea has strong presence at the leading edge of HPC – and has earned a reputation as a successful innovator. The company's R&D team are actively addressing the current needs of the HPC and related sectors and future challenges in upcoming systems in the 2017-2019 timeframe. This is through both H2020 and other sponsored initiatives with leading customers.

Its customer base includes the largest HPC systems in Europe, America and, most recently, Japan –as a partner in the ITHACA project Allinea brings unique capability and insight into future scalability needs of mainstream and extreme scale computing.

Role and responsibilities:

Within the ITHACA project, Allinea intends to address some of the next generation of challenges – that will be important in the 2018-2021 time-frame.

Our contribution is to bring develop required tools technology for WP3 and WP6, and to apply existing and the developed technology to the applications and their users within WP7 in order to benchmark and optimize applications to the ITHACA system.

Within WP3 Allinea will build on its background IP to support new or emerging standards that extend the reach and capability of its tools related to the programming models of WP3. Within WP7 we extend the Allinea tools to measure and report energy, time, and I/O demands of applications. Further, we assist Fraunhofer to support their GPI framework within tools.

Key Personnel:

Dr David Lecomber (male) holds a DPhil from Oxford University in Parallel Computation, and is CEO at Allinea Software. He is actively involved in the software, having led the team as CTO for the first 10 years of Allinea.

Dr Jonathan Byrd (male), holds a PhD from Warwick University in Markov methods, and is a Senior Developer at Allinea Software. Jonathan's primary focus is on performance profiling methods and scalability.

Publications and Products:

- **Allinea Forge:** is the complete toolsuite for software development - with everything needed to debug, profile, optimize, edit and build C, C++ and Fortran applications on Linux for high

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

performance - from single threads through to complex parallel HPC codes with MPI, OpenMP, threads or CUDA.

- **Allinea Performance Reports**: are the most effective way to characterize and understand the performance of HPC application runs. One single-page HTML report elegantly answers a range of vital questions for any HPC site:
    - Is this application well-optimized for the system and the processors it is running on?
    - Does it benefit from running at this scale?
    - Are there I/O, networking or threading bottlenecks affecting performance?
    - Which hardware, software or configuration changes can we make to improve performance further.
    - How much energy did this application use?

Relevant Projects or Activities:

- **SAGE**
    - Within the project Allinea extends support for I/O and the SAGE off-load framework to enable Allinea's tools to reach workloads adjacent to HPC such as data-driven analysis from particle accelerator community.

- **ExaNEST**
    - Developing initial support for ARM based large-scale HPC servers and providing performance analysis for the project partners porting their codes to the initial platform.

- **NextGEN-IO**
    - Researching and preparing for the upcoming world of NVRAM and SCM technologies and their impact on software; measuring workload patterns to inform the connection between storage and intelligent schedulers. The net result is anticipated to be improved system efficiency.

- **Compat-Multiscale**
    - This applications-focussed project seeks to create a framework for multiscale codes – those that do not follow the fixed-size large MPI code pattern – but that bring simulations at multiple levels to simulate complex real world situations with varied degrees of parallelism. Within this project Allinea is creating tools able handle many short-running and potentially related simulations, providing tool support and performance expertise for the applications both initially and throughout the project.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

10. SURFsara:

Short Name: **SURFsara**

Description of the Organisation:

SURFsara is the National Supercomputing and e-Science Support Center in the Netherlands. Among SURFsara's customers are all of the Dutch Universities, a number of large research, educational and government institutions, and the business community. The mission of SURFsara is to support research in the Netherlands by the development and provision of advanced ICT infrastructure, services and expertise. SURFsara provides expertise and services in the areas of High Performance Computing, e-Science & Cloud Services, Data Services, Network support, and Visualisation. SURFsara hosts the large national infrastructure services, i.e. the Dutch national Supercomputer service (Bull bullx system, 1.5 Pflop/s since the end of 2014), the National Compute Cluster (Dell cluster, 7856 cores, 149 Tflop/s), large data storage facilities and services and all important national grid services. This includes also a large part of the BiG Grid infrastructure (the Dutch e-Science grid that is a Tier-1 site for CERN LCG). SURFsara participates in a number of national and international HPC, e-science and grid activities. SURFsara is currently partner in the large European e-Infrastructure projects PRACE-4IP, EUdat, and Fortissimo. SURFsara has a long history and a proven track record in providing HPCN services to the Dutch research community. Important focus is on supporting users to enable grand challenge applications. SURFsara has built broad expertise in the implementation and escience support of grid infrastructures for the scientific research community. SURFsara provides also HPC-Cloud and Hadoop services to support research.

Role and responsibilities:

SURFsara has a long tradition in porting and scaling of applications on PRACE (PaRtnership for Advanced Computing in Europe) systems, and in performance modelling with relevant tools. In this role we have built a lot of know-how and skills that will now be used for the benefit of this project. Scaling and porting real-life applications on the ITHACA interconnect will generate tuned applications and feedback that will be used during the design stage of the interconnect. Our experiences with HPC ecosystems will be used to test and provide feedback on the ecosystem design for the ITHACA interconnect. SURFsara will perform tasks in WP7 and provide deliverables in the form of reports and tuned applications.

Key Personnel:

John Donners (male) studied physics at the Radboud University in Nijmegen, graduating with an MSc in computational physics. He started a PhD in oceanography at the Royal Netherlands Meteorological Institute. Following up, he joined the University of Reading to develop high-resolution climate simulations on the Earth Simulator in Japan. During this period, he got a broad experience with scientific data management, scientific file formats (mostly NetCDF, text and HDF), parallel programming, debugging and international collaboration. He joined SURFsara in 2008 and has been involved in HPC-Europa and DEISA and still is actively involved in PRACE and Fortissimo.

Valeriu Codreanu (male) is a supercomputing consultant at SURFsara. He received his master degree in electrical engineering in 2008 from the Faculty of Electrical Engineering and Information Theory of University Politehnica of Bucharest. He received his PhD from the same faculty in 2011. From 2011 till 2013 he worked at the University Groningen as a post- doctoral fellow in the project GPSME: A General Toolkit for GPUtilisation in SME applications. From 2013 till 2014 he worked as a post-doctoral fellow at Technical University of Eindhoven on the project, Best ENergy EFficiency solutions for heterogeneous multI-core Communicating systems

Walter Lioen (male) has a background in numerical mathematics (MSc) and worked from the mid-eighties as a scientific programmer on many different supercomputers. In 2007 he joined SURFsara as

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

a senior HPC consultant where he became group leader of the supercomputing group in 2008. Walter was and still is actively involved in DEISA and PRACE.

Publications and Products:

- **A pencil distributed finite difference code for strongly turbulent wall-bounded flows.** Erwin P. van der Poel, Rodolfo Ostilla-Mónico, John Donners, and Roberto Verzicco. Computers & Fluids, Volume 116, Pages 10–16, August 15, 2015. doi: 10.1016/j.compfluid.2015.04.007
- **CLTune: A Generic Auto-Tuner for OpenCL Kernels.** Cedric Nugteren and Valeriu Codreanu. IEEE 9th International Symposium on Embedded Multicore Many-core Systems-on-Chip (MCSoC-15), Turin, September 23-25, 2015. doi: 10.1109/MCSoC.2015.10
- **Performance analysis of EC-EARTH 3.1**, Donners et. al., PRACE whitepaper:http://www.prace-ri.eu/IMG/pdf/Performance_Analysis_of_EC-EARTH_3-1-2.pdf

Relevant Projects or Activities:

- **Fortissimo and Fortissimo-2** are collaborative projects that will enable European SMEs to be more competitive globally through the use of simulation services running on a High Performance Computing cloud infrastructure. SURFsara is working with ISVs and SMEs in four experiments in this project.

- **PRACE-PP, PRACE-1IP, PRACE-2IP, PRACE-3IP and PRACE-4IP** are EC-funded projects that complement the national investments to accelerate the deployment of the PRACE service to users from academia and industry. SURFsara contributes to these projects by porting, tuning and benchmarking applications.

- **DECI** (Distributed European Computing Initiative) is an European single-project HPC access scheme which is now supported by PRACE-4IP. Previously, DECI which stood for DEISA Extreme Computing Initiative, was conceived of and supported by the DEISA projects which issued six DECI-calls between 2004 and ??

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

11. *Institut National de Recherche en Informatique et Automatique*

Short Name: **INRIA**

Description of the Organisation:

Inria, the French National Institute for computer science and applied mathematics, promotes "scientific excellence for technology transfer and society". Graduates from the world's top universities, Inria's 2,700 employees rise to the challenges of digital sciences. With its open, agile model, Inria is able to explore original approaches with its partners in industry and academia and provide an efficient response to the multidisciplinary and application challenges of the digital transformation. Committed to assisting innovators, Inria provides the ideal conditions for fruitful relations between public research, private R&D and industry. Inria transfers its expertise and research results to startups, SMEs and major groups in fields as diverse as healthcare, transport, energy, communications, security and privacy protection, smart cities and the factory of the future. Inria has also fostered an entrepreneurial culture that has led to the creation of 120 startups.

Role and responsibilities:

Inria has a long linage of research teams focused on many different aspects of High Performance Computing. In the context of ITHACA, Inria will mainly involve its DataMove team focused on data aware large scale computing. Inria will investigate novel strategies for optimizing data movements for HPC applications, relying on ITHACA networking and storage developments. Inria will in particular focus on in situ processing strategies with the FlowVR framework it develops. Inria will perform tasks in WP3 and WP9 and provide large-scale test applications based on molecular dynamics simulations.

Key Personnel:

Bruno Raffin (male) is Research Director at INRIA and leader of the DataMove team. He led the development of the FlowVR middleware for large-scale data-flow oriented parallel applications, used for scientific visualization and computational steering. He recently retargeted FlowVR at in situ analytics for large-scale parallel application. He also worked on parallel algorithms and cache-efficient parallel data structures (cache oblivious mesh layouts, parallel adaptive sorting), strategies for task-based programming of multi-CPU and multi-GPU machines. Bruno Raffin accounts for more than 60 international publications, adviced 16 PhD students and 3 postdocs. He was responsible for INRIA of more than 15 national and European grants, and was the co-founder of the Icatis startup company (2004-2008). Bruno Raffin has been involved in more than 30 program committees of international conferences. He is the chair of the steering committee of the Eurographics Symposium on Parallel Graphics and Visualisation. Today his research is focused on in situ processing at large scale.

Pierre-François Dutot (male) is Associate Professor at University Grenoble Alpes since 2006. He received the PhD in Computer Science from Grenoble INP in August 2004, and MS and BS in Computer Science from the Ecole Normale Supérieure de Lyon. From 2005 to 2006 he was temporary assistant professor at the Université Henry Poincaré, Nancy 1. His research interests include parallel models, approximation algorithms and multi-objective scheduling, including data movement optimization strategies. He has published nearly 30 articles in international conferences and journals. He co-advised 2 PhD students. He is general co-chair for Euro-Par 2016 (http://europar2016.inria.fr/). He has also been in the program committee of several conferences (most recent ones: HeteroPar 2005-2013, IPDPS 2013-2014, 2016, EuroEDUPAR 2015, HCW 2016), and reviewer for prestigious journals (IEEE TPDS, IEEE TC, ParCo, etc.).

Frédéric Wagner is Associate professor at Ensimag, University Grenoble Alpes since 2006. His research activities focus on scheduling algorithms for complex platforms, more specif- icaly offline scheduling of communications, online algorithms improving data locality. He co-authored 5 papers in international journals and 12 international conferences.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Publications and Products:

- **Lessons Learned from Building In Situ Coupling Frameworks.** Matthieu Dorier, Matthieu Dreher, Tom Peterka, Gabriel Antoniu, Bruno Raffin, Justin M. Wozniak. *ISAV 2015 (in conjunction with SC15)*, Nov 2015.

- **Design and analysis of scheduling strategies for multi-CPU and multi-GPU architectures**. João V.F. Lima, Thierry Gautier, Vincent Danjean, Bruno Raffin and Nicolas Maillard. Parallel Computing, 44:37-52, 2015.

- **ExaViz: a Flexible Framework to Analyse, Steer and Interact with Molecular Dynamics Simulations.** Matthieu Dreher, Jessica Prevoteau-onquet, Mikael Trellet, Marc Piuzzi, Marc Baaden, Bruno Raffin, Nicolas Férey, Sophie Robert, Sébastien Limet. *Faraday Discussions of the Chemical Society*, Royal Society of Chemistry, 2014, Molecular simulations and visualization, 169, pp.119-142.

- **A Flexible Framework for Asynchronous In-situ and In Transit Analytics for Scientific Simulations**. Matthieu Dreher and Bruno Raffin. 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID'14), Chicago, 2014.

- **Approximating the discrete resource sharing scheduling problem**. M. Bougeret, P-F. Dutot, A. Goldman, Y. Ngoko, and D. Trystram. International Journal of Foundations of Computer Science, 22(3), 2011.

- **WSCOM: Online Task Scheduling with Data Transfers**. Jean-Noël Quintin, Frédéric Wagner., CCGRID, 2012, Ottawa, Canada. pp.344-351, 2012.

- http://flowvr.sf.net - **FlowVR, an open source framework for in situ analytics**.

Relevant Projects or Activities:

- **Avido** (2015-2018) is a collaborative research project funded by the French science foundation. Avido focuses on in situ analysis and visualization for large-scale numerical simulations.

- **VelaSSco** (2014-2016) is a FP7 research program that developed a petabyte scale query based scientific visualization infrastructure relying on Big Data technologies and supporting in situ data injection from running parallel simulations.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

12. ARM

Short Name: **ARM**

Description of the Organisation:

ARM Ltd (www.arm.com) is a world-renowned semiconductor IP company with around 4000 employees and headquartered in Cambridge, UK. Forbes has named ARM as the world's fifth most innovative company, using the "Innovation Premium", and is one of only two companies in their "semiconductor" category in the top 100. More recently, ARM and UNICEF have a multi-year partnership agreement to work on positive impacts of wearable and mobile technology to transform children's lives in the Third World.
ARM designs and licenses its IP (e.g. low-end 32-bit and high-end 64-bit CPUs, mobile GPUs, various peripherals, physical libraries etc.) to its chip manufacturing Partners. Energy efficiency is the DNA of its technology. ARM is the architecture of choice for more than 80% of the high-performance embedded products in design, and for more than 90% of the mobile phones. By licensing, rather than manufacturing and selling its chip technology, it established a new business model that has redefined the way microprocessors are designed, produced and sold. Today, there are more than 1000 partners in the ARM Connected Community. As the foundation of the company's global technology community, these Partners have played a pivotal role in the widespread adoption of the ARM architecture.
Partners utilize ARM's low-cost, power-efficient core designs to create and manufacture microprocessors, peripherals and SoC solutions. ARM Powered microprocessors are pervasive in the electronic products, driving key functions in a variety of applications in diverse markets, including automotive, consumer entertainment, imaging, microcontrollers, IoT and wearable devices, networking, storage, automotive, medical, security, wireless, smartphones, and tablet computers. To date, ARM Partners have shipped more than 70 billion ARM microprocessors. ARM Development Solutions Group (DSG) will participate in the project. The mission of DSG is to ensure that ARM has the best development tools available, wherever they come from. DSG develops integrated software products that run on ARM processors, these products include simulation models and tools, compilers, debuggers, performance analysers and IDEs. DSG works with the open source community to ensure OSS tools and compilers support the ARM architecture.

Role and responsibilities:

ARM's role within the ITHACA Horizon2020 project is to make the new systems, software and programming models developed under the ITHACA project accessible to software developers through the new techniques that ARM will develop and the existing tools that ARM will enhance with new functionality. Furthermore, these new tools and capabilities developed by ARM under ITHACA are intended to allow developers without specialist, in-depth knowledge of the new systems to be productive in developing efficient, high-performance applications that make best and most efficient use of those systems. This will both widen the audience that benefits from the results of the ITHACA bid and allow that audience to make best use of those results.

Key Personnel:

Geraint North (male) is ARM's Distinguished Engineer for Server and HPC Tools. He joined ARM in 2014 as a founding member of ARM's software development centre in Manchester, UK, which performs HPC-specific activity on compilers, libraries and performance tools, including support for ARM's recently announced HPC-focused Scalable Vector Extensions (SVE).
Previously a Master Inventor at IBM, Geraint was the storage architect for IBM's Cloud Systems Software group, designing products built on OpenStack. He has also worked on a wide range of products and research areas across enterprise storage, POWER systems hardware, AIX and BlueGene. Prior to IBM, Geraint was a Principal Engineer at Transitive Corporation (a spin-out from Manchester University), developing dynamic binary translation technology for Apple, Silicon Graphics, IBM and others. Geraint holds over twenty patents in the fields of dynamic binary translation, enterprise storage and microprocessor high-availability.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Travis Walton (male) is a Staff Software Engineer working at ARM. Travis joined ARM in 2015 and works within the Advanced Product Development team assigned to the Manchester Design Centre. He spends his time investigating future technologies relevant to the business and creating innovative new software tools. Travis has worked in the software industry for twenty years with the majority of that time spent developing new technologies for forward looking technology startups or working on the research teams of larger more established companies.

Publications and Products:

- Eric Van Hensbergen, Marc Snir, et.al. "Addressing Failures in Exascale Computing", International Journal of High Performance Computing Applications 1094342014522573, first published on March 21, 2014
- Eric Van Hensbergen, "From Sensors to Supercomputers, Big Data Begins with Little Data", International Advanced Research Workshop on High Performance Computing, Cetraro, Italy, 2014
- Nikola Rajovic, Alejandro Rico, Filippo Mantovani et al. "The Mont-Blanc Prototype: An Alternative Approach for HPC Systems" In SC¹16: International Conference for High Performance Computing, Networking, Storage and Analysis (Best Paper Nominee), November 2016
- Thomas Grass, Alejandro Rico, Marc Casas, Miquel Moreto, Eduard Ayguadé "TaskPoint: Sampled Simulation of Task-Based Programs" In ISPASS '16: International Symposium on Performance Analysis of Systems and Software, April 2016

Relevant Projects or Activities:

ARM has been an active and major participant in the European drive towards Exascale (see http://exascale-projects.eu/EuroExaFinalBrochure_v1.0.pdf) through European projects such as Mont-Blanc, Mont-Blanc 2 and Mont-Blanc 3. The Mont-Blanc family of projects is a push to develop energy efficient HPC systems.
Mont-Blanc 1 is a project to build a prototype ARM-based supercomputer using low-power commercially available embedded technology and to port large scale HPC applications to that prototype.
Mont-Blanc 2 is a project to build on the work of Mont-Blanc 1 and produce a first definition of the Mont-Blanc Exascale Architecture, exploring different alternatives for the compute node and its implications for the rest of the system. Under Mont-Blanc 2, the Mont-Blanc software stack was enhanced, focusing on 64Bit support and programming tools.
Mont-Blanc 3 will see the creation of a new, high-end HPC platform (SoC and node) that is able to deliver a new level of performance/energy ratio whilst executing real applications.
This new platform will use the latest ARM processors and will develop the necessary software ecosystem to support them.
ARM is an active participant in the embedded HPC community and in the **HIPEAC projects** (HIPEAC 1,2,3,4) and has hosted a number of interns during that time. The purpose of the HIPEAC project is to gather, coordinate and promote academic research efforts in the design and implementation of high performance commodity computing devices on the 10+ year horizon and to establish tight partnerships with leading European embedded systems manufacturers.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

13. *University of Cologne*

Short Name: UKOELN

Description of the Organisation:

University of Cologne (www.uni-koeln.de) is a public university. With more than 48.000 students it is one of the largest universities in Germany, located in the state of North Rhine-Westphalia. The university is structured in six faculties, namely Arts and Humanities; Management, Economics and Social Sciences; Mathematics and Natural Sciences; Medicine; Law and Human Sciences. The university became one of the German Excellence Universities within the Excellence Initiative. It has installed two life sciences related clusters of excellence, the CECAD (Cluster of Excellence for Cellular Stress Responses in Aging-Associated Diseases) and CEPLAS (Cluster of Excellence on Plant Sciences). Furthermore, two graduate schools where founded, the a.r.t.e.s. (Graduate School for Humanities Cologne) as well as bcgs (Bonn-Cologne Graduate School).
Contributions to this project will come from the Regional Computing Centre at the University of Cologne (RRZK) (www.rrzk.uni-koeln.de). RRZK provides IT-Services for the University of Cologne, which covers the operation of the campus network, as well as data storage, backup and archiving infrastructures. High performance computing (HPC) services range from the operation of an HPC Cluster to consulting and services for scientific users from the university and the state of North Rhine-Westphalia. As a result of the excellence initiative customers from the life sciences receive extended support in optimizing their applications for modern HPC architectures including early access to many coming technologies. Within the German Ministry of Research funded SuGI-Project, RRZK founded 2008 a national virtual organization for life sciences. RRZK had the co-lead in MoSGrid, a D-Grid project investigating grid technologies for users of molecular simulation tools.

Role and responsibilities:

RRZK will be mainly focused on WP7. Our group has a large expertise in the field of genetic pipelines on HPC-Systems and will contribute this knowledge to help develop middleware to enable modern supercomputers to efficiently drive such use cases.
Together with Fraunhofer ITWM the University of Cologne will analyze selected genetic pipelines and build a workflow generator with a customized API. In WP3 UKOELN will supply requirements of the considered genetic use-cases and will provide performance data and functionality as feedback to the ITHACA partners. All requirements of the genomics application will be collected. These include the requirements towards fault tolerance and parallelism, that GPI-Space has to deliver. This will form the basis for designing and implementing resilience techniques in WP6. By enabling the applications in the pipeline to use a common memory space, IO driven by staging will be reduced to a minimum, effectively reducing the required bandwidth and enabling much higher quantities of genomes being processed in an HPC-Environment. The improvements of the FlaGAS hardware approach alone and the FlaGAS plus GPI-Space approach will be compared.

Key Personnel:

- **Prof. Dr. Ulrich Lang** (male) is Professor in the Department of Computer Science at the University of Cologne. He is the former deputy director of the HLRS, one of the German national supercomputing centers. He initiated the development of COVISE, a collaborative visualization and simulation environment, when he was at the University of Stuttgart. He is the director of the Regional Computing Centre (RRZK) at the University of Cologne. He has a focus on high performance and grid computing and applying/optimizing these technologies to computation intensive tasks arising during the analysis of next-generation sequencing data. Additionally, he continues his involvement in visualization and virtual reality with a focus on large data sets.

- **Viktor Achter** (male) received a Diploma in economics and computer science at the University of Cologne. He leads the HPC Group of the Regional Computing Centre (RRZK) at the University of

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Cologne. Together with Prof. Lang he led the BMBF funded project NGSgoesHPC. Currently, together with Prof. Lang he is leading the Cologne part of the two BMBF funded projects SMOOSE and FaST.

- **Lech Nieroda** (male) received a Diploma in economics and computer science at the University of Cologne. He works in the HPC group of the University of Cologne. His focus is on the design and development of parallel and distributed applications. His key interests are high performance parallel processing and life science algorithms.

- **Dr. Martin Peifer** (male) leads a research group on computational cancer genomics. He has a track record in computational biology and bioinformatics with a focus on cancer genomics and the analysis of large-scale sequencing data. In particular, he developed an analysis framework to detect biologically relevant alterations in cancer genome sequencing data. This has let to several discoveries in lung cancer and neuroblastoma.

- **Dr. Susanne Motameny** (female) received a Diploma in mathematics and computer science at the Friedrich Schiller University Jena and a PhD in applied mathematics at the University of Cologne. She is part of the Bioinformatics group of the Cologne Center for Genomics (CCG) and the main developer of the NGS data analysis pipelines which are used for gene-panel, exome, and genome analyses at CCG. The pipelines are executed on the HPC clusters of the RRZK.

Publications and Products:

- **Nieroda L., Peifer M., Achter V.**, Velder J., **Lang U.**: Application of iRODS metadata management for cancer genome analysis workflow. iRODS UGM 2016 June 8-9, 2016, Chapel Hill, NC.
- **Peifer M.**, Hertwig F., Roels F., Dreidax D., Gartlgruber M., Menon R., Krämer A., …, **Achter V., Lang U., Peifer M.**, et al.: Telomerase activation by genomic rearrangements in high-risk neuroblastoma. Nature 526:700-704, 2015
- George J., Lim J.S., Jang S.J., Cun Y., Ozretic L., Kong G., Leenders F., …, **Peifer M, Achter V., Lang U.**, et al.: Comprehensive genomic profiles of small cell lung cancer. Nature 524:47-53, 2015
- Kawalia, A.; **Motameny, S.**; Wonczak, S.; Thiele, H.; **Nieroda, L.**; Jabbari, K.; Borowski, S.; Sinha, V.; Gunia, W.; **Lang, U.**; **Achter, V.**; Nürnberg, P.: Leveraging the Power of High Performance Computing for Next Generation Sequencing Data Analysis: Tricks and Twists from a High Throughput Exome Workflow. In: PLoS ONE 10 (2015), 5. http://dx.doi.org/10.1371/journal.pone.0126321. – DOI 10.1371/journal .pone.0126321
- Meusemann K., von Reumont B.M., Simon S., Roeding F., Strauss S., Kück P., Ebersberger I., Walzl M., Pass G., Breuers S., **Achter V**., von Haeseler A., Burmeister T., Hadrys H., Wägele J.W., and Misof B. A phylogenetic approach to resolve the arthropod tree of life. ***Mol. Biol. Evol. 27***: 2451-64, (2010).
- **Achter V., Lang U.**, Reuther B., Müller P.. SuGI-Einsatz von Grid in mittleren und kleineren Rechenzentren, 1. DFN Forum Kommunikationstechnologien, 2008, ***GI Lecture Notes in Informatics***, Bonn: Köllen; ISBN 978-3-88579-224-6
- Hauke J., Schild A., Neugebauer A., Lappa A., Fricke J., Fauser S., Rösler S., Pannes A., Zarrinam D., Altmüller J., **Motameny S.**, Nürnberg G., Nürnberg P., Hahnen E., and Beck B.B. A novel large in-frame deletion within the CACNAF1 gene associates with a cone-rod dystrophy 3-like phenotype. PLoS ONE 8 (2013), 10. http://dx.doi.org/10.1371/journal.pone.0076414 - DOI 10.1371/journal.pone.0076414.
- Fernandez-Cuesta L., **Peifer M.**, Lu X., Sun R., Ozretc L, …, **Achter V., Lang U.**, et al.: Frequent mutations in chromatin-remodeling genes in pulmonary carcinoids. Nature Communications 5:1-7, 2014.

Relevant Projects or Activities:

RRZK participated to cooperative projects connected to the subject of this proposal:

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

- **NGSgoesHPC (Skalierbare HPC Lösungen zur effizienten Genomanalyse)** (2011-2014) was a German Ministry of Research (BMBF) funded project. The goal of the project was to optimize algorithms and porting them on modern HPC-infrastructures. In this project the RRZK had the focus on optimizing typical life science assembly and alignment codes to improve their usability on modern and future hardware architectures.

- **SMOOSE (Systemische Analyse von Modulatoren der onkogenen Signalübertragung)** (2014-2016, 2017-2018) is a BMBF funded project. The SMOOSE project propose a multi-disciplinary approach - involving computer science and computational biology, molecular biology, molecular cancer biology and (epi-) genomics, genetically manipulated mice, chemical and structural biology, molecular pathology and medical oncology. In the SMOOSE project RRZK provide expertise and infrastructure for data management and analysis. Computational and database infrastructures are required to ensure efficient data management of large genomic datasets such as genome sequencing. The processing of large genomic datasets involves optimized algorithms and analysis pipelines. The analysis of cancer genomes is usually performed on large scale high-performance computers (HPCs). HPCs allow parallelized processing of the analysis tasks.

- **FaST (Find a Suitable Topology for Exascale Applications)** (2014-2016) is a "Find a Suitable Topology for Exascale Applications" (FaST) is a research project funded by the German Ministry of Education and Research. It deals with the temporal and spatial placement of processes on high performance computers of the future. It is widely assumed that the current trend in hardware development will continue and that the CPU performance will therefore grow considerably faster than the I/O performance. In order to prevent that these resources become bottlenecks in the system, FaST develops a new scheduling concept which monitors the system resources and locally adapts the distribution of the jobs. For monitoring the system a new agent-based system will be developed. The adaptations to the schedule will be realized by process migration. The effectiveness of the concept will be demonstrated in a prototype implementation using applications like LAMA and mpiBLAST.

- **CancerSysDB (The Cancer Systems Biology Database) (**2013–2017**)** is a project funded by the German Research Community. It aims to develop a database for integrative data analysis of molecular cancer datasets. Here, data integration is meant to be cross-data type, cross-cancer type as well as cross-study integration of the data. It will be available as a public instance containing published data from The Cancer Genome Atlas (TCGA) and also as a downloadable software which enables scientists to link a private instance of the CancerSysDB to their own cancer genomics pipelines and thus set up their local infrastructure for data organization and integrative analysis. This will finally speed up the computational procedures in clinical cancer research and open scientists enhanced perspectives on their data. As a development and research partner the RRZK contributed to the following topics: design and development of the software system, acquiring resources for the public instance of the software.

- **HD(CP)²** (**High Definition Clouds and Precipitation for Advancing Climate Prediction)** (2016-2019): The project is a research initiative, funded by the German Federal Ministry of Research, to improve our understanding of atmospheric processes such as humidity and cloud formation. Using a wide range of high quality observation data and new developed high-resolution simulations a more precise climate prediction should be possible. The observation and simulation domain of HD(CP)² is concentrated on Germany and the Netherlands. An important goal for the HD(CP)² module integration is the organization of these observation data in a public accessible data archive. Establishing the technical infrastructure and developing service tools for metadata editing, quality checking and statistical analysis is the contribution of the Regional Computing Centre in Cologne.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

4.2. Third parties involved in the project (including use of third party resources)

1. Bull

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	N
Does the participant envisage that part of its work is performed by linked third parties[1]	N
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N

2. Seagate

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	**Y**
Does the participant envisage that part of its work is performed by linked third parties[1]	N
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N
Seagate Systems UK staff will perform all key research and evaluation activities within the project. Seagate has however built up a team of experts in the development of software specifically skilled in high performance parallel data storage. Some of these individuals are employed through an agency based in France: EURL Tweag, based on 4 Allée de l'Alboni, Ville d'Avray, 92410 France). They will be employed to perform some of the software development activities in Task 4.2 of WP4 and supporting the integration of this with hardware and other software components in ITHACA. As the team is highly skilled in these specialised fields it will be extremely difficult to use alternative individuals and still achieve the project goals.	

3. CEA

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	N
Does the participant envisage that part of its work is performed by linked third parties[1]	N
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N

4. UPV

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	N
Does the participant envisage that part of its work is performed by linked third parties[1]	N
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N

[1] A third party that is an affiliated entity or has a legal link to a participant implying collaboration not limited to the action. (Article 14 of the Model Grant Agreement).

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

5. *UCLM*

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	**Y**
Does the participant envisage that part of its work is performed by linked third parties[1]	N
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N
A subcontract part is provided for the making of videos for the subtask sT2.3.3 of WP2.	

6. *DKRZ*

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	N
Does the participant envisage that part of its work is performed by linked third parties[1]	N
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N

7. *Fraunhofer ITWM:*

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	N
Does the participant envisage that part of its work is performed by linked third parties[1]	N
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N

8. *BSC:*

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	N
Does the participant envisage that part of its work is performed by linked third parties[1]	N
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	**Y**
Some of the work carried out at the Barcelona Supercomputing Center – Centro Nacional de Supercomputación will be contributed free of charge by Third Parties: Universitat Politècnica de Catalunya (UPC), and the Spanish Council for Scientific Research (CSIC). *	

* The BSC is a consortium that is composed of the following member institutions: Universitat Politècnica de Catalunya (UPC), Spanish Council for Scientific Research (CSIC), as well as the Spanish and the Catalan governments. Both UPC and CSIC contribute in kind by making human resources available to work on projects. The relationship between BSC and CSIC / UPC (respectively) is defined in an agreement with each institution that was established prior to the start of this project.

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Universitat Politècnica de Catalunya (UPC)

The High Performance Computing research group of the Computer Architecture Department at the Universitat Politècnica de Catalunya (UPC) is the leading research group in Europe in topics related to high performance processor architectures, runtime support for parallel programming models, performance tuning applications for supercomputing and Cloud Computing.

Directly derived from the research effort at the Computer Architecture Department, the CEPBA (European Center for Parallelism in Barcelona) was founded in 1991 to offer supercomputing resources to the research community and as a development center for industrial computing technology products. In 2000, IBM joined forces with CEPBA to form the CIRI (CEPBA-IBM Research Institute Joint Lab) in Barcelona in order to strengthen relationships between IBM and UPC researchers in computer architecture.

In 2005, the Spanish and Catalan governments signed an agreement with IBM to buy the 4th supercomputer in the world and extend the operations of CIRI to become the Barcelona Supercomputing Center (BSC).

The High Performance Computing research group at the UPC shares many key resources with the BSC, including several key personnel that will be dedicated to this project. There is a signed Collaboration Agreement between the UPC and the BSC establishing the framework of the relationship between these two entities. According to this agreement, several professors of the UPC are made available to the BSC to work on projects.

Consejo Superior de Investigaciones Científicas (CSIC)

The objective of the Spanish Council for Scientific Research (CSIC) is to promote, coordinate, develop and disseminate the scientific and technology research of a multidisciplinary nature, with the aim of contributing to the pursuit of higher knowledge as well as economic, social and cultural development in Spain. It also promotes training of personnel and consultancy to public and private institutions.

CSIC researchers carry out their work at universities and research centers based in Spain, institutions with which CSIC actively collaborates. This collaboration takes place within the framework of long-term agreements, ensuring that CSIC researchers are fully integrated into teams and research projects. CSIC has signed collaboration agreements with several entities, including the BSC. The CSIC researchers at the BSC will provide expertise in the areas of programming models and performance analysis and modelling of applications to this project.

Dra. Rosa M. Badia is a CSIC researcher affiliated with the BSC. He/she carries out his/her research in association with the Barcelona Supercomputing Center - Centro Nacional de Computación on the BSC premises.

9. Allinea:

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	**Y**
Does the participant envisage that part of its work is performed by linked third parties[1]	**Y**
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N
Some parts of WP6 and WP7 will be performed by the affiliate Allinea Software GMBH agency. (100% owned by Allinea's UK HQ).	

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

10. *SURFsara:*

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	N
Does the participant envisage that part of its work is performed by linked third parties[1]	N
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N

11. *INRIA:*

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	N
Does the participant envisage that part of its work is performed by linked third parties[1]	**Y**
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N
In this Project, INRIA represents the Team DATAMOVE DATAMOVE Team is a Joint Research Unit for which INRIA will represent Université Grenoble Alpes (UGA), a Third Party linked to Inria in future Grant Agreement and Consortium Agreement: UGA as a Third Party linked to Inria, will carry out part of the work attributed by the future Grant Agreement. They will fill Third Party's Financial Reports with their own costs. UGA may charge costs related to the expenses of Pierre François Dutot.	

12. *ARM:*

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	N
Does the participant envisage that part of its work is performed by linked third parties[1]	N
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N

13. *University of Cologne:*

Does the participant plan to subcontract certain tasks (please note that core tasks of the project should not be sub-contracted)	N
Does the participant envisage that part of its work is performed by linked third parties[1]	N
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	N

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Section 5: Ethics and Security

5.1 Ethics

The research carried out as part of ITHACA does not enter any ethics issues in the ethical issue table in the administrative proposal forms.

Sex/gender analysis:
We provide one example of how the project's sex/gender analysis is addressed by ITHACA coordinator, Bull (France) in its programs. The industrial and research partners in ITHACA have similar programs.

Bull

The vast majority of the Bull workforce is in Europe. It essentially operates within a highly structured social, cultural and legislative environment, with strict standards when it comes to human resources. Bull adheres to national and international principles and recommendations in the areas of human rights and labor law.
Bull's social contract is built on three pillars: the recruitment of young people whom the group wants to train in the latest technology and equip with the right skills so that they can develop their capacity for innovation; opportunities for varied and lifelong career development prioritizing workplace wellness to facilitate commitment and entrepreneurial spirit.
The value added offered by Bull's solutions partly derive from the group's ability to design innovative solutions based on its technological and functional expertise. Bull is therefore developing a policy built on gender diversity, the employment of older workers and the integration of employees with disabilities. It is committed to employing individuals with disabilities, with an employment rate well above the average among IT companies, especially in France, with a particular focus on people who are blind or visually impaired.
Two policy components were instituted to promote gender diversity, the first guarantees fair treatment from the standpoint of remuneration in the group, as was the case in France, since 2012. The second was the signature of an agreement with the social partners on professional equality between women and men.
Lastly, Bull is committed to a policy of employing older workers, again in association with its social partners. Apart from signing a company-wide agreement, the action taken to assist people in the later stages of their careers guarantees equivalent career paths regardless of age, especially in terms of training and pay.

5.2 Security[2]

Please indicate if your project will involve:

- activities or results raising security issues: **NO**
- 'EU-classified information' as background or results: **NO**

The research carried out as part of ITHACA does not involve activities or results raising security issues; nor 'EU-classified information' as background or results.

[2] Article 37.1 of Model Grant Agreement. *Before disclosing results of activities raising security issues to a third party (including affiliated entities), a beneficiary must inform the coordinator — which must request written approval from the Commission/Agency;* Article 37. *Activities related to 'classified deliverables' must comply with the 'security requirements' until they are declassified; Action tasks related to classified deliverables may not be subcontracted without prior explicit written approval from the Commission/Agency.; The beneficiaries must inform the coordinator — which must immediately inform the Commission/Agency — of any changes in the security context and — if necessary —request for Annex 1 to be amended (see Article 55)*

This proposal version was submitted by **bruno FARCY** on **27/09/2016 16:27:49** Brussels Local Time. Issued by the Participant Portal Submission Service.

Digitally signed by sealingservice.grants.ec.europa.eu
Date: 2016.09.27 16:28:15 CEST
Reason: Acknowledgment of Receipt

This electronic receipt is a digitally signed version of the document submitted by your organisation. Both the content of the document and a set of metadata have been digitally sealed.

This digital signature mechanism, using a public-private key pair mechanism, uniquely binds this eReceipt to the modules of the Participant Portal of the European Commission, to the transaction for which it was generated and ensures its full integrity. Therefore a complete digitally signed trail of the transaction is available both for your organisation and for the issuer of the eReceipt.

Any attempt to modify the content will lead to a break of the integrity of the electronic signature, which can be verified at any time by clicking on the eReceipt validation symbol.

More info about eReceipts can be found in the FAQ page of the Participant Portal. (http://ec.europa.eu/research/participants/portal/page/faq)