

What do we know about BSC infrastructure?

Authors: Núria Pérez-Zanón and An-Chi Ho (December 2021)
Updated by Victòria Agudetse (May 2024)



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

What do we know about BSC-ES infrastructure?

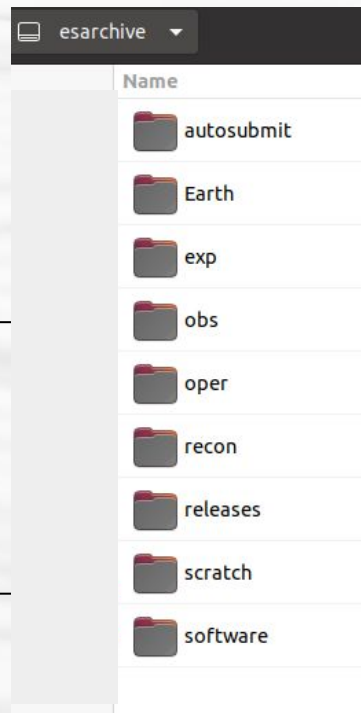
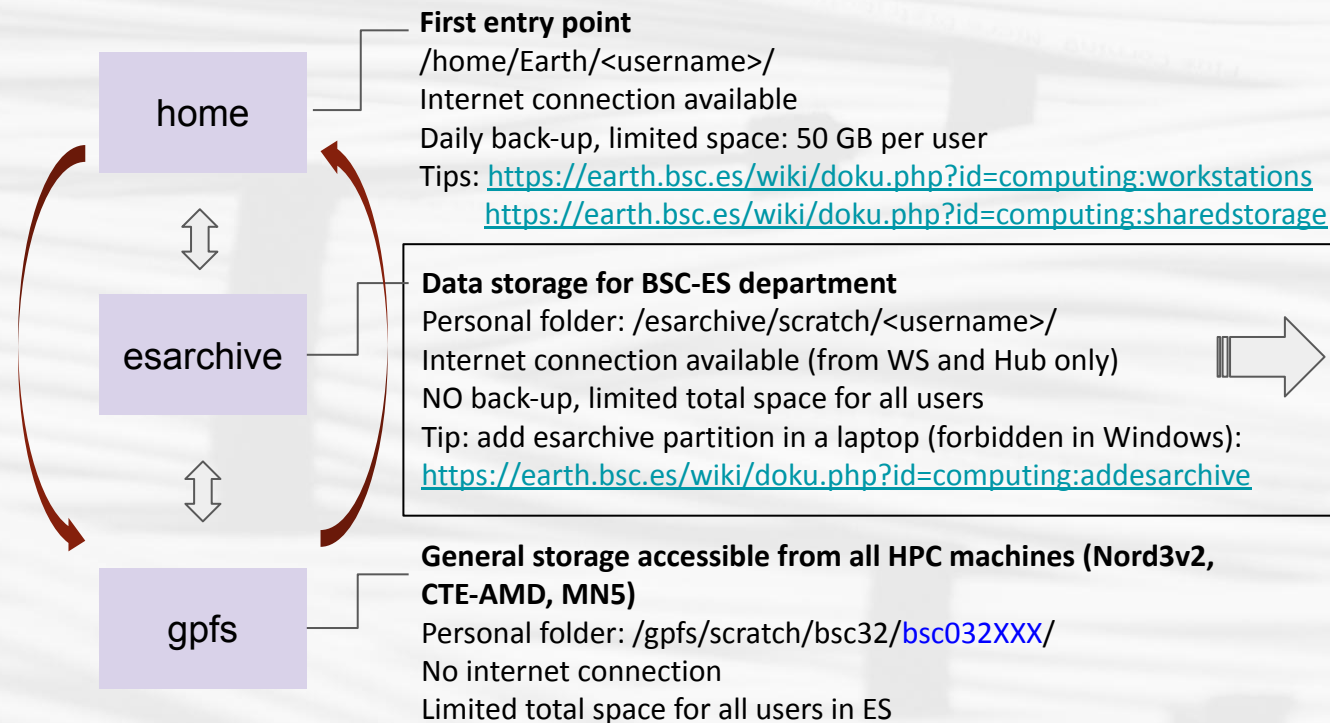
Aside from the data and software in our personal laptops, we all have access to common BSC infrastructure.

We access the BSC infrastructure:

- ★ When we connect to the [BSC-ES Hub](#)
- ★ When we use the [workstations](#) in the office
- ★ When we connect remotely via ssh to a workstation (bscearthXXX.int.bsc.es)
 - To ssh from windows:
<https://earth.bsc.es/wiki/doku.php?id=computing:sshwindows>
 - To set up passwordless ssh connection:
<https://earth.bsc.es/wiki/doku.php?id=computing:sshkeyautologon>
- ★ When we connect to one of the servers or HPC machines in BSC (MN5, Nord3v2, etc.)

What do we know about BSC-ES infrastructure?

When we connect to the BSC infrastructure, we find several **partitions**. A disk partition, or simply 'partition', is a segment of a hard drive that is separate and independent from other segments. Each partition serves a different purpose and is accessible from different machines.



What do we know about BSC-ES infrastructure?

It is also possible to connect to BSC infrastructure through **servers** (physical machines), which have different uses:

- ★ **bscearth000.int.bsc.es and bscearth001.int.bsc.es**
 - Download data
 - run the automatic package tests (GitLab CI/CD, see e.g.: <https://earth.bsc.es/gitlab/es/s2dv/-/pipelines>)
- ★ **transfer1.bsc.es (formerly dt01.bsc.es and dt02.bsc.es)**
 - Internal transfer of data, e.g. from esarchive to GPFS and vice versa.
- ★ **bscesshiny01.bsc.es**
 - Shiny server, hosts shiny apps.
- ★ **bscesftp.bsc.es**
 - Share files externally, see: https://earth.bsc.es/wiki/doku.php?id=computing:public_ftp
- ★ **bscesautosubmit01.bsc.es and bscesautosubmit02.bsc.es**
 - Launch workflows with the Autosubmit workflow manager <https://earth.bsc.es/wiki/doku.php?id=tools:autosubmit>

What do we know about BSC-ES infrastructure?

A **software stack** is the collection of programs and modules (including the operating system, architectural layers, protocols, runtime environments, ...) that are installed in a machine.

- ★ The software stack at BSC can be different among different machines and departments
- ★ We have access to:
 - BSC software stack (not managed by CES)
 - BSC-ES software stack (managed by CES)
 - Workstations, Nord3v2 and CTE-AMD already using it
 - Hub has a slightly different software stack (more updated, but on testing status)
 - In some machines, we should edit the **bashrc** to use it (instructions are always in the wiki: <https://earth.bsc.es/wiki/doku.php?id=library:computing>)
 - It is built on **modules**, some useful commands are:
 - *module list* # show all loaded modules
 - *module load ** # load the '*' module
 - *module av ** # show all available modules matching '*'
 - other software programs like mendeley can be open in the workstation:
/shared/earth/software/mendeley/latest/bin/mendeleydesktop
- ★ Open an issue in [the Requests GitLab](#) to ask for new software or R packages

What do we know about BSC-ES infrastructure?

What information do we need to know for each machine?

- Does it have BSC-ES software?
- is /esarchive/ mounted?
- Internet access?
- Job scheduler: slurm, Isf...?
- Memory per node, cores per node....

Hub
Workstations (WS)
Marenostrum 5
AMD cluster
Nord3_v2

Find the information here: <https://earth.bsc.es/wiki/doku.php?id=library:computing>

What do we know about BSC-ES infrastructure?

Workstations

- R/4.1.2
- To be used for debugging code (small data) or running startR workflows in remote machines
- Internet connection
- BSC-ES software stack
- /esarchive is mounted

Hub (testing phase)

- R/4.2.1
- To be used for debugging code (small data) or running small jobs. Will replace workstations.
- Internet connection
- BSC-ES software stack
- /esarchive is mounted

Nord3_v2

- R/4.1.2
- To be used to run more memory-intensive jobs
- job scheduler: **slurm**
- No internet connection
- BSC-ES software stack
- /esarchive is mounted
- **will be decommissioned (when?)**

Marenostrum 5

- 'Pre-pre-production' status
- To be used to run more memory-intensive jobs
- **BSC-ES software stack currently not available**, conda environments can be installed
- internet access in login node 4
- no access to /esarchive (non-negotiable)

CTE-AMD

- R/4.1.2 or R/4.3.3 (for R-INLA)
- To be used to run more memory-intensive jobs
- job scheduler: **slurm**
- BSC-ES software stack
- no access to /esarchive (for now?)

Nord4

- Coming soon?
- /esarchive?

What do we know about BSC-ES infrastructure?

Recommendations

★ Save your scripts in GitLab (intermediate and final versions)

- In an existing GitLab project
- In a personal project
- Documentation: <https://earth.bsc.es/wiki/doku.php?id=library:computing#git>
- If you have internet connection, you can source your code directly from GitLab
- Clone repositories under /esarchive/scratch/<username>/
 - You will have internet connection to push your changes
 - The code will be accessible from workstations, hub and Nord3v2
 - There is no back-up copy of /esarchive (another good reason to use gitlab)

★ Don't install local versions of R packages

- If you do, we cannot debug the code and reproduce the errors
- Better to open an issue in Requests to ask for the installation: it's easier to debug and everyone can use it

★ Infrastructure in the wiki:

https://earth.bsc.es/wiki/doku.php?id=library:best_practices#network_infrastructure

What do we know about BSC-ES infrastructure?

Q&A: What else do we need to know? What questions do we have?

- Q: When will we be able to use Nord4?

A: There is no official date yet.

- Q: Will Nord4 and/or CTE-AMD have /esarchive mounted?

A: It is currently being negotiated, it is likely that at least one of the two may have it, but we do not know for sure.

- Q: Can we use RStudio in the Hub?

A: Not right now, but CES is working on it. Requests issue:

<https://earth.bsc.es/gitlab/es/requests/-/issues/2154>