Efficient data handling in the ClimateDT ecosystem

Ivan Alsina Ferrer - BSC

Earth Sciences Department

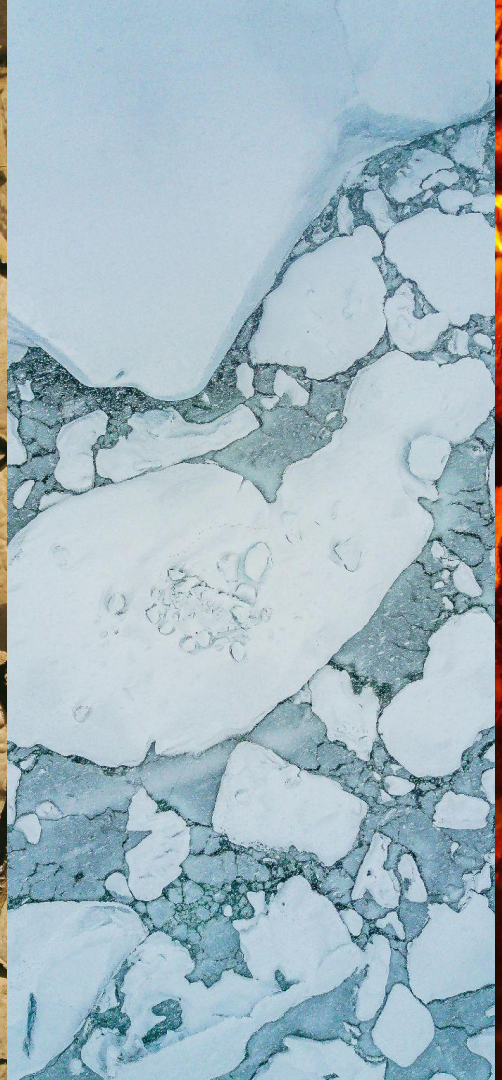
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

Destination Earth

Our workflow (manager)

One-Pass Algorithm (OPA)

Destination Earth

Flagship initiative of the European Commision

Destination Earth

Flagship initiative of the European Commision to develop a highly accurate digital twin of the Earth.

Destination Earth

Flagship initiative of the European Commision to develop a highly accurate digital twin of the Earth.

Implemented by:

Partnering with more than 90 institutions in 25 countries

Earth Sciences Department | BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación

Destination Earth

Destination Earth

Support climate change adaptation and mitigation strategies

Destination Earth

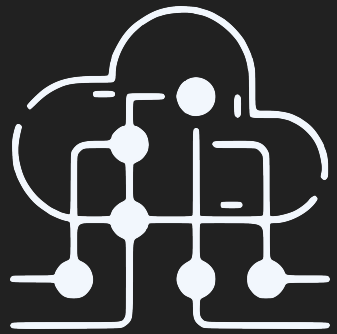
Support climate change adaptation and mitigation strategies by providing highly accurate and interactive simulations

Destination Earth

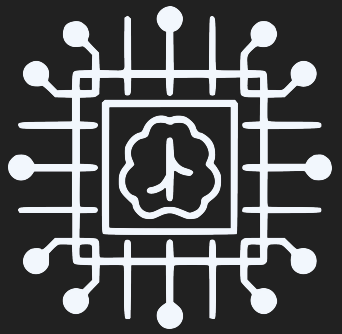
Support climate change adaptation and mitigation strategies by providing highly accurate and interactive simulations of the Earth system and of the impact of human activities on our planet.

Components

DIGITAL TWIN &
DIGITAL TWIN ENGINE



DATALAKE



CORE SERVICE
PLATFORM

DIGITAL TWIN &
DIGITAL TWIN ENGINE

Detail

Quality

Interactivity

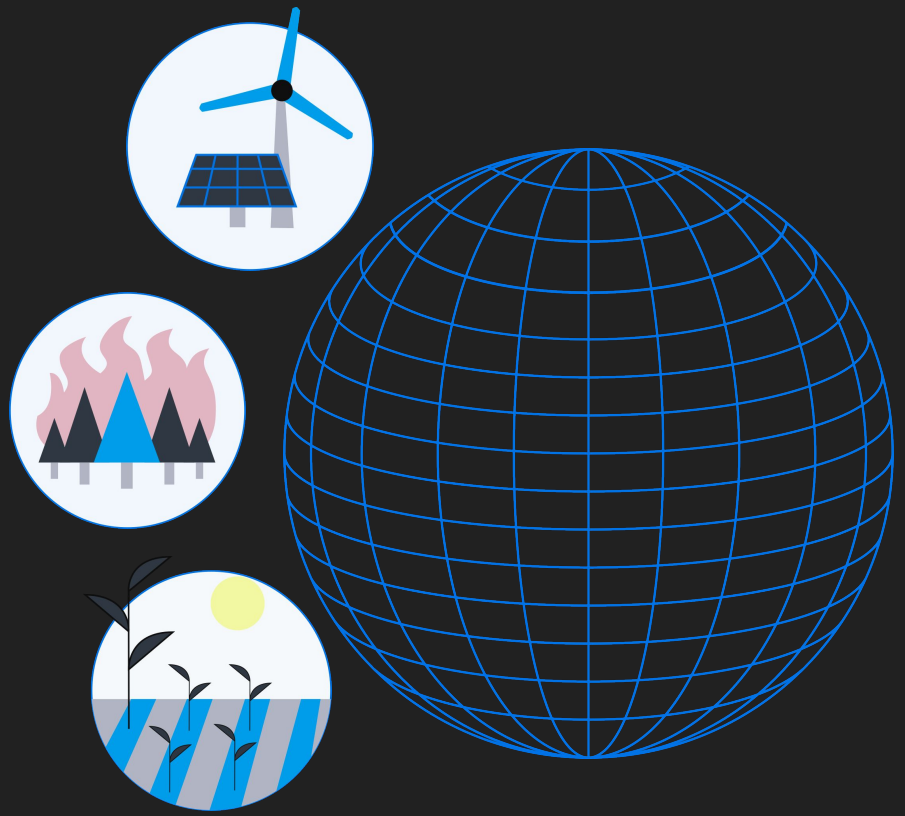
Earth Sciences Department

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

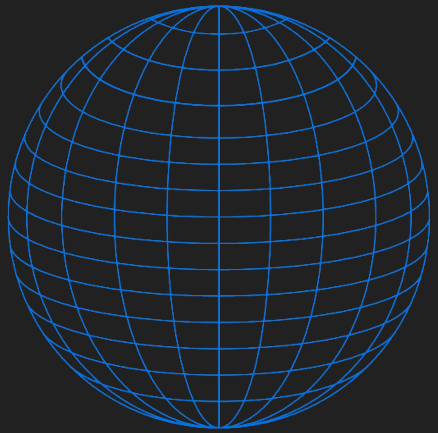
Destination Earth

Digital Twin Engine

Co-design approach with selected data consumers



Digital Twin Engine

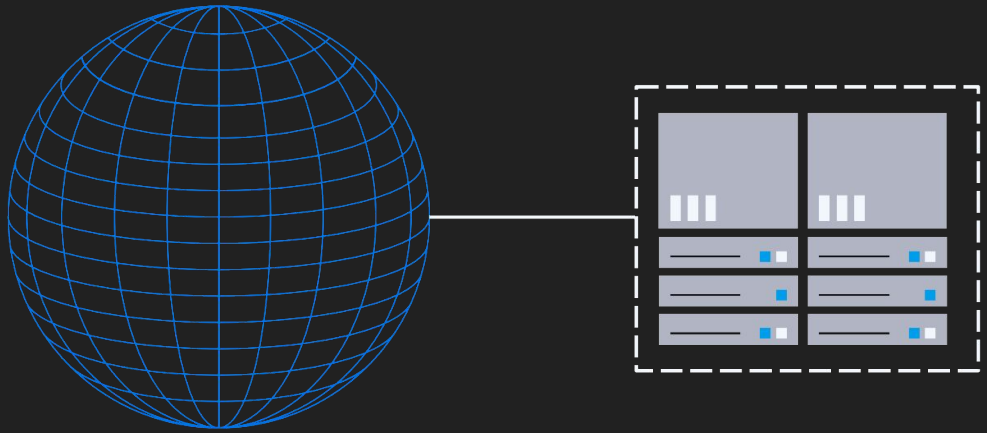


Earth Sciences Department

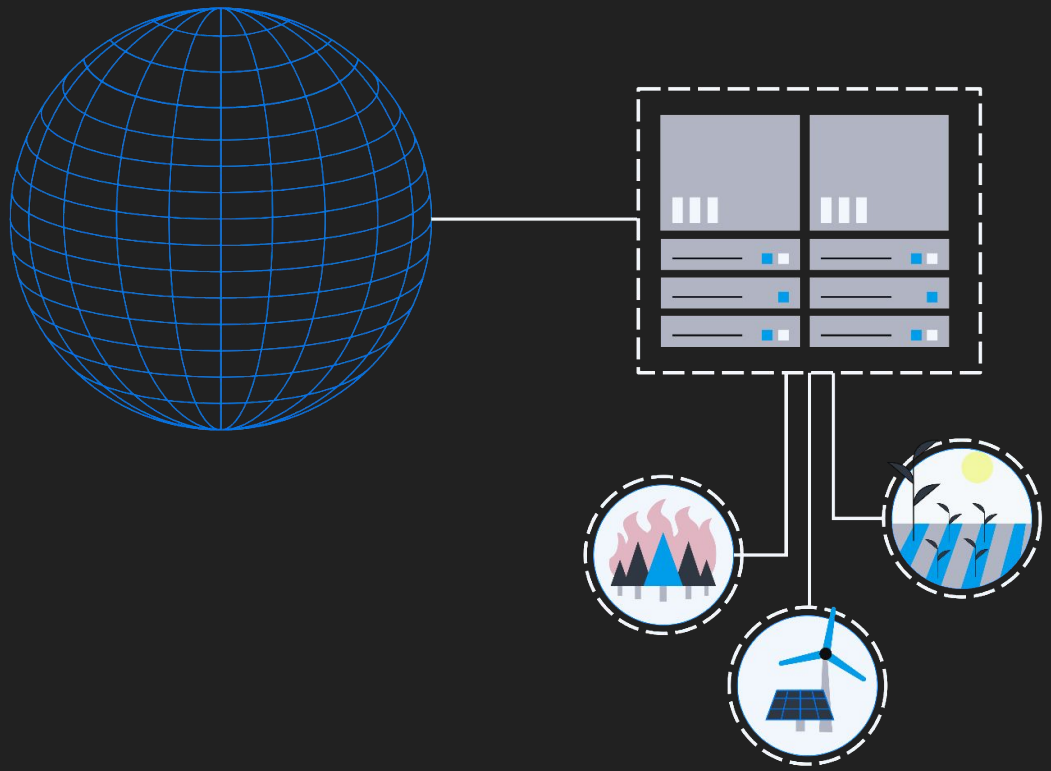
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

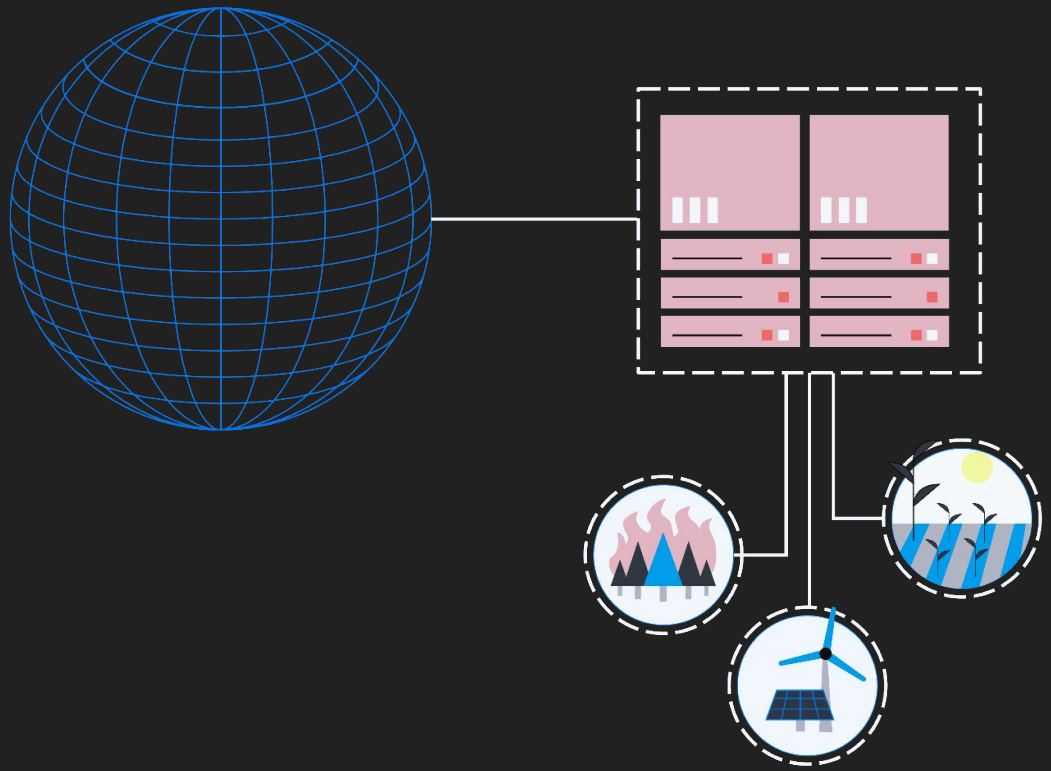
Digital Twin Engine



Digital Twin Engine



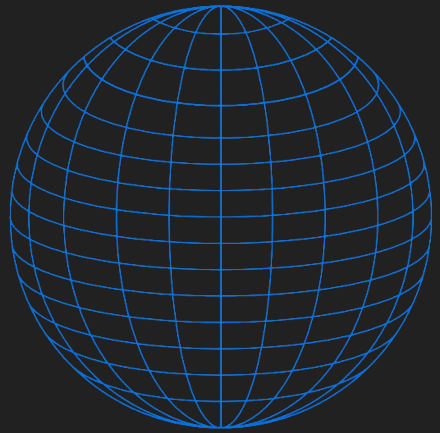
Digital Twin Engine



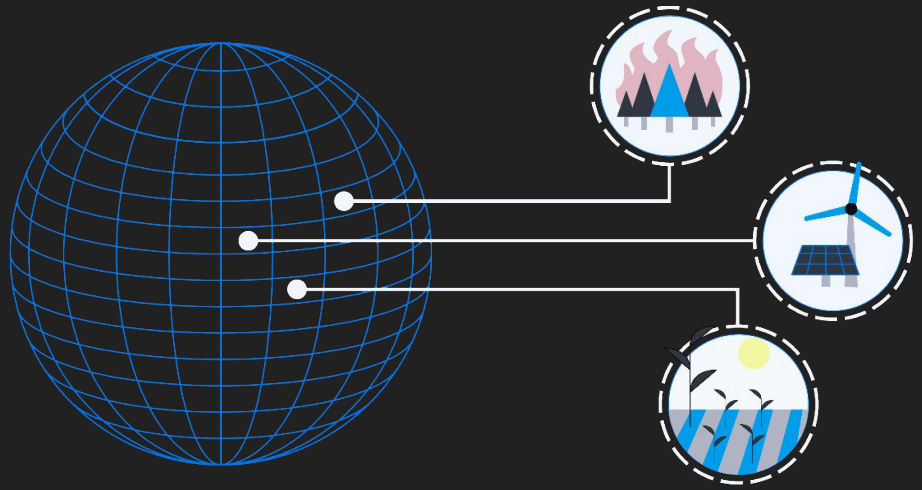
Data sizes

	DAY	YEAR
Reduced portfolio	10 GB	3.5 TB
Full portfolio (IFS-based)	110 GB	39.2 TB
Full portfolio (ICON)	80 GB	28.5 TB

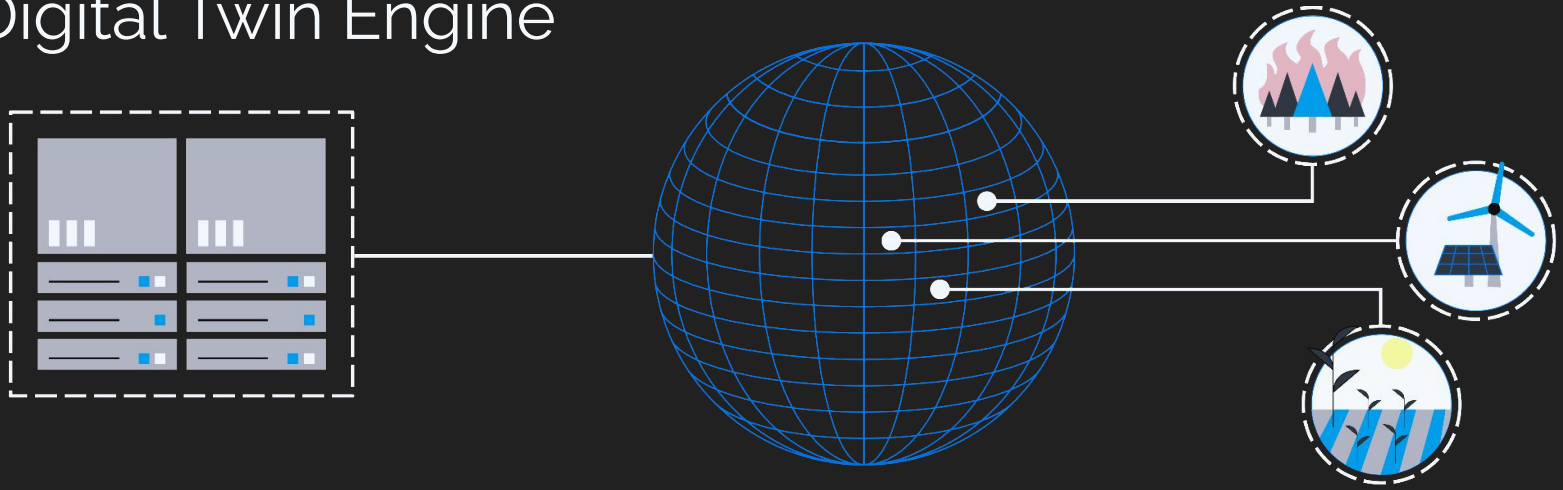
Streaming concept



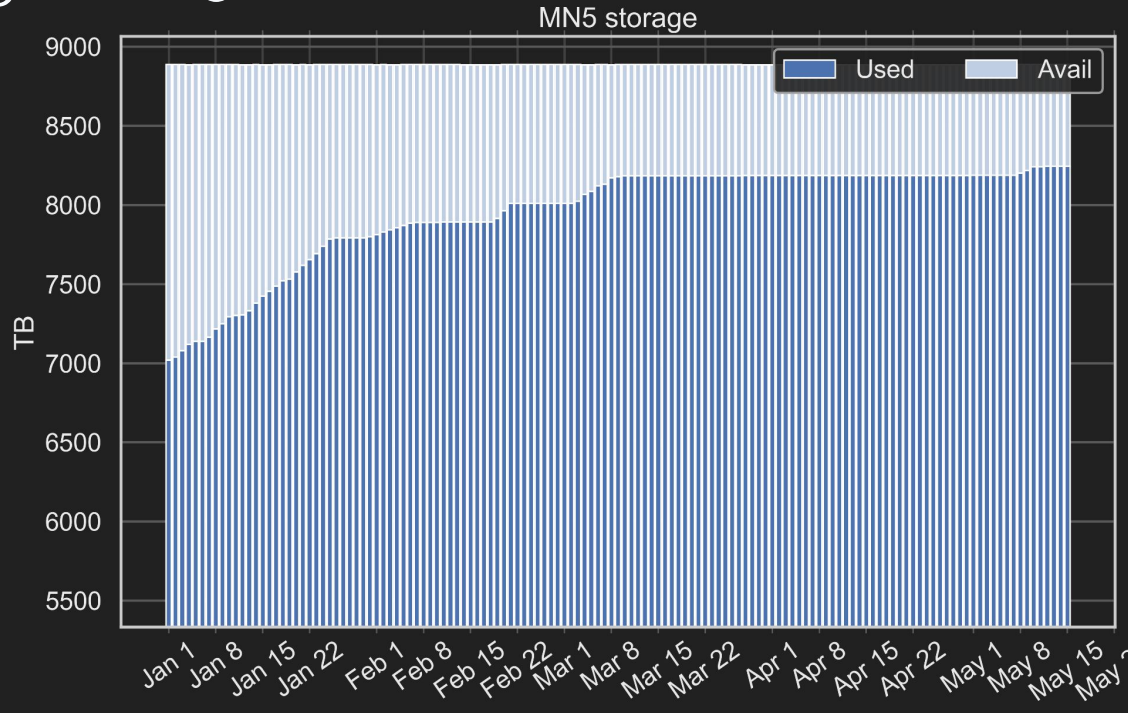
Streaming concept



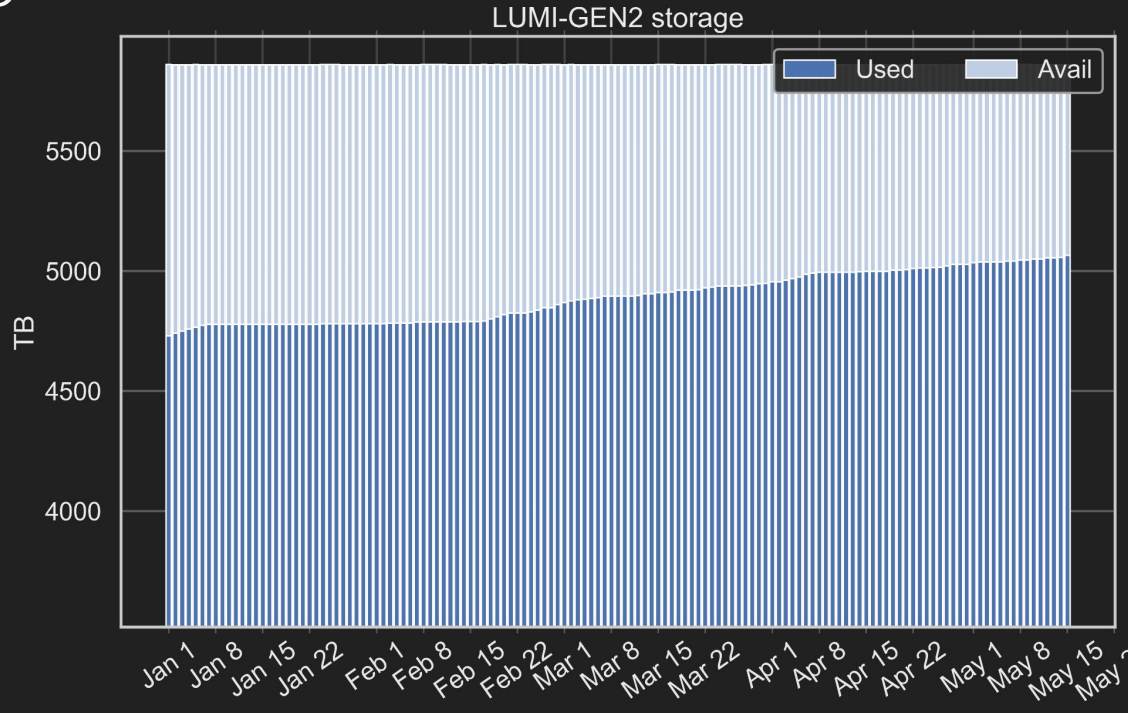
Digital Twin Engine



Data usage: MN5



Data usage: LUMI



Our workflow engine

AUTOSUBMIT

Earth Sciences Department

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

Our workflow engine

AUTOSUBMIT

Complex setup, ***structured configuration***

DEFINITION
PARAMETERIZATION
REPRODUCIBILITY

Our workflow engine

AUTOSUBMIT

Complex setup, ***structured configuration***

DEFINITION
PARAMETERIZATION
REPRODUCIBILITY

Many platforms, ***centralized execution engine***

CONTROL
LOGGING
TRACEABILITY

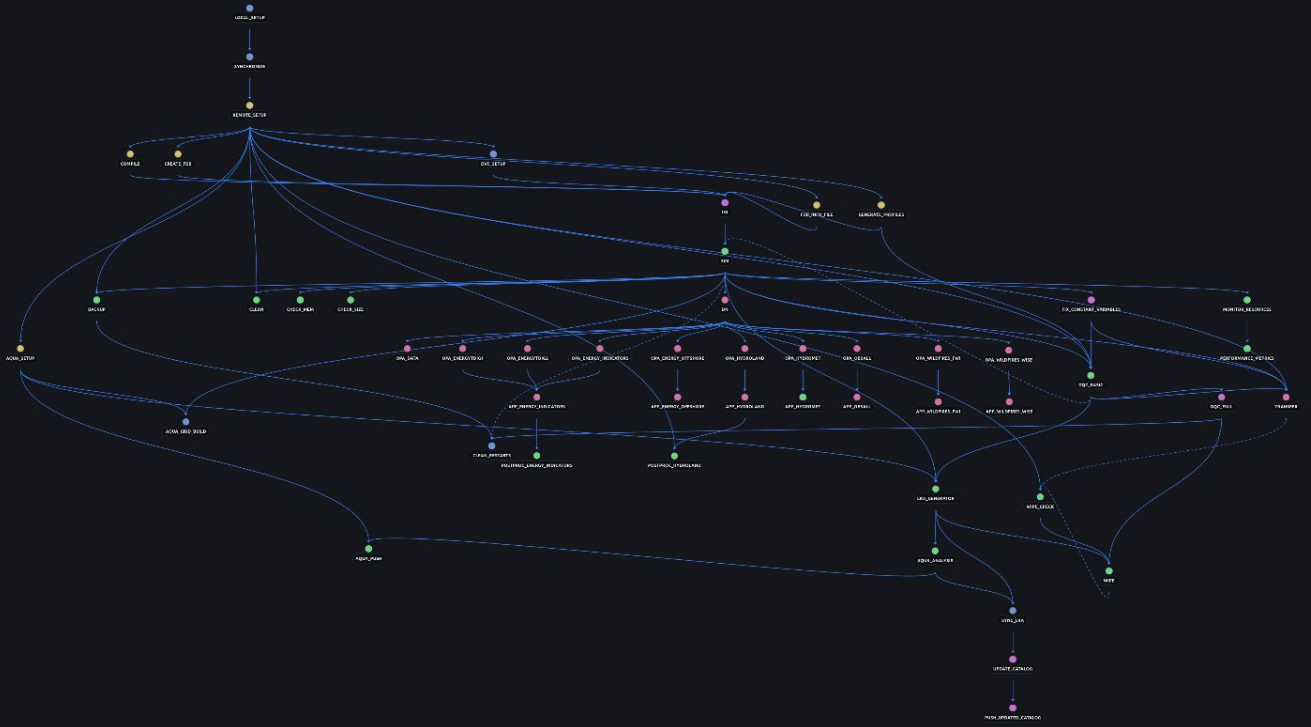
Our workflow engine

AUTOSUBMIT

Complex setup, ***structured configuration***	Many platforms, ***centralized execution engine***	Long runs, ***clear monitoring and analysis***
DEFINITION PARAMETERIZATION REPRODUCIBILITY	CONTROL LOGGING TRACEABILITY	VISUALIZATION STATISTICS PERFORMANCE METRICS

Our workflow

Our workflow

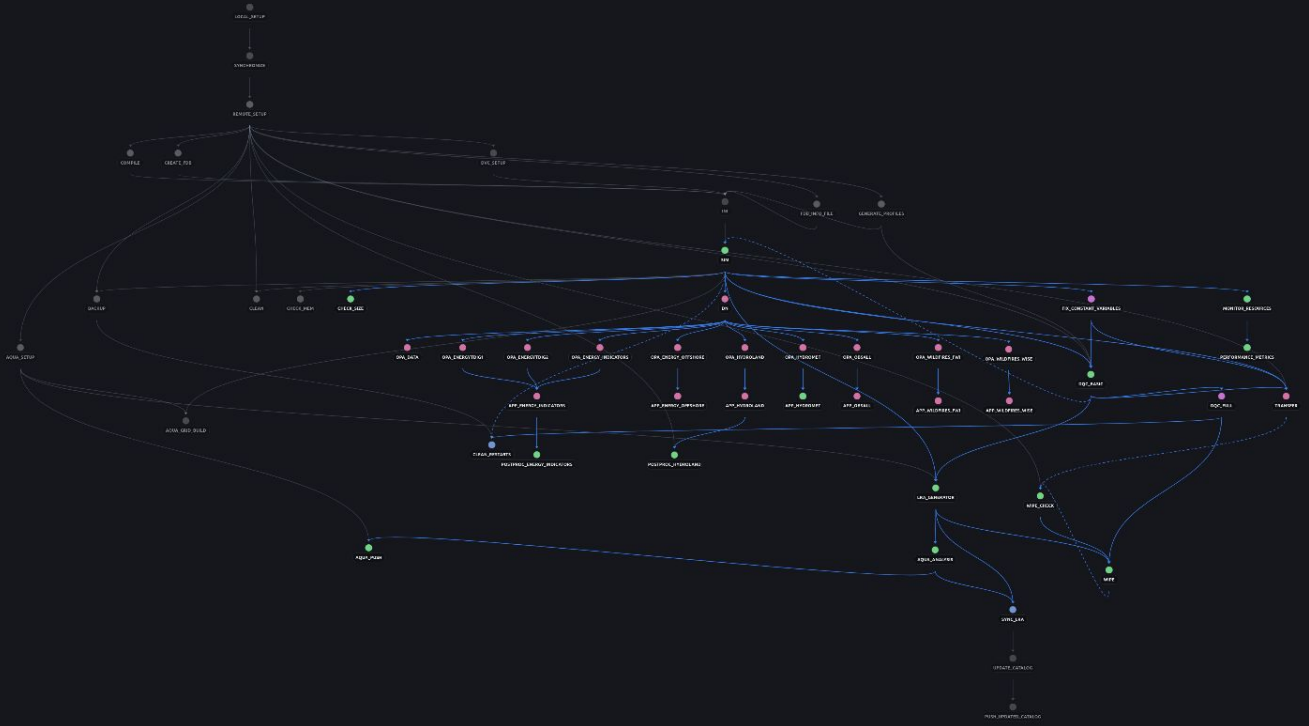


Earth Sciences Department

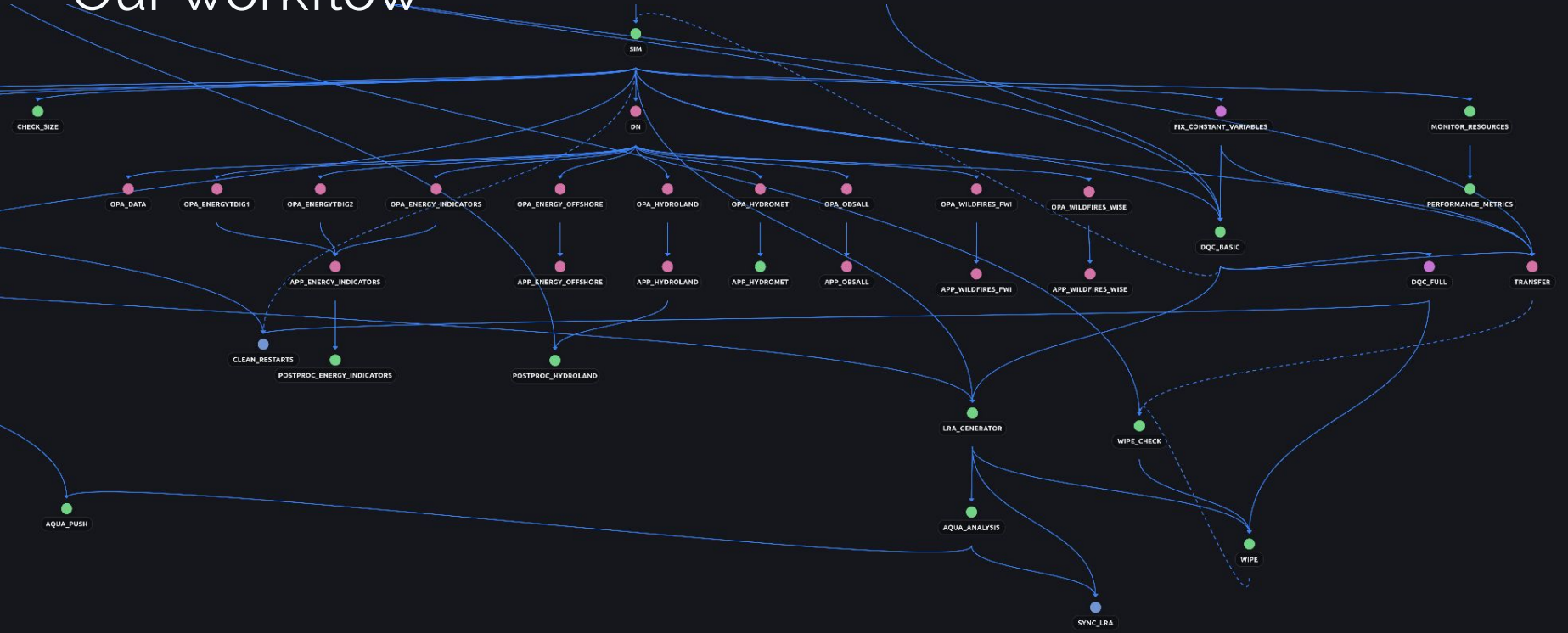
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

Our workflow



Our workflow

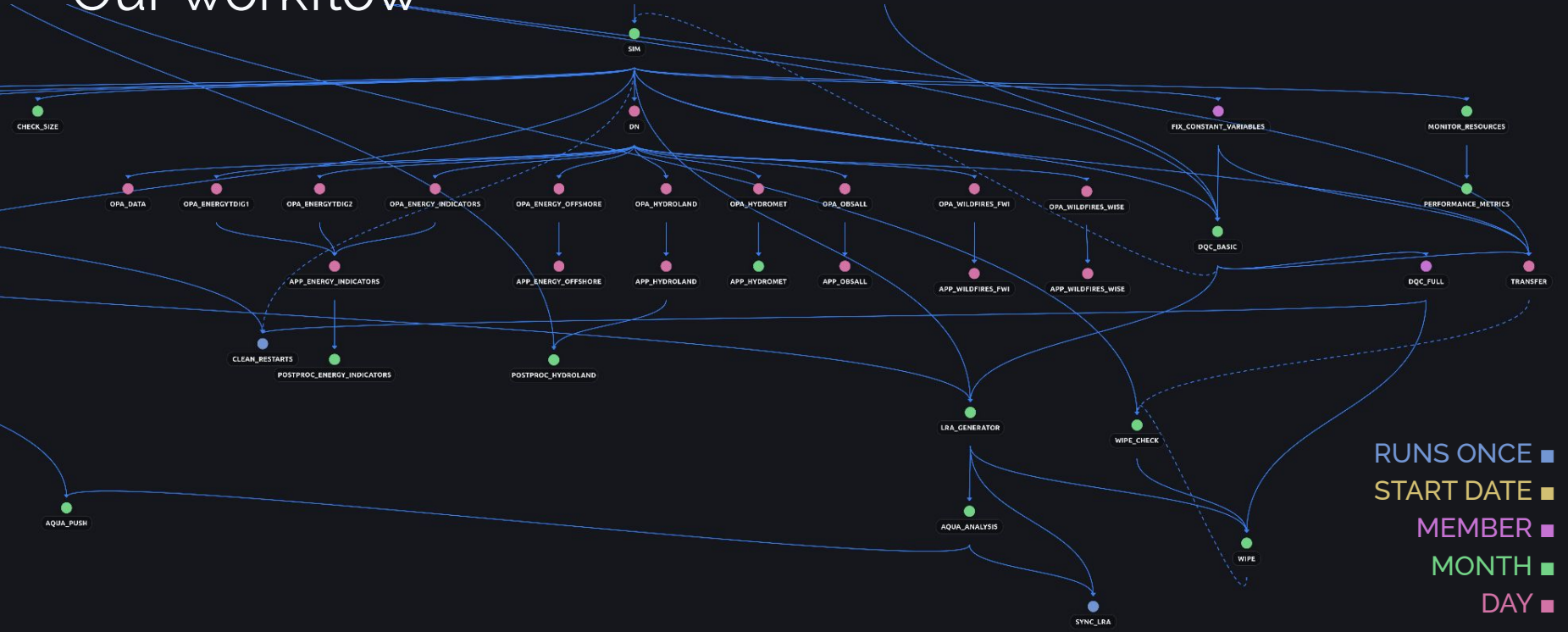


Earth Sciences Department

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

Our workflow

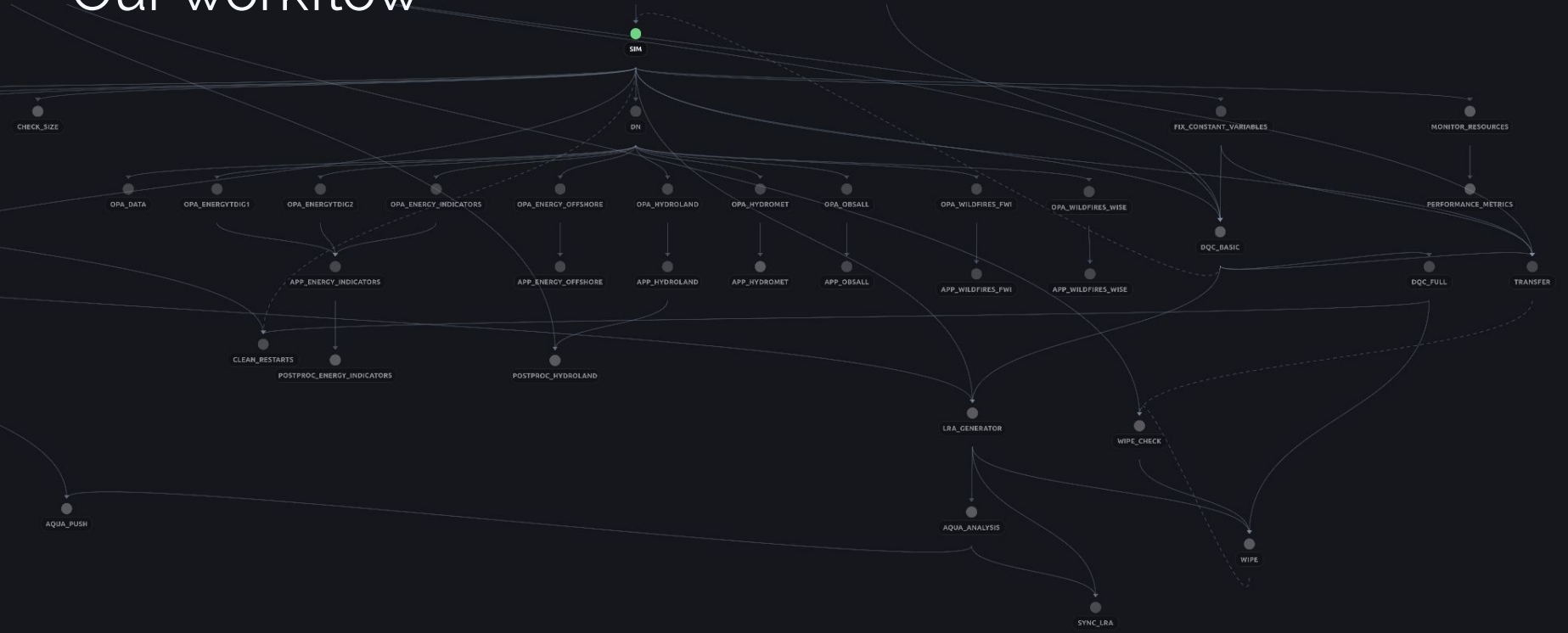


Earth Sciences Department

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

Our workflow

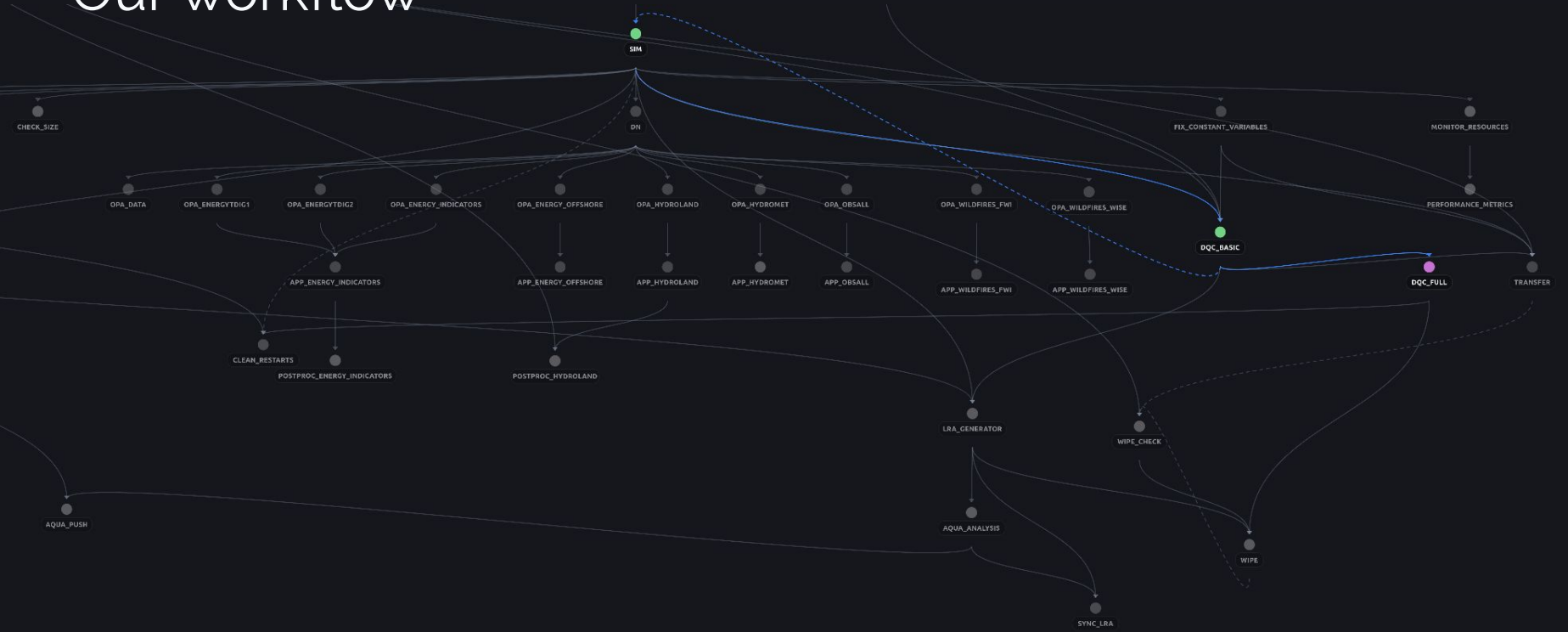


Earth Sciences Department

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

Our workflow

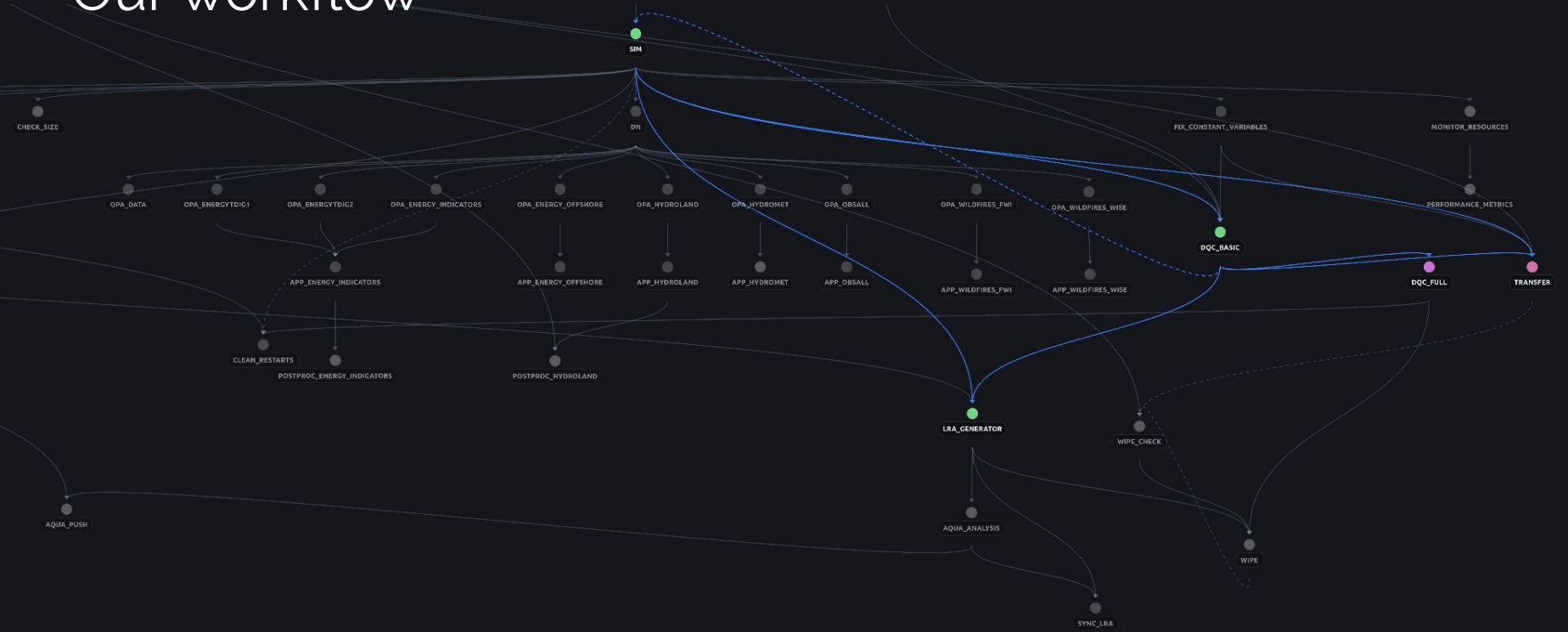


Earth Sciences Department

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

Our workflow

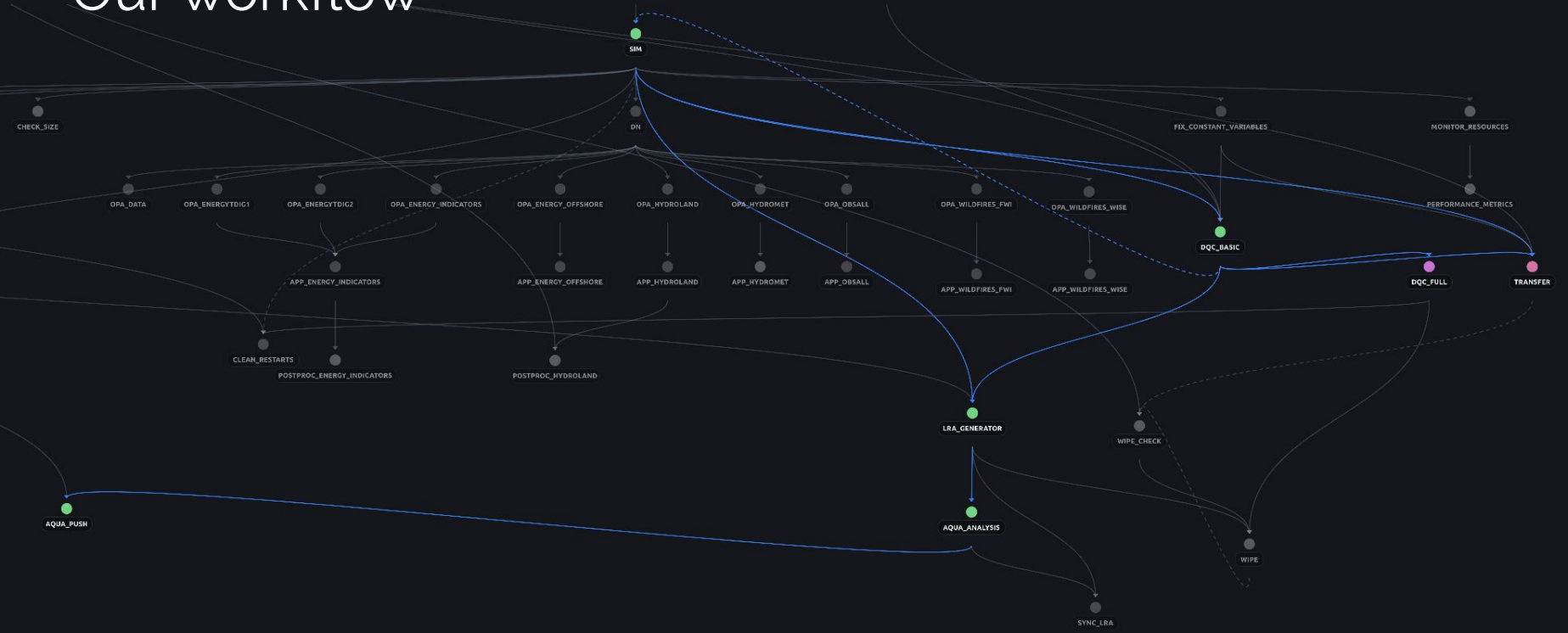


Earth Sciences Department

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

Our workflow

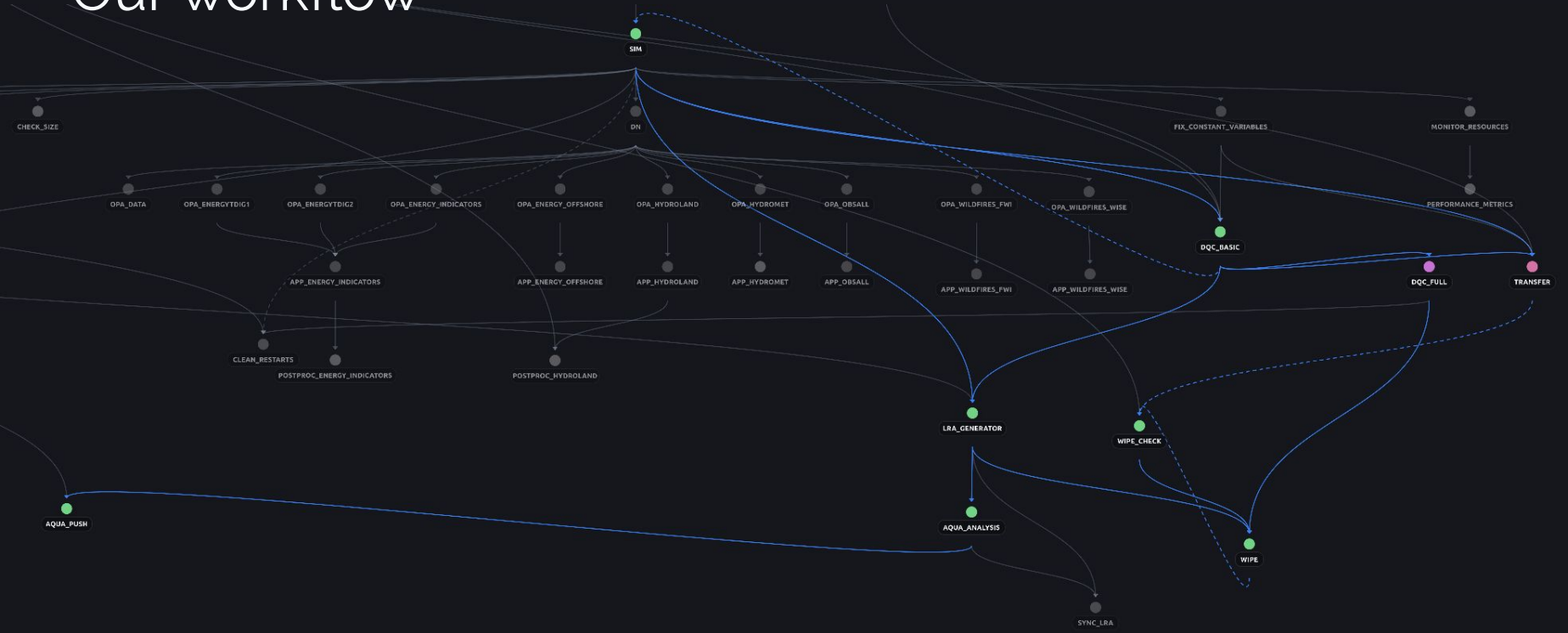


Earth Sciences Department

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

Our workflow

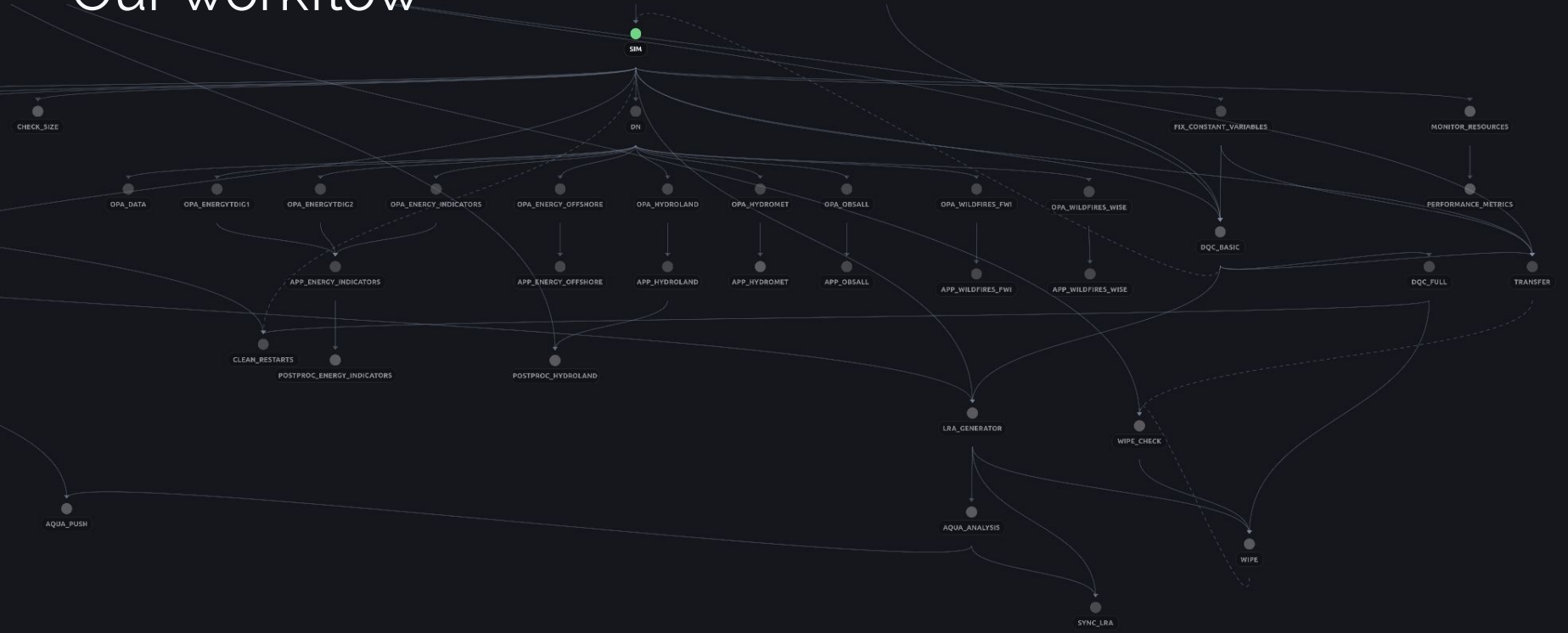


Earth Sciences Department

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

Our workflow

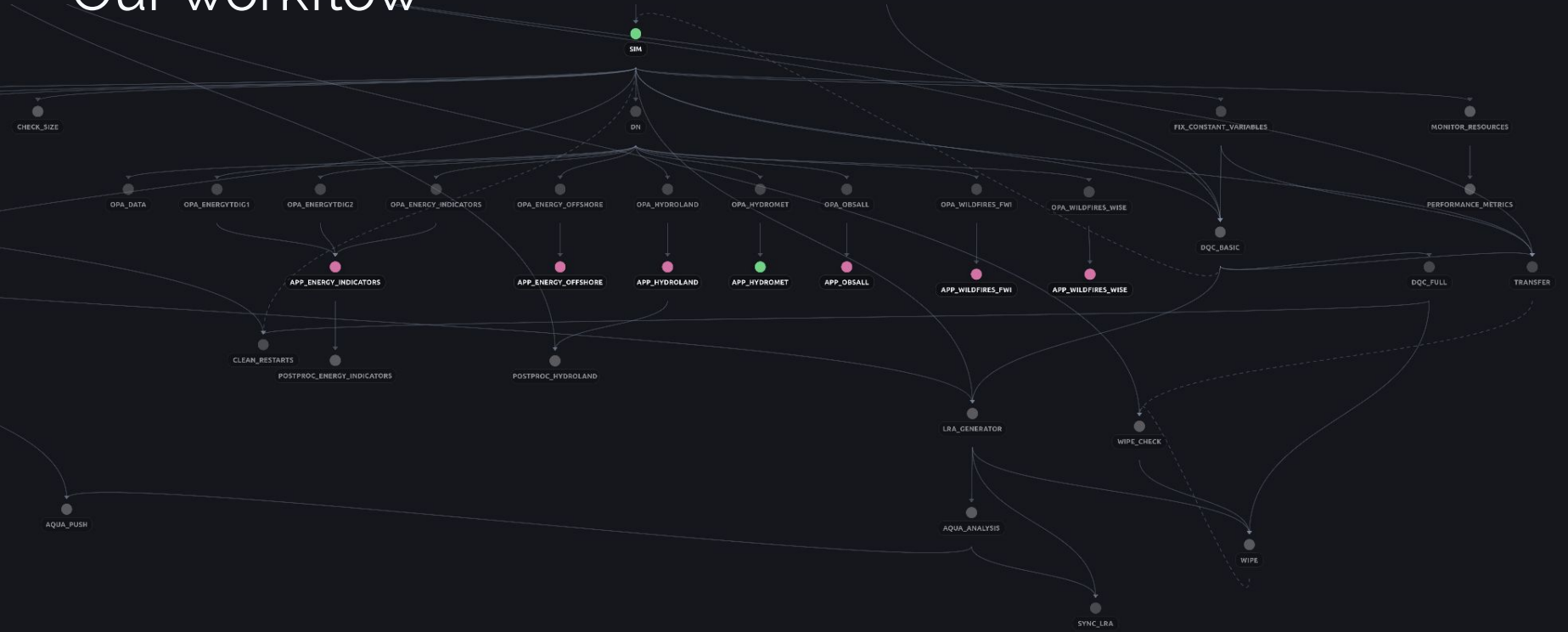


Earth Sciences Department

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

Our workflow

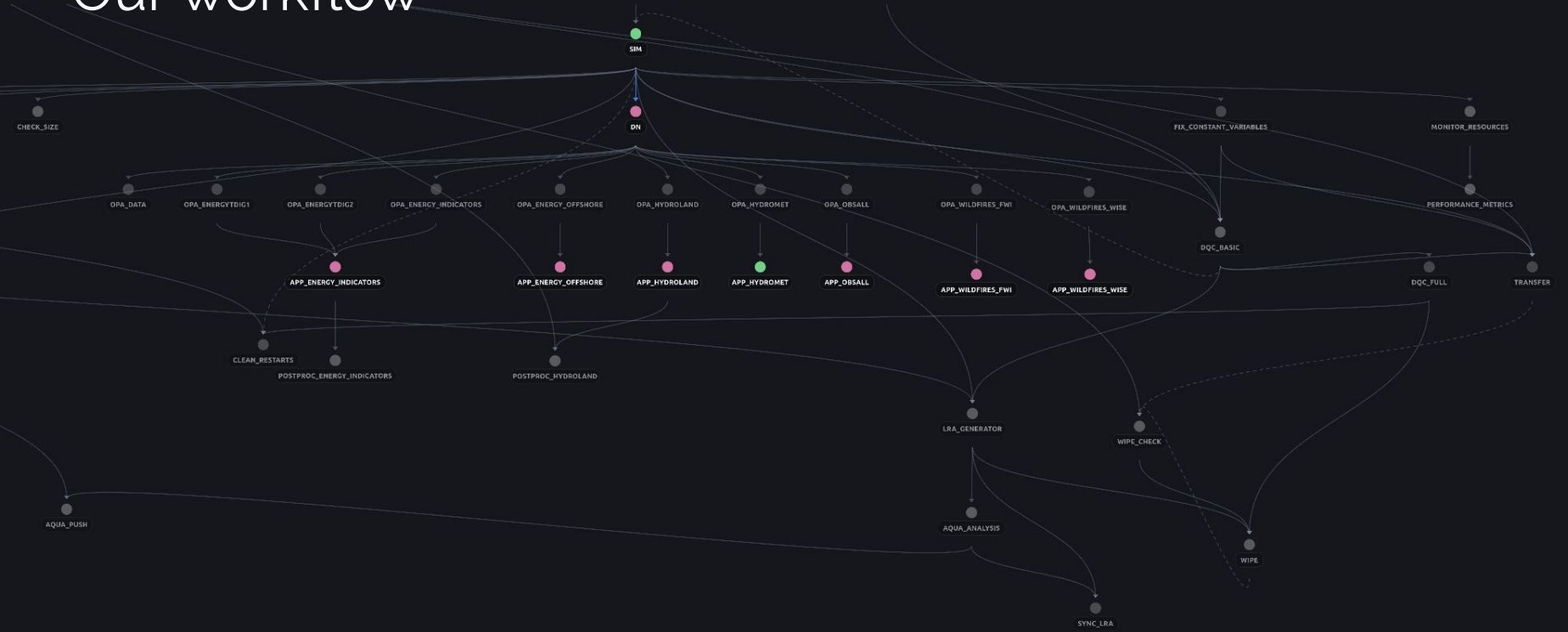


Earth Sciences Department

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

Our workflow

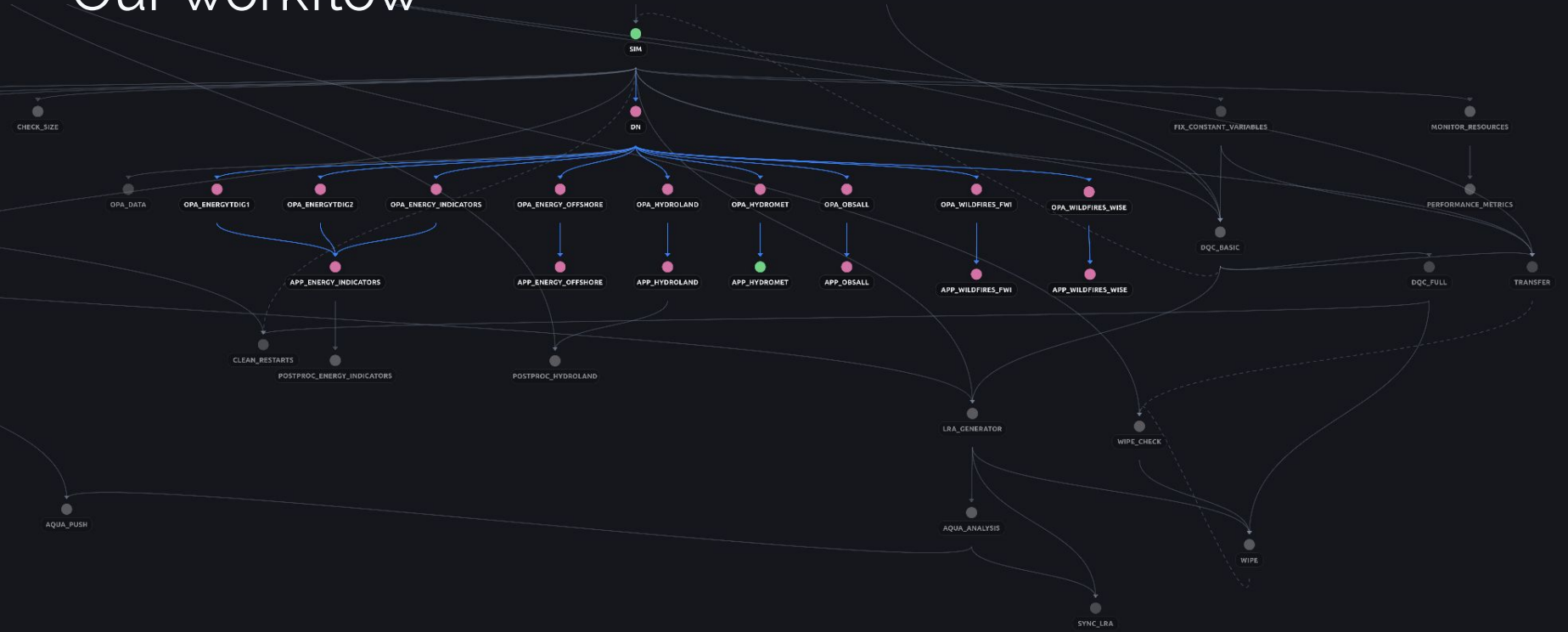


Earth Sciences Department

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

Our workflow



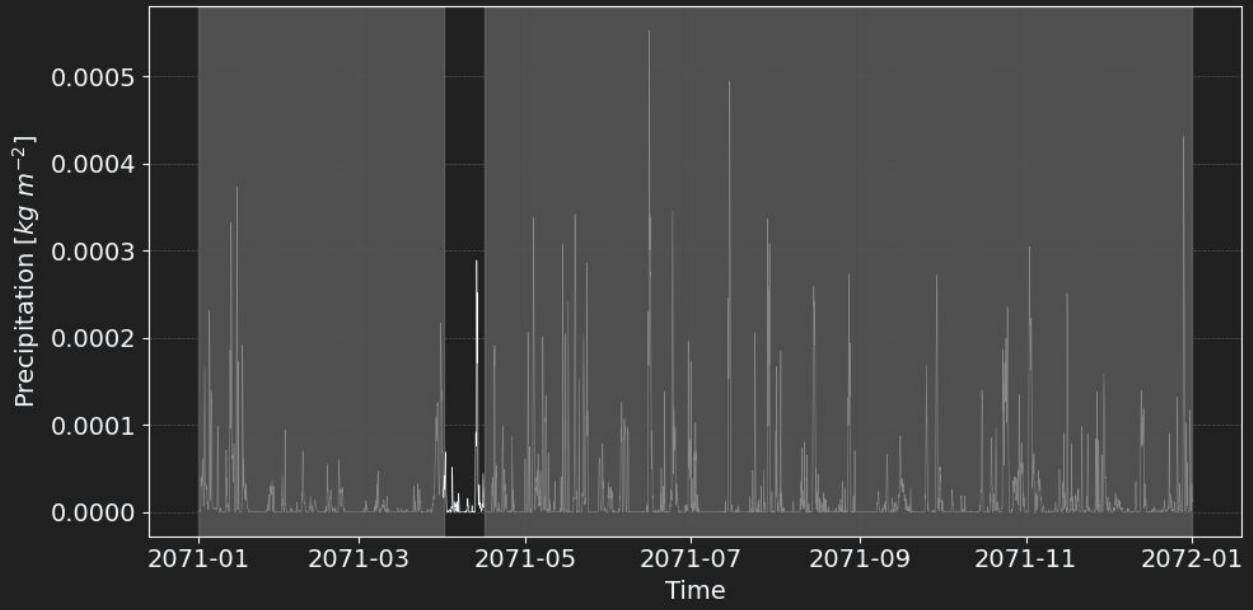
Earth Sciences Department

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Destination Earth

One-Pass Algorithm (OPA)

One-Pass Algorithm (OPA)



One-Pass Algorithm (OPA)

Approximate a function *f* to an object *S*

$$f(X_n) \approx S_n$$

One-Pass Algorithm (OPA)

Approximate a function *f* to an object *S*

$$f(X_n) \approx S_n$$

Where *S* is computed recursively with incoming chunked data

$$S_n = g(S_{n-w}, X_w)$$

One-Pass Algorithm (OPA)

Approximate a function *f* to an object *S*

$$f(X_n) \approx S_n$$

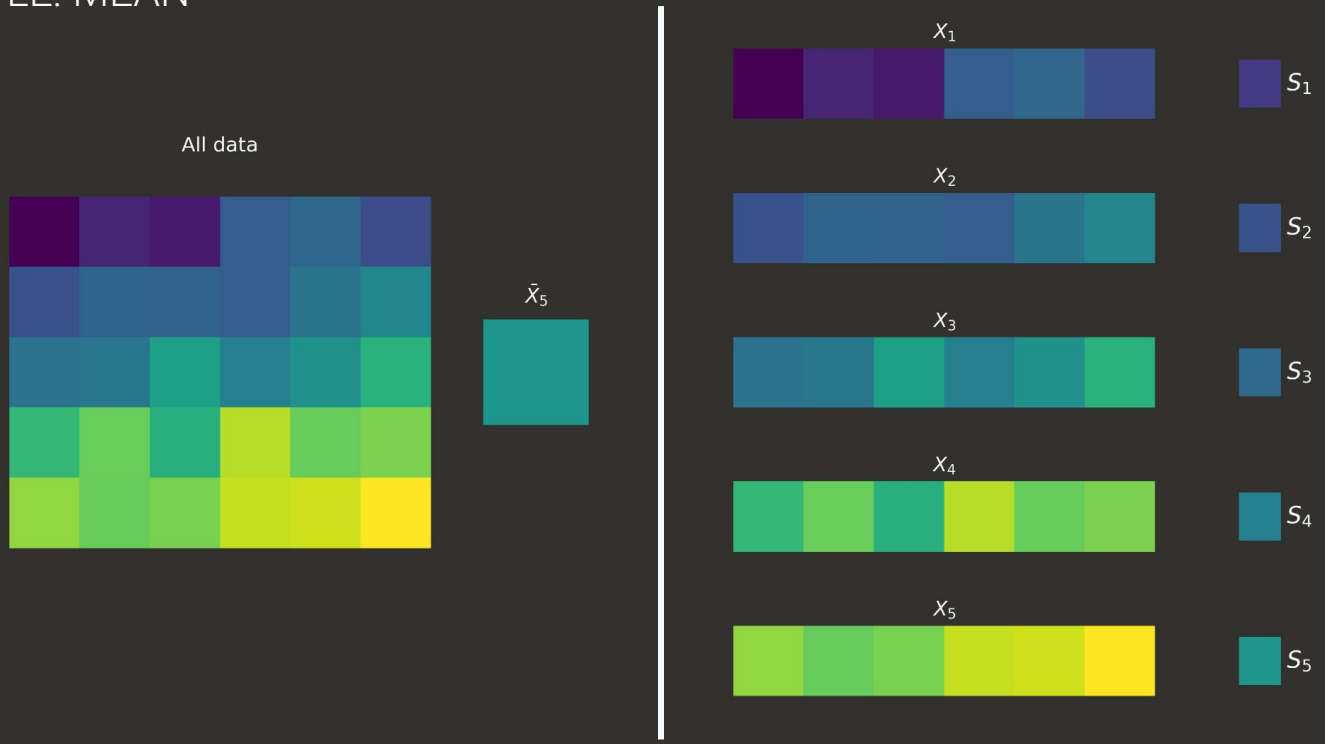
Where *S* is computed recursively with incoming chunked data

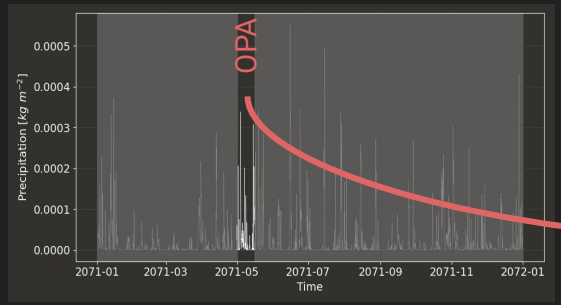
$$S_n = g(S_{n-w}, X_w)$$

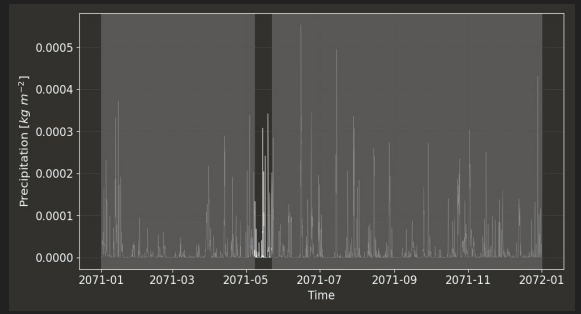
EXAMPLE: MEAN

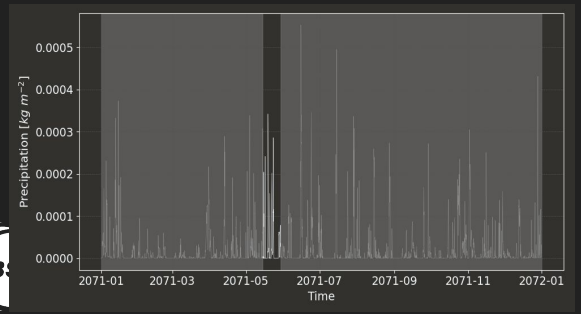
$$g(S_{n-w}, X_w) = S_{n-w} + w \times \frac{\bar{X}_w - S_{n-w}}{n}$$

EXAMPLE: MEAN



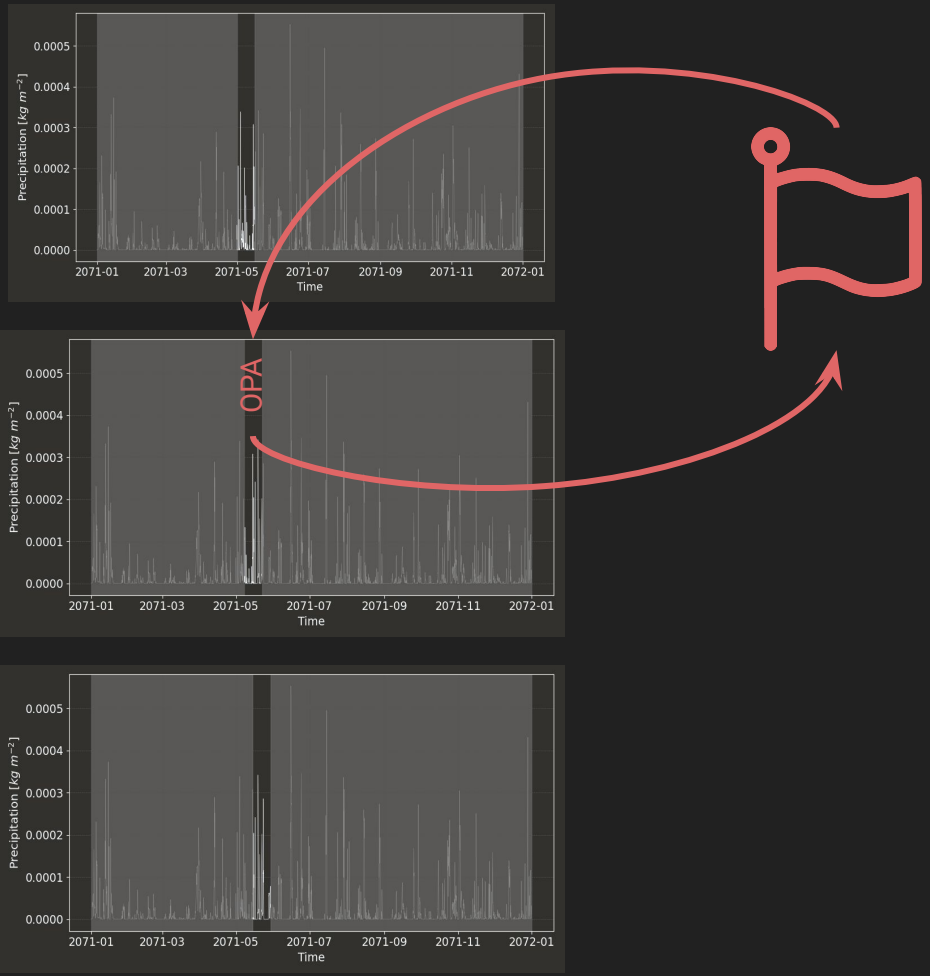






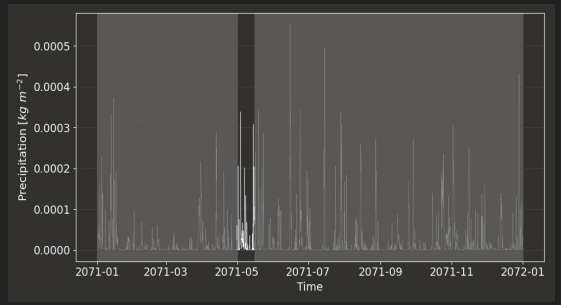
Earth Sciences Department

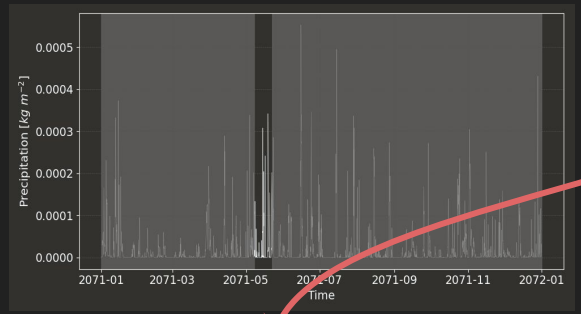
Destination Earth

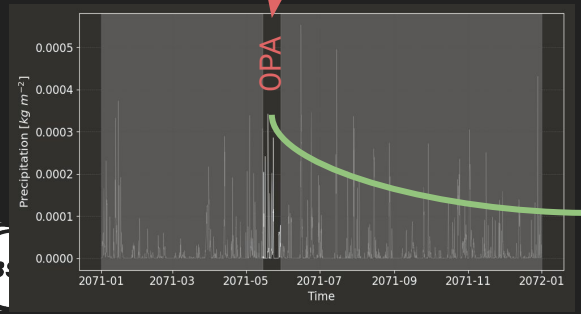


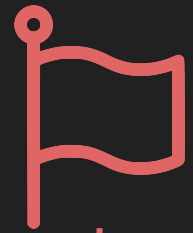
Earth Sciences Department

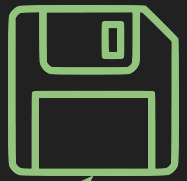
Destination Earth





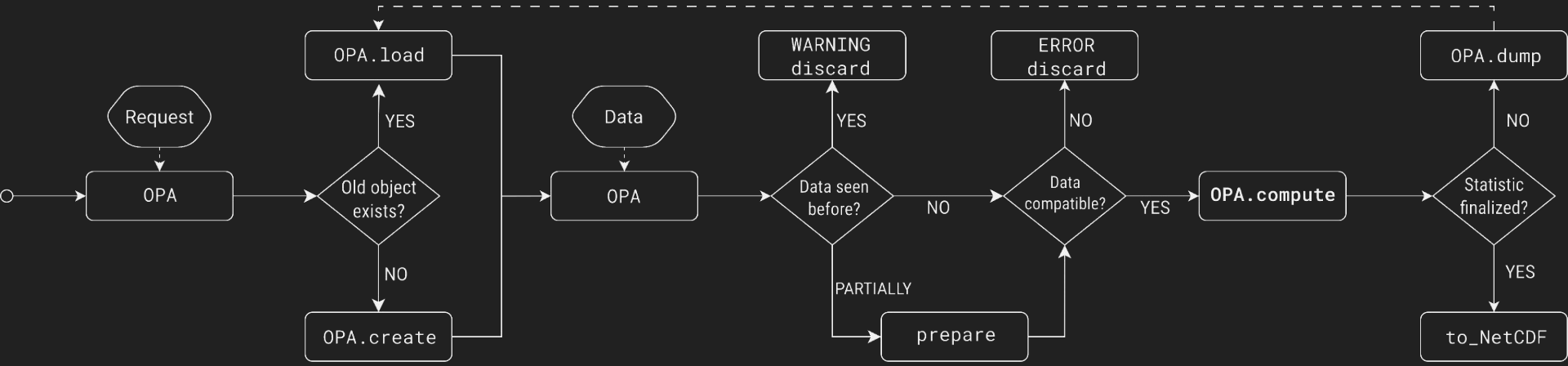






Earth Sciences Department

Destination Earth


Request
OPA
Old object exists?
YES
NO
OPA.load
OPA.create
Data
OPA
WARNING discard
Data seen before?
YES
NO
PARTIALLY
ERROR discard
Data compatible?
NO
YES
prepare
OPA.compute
OPA.dump
Statistic finalized?
NO
YES
to_NetCDF

OPA request

MEAN Mean

VAR Variance

STD Standard deviation

MIN Minimum

MAX Maximum

THRE Threshold exceedance

MDC Mean diurnal cycle

*HIST Histogram**

*PERC Percentile**

OPA request

*Approximation via the
t-digest algorithm

Dunning, T., & Ertl, O. (2019). Computing Extremely Accurate Quantiles Using t-Digests. *ArXiv, abs/1902.04023*.

HIST *Histogram**

PERC *Percentile**

OPA request

( , MEAN, DAILY)

( , VAR, WEEKLY)

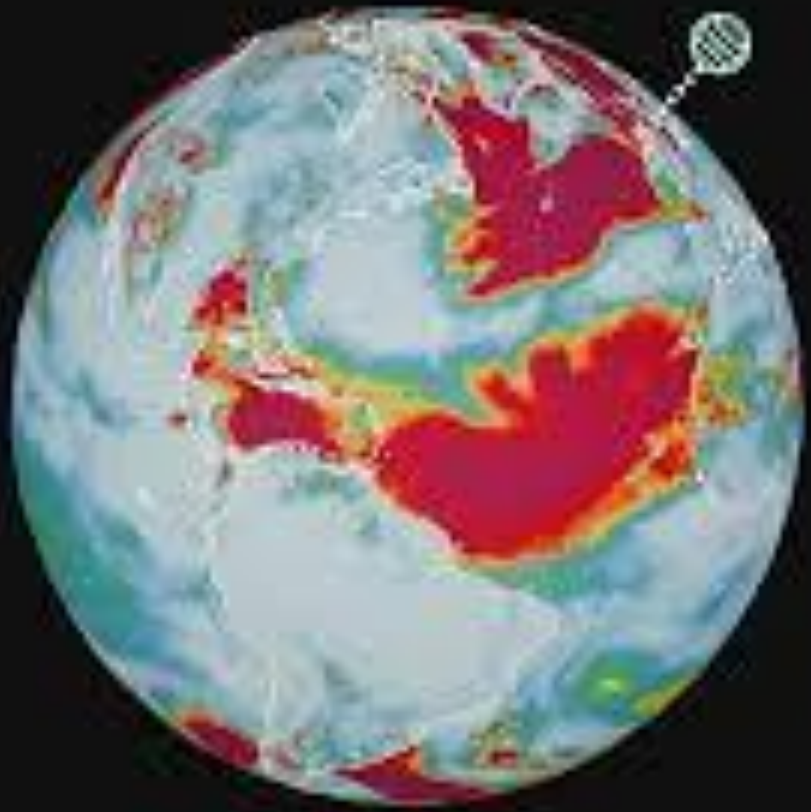
( , MIN, MONTHLY)

( , MAX, 3MONTHLY)

( , MEAN, YEARLY)

( , HIST, WEEKLY)

OPA request



Earth Science Department

Destination Earth