In-memory analytics workflows on large scale climate datasets

Eng. Alessandro D'Anca
Dr. Sandro Fiore and Prof. Giovanni Aloisio
CMCC Foundation
Advanced Scientific Computing Division

on behalf of the Ophidia Team

Understanding I/O Performance Behavior
UIOP Workshop
March 22nd
DKRZ, Hamburg, Germany

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

cmcc
Centro Euro-Mediterraneo sui Cambiamenti Climatici

Outline

- Analytics requirements and needs in the climate context
- Ophidia
  - **Architecture v1.0**
    - Primitives
    - Data and metadata operators
  - **Architecture v2.0**
    - In memory analysis
    - Workflow support
- Climate indicator processing examples
- Useful link & documentation
  - Website, github, youtube, pypi, …

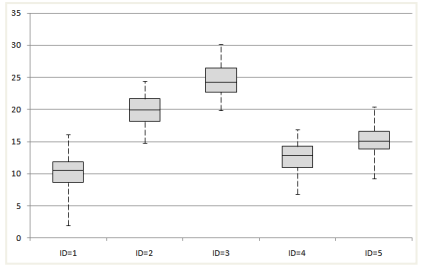
Data analytics requirements and use cases

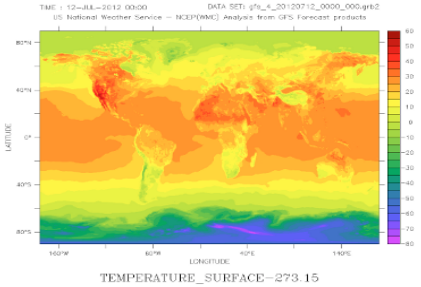
Requirements and needs focus on:

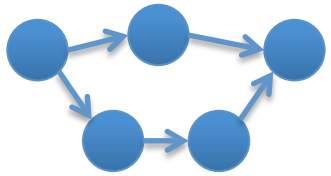
- *Time series analysis*
- *Data subsetting*
- *Model intercomparison*
- *Multimodel means*
- *Massive data reduction*
- *Data transformation (through array-based primitives)*
- *Param. Sweep experiments (same task applied on a set of data)*
- *Maps generation*
- *Ensemble analysis*
- *Data analytics workflow support*

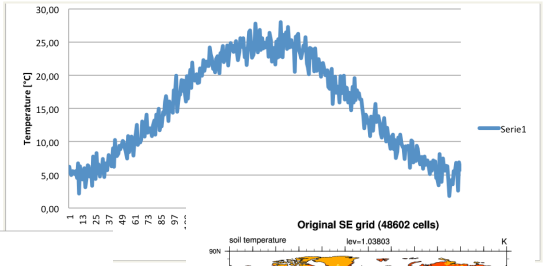
But also…

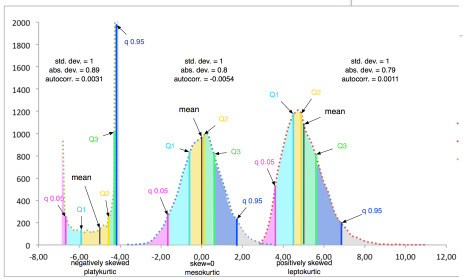
- *Performance*
- *re-usability*
- *extensibility*

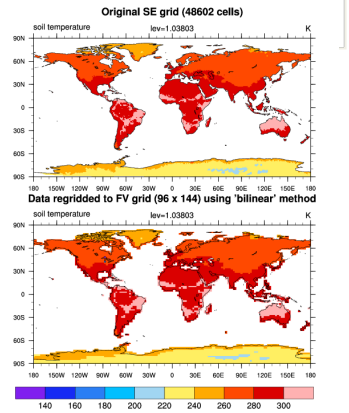












The Ophidia project

Ophidia (http://ophidia.cmcc.it) is a CMCC Foundation research project addressing big data challenges for eScience

It provides support for declarative, parallel, server-side data analysis exploiting parallel computing techniques and database approaches

Exploits a multidimensional data model providing the data cube abstraction for access and analysis of scientific n-dimensional data

Ophidia in a nutshell

- ✔ ***Big data stack for scientific data analysis***
- ✔ ***Features:*** *time series analysis (array-based analysis), data subsetting (by value/index), data aggregation, model intercomparison, OLAP, etc.*
- ✔ *Use of parallel operators and parallel I/O*
- ✔ ***Support for complex workflows / operational chains***
- ✔ *Extensible:* ***simple API*** *to support framework extensions like new operators and array-based primitives*
  - ✔ *currently 50+ operators and 70+ primitives provided*
- ✔ ***Multiple interfaces*** *available (WS-I, GSI/VOMS, OGC-WPS).*
- ✔ *Programmatic access via C and* ***Python APIs***
- ✔ *Support for both* ***batch & interactive*** *data analysis*
- ✔ *Command line interpreter for submitting operators.*

Ophidia original architecture

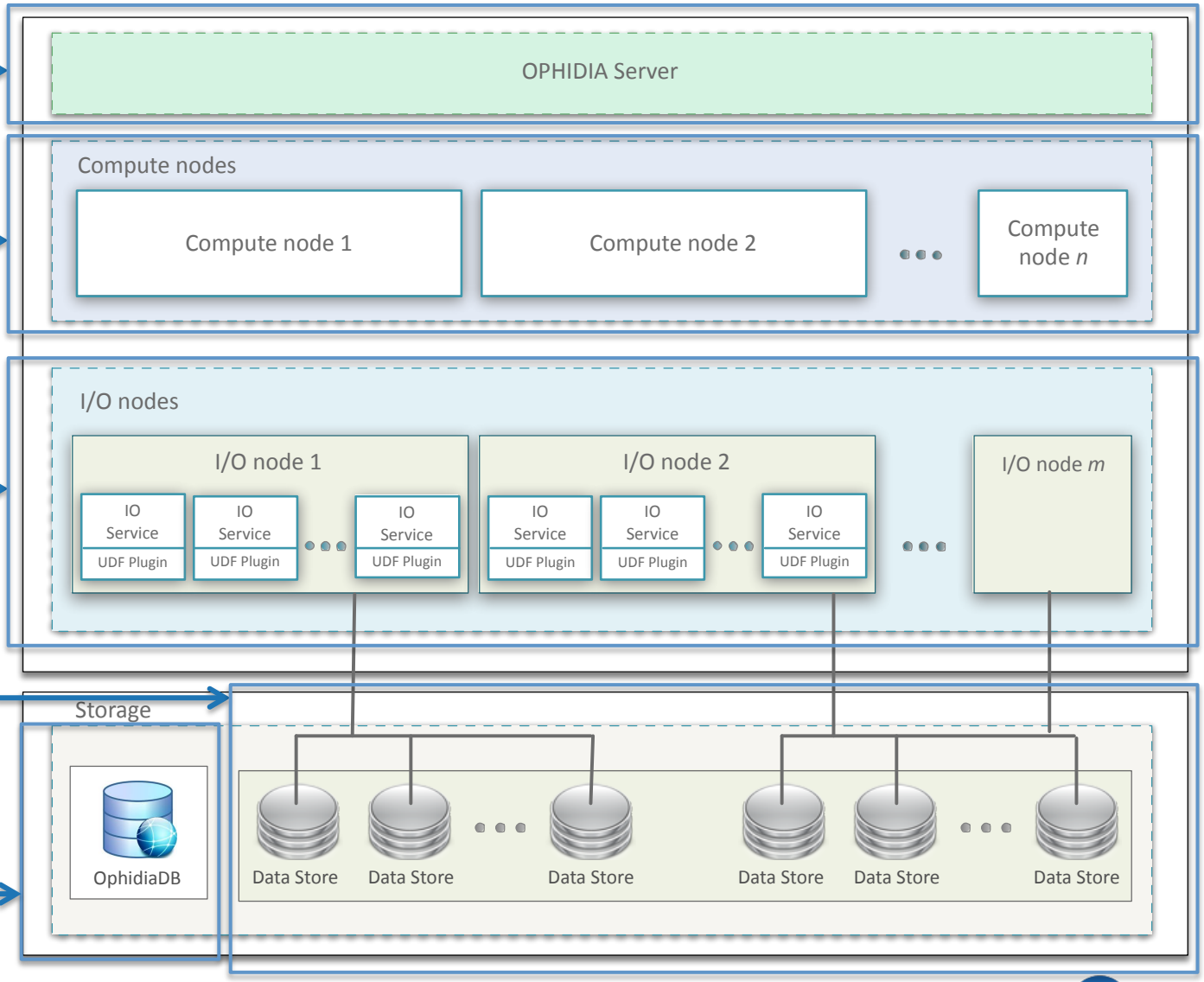
Multi-interface server front-end & job/ workflow management

Analytics framework for the execution of parallel MPI-based (data cube) operators

Multiple I/O servers (MySQL) run array-based primitives (UDF) on data

Distributed hardware resources to manage data storage

OphidiaDB maps data fragmentation and tracks metadata



Array based primitives (about 70)

- Ophidia provides a **wide set of array-based primitives** to perform data summarization, sub-setting, predicates evaluation, statistical analysis, compression, etc.
- Primitives come as plugins and are applied on a single datacube chunk (fragment)
- **Primitives can be nested** to get more complex functionalities
- **Compression is a primitive too!**
- New primitives can be easily integrated as additional plugins

Datacube abstraction and operators (about 50)

OPERATOR NAME	OPERATOR DESCRIPTION
Operators "Data processing" – Domain-agnostic	
OPH_APPLY(*datacube_in*, *datacube_out*, *array_based_primitive*)	Creates the *datacube_out* by applying the *array-based primitive* to the *datacube_in*
OPH_DUPLICATE(*datacube_in*, *datacube_out*)	Creates a copy of the *datacube_in* in the *datacube_out*
OPH_SUBSET(*datacube_in*, *subset_string*, *datacube_out*)	Creates the *datacube_out* by doing a sub-setting of the *datacube_in* by applying the *subset_string*
OPH_MERGE(*datacube_in*, *merge_param*, *datacube_out*)	Creates the *datacube_out* by merging groups of *merge_param* fragments from *datacube_in*
OPH_SPLIT(*datacube_in*, *split_param*, *datacube_out*)	Creates the *datacube_out* by splitting into groups of *split_param* fragments each fragment of the *datacube_in*
OPH_INTERCOMPARISON (*datacube_in1*, *datacube_in2*, *datacube_out*)	Creates the *datacube_out* which is the element-wise difference between *datacube_in1* and *datacube_in2*
OPH_DELETE(*datacube_in*)	Removes the *datacube_in*

Data processing

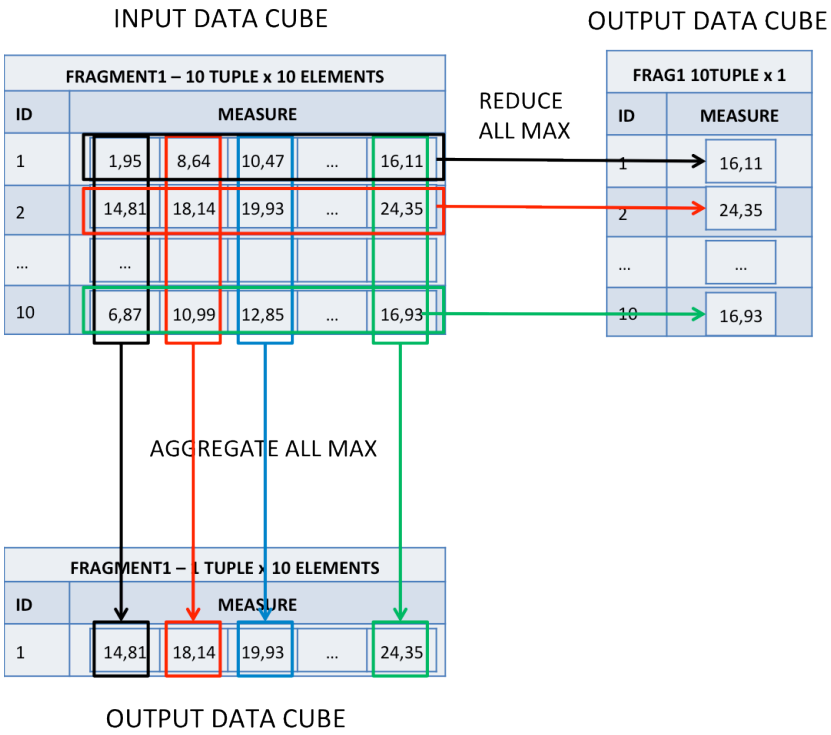
Import/Export

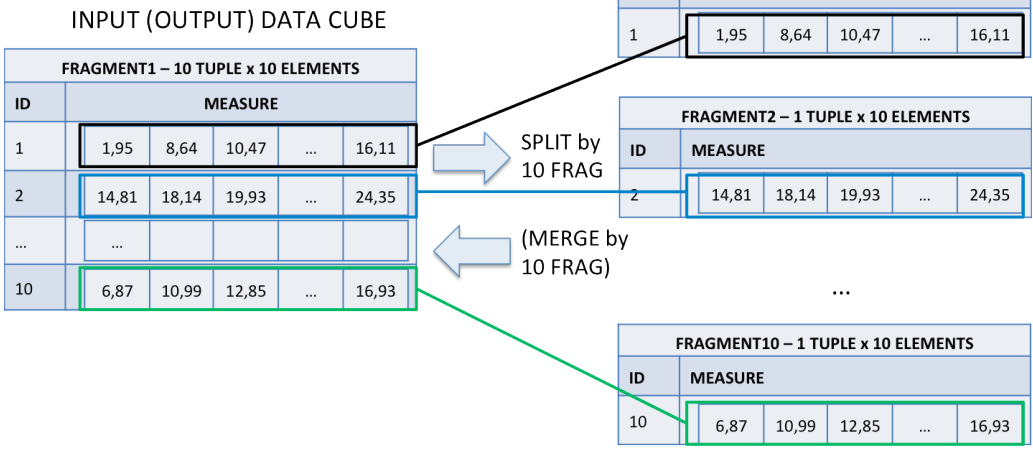
Data Access

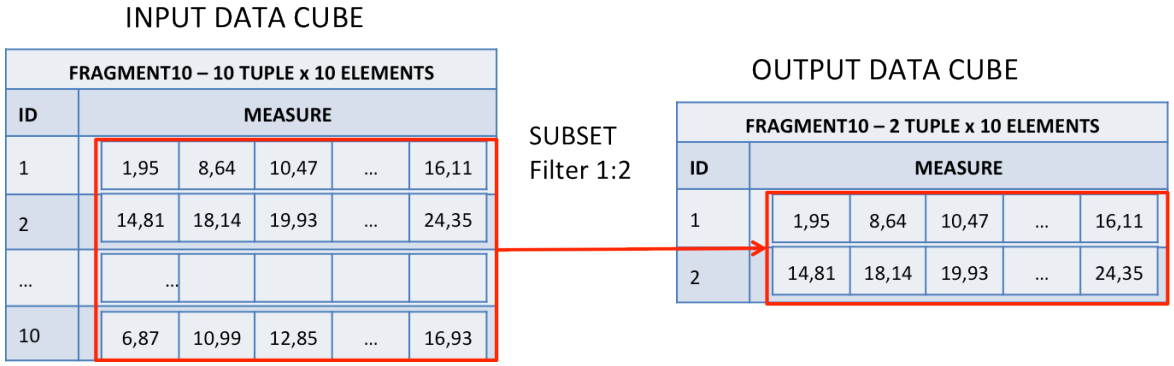
Metadata management

OPERATOR NAME	OPERATOR DESCRIPTION
Operators "Data processing" – Domain-oriented	
OPH_EXPORT_NC (*datacube_in*, *file_out*)	Exports the *datacube_in* data into the *file_out* NetCDF file.
OPH_IMPORT_NC (*file_in*, *datacube_out*)	Imports the data stored into the *file_in* NetCDF file into the new *datacube_in* datacube
Operators "Data access"	
OPH_INSPECT_FRAG (*datacube_in*, *fragment_in*)	Inspects the data stored in the *fragment_in* from the *datacube_in*
OPH_PUBLISH(*datacube_in*)	Publishes the *datacube_in* fragments into HTML pages
Operators "Metadata"	
OPH_CUBE_ELEMENTS (*datacube_in*)	Provides the total number of the elements in the *datacube_in*
OPH_CUBE_SIZE (*datacube_in*)	Provides the disk space occupied by the *datacube_in*
OPH_LIST(void)	Provides the list of available datacubes.
OPH_CUBEIO(*datacube_in*)	Provides the provenance information related to the *datacube_in*
OPH_FIND(*search_param*)	Provides the list of datacubes matching the *search_param* criteria

The analytics framework: “data” operators





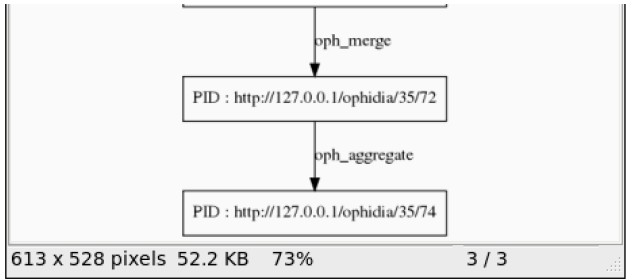


The analytics framework: “metadata” operators

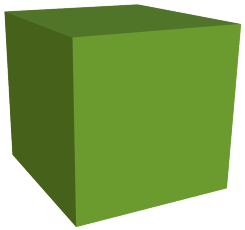
```
[37..4416] >> oph_cubeschema cube=http://127.0.0.1/ophidia/35/67;
[Request]:
operator=oph_cubeschema;cube=http://127.0.0.1/ophidia/35/67;sessionid=http://127.0.0.1/ophidia/sessions/374383780832141666641463737283924416/experiment;exec_mode=sync;ncores=1;c
wd=/;
[37..4416] >> oph_explorecube cube=http://127.0.0.1/ophidia/35/67;subset_dims=lat|lon|time;subset_filter=39:42|15:19|1:275;show_time=yes;
[Request]:
operator=oph_explorecube;cube=http://127.0.0.1/ophidia/35/67;subset_dims=lat|lon|time;subset_filter=39:42|15:19|1:275;show_time=yes;sessionid=http://127.0.0.1/ophidia/sessions/3
74383780832141666641463737283924416/experiment;exec_mode=sync;ncores=1;cwd=/;

[JobID]:
http://127.0.0.1/ophidia/sessions/374383780832141666641463737283924416/experiment?106#224

[Response]:
tos
---
```

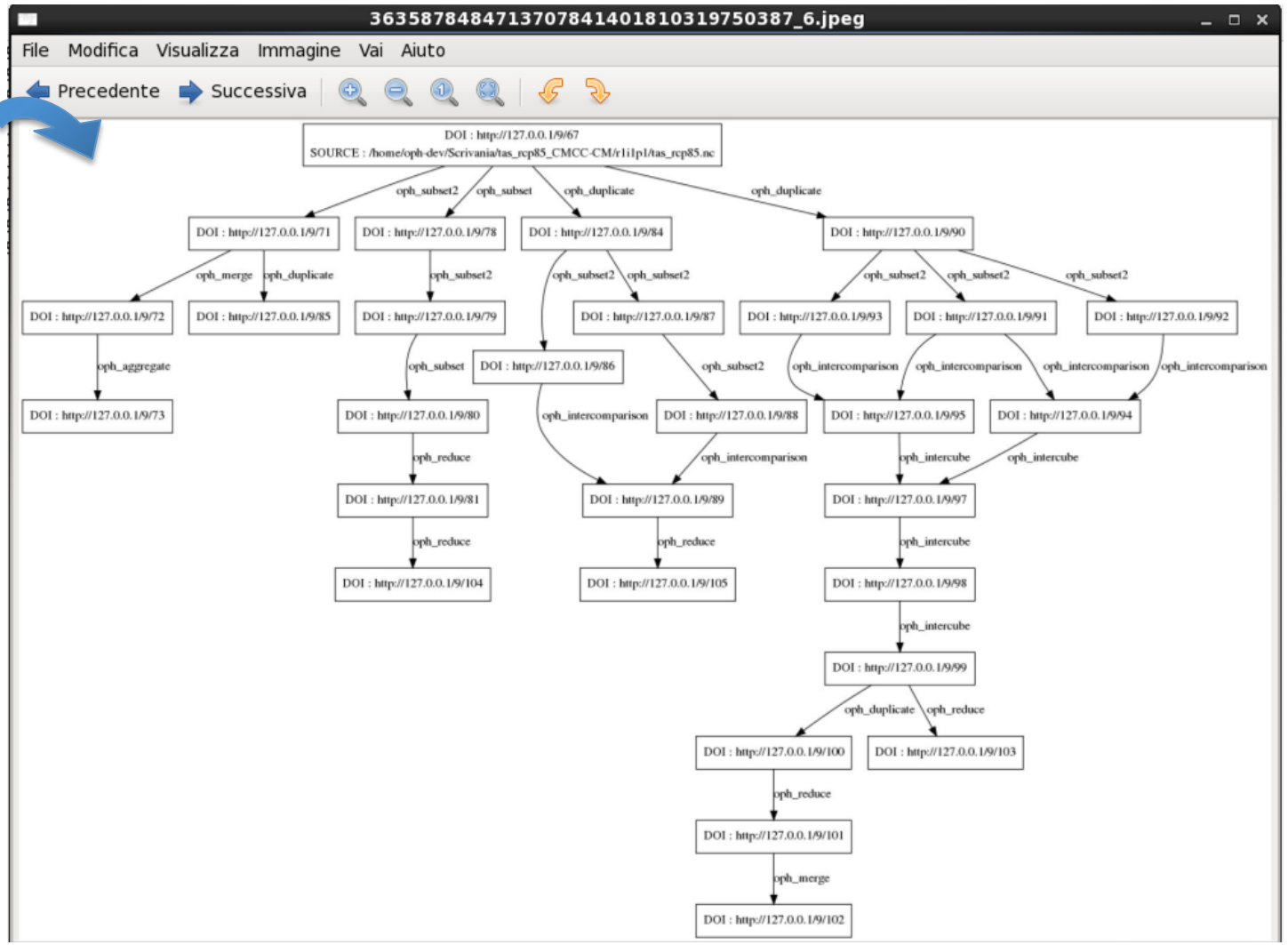
lat	lon	tos
39.500000	15.000000	1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20
39.500000	17.000000	287.3930664062, 286.8287048340, 286.5860595703, 286.9228210449, 288.5254516602, 292.3968200684, 295.8656921387, 297.2062072754, 295.7126464844
39.500000	19.000000	287.6926879883, 287.0508117676, 286.7896118164, 287.0781555176, 288.6802062988, 292.6882629395, 296.4769287109, 297.6632385254, 296.3418273926
40.500000	15.000000	1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20
40.500000	17.000000	287.1098632812, 286.5683593750, 286.2949829102, 286.5216674805, 288.0316772461, 291.7698974609, 295.4139709473, 296.8489685059, 295.4132995605
40.500000	19.000000	287.4010009766, 286.7818298340, 286.4914245605, 286.7260742188, 288.3006286621, 292.1842346191, 296.0237731934, 297.2694702148, 295.9751892090
41.500000	15.000000	1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20, 1.00000002e+20
41.500000	17.000000	286.5835876465, 286.0175781250, 285.7146911621, 285.9142761230, 287.4476623535, 291.1032104492, 294.7090454102, 296.0852355957, 294.7053222656
41.500000	19.000000	286.9717712402, 286.3946838379, 286.0617675781, 286.1446228027, 287.6101989746, 291.2955017090, 295.2700195312, 296.5146179199, 295.3194274902

```
Summary
-------
Selected 9 rows out of 9
digraph DG {
        node    [shape=box]

        0       [label="PID : http://127.0.0.1/ophidia/35/74\n"]
        1       [label="PID : http://127.0.0.1/ophidia/35/72\n"]
        2       [label="PID : http://127.0.0.1/ophidia/35/71\n"]
        3       [label="PID : http://127.0.0.1/ophidia/35/70\n"]
        4       [label="PID : http://127.0.0.1/ophidia/35/66\nSOURCE : /repo/tos_01_2002-2003.nc\n"]
        5       [label="PID : http://127.0.0.1/ophidia/35/67\nSOURCE : /repo/tos_01_2001-2002.nc\n"]

        1->0    [label="oph_aggregate"]
        2->1    [label="oph_merge"]
```



Provenance management (PID-based)





Ophidia architecture extensions

The architecture has evolved in order to enhance performances and scalability and achieve decoupling from the underlying storage.

- enhance **performance** executing analytic tasks **in-memory**
- **decouple** the **analytical framework** from the **storage features** via unique API
- supports **different** types of **I/O server**
- **interoperability** with different types of **storage systems**

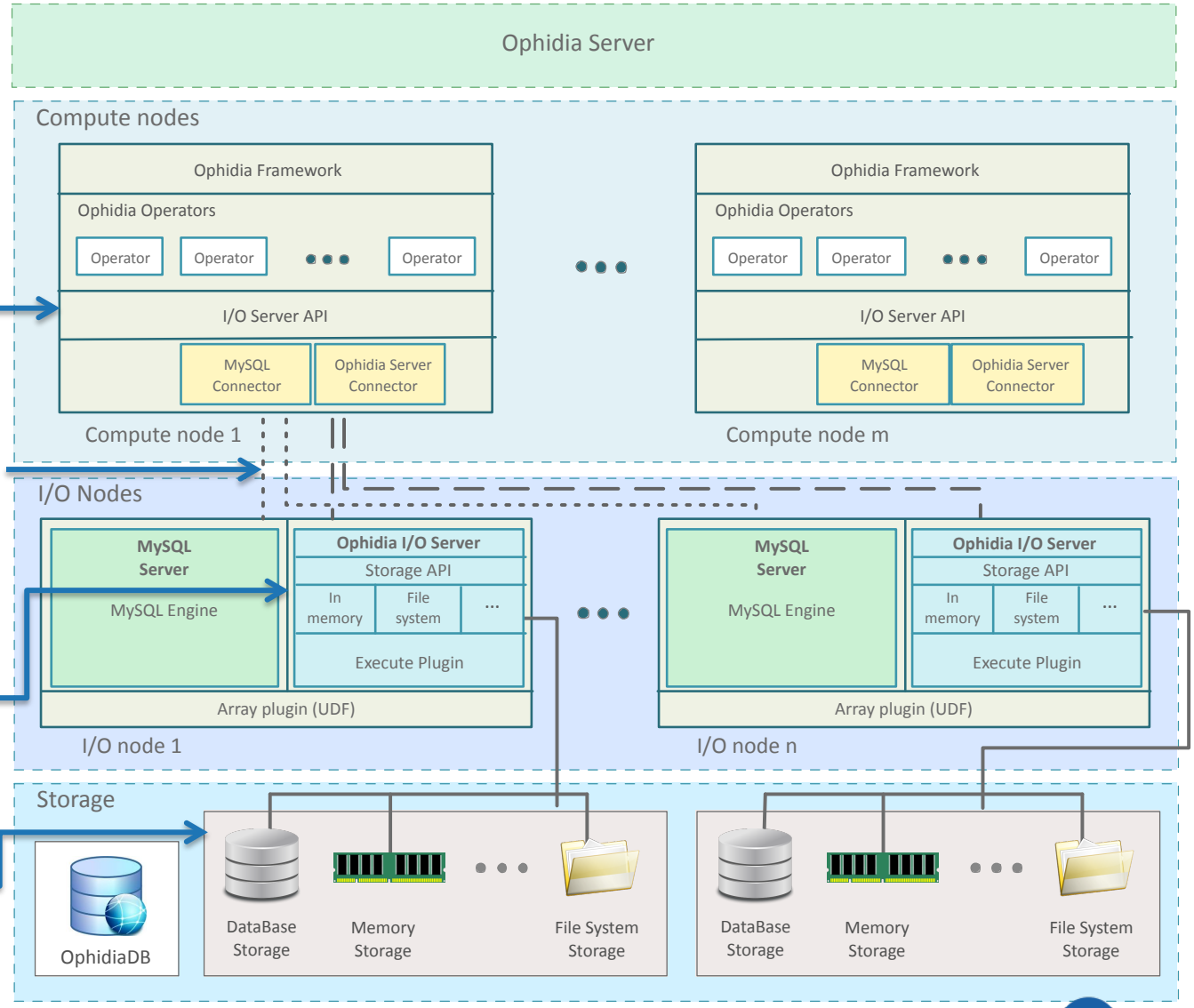
Ophidia architecture v2.0

Higher degree of decoupling among framework and I/O components

Support different I/O servers

Native I/O server with parallel execution engine

Multiple storage systems supported



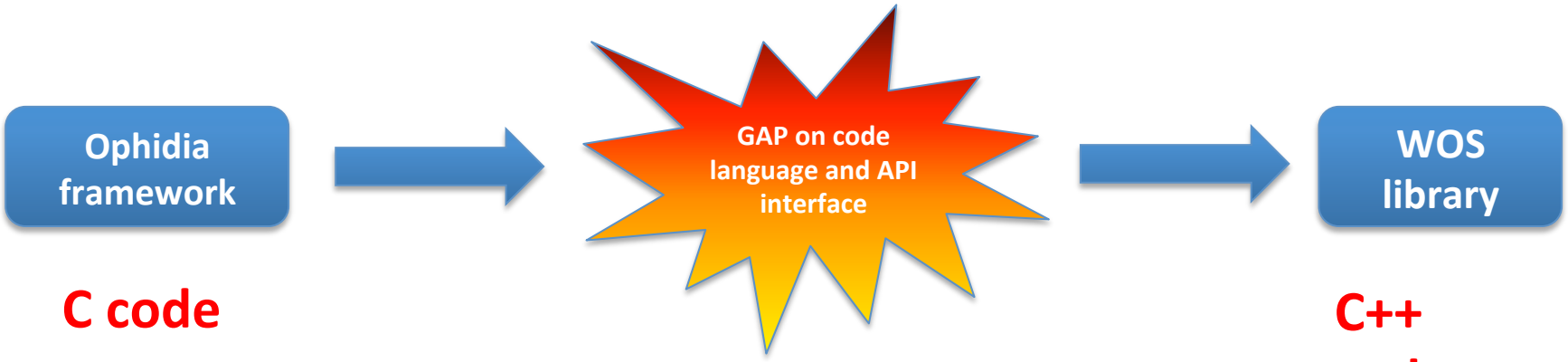
Ophidia I/O server: requirements

The I/O server provides a native solution for the scientific domain applications. The requirements for the Ophidia I/O server are:

- run **data analytics tasks in-memory** taking advantage of the lower latency
- **binary array-oriented engine** to efficiently process scientific multidimensional data
- interact directly with the storage layer to **exploit data locality**
- exploit **parallelism at the array-level**
- **NoSQL approach** based on a key-value store, providing a **declarative query language** (SQL-like)
- guarantee extensibility and interoperability of the I/O server to **support multiple storage back-ends**

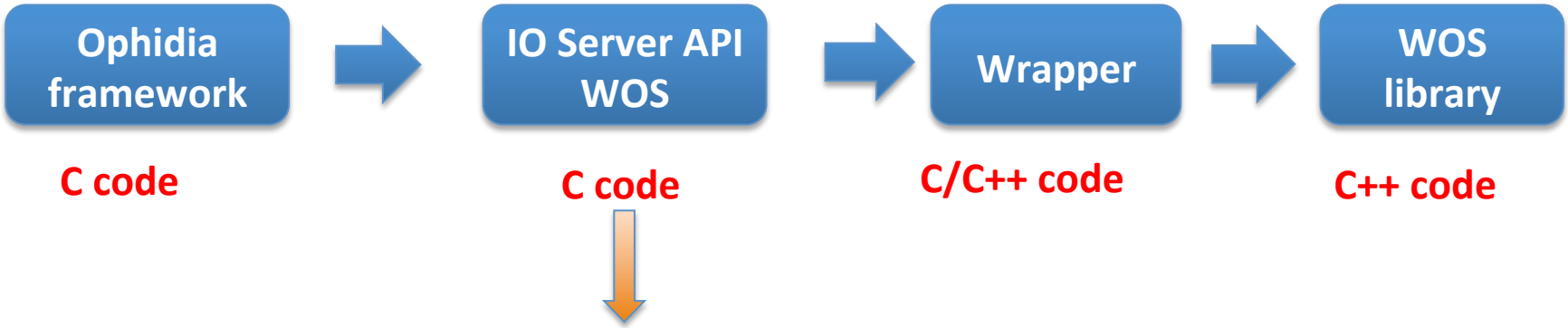
WOS Integration in Ophidia (I)

- ✔ *WOS semantics have been simplified to **PUT, GET, and DELETE**.*
- ✔ *The C++ API provides the following functionality available in both **blocking and non-blocking** forms:*
  - *Connect to cluster*
  - *Create WOS object*
  - *Put, Get, Delete object*
  - *Reserve and PutOID*
  - *Streaming*



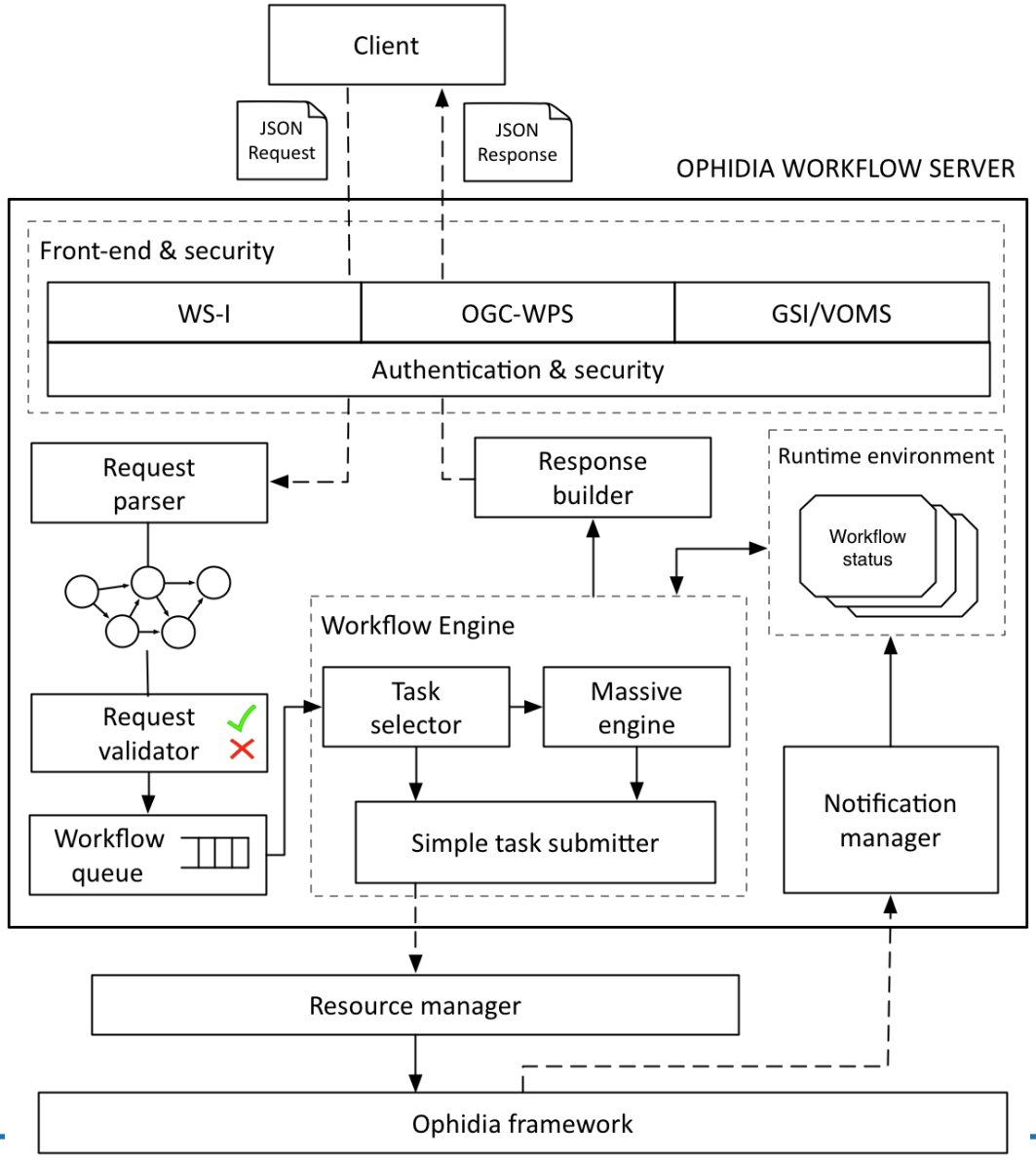
WOS Integration in Ophidia (II)

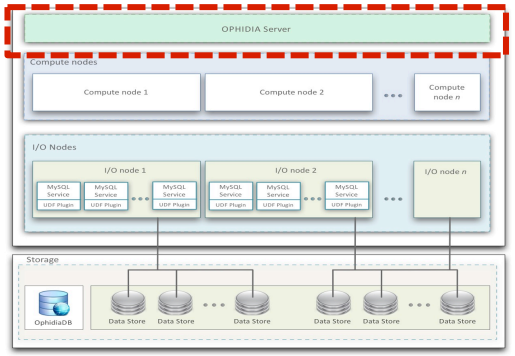
- ✔ ***First step**: define a **Wrapper** to connect C++ API to C functions*
- ✔ ***Second step:** implementation of **Ophidia IO Server API for WOS***



FUNCTION NAME	INPUT
Connect	oph_iostore_handler *handle
Get	oph_iostore_handler * handle, oph_iostore_resource_id *res_id, oph_iostore_frag_record_set **frag_record
Put	oph_iostore_handler * handle, oph_ioserver_frag_record_set *frag_record, oph_iostore_resource_id **res_id
Delete	oph_iostore_handler* handle, oph_iostore_resource_id *res_id

The Ophidia Server





The workflow runtime engine is the core component of the Ophidia Server:

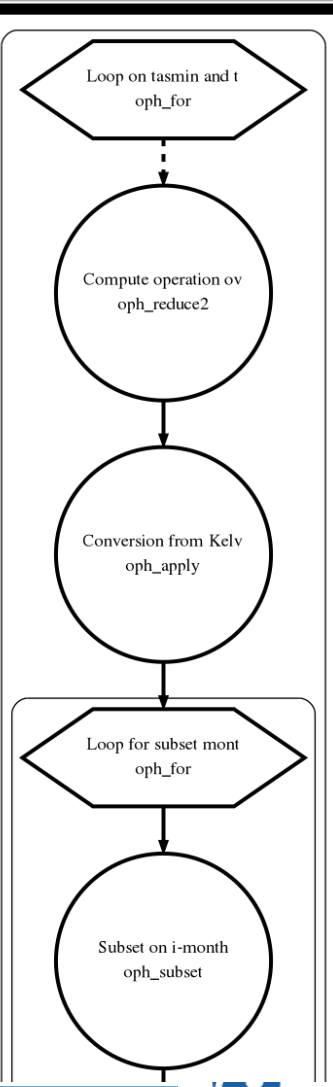
- it formats the commands for the analytics framework;
- submits the tasks to the resource manager;
- checks for task status updates in the runtime environment and
- provides the proper response messages.

Workflow JSON representation

ophrpm@ophidiarpm:~/workflow

ophrpm@ophidiarpm:~/workflow

```
"tasks": [
    {
        "name": "Loop on tasmin and tasmax cubes",
        "operator": "oph_for",
        "arguments": [ "name=cube","counter=1:2", "values=${1}|${2}", "parallel=yes" ]
    },
    {
            "name": "Compute operation over time",
            "operator": "oph_reduce2",
            "arguments": [
                    "cube=@{cube}",
                    "dim=time",
                    "concept_level=M",
                    "midnight=00",
                    "operation=$3",
                    "container=tmp"
            ],
            "dependencies": [
                    { "task": "Loop on tasmin and tasmax cubes" }
            ]
    },
    {
            "name": "Conversion from Kelvin to Celsius degrees",
            "operator": "oph_apply",
            "arguments": [
                    "query=oph_sum_scalar('oph_float','oph_float',measure,-273.15)"
            ],
            "dependencies": [{
                    "task": "Compute operation over time",
                    "type": "single"
            }]
    },
    {
            "name": "Loop for subset months",
            "operator": "oph_for",
            "arguments": [ "name=index","counter=1:12", "values=Jan|Feb|Mar|Apr|May|Jun|Jul|Aug|Sep|Oct|N
            "dependencies": [ { "task":"Conversion from Kelvin to Celsius degrees", "type": "single" } ]
    },
            {
                    "name": "Subset on i-month",
                    "operator": "oph_subset",
                    "arguments": [
                            "subset_dims=time",
                            "subset_filter=&index:12:end"
                    ],
                    "dependencies": [
--More--(65%)
```



Workflow submission

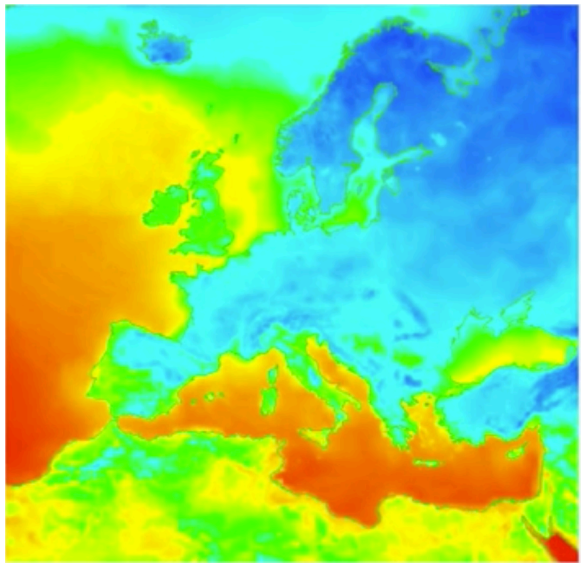
ophrpm@ophidiarpm:~/devel/oph-client/res

ophrpm@ophidiarpm:~/workflow

```
[37..6380] >>
[37..6380] >> ./Tind_loop.json http://193.204.199.174/ophidia/29/2046 http://193.204.199.174/ophidia/30/2047 max
[JobID]:
http://193.204.199.174/ophidia/sessions/376699238311302232511449455166146380/experiment?247#3144

[37..6380] >> view 247
[247] ./Tind_loop.json http://193.204.199.174/ophidia/29/2046 http://193.204.199.174/ophidia/30/2047 max [http://193.204.199.174/ophidia/sessions/376699238311302
232511449455166146380/experiment?247#3144]

[Response]:
Workflow Status
---------------

OPH_STATUS_COMPLETED

Workflow Progress
-----------------

+==========================+======================+
| NUMBER OF COMPLETED TASKS | TOTAL NUMBER OF TASKS |
+==========================+======================+
| 82                       | 82                   |
+==========================+======================+

Workflow Task List
------------------

+=========================================================+======================+========+=======+===========+============================+=====+=============+
| OPH JOB ID                                              | SESSION CODE         | WORKFL | MARKE | PARENT MA | TASK NAME                  | TYP | EXIT STATUS |
|                                                         |                      | OW ID  | R ID  | RKER ID   |                            | E   |             |
+=========================================================+======================+========+=======+===========+============================+=====+=============+
| http://193.204.199.174/ophidia/sessions/37669923831130  | 37669923831130223251 | 247    | 3145  | 3144      | Loop on tasmin and tasmax  | SIM | OPH_STATUS_ |
| 2232511449455166146380/experiment?247#3145              | 1449455166146380     |        |       |           | cubes                      | PLE | COMPLETED   |
|---------------------------------------------------------|----------------------|--------|-------|-----------|----------------------------|-----|-------------|
| http://193.204.199.174/ophidia/sessions/37669923831130  | 37669923831130223251 | 247    | 3146  | 3144      | Compute operation over ti  | SIM | OPH_STATUS_ |
| 2232511449455166146380/experiment?247#3146              | 1449455166146380     |        |       |           | me (1)                     | PLE | COMPLETED   |
|---------------------------------------------------------|----------------------|--------|-------|-----------|----------------------------|-----|-------------|
| http://193.204.199.174/ophidia/sessions/37669923831130  | 37669923831130223251 | 247    | 3147  | 3144      | Compute operation over ti  | SIM | OPH_STATUS_ |
| 2232511449455166146380/experiment?247#3147              | 1449455166146380     |        |       |           | me (2)                     | PLE | COMPLETED   |
|---------------------------------------------------------|----------------------|--------|-------|-----------|----------------------------|-----|-------------|
| http://193.204.199.174/ophidia/sessions/37669923831130  | 37669923831130223251 | 247    | 3148  | 3144      | Conversion from Kelvin to  | SIM | OPH_STATUS_ |
| 2232511449455166146380/experiment?247#3148              | 1449455166146380     |        |       |           | Celsius degrees (1)        | PLE | COMPLETED   |
|---------------------------------------------------------|----------------------|--------|-------|-----------|----------------------------|-----|-------------|
| http://193.204.199.174/ophidia/sessions/37669923831130  | 37669923831130223251 | 247    | 3149  | 3144      | Conversion from Kelvin to  | SIM | OPH_STATUS_ |
| 2232511449455166146380/experiment?247#3149              | 1449455166146380     |        |       |           | Celsius degrees (2)        | PLE | COMPLETED   |
|---------------------------------------------------------|----------------------|--------|-------|-----------|----------------------------|-----|-------------|
```

Climate indicators processing (I)

- ✔ *In the CLIPC project, processing chains for data analysis are being implemented with Ophidia to compute **climate indicators***
- ✔ ***Parallel approach***
  - ✓ Inter-parallelism: Multiple branches are executed in parallel
  - ✓ Intra-parallelism: data analysis operators have been parallelized too (e.g. MPI)
- ✔ ***First set of indicators** includes: TNn, **TNx**, TXn, **TXx***
  - ✔ *Input files: 12GBs (TasMin & TasMax)*

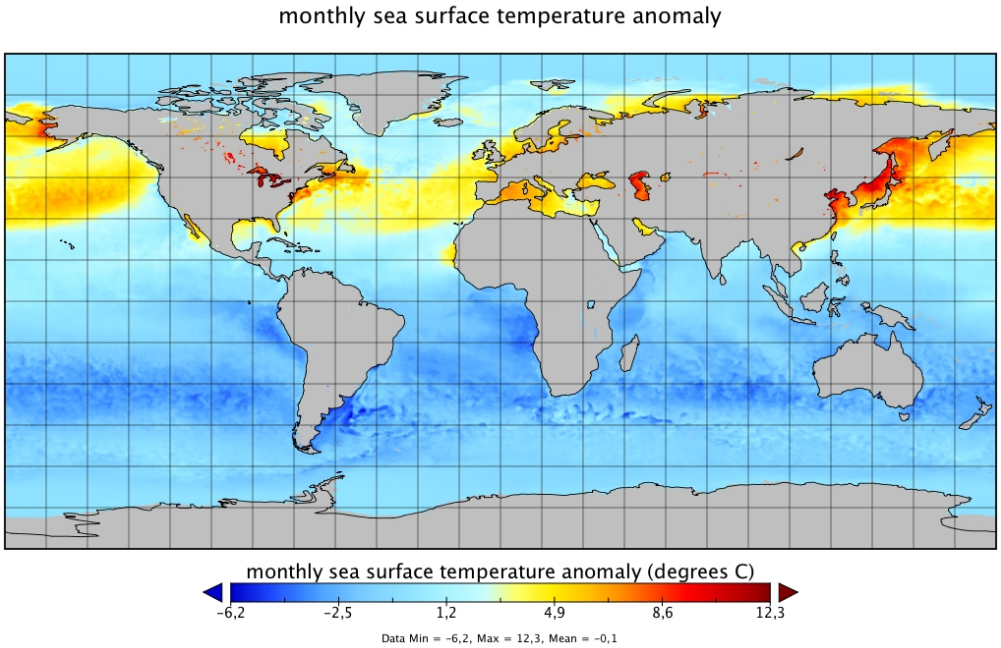


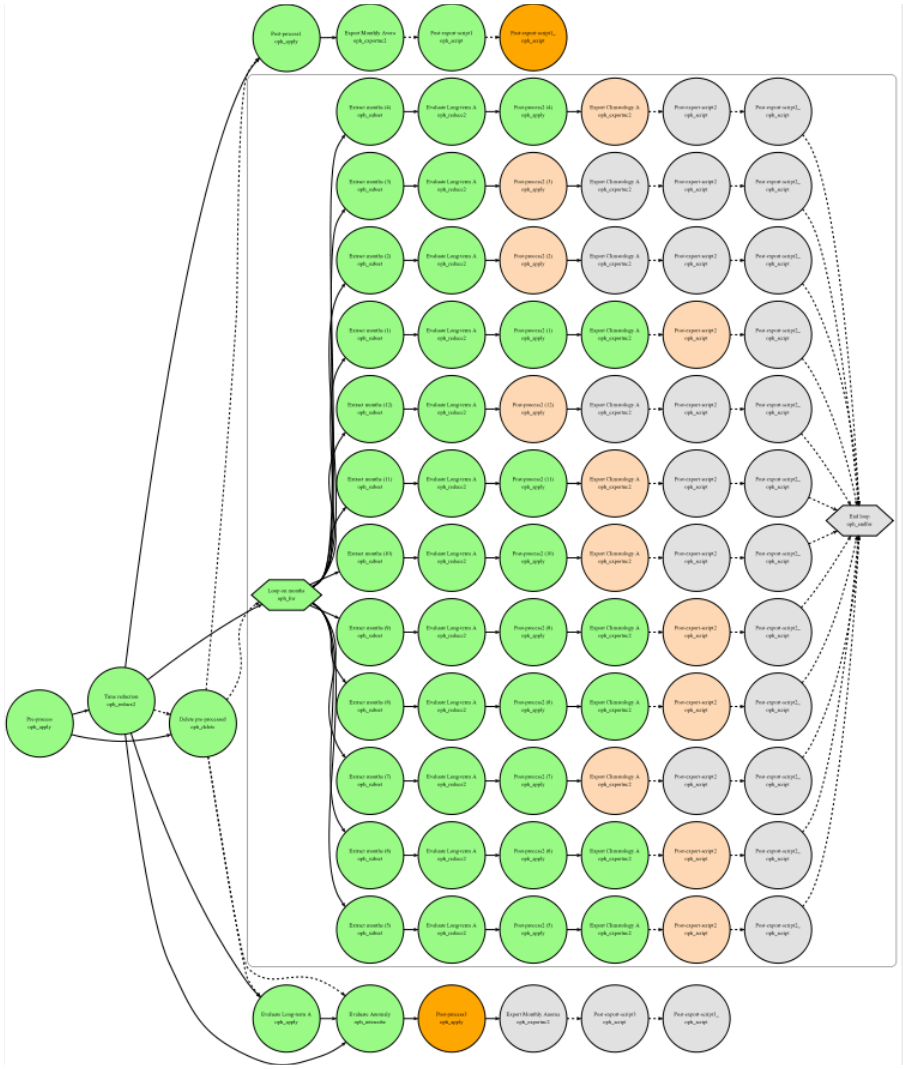
Climate indicators processing (II)

Climate Information Portal

SST mean, anomaly, climatology mean

- ✔ *Dataset time range: 1991-2010*
- ✔ ***7062** nc files*
- ✔ ***350GB** of input data*
- ✔ ***87 tasks** performed*
- ✔ *12x51MB + 2x12GB of output files*

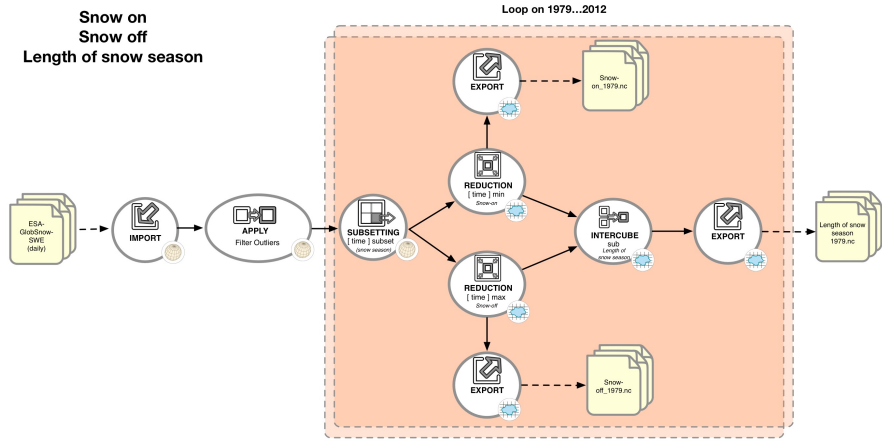


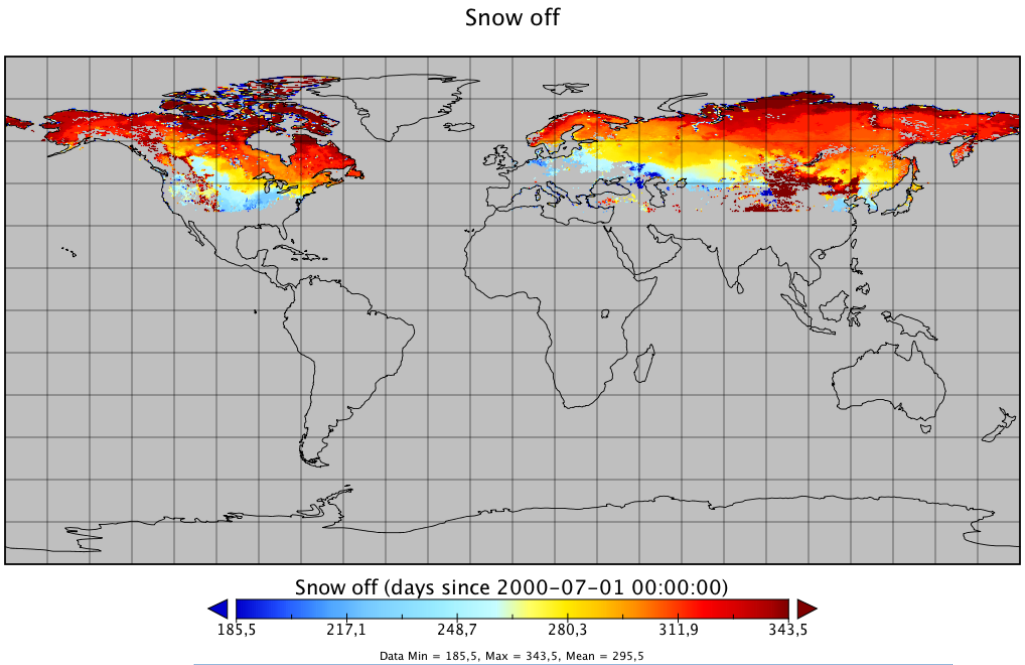


Climate indicators processing (III)

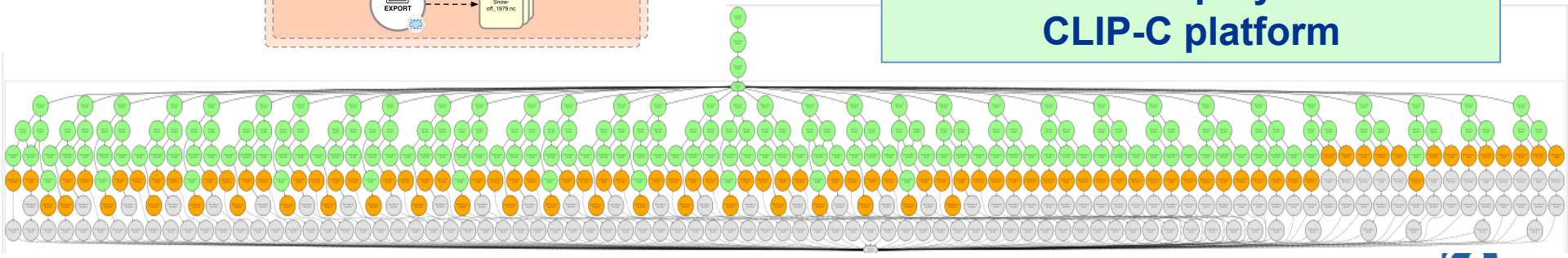
Snow on/off – Length of snow season

- ✔ *Dataset time range: 1979-2012*
- ✔ ***6341** nc files*
- ✔ ***50 GB** of input data*
- ✔ ***599 tasks** performed*
- ✔ ***99** NetCDF output files (6MB each)*
- ✔ ***21** tasks in the exp. description*



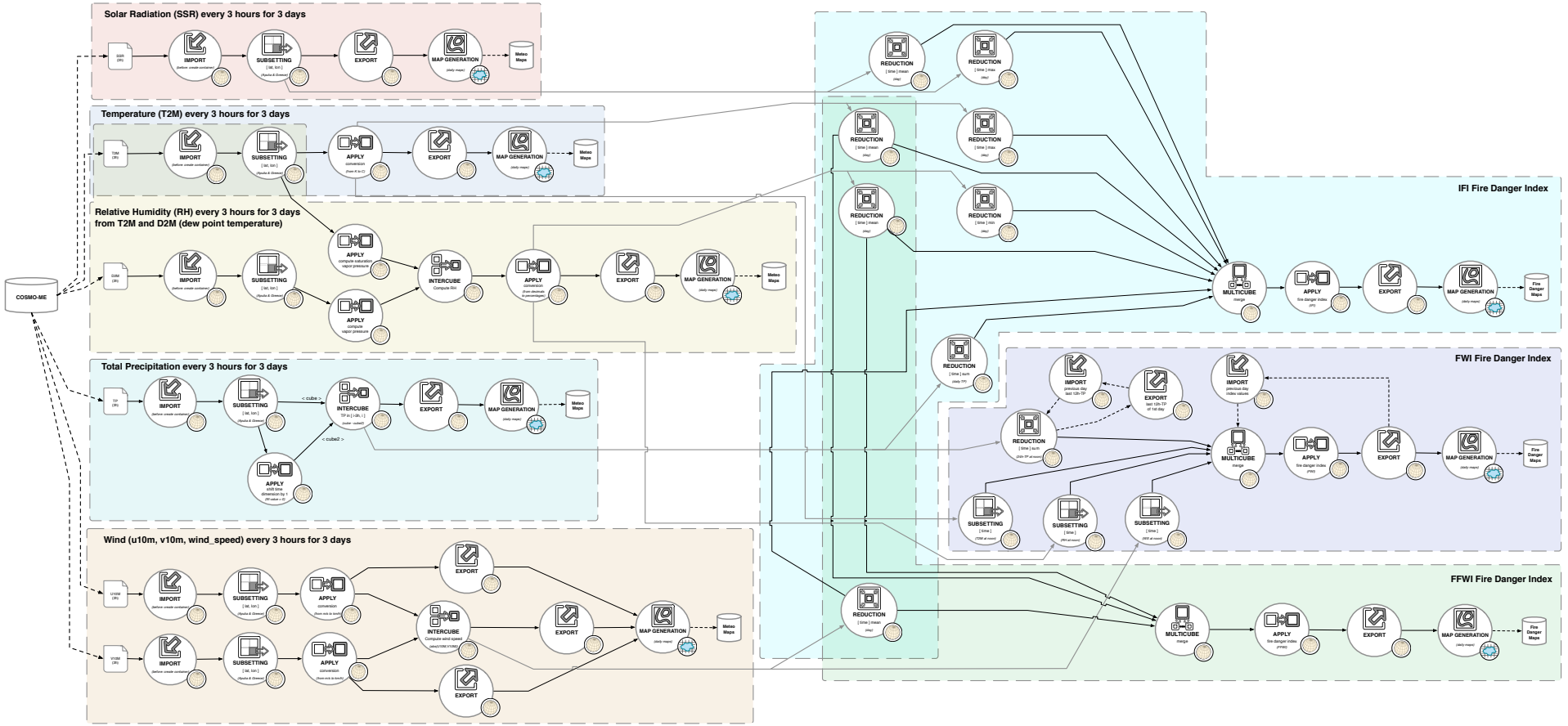


CMCC produced indicators have been deployed in the CLIP-C platform

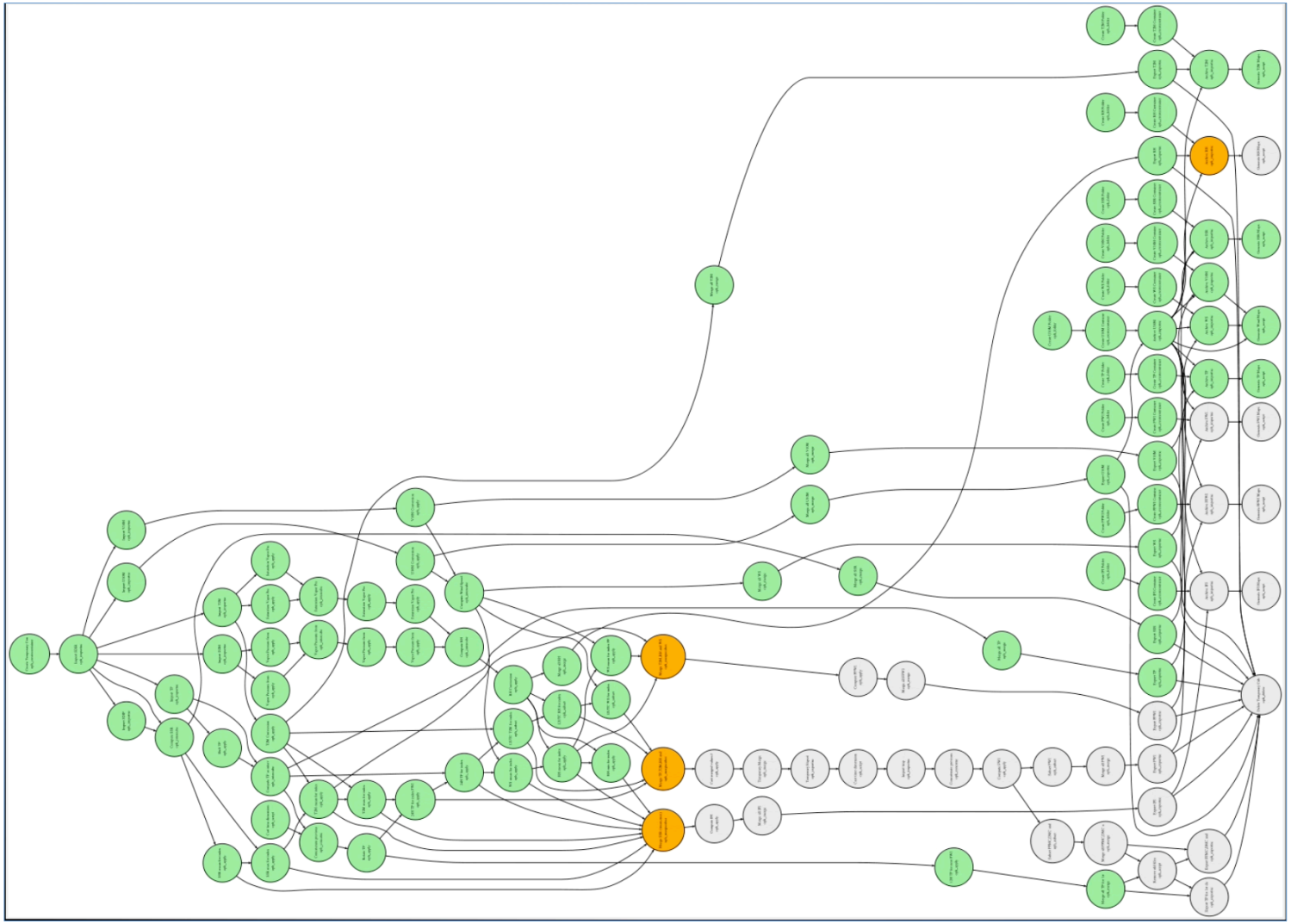


Workflows in Ophidia

OFIDIA main objective is to build a **cross-border operational fire danger prevention infrastructure** that advances the ability of regional stakeholders across Apulia and Ioannina Regions to **detect** and **fight forest wildfires**



Workflow runtime execution (fire danger analysis)



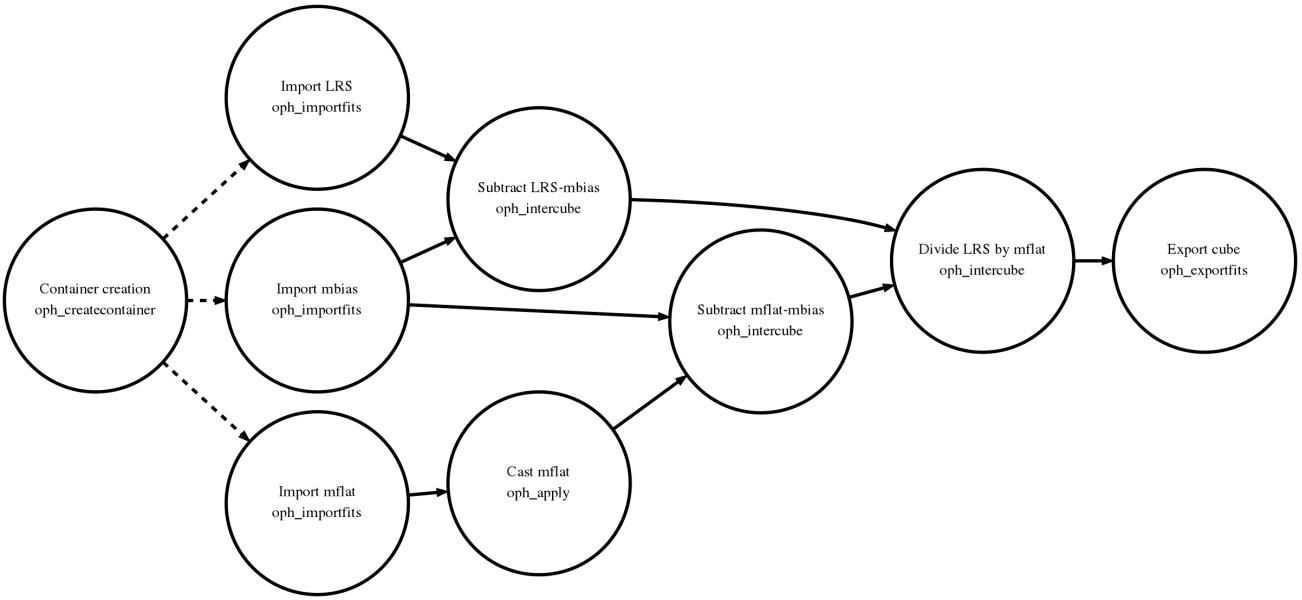
https://www.youtube.com/watch?v=vxbYF1Zhpuc&feature=youtu.be

Images Calibration Workflow

Ophidia has been used to **calibrate astronomic LRS images**.

Input and output datasets are in **FITS** (Flexible Image Transport System) format.

Workflow involves 9 tasks.



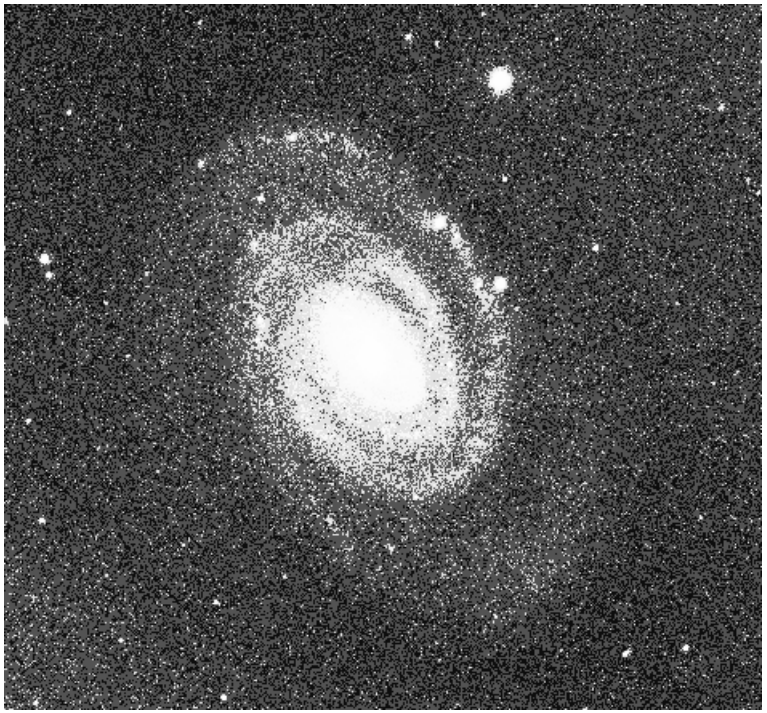
- Creation of the Ophidia container
- Import of the 3 input FITS files: LRS, master flat and master bias
- Cast of the master flat dataset to allow comparison
- Subtraction LRS – mbias
- Subtraction mflat – mbias
- Final Image calibration (LRS-mbias)/(mflat-mbias)
- Export as FITS file

Images Calibration Output

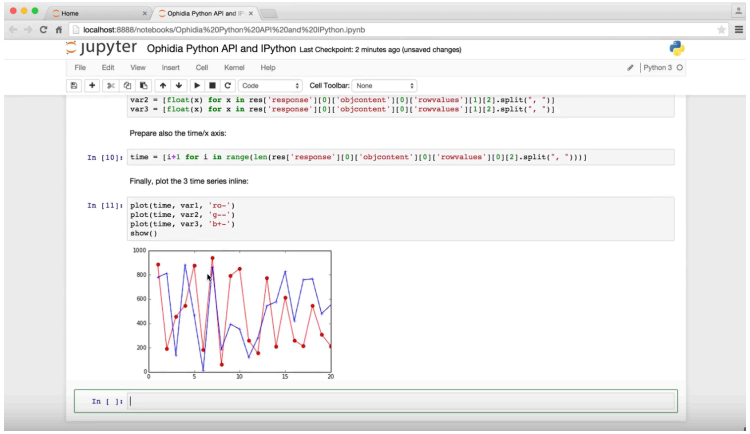
The **Ophidia terminal** is used to submit the workflow and to check the progress status of the execution.

Output data can be displayed in a tabular form using the **explorecube** operator and exported for visualization.

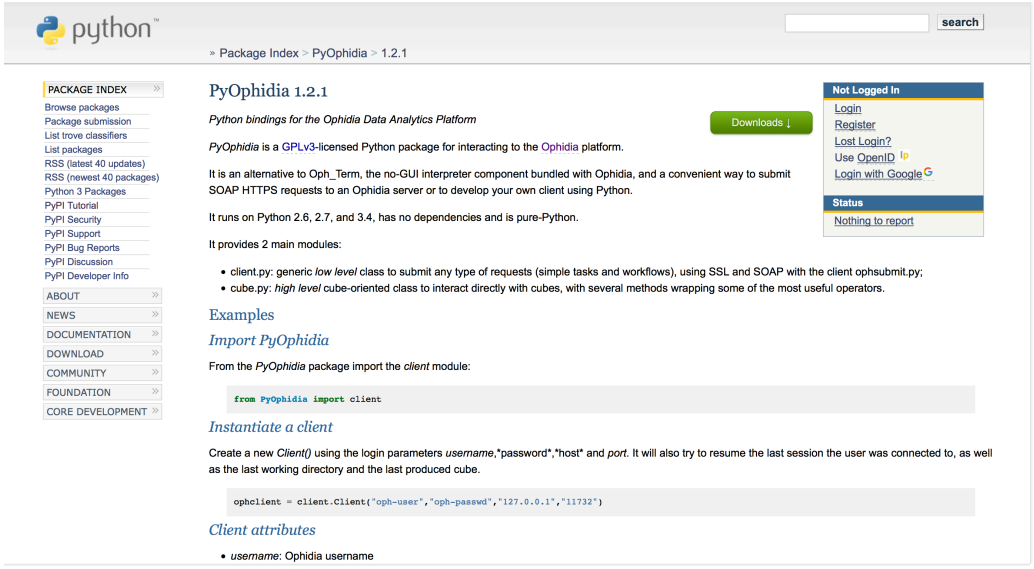
```
[24..2637] >> oph_explorecube limit_filter=5;subset_dims=NAXIS2;subset_filter=1:5;
[Request]:
operator=oph_explorecube;limit_filter=5;subset_dims=NAXIS2;subset_filter=1:5;sessionid
=http://127.0.0.1/ophidia/sessions/247221203340147893451485962378926 37/experiment;exec
_mode=sync;cube=http://127.0.0.1/ophidia/13/206;cwd=/;

[JobID]:
http://127.0.0.1/ophidia/sessions/247221203340147893451485962378926 37/experiment?130#5
28

[Response]:
mymeasure
---------

+========+=================================================================================+
| NAXIS1 | mymeasure                                                                       |
+========+=================================================================================+
| 1      | -0.1520704949, -0.1246676314, -0.1231974343, -0.1362365279, -0.1255137633       |
|--------|---------------------------------------------------------------------------------|
| 2      | 0.0000000000, 71.0000000000, -0.4791666667, -0.2941176471, -0.2807017544        |
|--------|---------------------------------------------------------------------------------|
| 3      | -0.1172566372, -0.1265822785, -0.1584905660, -0.1275797373, -0.1256931608       |
|--------|---------------------------------------------------------------------------------|
| 4      | -0.1373873874, -0.1339449541, -0.0925589837, -0.1188679245, -0.1350844278       |
|--------|---------------------------------------------------------------------------------|
| 5      | -0.1235955056, -0.1266055046, -0.1528301887, -0.1299435028, -0.1851145038       |
+========+=================================================================================+
```



Programmatic access through the PyOphidia class



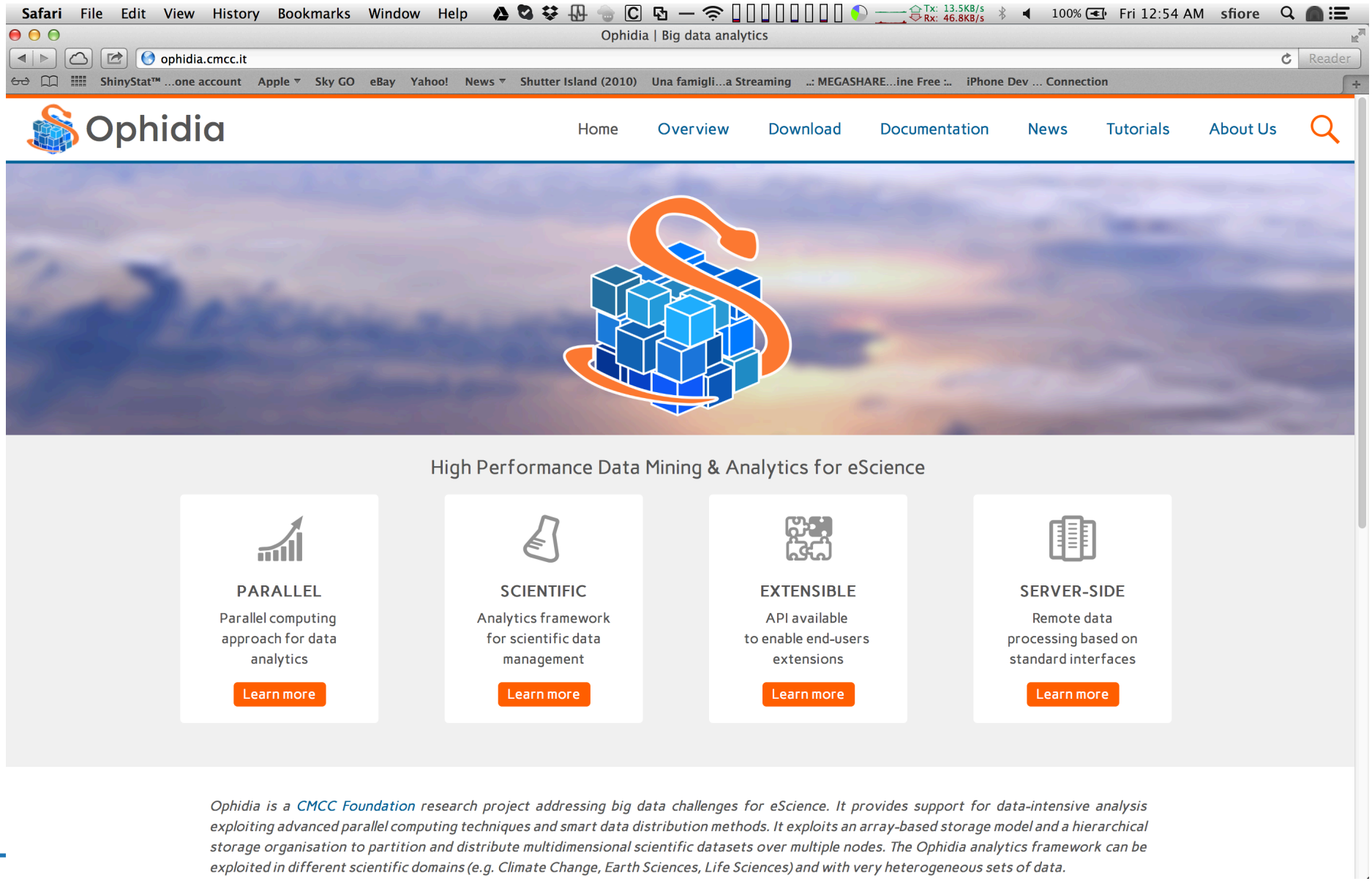
https://www.youtube.com/watch?v=8pcrBXboF6U&feature=youtu.be



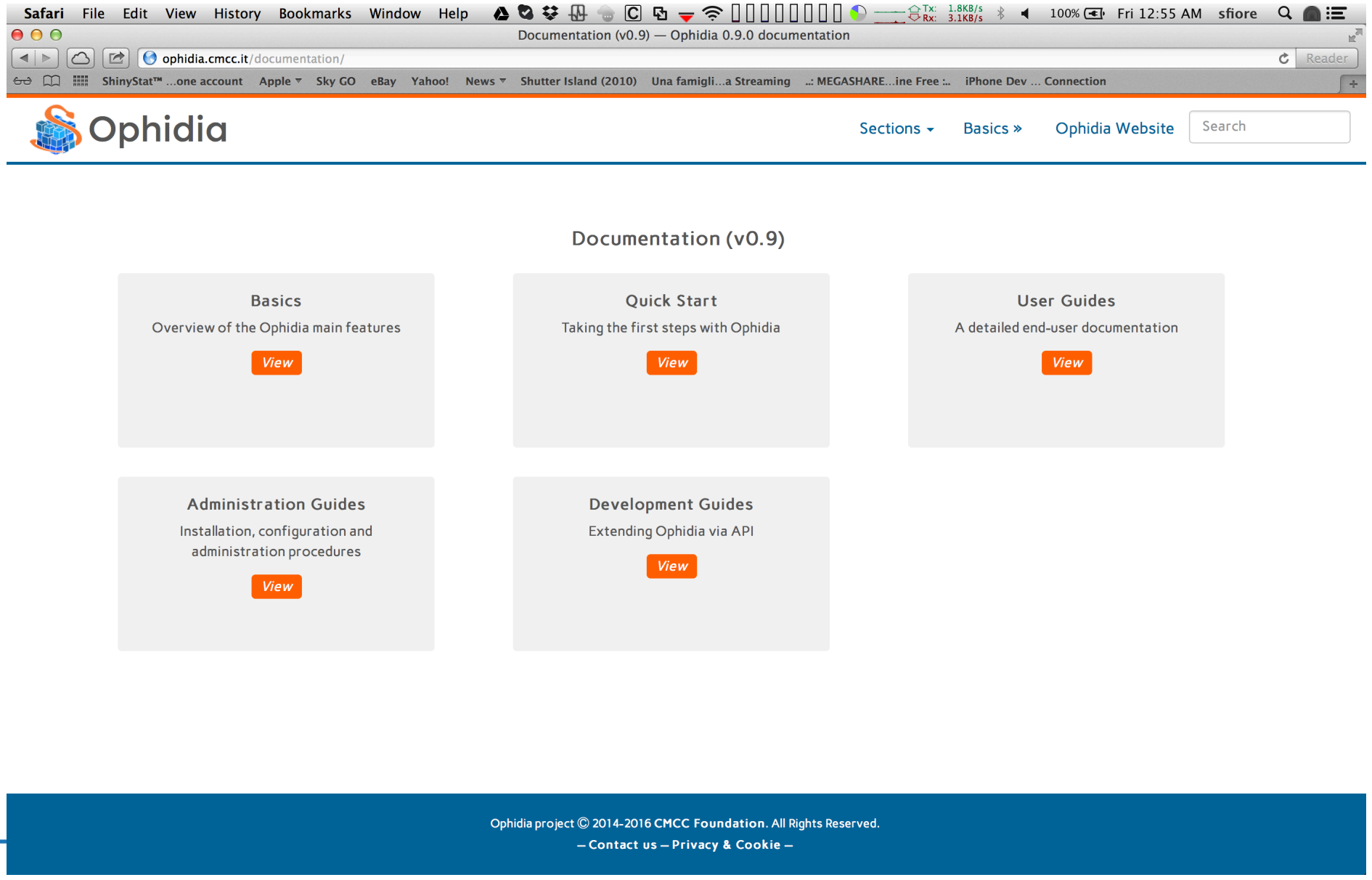
https://pypi.python.org/pypi/PyOphidia/1.2.1

- ✓ **PyOphidia** provides a Python interface to submit commands to the Ophidia Server and to retrieve/deserialize the results
- ✓ Two classes implemented:
  - ✓ **Client class**: connect to the server, navigate into the ophidia file system, submit workflows, manage sessions, etc.
  - ✓ **Cube class**: manipulate cubes (reduce, subset, operations between cubes, intercomparison, etc.), get information on cubes (schema, dimensions, metadata, etc.)

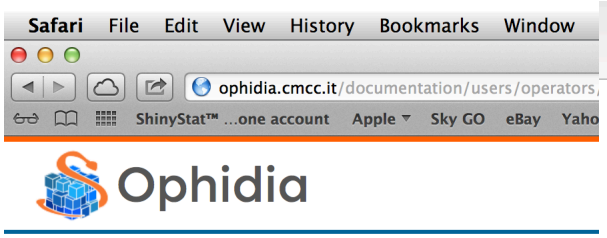
Documentation – http://ophidia.cmcc.it

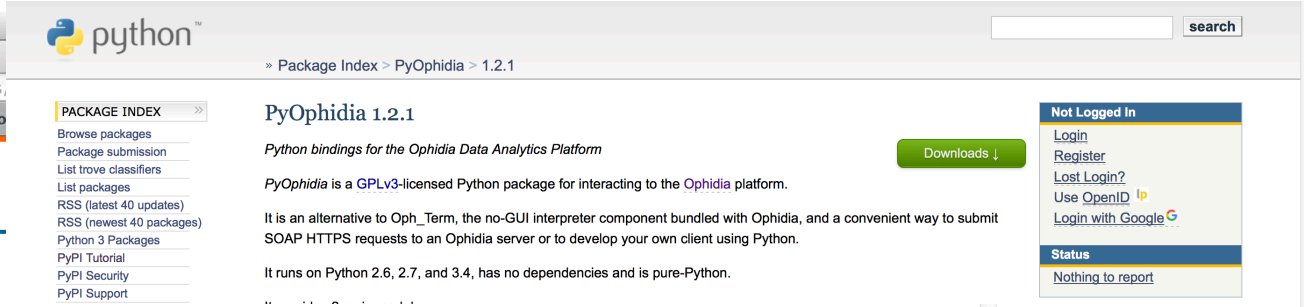


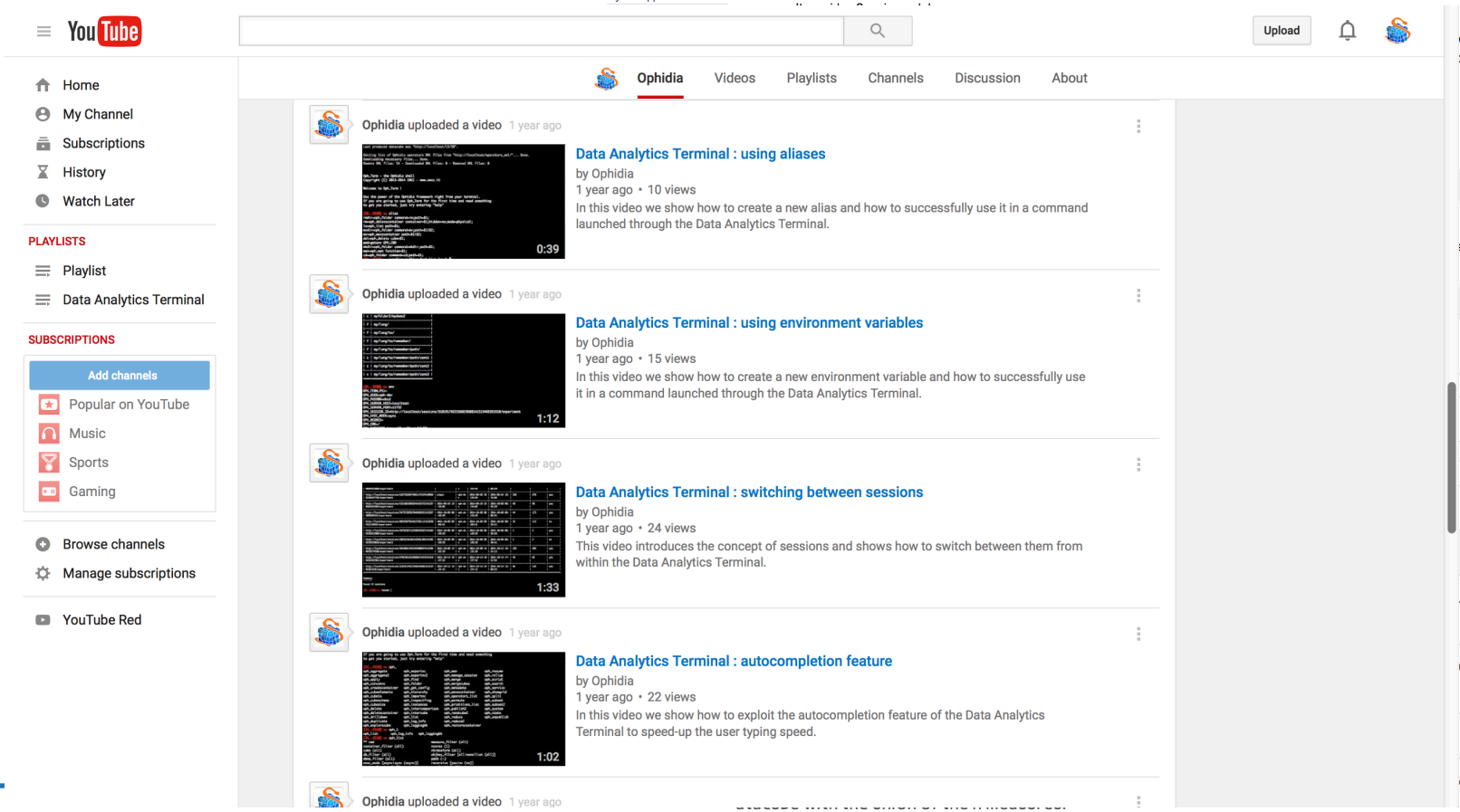
User, administration and devel guides



Operators manual







Conclusions

- ✔ Ophidia is a big data analytics framework for eScience
- ✔ OLAP approach for big data – multidimensional data model
- ✔ Multiple use cases for data analysis in different domains/contexts have been implemented
- ✔ It provides access via CLI (end-users) and API (devel users)
- ✔ Official Release available from February 2016 on github

Future activities will regard:

- Automatic recovery of data from node failures is under testing.
- Implementation of plugins for additional storage systems (e.g. HDFS, WOS).
- Integration of check-point mechanisms to manage data cubes over a two-layer (persistent/memory) storage system

Ophidia – Useful Resources

- *Website: http://ophidia.cmcc.it*
- *Doc : http://ophidia.cmcc.it/documentation*
- *The Ophidia code is available on GitHub under GPLv3 license at https://github.com/OphidiaBigData*
- *RPMs are also available for CentOS7 and Ubuntu14 at the following repo: http://download.ophidia.cmcc.it/*
- *Youtube Channel https://www.youtube.com/user/OphidiaBigData/*
- *A Virtual Machine Image (OVA format) is also available at https://download.ophidia.cmcc.it/vmi_desktop/ to get started in a few minutes with Ophidia*

Thanks

http://ophidia.cmcc.it

@OphidiaBigData

www.youtube.com/user/OphidiaBigData