cmcc
Centro Euro-Mediterraneo
sui Cambiamenti Climatici

www.cmcc.it

Data analytics workflows with Ophidia

Donatello Elia, Sonia Scardigno, Fabrizio Antonio
Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Lecce, Italy

ESiWACE2 HPDA & Vis Training 2022
6th-9th September 2022, Virtual Event

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

Session outline

Introduction to scientific workflows and motivations

Data analytics workflows in Ophidia

Real-world examples of analytics workflow with the Ophidia framework

Workflows core concepts: workflow constructs, execution monitoring

DEMO: Tutorial about workflow creation and execution with Ophidia

HANDS-ON: Data analytics workflows examples

Large-scale climate analysis

Complexity of the analysis leads to the need for ***end-to-end workflow support***

- Typical approaches (mostly based on bash-like scripts) requires climate scientists to take care of implement and replicate workflow-like control logic
- Analyses can require the execution of *tens/hundreds of analytics operators*
  - *Efficient orchestration of the tasks is critical*
  - *Parallelism must be handled both at intra-task and inter-task level*
  - *Task failure should also be considered*

Workflows can represent a way to define ***portable*** and ***re-usable*** analyses (targeting FAIR principles)

Session outline

Introduction to scientific workflows and motivations

Data analytics workflows in Ophidia

Real-world examples of analytics workflow with the Ophidia framework

Workflows core concepts: workflow constructs, execution monitoring

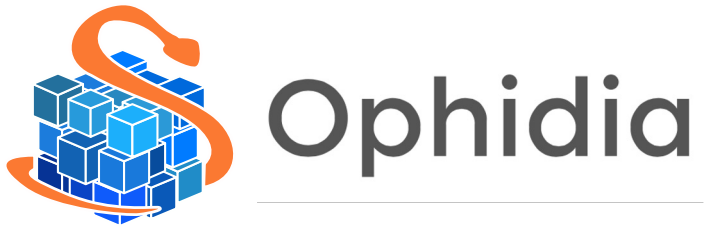
DEMO: Tutorial about workflow creation and execution with Ophidia

HANDS-ON: Data analytics workflows examples

Ophidia HPDA framework

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

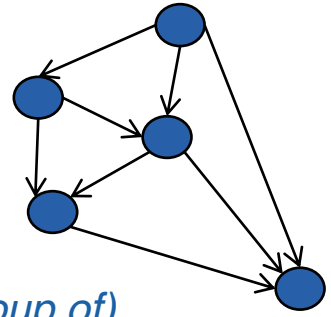
- A **HPDA framework** for multi-dimensional scientific data joining HPC paradigms with scientific data analytics approaches
- **In-memory** and **server-side** data analysis exploiting parallel computing techniques
- Multi-dimensional, array-based, storage model and partitioning schema for scientific data leveraging the **datacube** abstraction
- End-to-end mechanisms to support **interactive analysis, complex experiments** and **large workflows** on scientific data



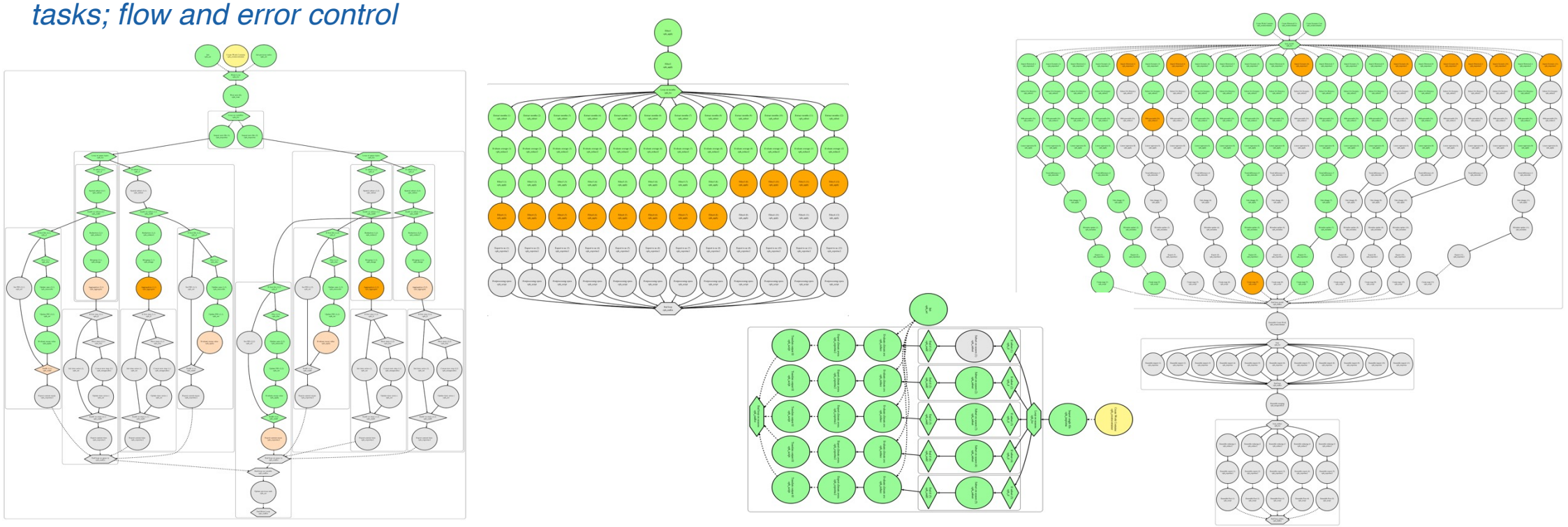
S. Fiore, D. Elia, C. Palazzo, F. Antonio, A. D'Anca, I. Foster, G. Aloisio, "Towards High Performance Data Analytics for Climate Change", ISC High Performance 2019, LNCS Springer, 2019

Analytics workflows

Ophidia supports the execution of complex workflows of operators.

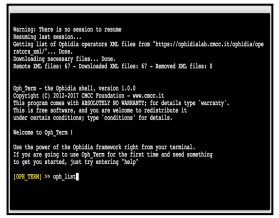


- Defines a **JSON representation** for the workflow DAG specification
- Supports different constructs: *dependencies; massive tasks; iterative (group of) tasks; parallel (group of) tasks; flow and error control*



C. Palazzo, A. Mariello, S. Fiore, A. D'Anca, D. Elia, D. N. Williams, G. Aloisio, "A Workflow-Enabled Big Data Analytics Software Stack for eScience", HPCS 2015, pp. 545-552

Ophidia architecture: front-end layer



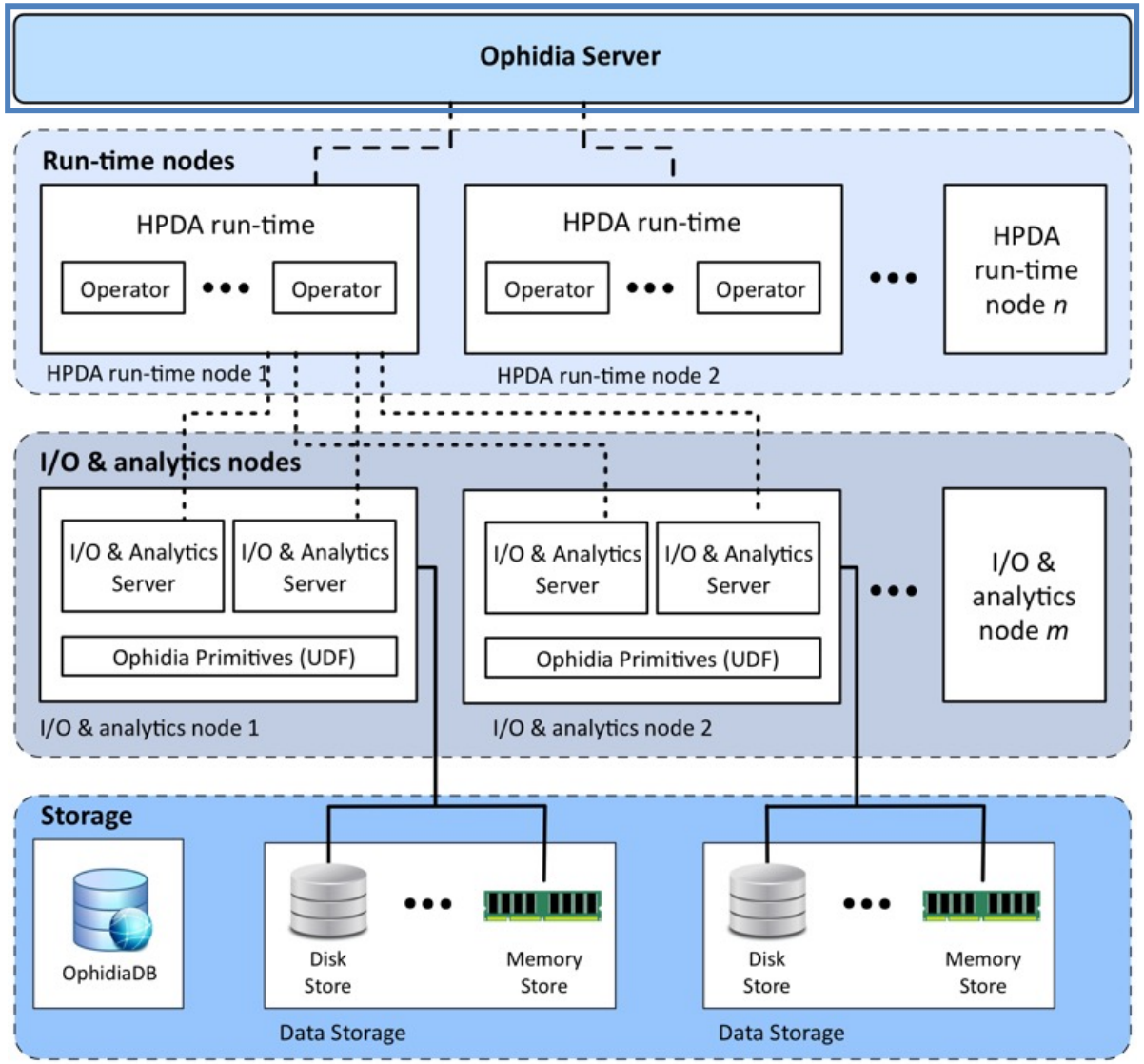
*The **Ophidia Server** is the **multi-interface** server front-end*

*Manages user **authN/authZ, sessions** and enables server-side computation*

*Handles **single task** and **workflows** execution and monitors their execution on the server side*

Remote interactions with:

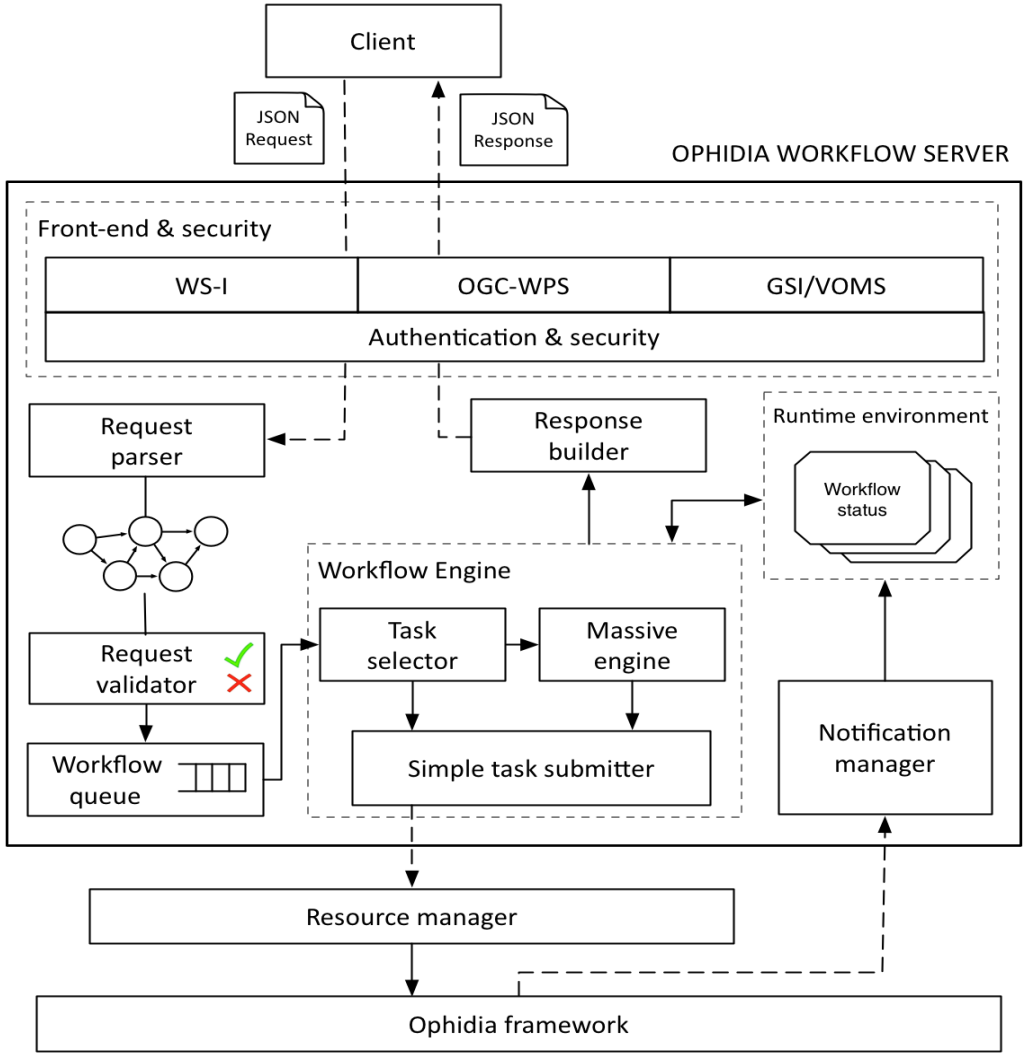
- *Terminal CLI*
- *Python APIs*
- *WPS clients*



The Ophidia Server

The **workflow management system (WMS)** is a core component of the Ophidia Server:

- *manages **user request***
- ***formats the commands** for the analytics framework*
- *handles task **dependencies** and **execution** flow*
- ***submits** the **tasks** to the resource manager*
- ***manages task status** updates*
- ***monitors the workflow** execution*
- *provides the proper **response messages***



C. Palazzo, A. Mariello, S. Fiore, A. D'Anca, D. Elia, D. N. Williams, G. Aloisio, "A Workflow-Enabled Big Data Analytics Software Stack for eScience", HPCS 2015, pp. 545-552

Ophidia Terminal

The **Ophidia Terminal**, a CLI bash-like client for the Ophidia HPDA Framework:

- Executing *interactive* data analytics sessions;
- Submit *batch* data analytics tasks of *workflows*;
- Experiment and operators *debugging*;
- *File system exploration* and *environment management*.

```
[11..4495] >> oph_list level=2;
[Request]:
operator=oph_list;path=;level=2;sessionid=http://127.0.0.1/ophidia/sessions/1112
3869522950595227155862181815 4495/experiment;exec_mode=sync;cdd=/;

[JobID]:
http://127.0.0.1/ophidia/sessions/1112386952295059522715586218181544 95/experiment?2#45

[Response]:
Ophidia Filesystem: /
---------------------
+===+====================+================================================+=============+
| T | PATH               | DATACUBE PID                                   | DESCRIPTION |
+===+====================+================================================+=============+
| f | testFolder/        |                                                |             |
|---|--------------------|------------------------------------------------|-------------|
| c | test               | http://127.0.0.1/ophidia/2917/374976           |             |
+===+====================+================================================+=============+
```

Session outline

Introduction to scientific workflows and motivations

Data analytics workflows in Ophidia

Real-world examples of analytics workflow with the Ophidia framework

Workflows core concepts: workflow constructs, execution monitoring

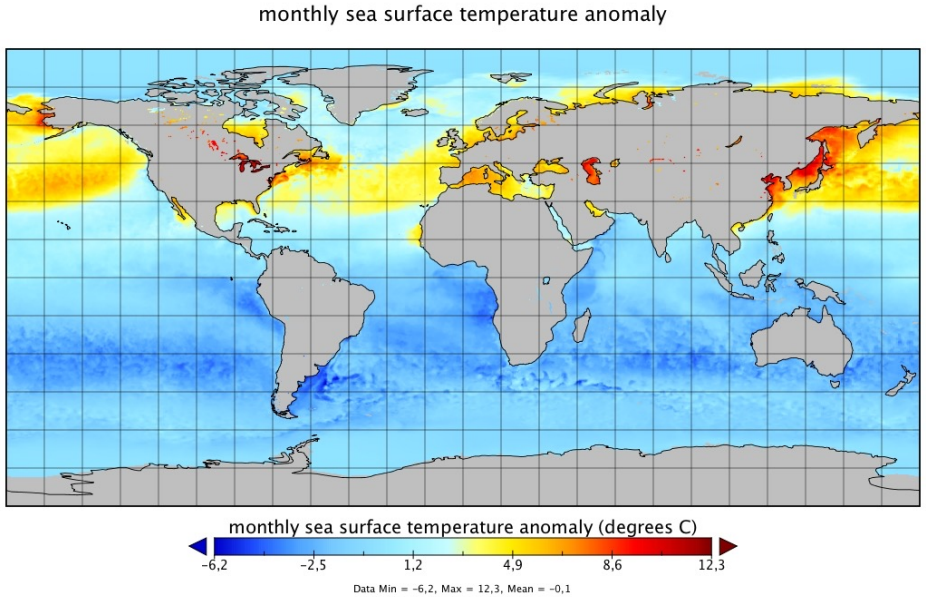
DEMO: Tutorial about workflow creation and execution with Ophidia

HANDS-ON: Data analytics workflows examples

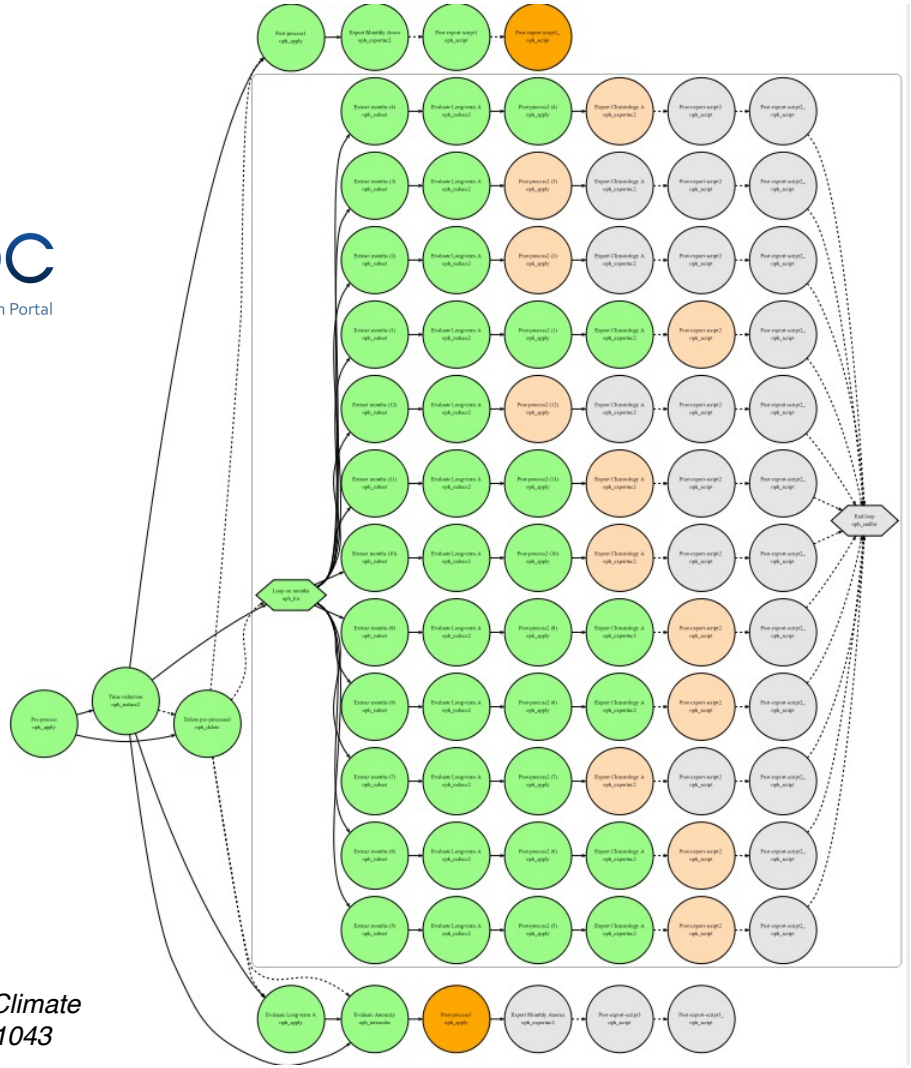
Workflow example I: climate indicators processing

SST (monthly) mean, anomaly, climatological mean

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



Climate Information Portal



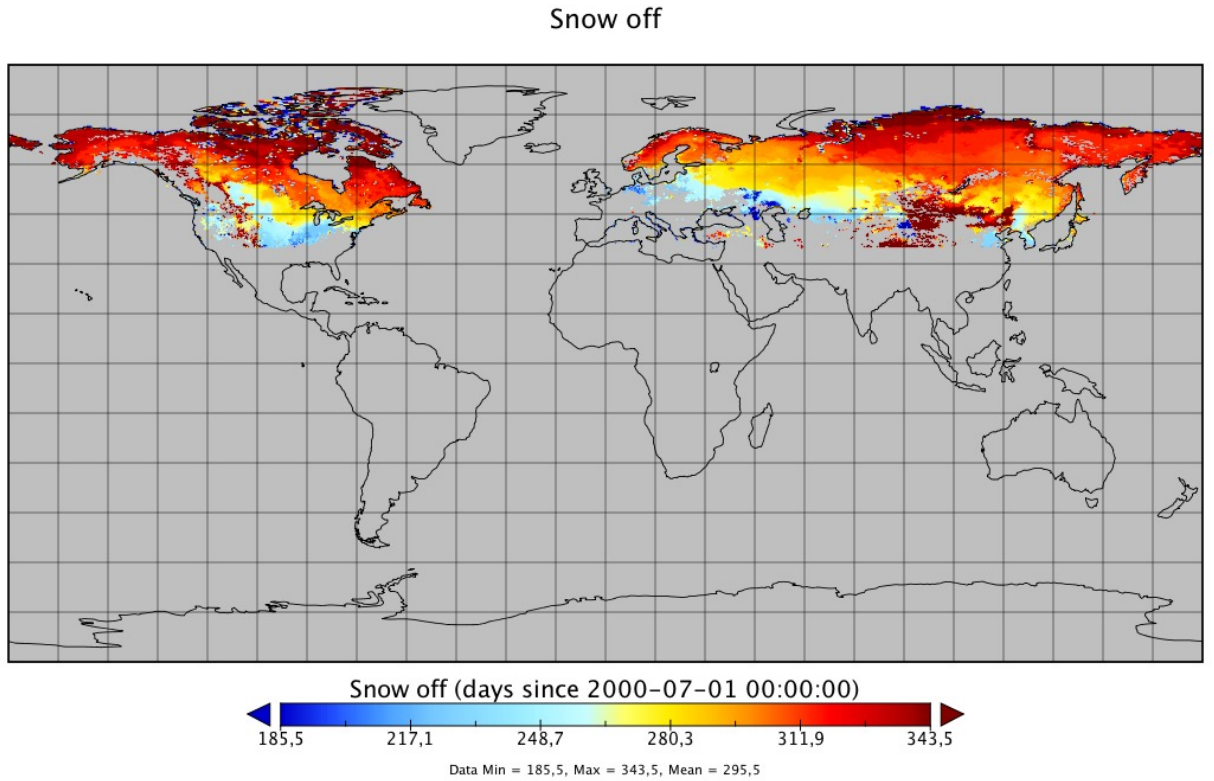
A. D'Anca, et al., "On the Use of In-memory Analytics Workflows to Compute eScience Indicators from Large Climate Datasets," 2017 17th IEEE/ACM Int. Symposium on Cluster, Cloud and Grid Computing (CCGRID), pp. 1035-1043

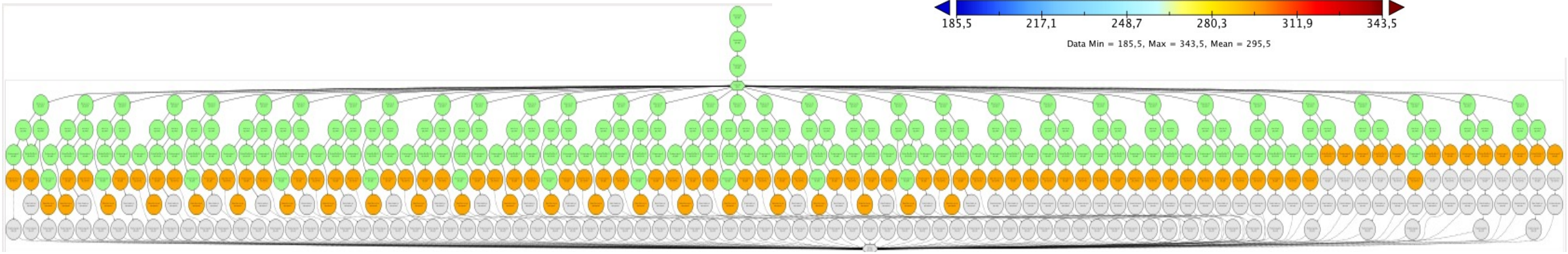
Workflow example II: climate indicators processing

Snow on/off – Length of snow season
(single workflow for 3 indicators)

- *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*

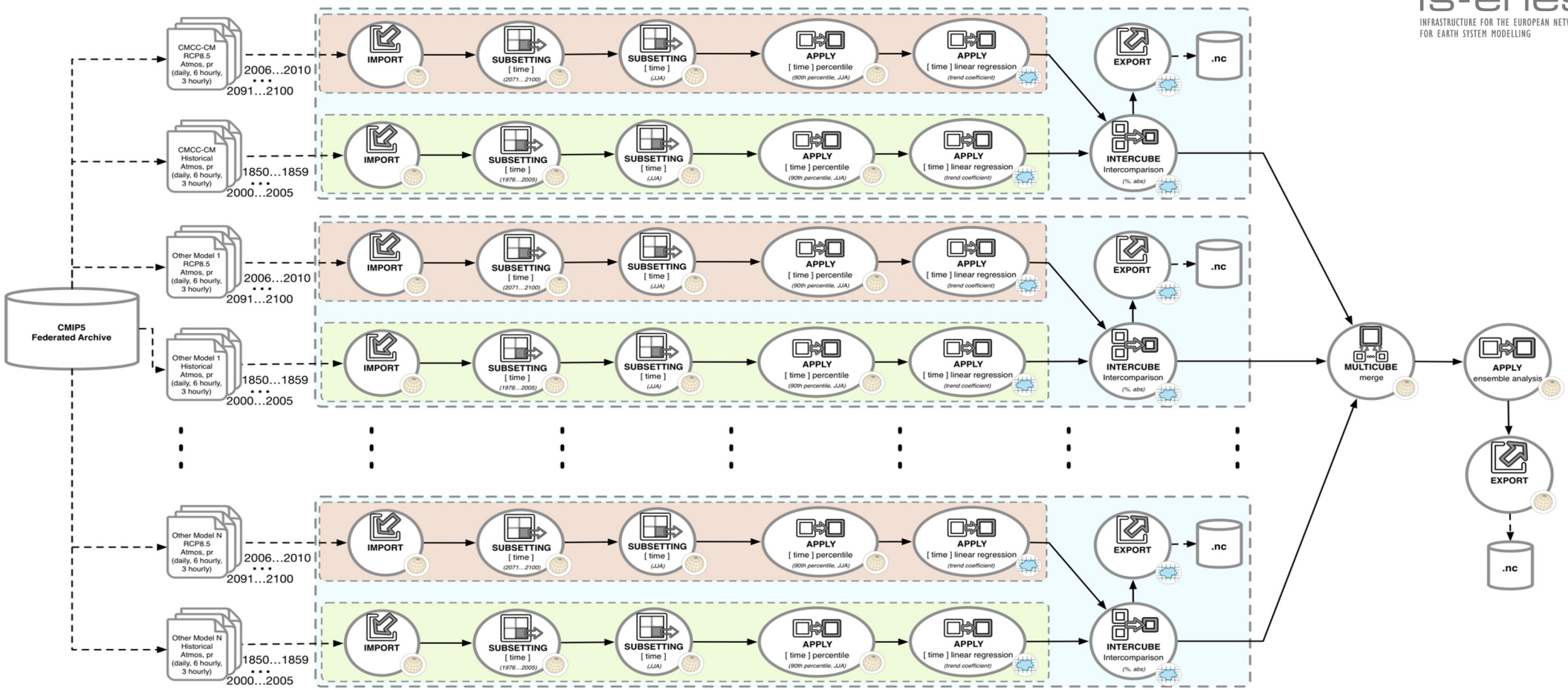
Climate Information Portal





Workflow example III: multi-model experiment design

Precipitation Trend Analysis use case implemented as an Ophidia workflow



S. Fiore, et al., "Distributed and cloud-based multi-model analytics experiments on large volumes of climate change data in the earth system grid federation eco-system". In Big Data (Big Data), 2016 IEEE Int. Conference on. IEEE, 2016. pp. 2911-2918

Workflow example III: multi-model experiment design

Precipitation Trend Analysis use case implemented as an Ophidia workflow

Single model precipitation trend analysis

S. Fiore, et al., "Distributed and cloud-based multi-model analytics experiments on large volumes of climate change data in the earth system grid federation eco-system". In Big Data (Big Data), 2016 IEEE Int. Conference on. IEEE, 2016. pp. 2911-2918

Workflow example III: multi-model experiment design

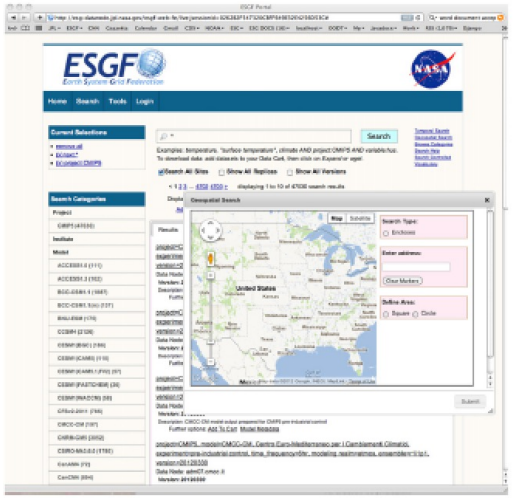
Precipitation Trend Analysis use case implemented as an Ophidia workflow

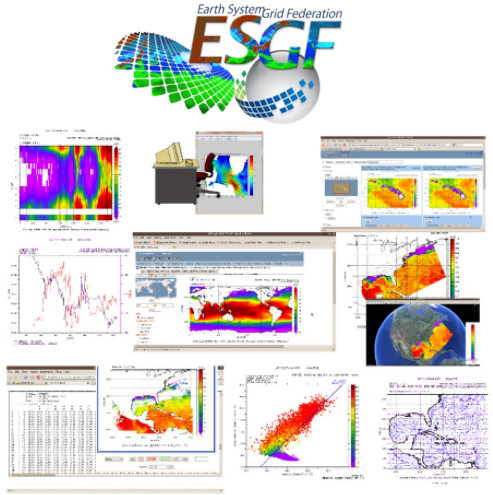
Single model precipitation trend analysis

S. Fiore, et al., "Distributed and cloud-based multi-model analytics experiments on large volumes of climate change data in the earth system grid federation eco-system". In Big Data (Big Data), 2016 IEEE Int. Conference on. IEEE, 2016. pp. 2911-2918

Multi-model experiment input data

ESGF[1] is a coordinated multiagency, international collaboration of institutions that continually develop, deploy, and maintain software needed to facilitate and empower the study of climate.



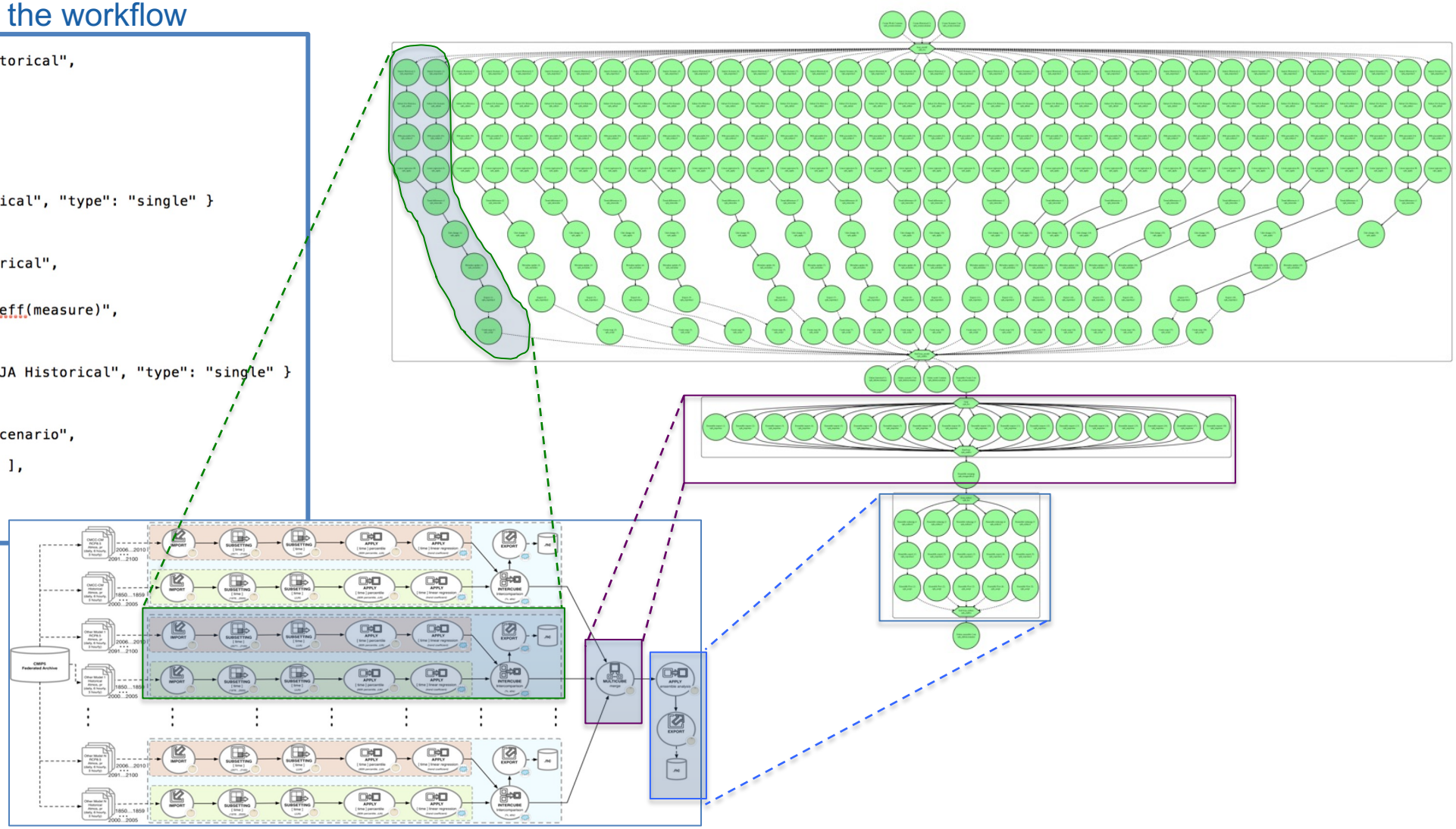


Model acronym	Model expansion	Institute
CCSM4	Community Climate System Model, v4	National Center for Atmospheric Research (NCAR)
CMCC-CESM	CMCC - Community Earth System Model	Euro-Mediterranean Center on Climate Change (CMCC)
CMCC-CMS	CMCC - Coupled Modeling System	Euro-Mediterranean Center on Climate Change (CMCC)
CMCC-CM	CMCC - Climate Model	Euro-Mediterranean Center on Climate Change (CMCC)
CNRM-CM5	CNRM - Coupled Global Climate Model, v5	Centre National de Recherches Météorologiques (CNRM)/Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS)
CSIRO Mk3.6.0	CSIRO Mark, v3.6.0	Commonwealth Scientific and Industrial Research Organisation (CSIRO) in collaboration with Queensland Climate-Change Centre of Excellence (QCCCE)
CanESM2	Second Generation Canadian Earth System Model	Canadian Centre for Climate Modelling and Analysis (CCCma)
GFDL-CM3	GFDL Climate Model, v3	National Oceanic and Atmospheric Administration (NOAA)/Geophysical Fluid Dynamics Laboratory (GFDL)
GFDL-ESM2G	GFDL Earth System Model with Generalized Ocean Layer Dynamics (GOLD) component	National Oceanic and Atmospheric Administration (NOAA)/Geophysical Fluid Dynamics Laboratory (GFDL)
GFDL-ESM2M	GFDL Earth System Model with Modular Ocean Model 4 (MOM4) component	National Oceanic and Atmospheric Administration (NOAA)/Geophysical Fluid Dynamics Laboratory (GFDL)
HadGEM2-CC	Hadley Centre Global Environment Model, v2 (Carbon Cycle)	Met Office (UKMO) Hadley Centre (HC)
HadGEM2-ES	Hadley Centre Global Environment Model, v2 (Earth System)	Met Office (UKMO) Hadley Centre (HC)
INM-CM4.0	INM Coupled Model, v4.0	Institute of Numerical Mathematics (INM)
IPSL-CM5A-MR	IPSL Coupled Model, version 5, coupled with NEMO, mid resolution	L'Institut Pierre-Simon Laplace (IPSL)
MIROC5	Model for Interdisciplinary Research on Climate, v5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology
MPI-ESM-MR	MPI Earth System Model, medium resolution	Max Planck Institute for Meteorology (MPI-M)
MRI-CGCM3	MRI Coupled Atmosphere - Ocean General Circulation Model, v3	Meteorological Research Institute (MRI)
NorESM1-M	Norwegian Earth System Model, v1 (intermediate resolution)	Norwegian Climate Centre (NCC)

Multi-model experiment implementation & execution

JSON implementation of the workflow

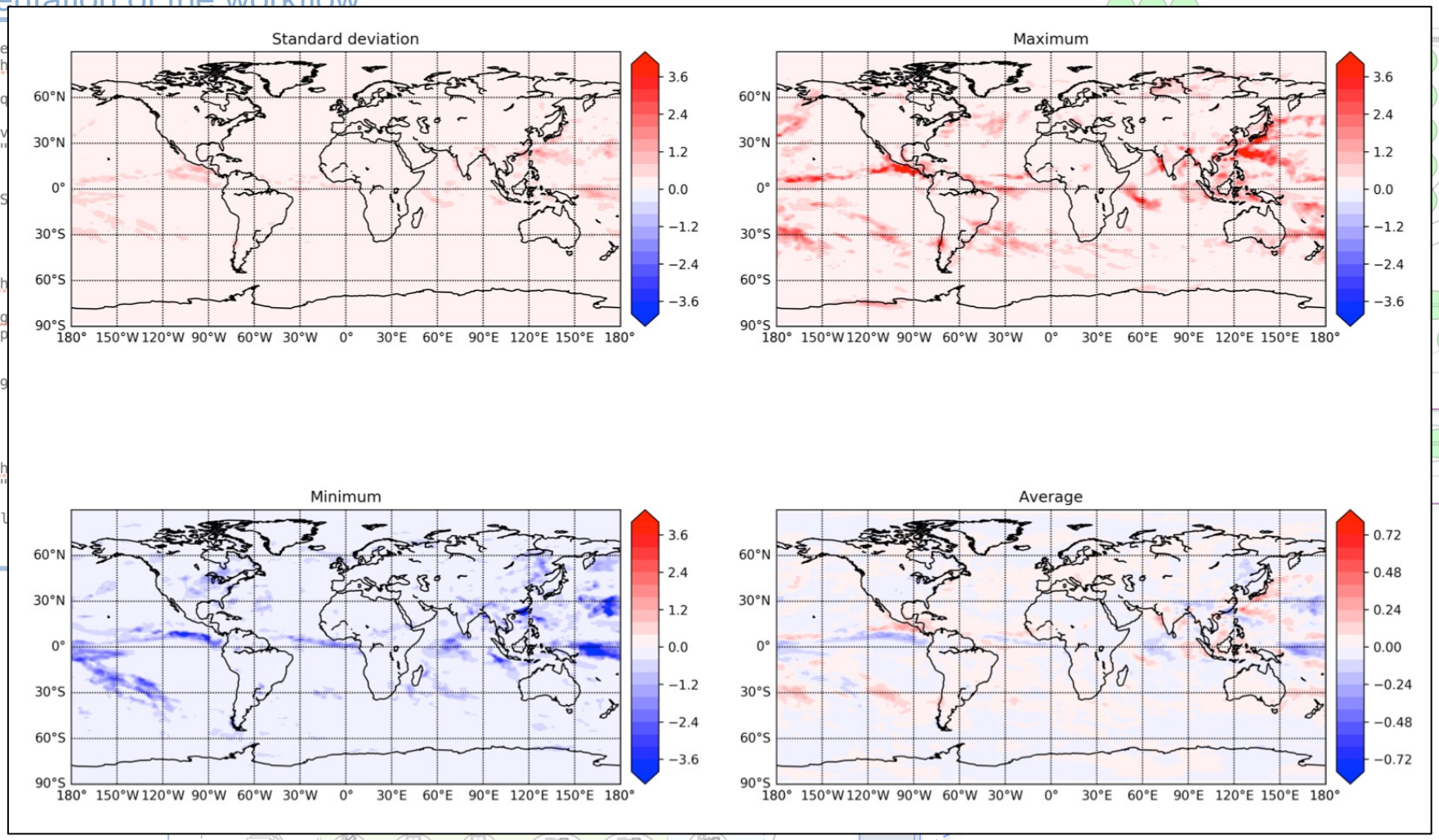
```
    {
        "name": "90th percentile JJA Historical",
        "operator": "oph_reduce2",
        "arguments": [
            "operation=quantile",
            "dim=time",
            "concept_level=y",
            "order=${5}"
        ],
        "dependencies": [
            { "task": "Subset JJA Historical", "type": "single" }
        ]
    },
    {
        "name": "Linear regression Historical",
        "operator": "oph_apply",
        "arguments": [
            "query=oph_gsl_fit_linear_coeff(measure)",
            "measure_type=auto"
        ],
        "dependencies": [
            { "task": "90th percentile JJA Historical", "type": "single" }
        ]
    },
    {
        "name": "Import Type Selection Scenario",
        "operator": "oph_if",
        "arguments": [ "condition=${10}" ],
        "dependencies": [
            { "task": "loop_model" }
        ]
    },
```

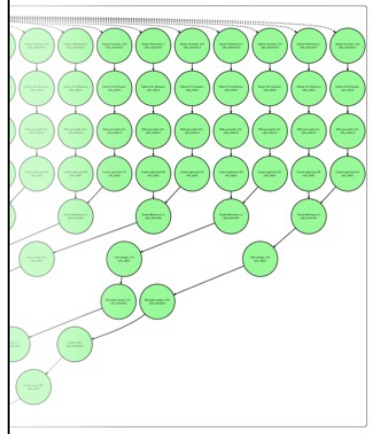


Multi-model experiment implementation & execution

JSON implementation of the workflow

```
{
    "name": "90th pe
    "operator": "oph
    "arguments": [
        "operation=q
        "dim=time",
        "concept_lev
        "order=${5}"
    ],
    "dependencies":
        { "task": "S
    ]
},
{
    "name": "Linear
    "operator": "oph
    "arguments": [
        "query=oph_q
        "measure_typ
    ],
    "dependencies":
        { "task": "9
    ]
},
{
    "name": "Import
    "operator": "oph
    "arguments": [ "
    "dependencies":
        { "task": "l
    ]
},
```





Session outline

Introduction to scientific workflows and motivations

Data analytics workflows in Ophidia

Real-world examples of analytics workflow with the Ophidia framework

Workflows core concepts: workflow constructs, execution monitoring

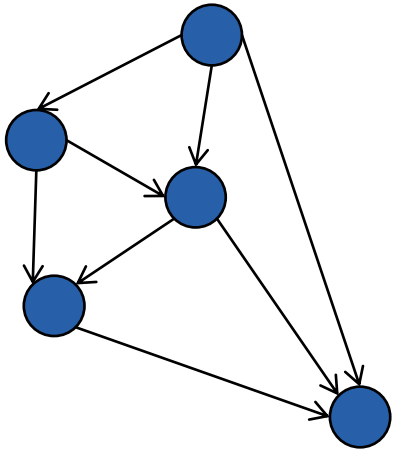
DEMO: Tutorial about workflow creation and execution with Ophidia

HANDS-ON: Data analytics workflows examples

Analytics Workflow Schema

Ophidia **workflows schema**:

- based on **JSON representation** for requests/responses
- defines application-level **semantic** and **syntactic rules**
- models scientific computations as **DAG**



Main supported abstractions:

- *Shared properties*
- *Flow/data dependencies*
- *Simple/massive tasks*
- *Iterative (group of) tasks*
- *Parallel (group of) tasks*
- *Flow and error control*
- *Interleaving and interactive tasks*
- *Execution checkpointing*

Behind the scene: workflow JSON representation

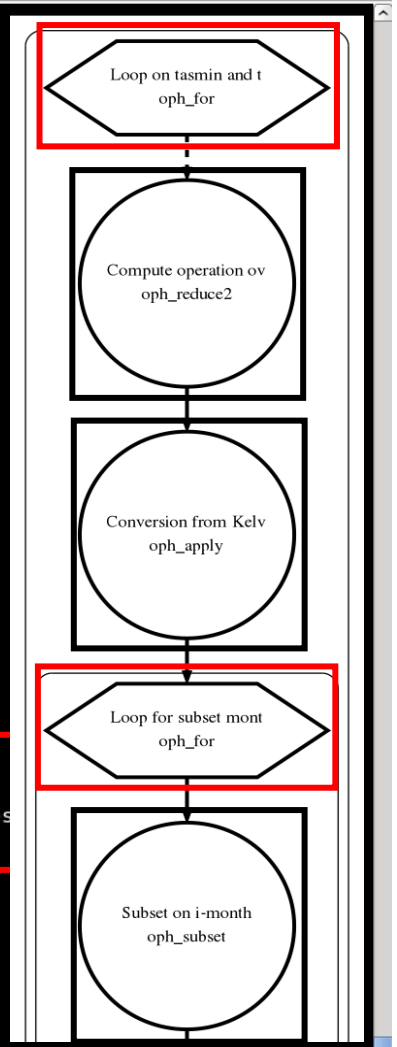
ophrpm@ophidiarpm:~/workflow | ophrpm@ophidiarpm:~/workflow

```
    "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|Nov|Dec", "parallel=yes" ],
                        "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%)
```

Behind the scene: workflow JSON representation

ophrpm@ophidiarpm:~/workflow

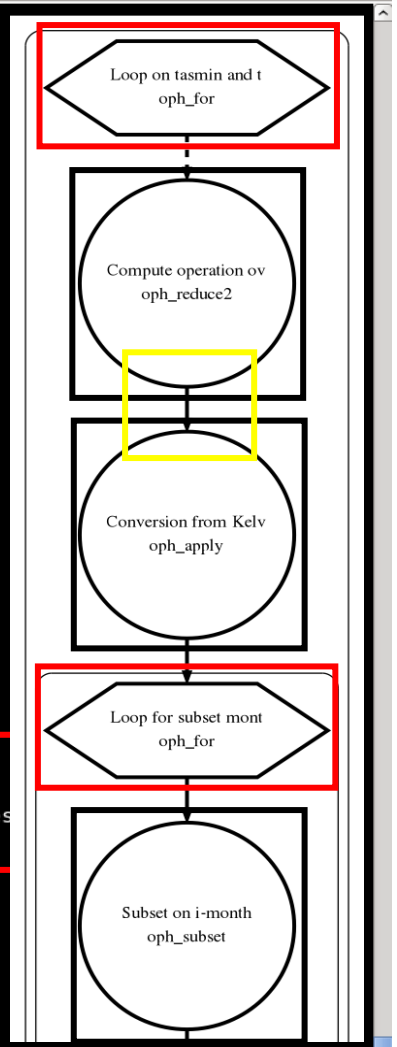
```
"tasks": [
    {
        "name": "Loop on tasmin and tasmax cubes",
        "operator": "oph_for",
        "arguments": [ "name=cube","counter=1:2", "values=${1}|${2}", "parallel=yes" ]
    },
    {
        "name": "Compute operation over time",
        "operator": "oph_reduce2",
        "arguments": [
                "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|Nov|Dec", "parallel=yes
        "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%)
```



Behind the scene: workflow JSON representation

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|Nov|Dec", "parallel=yes
        "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%)
```



Python module for workflow management

ESDM-PAV Client Python module provide an interface to model *Post-processing, Analytics and Visualisation* (PAV) experiments composed of multiple tasks (https://github.com/OphidiaBigData/esdm-pav-client):

- Developed in the context of the **ESiWACE2 project** for the ESDM-PAV software stack
- High-level API to **model experiments** with JSON-based schema
  - Experiments and tasks handled as Python objects
- Support for workflow **execution** and **monitoring**
- Native integration with **Jupyter Notebooks**
- Integrates Ophidia operators and can be **interfaced directly with the Ophidia WMS** and all its capabilities

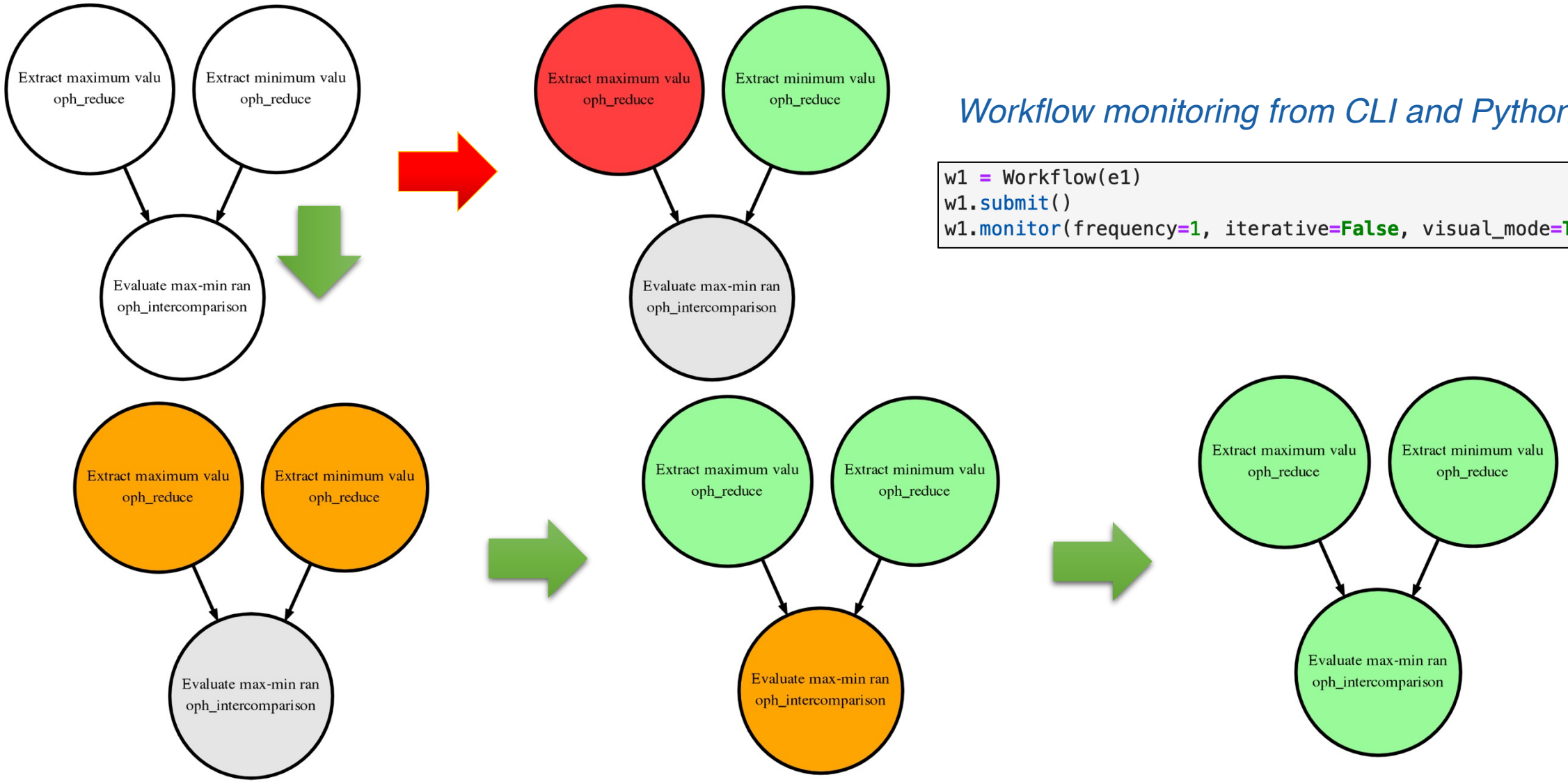
```python
from esdm_pav_client import Workflow, Experiment

e1 = Experiment(
    name="Example workflow",
    author="CMCC"
)
t1 = e1.newTask(name="Import",
                type="ophidia",
                operator='oph_importnc',
                arguments={'measure': 'tas',
                           'input': 'tas.nc',
                           'description': 'Imported data'})
t2 = e1.newTask(name="Subset arrays",
                type="ophidia",
                operator='oph_subset',
                arguments={'subset_filter': '1:365',
                           'subset_dims': 'time',
                           'description': 'First year'},
                dependencies={t1:'cube'})
t3 = e1.newTask(name="Average",
                type="ophidia",
                operator='oph_reduce',
                arguments={'operation': 'avg',
                           'description': 'Average cube'},
                dependencies={t2:'cube'})
t4 = e1.newTask(name="Export",
                type="ophidia",
                operator='oph_exportnc2',
                arguments={'output': 'out.nc'},
                dependencies={t3:'cube'})

e1.check()

w1 = Workflow(e1)
w1.submit()
w1.monitor(frequency=1, iterative=False, visual_mode=True)
w1.cancel()
```

Workflow status monitoring



Workflow monitoring from CLI and Python

```
w1 = Workflow(e1)
w1.submit()
w1.monitor(frequency=1, iterative=False, visual_mode=True)
```

Analytics workflows constructs

Workflow Management

This group includes a number of flow control operators that could be used within an Ophidia workflow to implement complex data processing in batch mode. In particular, they implement several advanced features: setting of run-time variables, iterative and parallel interface, selection interface, interactive workflows, interleaving workflows, etc.

NAME	DESCRIPTION
OPH_ELSE	Start the last sub-block of a selection block "if".
OPH_ELSEIF	Start a new sub-block of a selection block "if".
OPH_ENDFOR	Close a loop "for".
OPH_ENDIF	Close a selection block "if".
OPH_FOR	Implement a loop "for".
OPH_IF	Open a "if" selection block.
OPH_INPUT	It sends commands or data to an interactive task.
OPH_SET	Set a parameter in the workflow environment.
OPH_WAIT	Wait until an event occurs.

Flow control: iterative interface (“for loop”)

Allows **to repeat the execution of a block** of workflow **tasks** over different input data or over the result of the previous iteration.

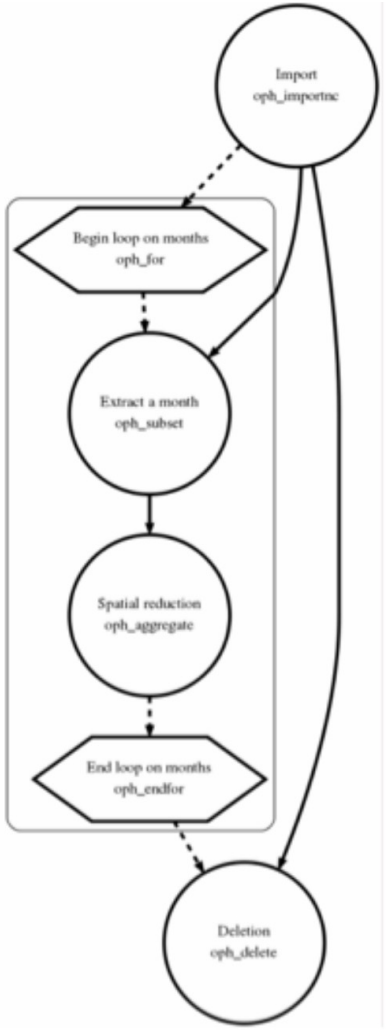
Selection interface operators:

- *OPH_FOR or FOR*
- *OPH_ENDFOR or ENDFOR*

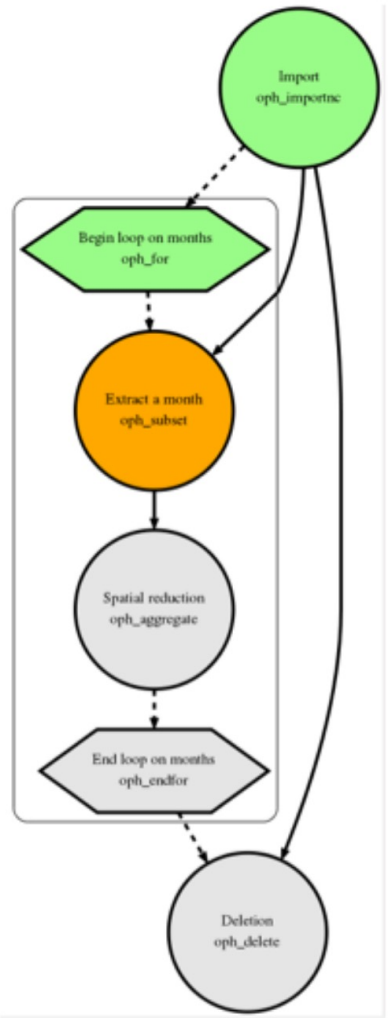
The statement can be used in nested fashion

```
{
        "name": "Begin loop on months",
        "operator": "oph_for",
        "arguments":
        [
                "name=index",
                "counter=1:12",
                "values=Jan|Feb|Mar|Apr|May|Jun|Jul|Aug|Sep|Oct|Nov|Dec"
        ]
},
```

AT DEFINITION TIME



AT RUNTIME



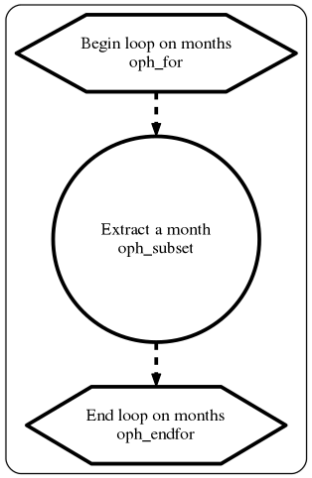
Workflow iterative interface documentation: http://ophidia.cmcc.it/documentation/users/workflow/workflow_for.html

Flow control: parallel interface

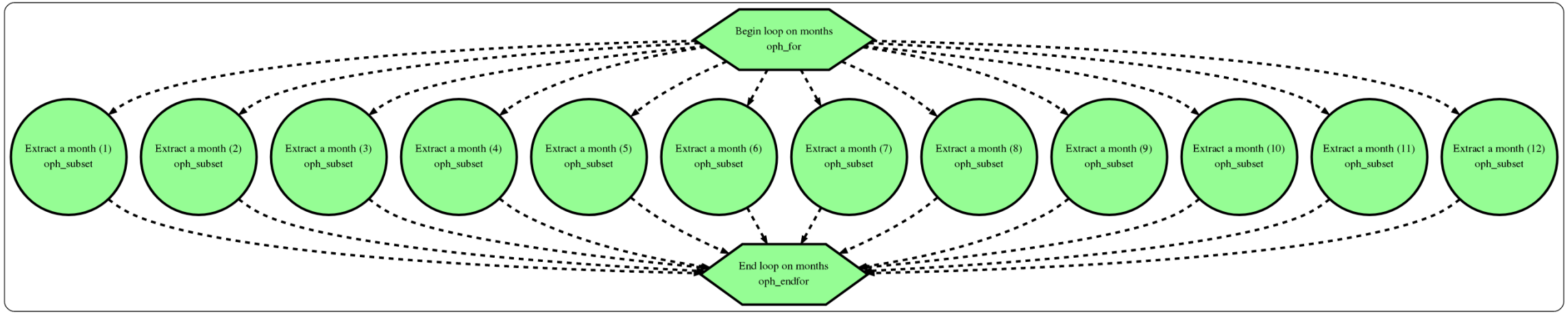
Extension of the OPH_FOR interface for parallel (**concurrent**) execution of the loop iterations.

```
{
        "name": "Begin loop on months",
        "operator": "oph_for",
        "arguments":
        [
                "parallel=yes",
                "name=index",
                "counter=1:12",
                "values=Jan|Feb|Mar|Apr|May|Jun|Jul|Aug|Sep|Oct|Nov|Dec"
        ]
}
```

AT DEFINITION TIME



AT RUNTIME



Workflow parallel interface documentation: http://ophidia.cmcc.it/documentation/users/workflow/workflow_for.html#parallel-interface

Flow control: selection Interface (“if/else”)

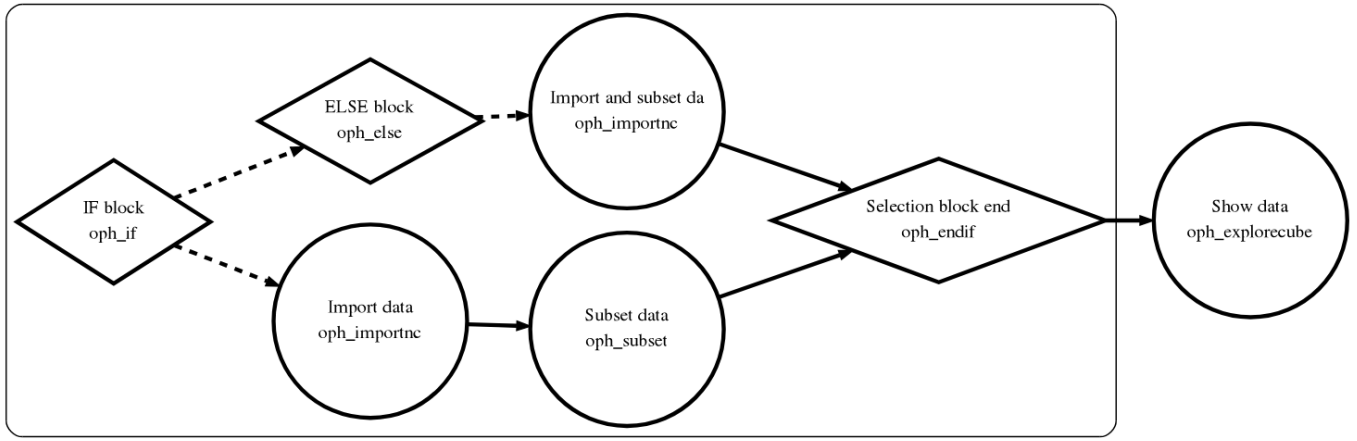
Enables the workflow manager to **dynamically execute a block of tasks** based on boolean conditions evaluated at run-time.

Selection interface operators:

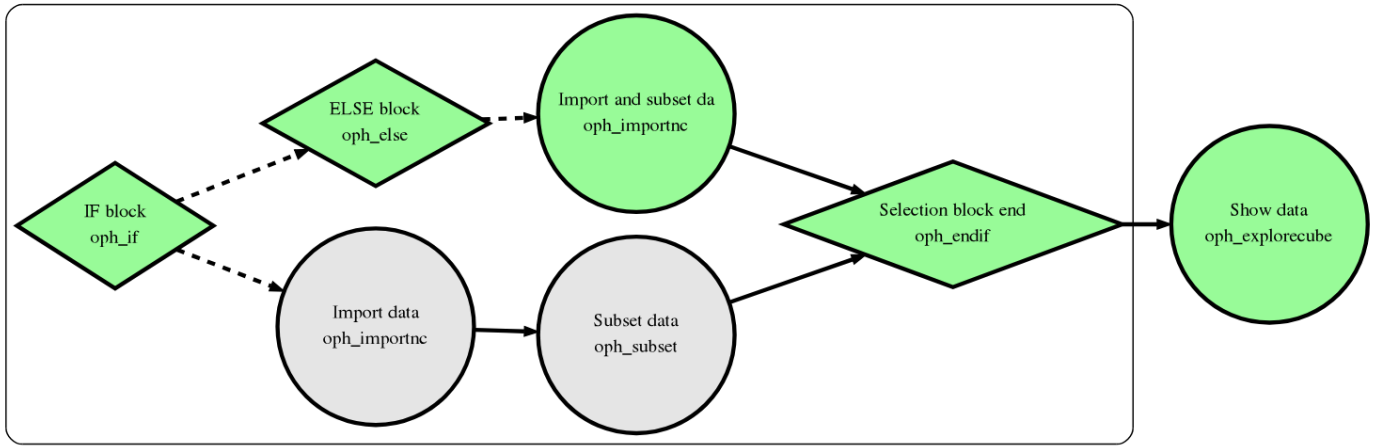
- *OPH_IF or IF*
- *OPH_ELSEIF or ELSEIF*
- *OPH_ELSE or ELSE*
- *OPH_ENDIF or ENDIF*

```
{
        "name": "IF block",
        "operator": "oph_if",
        "arguments": [ "condition=$1" ]
},
```

AT DEFINITION TIME



AT RUNTIME



Workflow selection interface documentation: http://ophidia.cmcc.it/documentation/users/workflow/workflow_if.html

Workflow error handling

In case of very large workflow executions **errors** in one of more **tasks** are likely.

Supported behaviours in case of task failure:

- *abort:* the whole workflow is interrupted (default)
- *skip:* the task is skipped and execution continues on the descendant tasks
- *continue:* the task and all depending task will be ignored, while other task will be executed
- *repeat N:* the task is re-executed N times, before aborting the whole execution. A custom *timeout interval* between resubmissions can be specified.

DEFINED AT GLOBAL WORKFLOW LEVEL

```
"name": "Example5",
"author": "Foo",
"abstract": "Simple workflow with automatic repetition
"exec_mode": "sync",
"ncores": "1",
"cube": "http://hostname/1/1",
"on_error": "repeat 2",
"tasks":
```

DEFINED AT TASK LEVEL (precedence)

```
{
        "name": "Extract maximum value",
        "operator": "oph_reduce",
        "arguments": [ "operation=max" ],
        "on_error": "repeat 5"
},
```

Workflow error handling documentation: http://ophidia.cmcc.it/documentation/users/workflow/workflow_advanced.html#handling-task-errors

Workflow submission

```
e1 = Experiment.load("Tind_loop.json")
w1 = Workflow(e1)
w1.submit("http://193.204.199.174/ophidia/29/2046",
          "http://193.204.199.174/ophidia/29/2047",
          "max")
```

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_ |
| 223251144945516614 6380/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_ |
| 223251144945516614 6380/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_ |
| 223251144945516614 6380/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_ |
| 223251144945516614 6380/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_ |
| 223251144945516614 6380/experiment?247#3149             | 1449455166146380     |        |       |           | Celsius degrees (2)        | PLE | COMPLETED   |
|---------------------------------------------------------|----------------------|--------|-------|-----------|----------------------------|-----|-------------|
```

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_ |
| 223251144945516614 6380/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_ |
| 223251144945516614 6380/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_ |
| 223251144945516614 6380/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_ |
| 223251144945516614 6380/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_ |
| 223251144945516614 6380/experiment?247#3149             | 1449455166146380     |        |       |           | Celsius degrees (2)        | PLE | COMPLETED   |
|---------------------------------------------------------|----------------------|--------|-------|-----------|----------------------------|-----|-------------|
```

Workflow submission

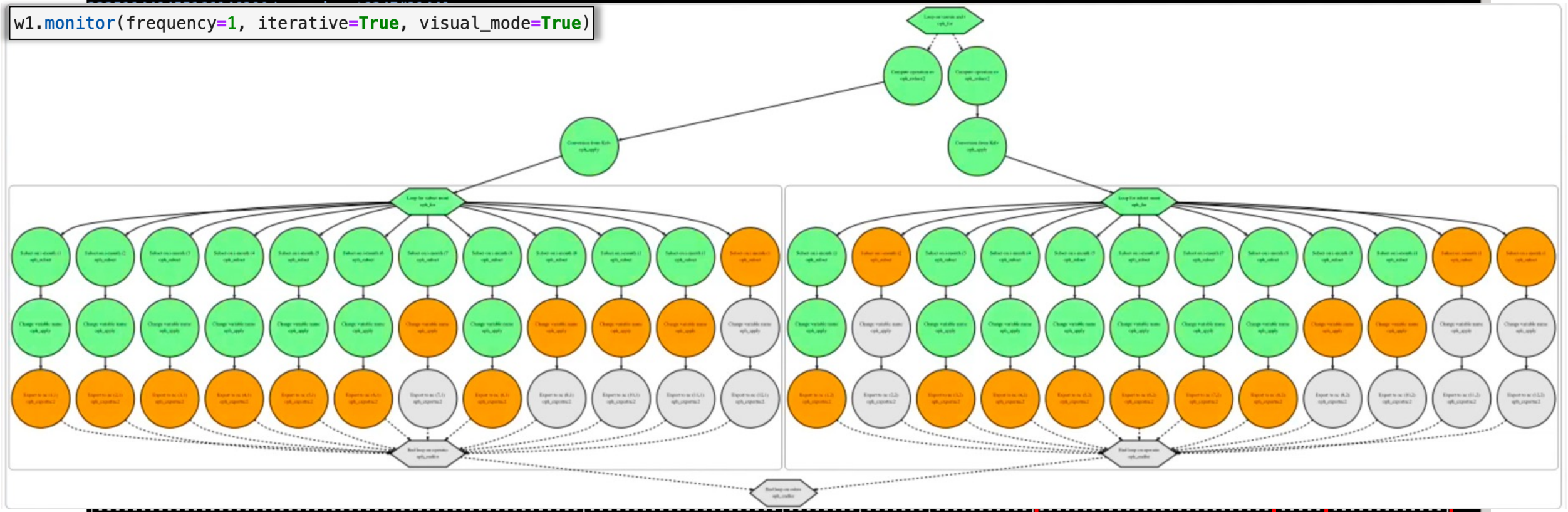
ophrpm@ophidiarpm:~/devel/oph-client/res

```
[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/37669923831130223251144945516614638O/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
```

ophrpm@ophidiarpm:~/workflow

```
w1.monitor(frequency=1, iterative=True, visual_mode=True)
```



http://193.204.199.174/ophidia/sessions/37669923831130 223251144945516614638O/experiment?247#3149	37669923831130223251 1449455166146380	247	3149	3144	Conversion from Kelvin to Celsius degrees (2)	SIMPLE	OPH_STATUS_COMPLETED

What have we learned so far?

Complex climate data analysis requires workflow support

Real case studies can be effectively modeled as (complex) workflows composed of hundreds of tasks

*The **Ophidia HPDA framework** provides **workflow management features:***

- *Target large-scale analysis and parallel execution of tasks*
- *Support for different constructs and workflow resiliency*
- *Integrated job orchestration, management and monitoring features*

The PAV Python module can help the workflow modeling, execution and monitoring

Next: Demo and hands-on of Ophidia workflows

References and further readings

- *E. Deelman, et al. (2018) 'The future of scientific workflows', The International Journal of High Performance Computing Applications, 32(1), pp. 159–175.*
- *S. Fiore, et al. (2013). Ophidia: Toward Big Data Analytics for eScience. ICCS 2013, volume 18 of Procedia Computer Science, pp. 2376-2385.*
- *S. Fiore, et al. (2014). "Ophidia: A Full Software Stack for Scientific Data Analytics", proc. of the 2014 Int. Conference on High Performance Computing & Simulation (HPCS 2014), pp. 343-350.*
- *S. Fiore, D. Elia, C. Palazzo, F. Antonio, A. D'Anca, I. Foster and G. Aloisio (2019), "Towards High Performance Data Analytics for Climate Change", ISC High Performance 2019. Lecture Notes in Computer Science, vol. 11887, pp. 240-257.*
- *D. Elia, S. Fiore and G. Aloisio, "Towards HPC and Big Data Analytics Convergence: Design and Experimental Evaluation of a HPDA Framework for eScience at Scale," in IEEE Access, vol. 9, pp. 73307-73326, 2021*
- *D. Elia, et al. (2016). "An in-memory based framework for scientific data analytics". In Proc. of the ACM Int. Conference on Computing Frontiers (CF '16), pp. 424-429.*
- *C. Palazzo, et al. (2015), "A Workflow-Enabled Big Data Analytics Software Stack for eScience", HPCS 2015, pp. 545-552*
- *A. D'Anca, et al. (2017), "On the Use of In-memory Analytics Workflows to Compute eScience Indicators from Large Climate Datasets," 2017 17th IEEE/ACM Int. Symposium on Cluster, Cloud and Grid Computing (CCGRID), pp. 1035-1043.*
- *S. Fiore, et al. (2016). "Distributed and cloud-based multi-model analytics experiments on large volumes of climate change data in the earth system grid federation eco-system". In Big Data (Big Data), 2016 IEEE Int. Conference on. IEEE. pp. 2911-2918.*
- *D. Elia, C. Palazzo, M. Musio, R. Budich, O. Heidmann, D. Spickermann, and J. Kunkel. (2021). "Deliverable D5.1 Report on the ESDM runtime extensions for parallel in-flight analytics". Zenodo. https://doi.org/10.5281/zenodo.5799437*

Questions?

ESiWACE2 *has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823988*

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

Website: https://www.esiwace.eu

More about Ophidia

 *http://ophidia.cmcc.it*

 *https://github.com/OphidiaBigData*

 *ophidia-info [at] cmcc.it*

 *https://twitter.com/OphidiaBigData*

cmcc