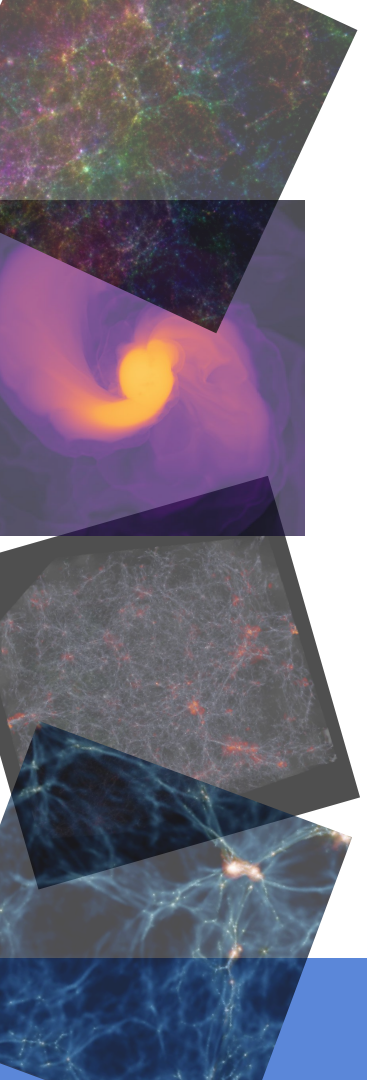


High Performance Visualization for Astronomy & Cosmology

Eva Sciacca
Fabio Vitello, Valentina Cesare, Nicola Tuccari, Giuseppe Tudisco

INAF
ISTITUTO NAZIONALE DI ASTROFISICA
OSSERVATORIO ASTROFISICO DI CATANIA

INAF USC VIII - Calcolo Critico
15-16 June 2023, Catania

USC VIII

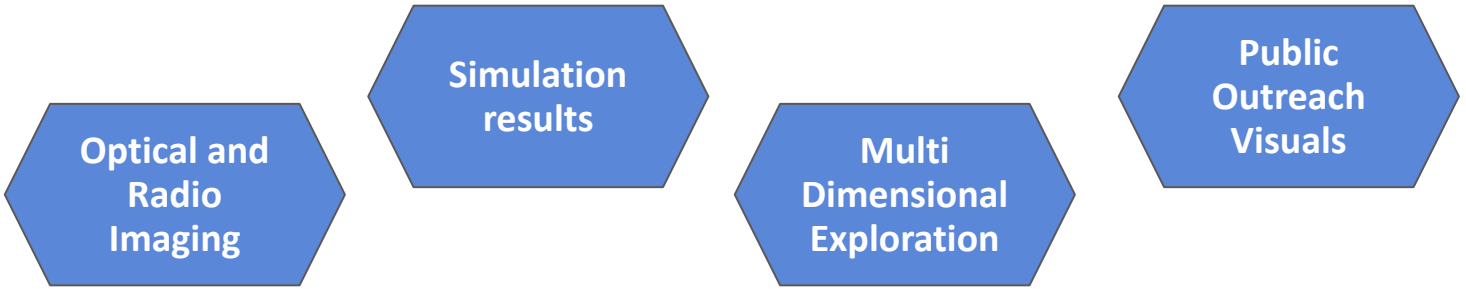
Background

Modern Astronomy and Cosmology produce massively large data volumes (order of PBs) coming from observations or simulation codes.

Such data volumes pose significant challenges for storage, access and data analysis.

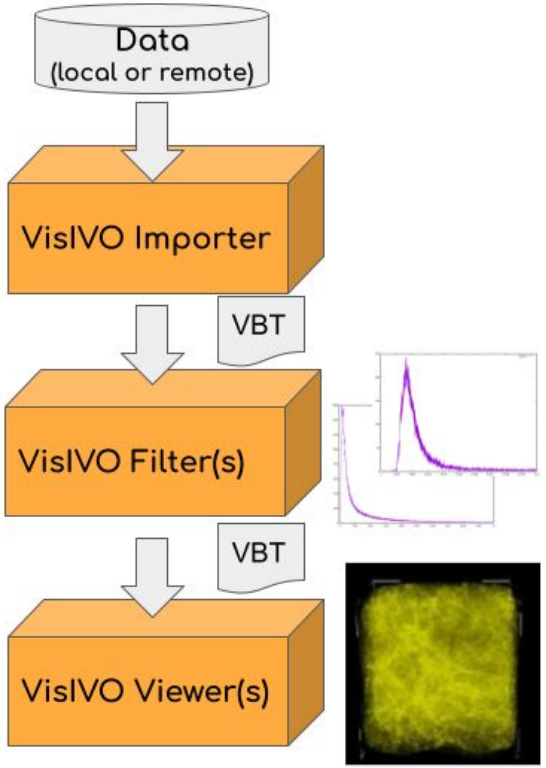
A critical aspect in understanding, interpreting, and verifying the outcome of automated analysis and data mining processes is the visualization of the scientific results.

Data visualization is a fundamental, enabling technology for knowledge discovery, and an important research field that covers a number of different topics such as:



Visualization Interface for the Virtual Observatory

- Developed and maintained by INAF OACT
- Adopt the Virtual Observatory standards
- Perform 3D and multi-dimensional data analysis and knowledge discovery of a-priori unknown relationships between multi-variate and complex astrophysical datasets.
- VisIVO is deployed in a variety of flavours as follows:
  - **VisIVO Server** - a CLI platform for high performance visualization,
  - **VisIVO Library** - for running complex workflows on DCI, clouds and HPC infrastructures.
  - **VisIVO ViaLactea Visual Analytics** exploits a combination of all new-generation surveys of the Galactic Plane to analyze star forming regions of the Milky Way.



See talk tomorrow by G. Tudisco on SKA

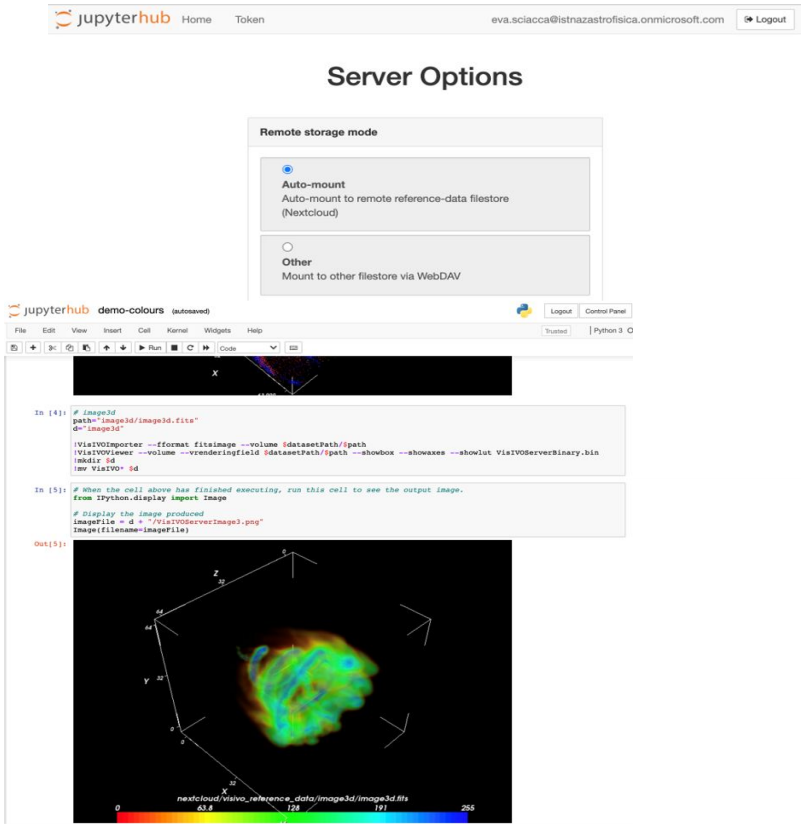
https://visivo.readthedocs.io/

Current Status

VisIVO has been already deployed using Science Gateways to access DCIs (including clusters, grids and clouds) using containerization and virtualization technologies.

It has been selected as pilot application deployed on the EOSCpilot infrastructure demonstrating that the tools can be accessed using gateways and cloud platforms, and

It has been deployed on EOSC, efficiently exploiting Cloud infrastructures and interactive notebooks applications.



The National context

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Spoke 1 “Future HPC & Big Data” is the technological pillar

- development of highly innovative hardware and software technologies for the supercomputers of the future
- creation of new laboratories as an integral part of a world-class national federated center with expertise in hardware and software co-design,
- strengthen Italian leadership in the EuroHPC Joint Undertaking and in the data infrastructure ecosystem for science and industry

Evolving VisIVO within two Flagship Projects:

- FL3 “Flagship on workflows, I/O, and HPC-cloud convergence”
- FL5 “HW-SW co-design, benchmarking, patterns, and microkernels”

The International context

Scalable Parallel Astrophysical Codes for Exascale (SPACE)

- EU Centre of Excellence focused on Astrophysical and Cosmological (A&C) applications
- Extensively re-engineer A&C codes for the efficient and effective exploitation of exascale computing capabilities
- High-performance data analysis of the data torrent produced by exascale A&C simulation applications with machine-learning and visualization tools.

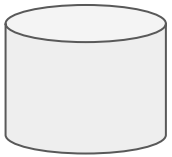
VisIVO is part of the suite of tools for "Extreme data processing and analysis" (WP3)

To be exploited and adapted for high performance visualisation (remote, in-situ, in-transit) of data generated on the (pre-)Exascale systems by the SPACE user-community and other A&C applications.

Foreseen showcases and computing needs

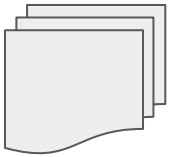
Showcase name	Codes involved
Magnetic Reconnection in plasmas	iPIC3D, PLUTO
Black Hole Accretion and jet launching - Compact-object binary mergers - Gravitational Waves	BHAC, WhiskyTHC
Large-scale galaxy cluster	OPEN Gadget, RAMSES, ChaNGa, Gasoline
Magneticum (www.magneticum.org)	OPEN Gadget
Protostellar collapse	RAMSES, PLUTO, ChaNGa, Gasoline
Isolated Galaxy	RAMSES, ChaNGa, Gasoline, OPEN Gadget
Galaxy formation and the circumgalactic medium	OPEN Gadget, RAMSES, ChaNGa, Gasoline
Protoplanetary discs	ChaNGa, Gasoline

OpenGADGET



700 GB → **15 PB**

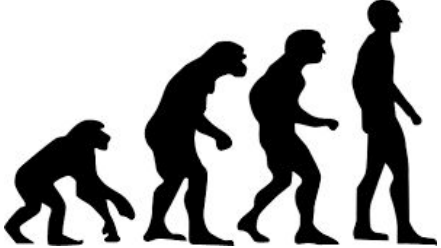
cutting edge experiments



300 → **120.000**

number of files

Evolution



The final aim of the VisIVO evolution direction and the related implementation activities will be tailored to pursue the following objectives:

1. *Optimize* and *parallelize* VisIVO modules to efficiently handle and process A&C data on *Exascale* computing resources,
2. Enhance the *portability* of the VisIVO modular applications and their resource requirements,
3. Foster *reproducibility* and *maintainability*,
4. Take advantage of a more flexible resource exploitation over *heterogeneous* HPC facilities (including also mixed HPC-Cloud resources),
5. Minimise data-movement overheads and *improve I/O* performances.

The VisIVO tools and related software will be provided by INAF Astrophysical Observatory of Catania (OACT). For this work plan we expect collaboration with UNITO, UNIPI and with CINECA.

HPC enabling

Importing modules parallelized for multi node/multi thread platforms using MPI.

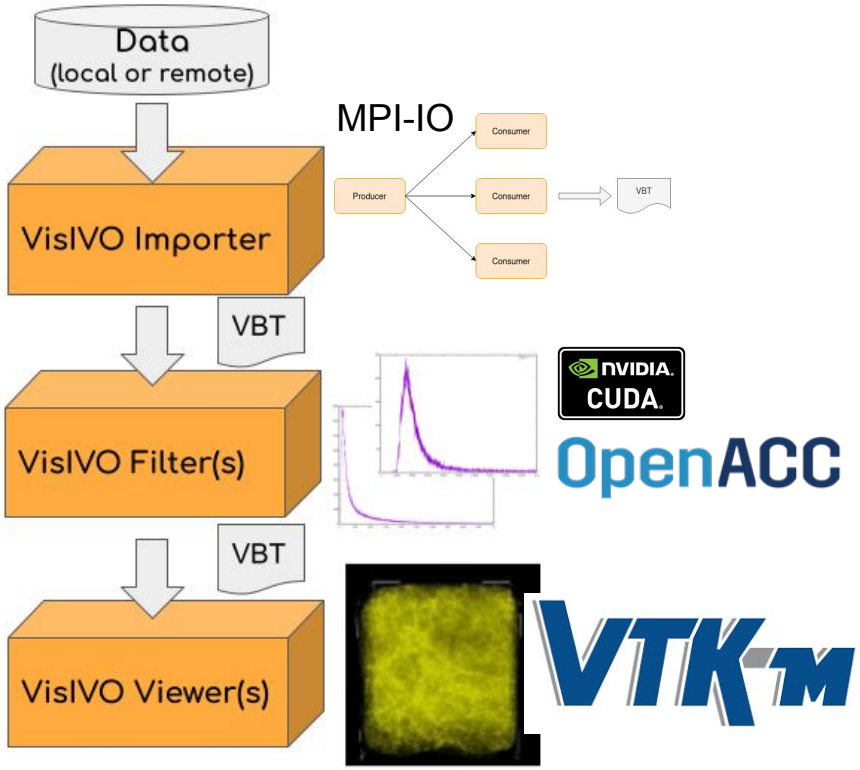
- MPI-IO to parallelize multiple reads and writes on common files and a Consumer-Producer approach, useful for load balancing.

Filtering modules extended to exploit multi GPU platforms investigating CUDA and OpenACC.

- Some of them may instead employ MPI(e.g. the filters to merge VBTs or add new tabular columns).

Viewer modules, based on VTK, already optimized for emerging processor architectures, will be tailored to support the fine-grained concurrency for data analysis and visualization algorithms required to drive extreme scale computing.

- Providing abstract models for data and execution that can be applied to a variety of algorithms across many different processor architectures.



HPC Request @ Pleiadi

Requested n. of nodes/cores:

8 nodes

Requested core/h:

100.000

Requested RAM/node:

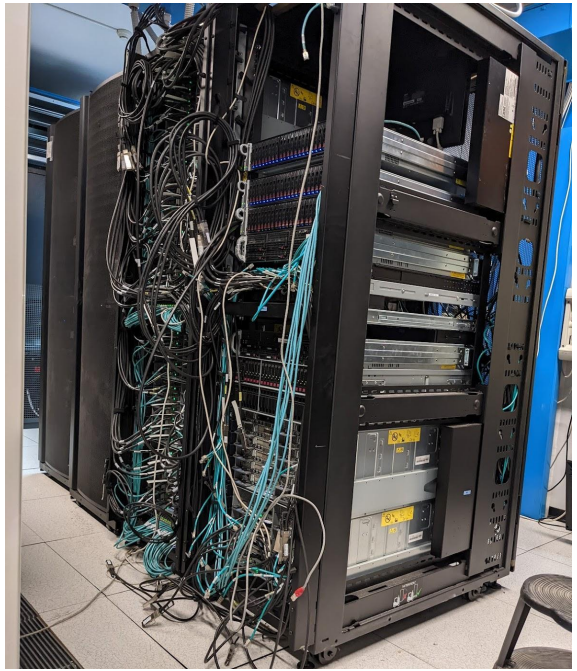
128 GB

Requested scratch storage:

1 TB

Start Date and End Date:

01/08/2023 - 31/01/2024



StreamFlow

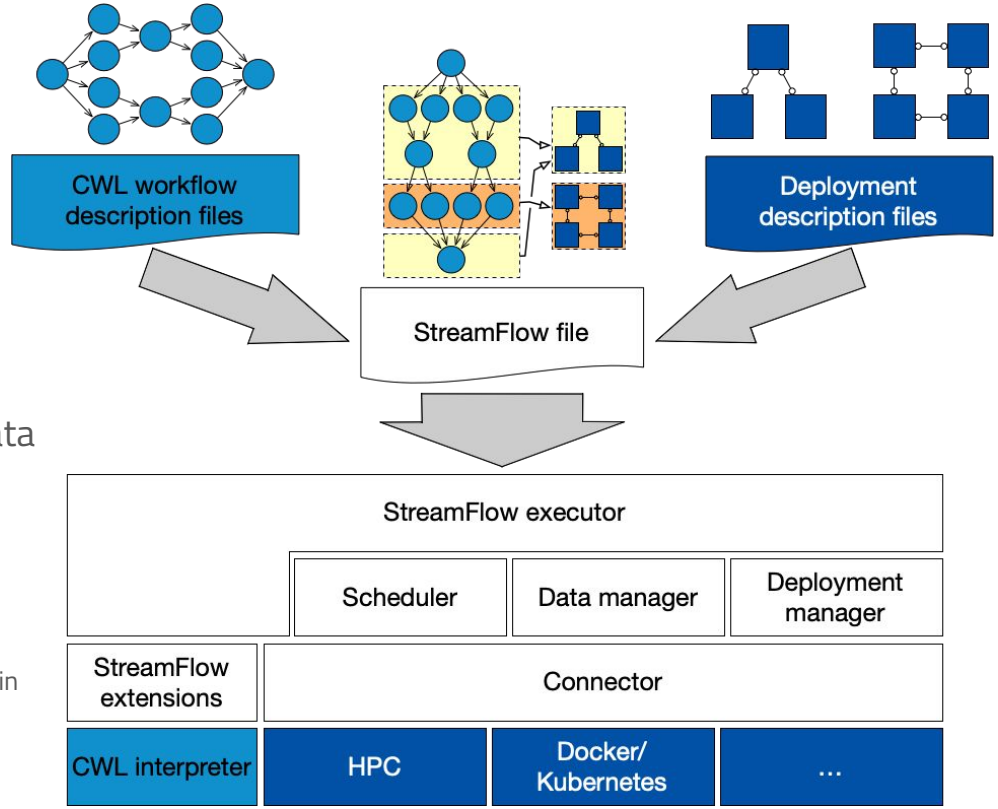
UNIVERSITÀ DI TORINO

StreamFlow
Towards Cloud-HPC Continuum

A container-native Workflow Management System based on the Common Workflow Language (CWL) standard and designed around two main principles:

- Allowing the execution of tasks in multi-container environments, in order to support concurrent execution of multiple communicating tasks in a multi-agent ecosystem;
- Relaxing the requirement of a single shared data space, in order to allow for hybrid workflow executions on top of hybrid cloud/HPC infrastructures.

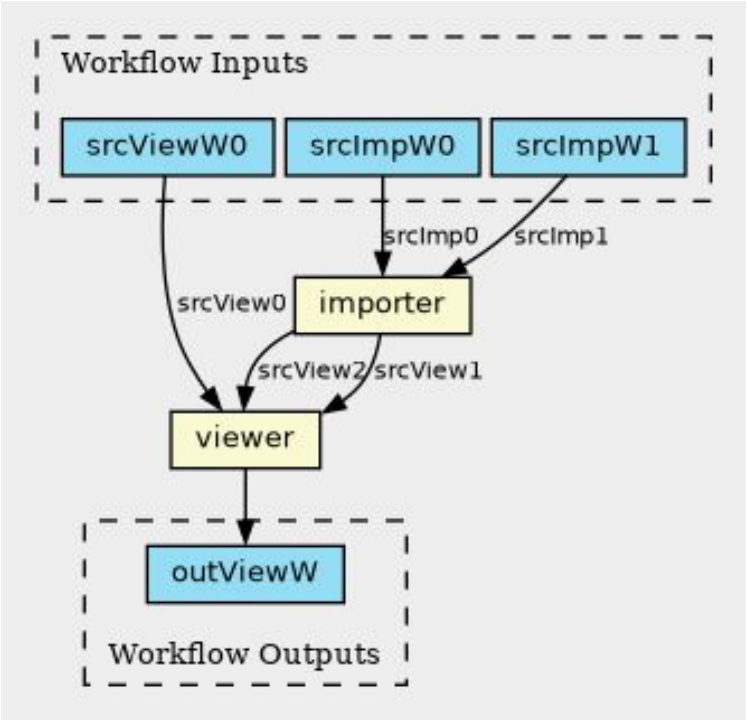
I. Colonnelli, B. Cantalupo, I. Merelli and M. Aldinucci, "StreamFlow: cross-breeding cloud with HPC," in IEEE Transactions on Emerging Topics in Computing, vol. 9, iss. 4, p. 1723-1737, 2021. doi: 10.1109/TETC.2020.3019202.



The Common Workflow Language (CWL)

```
1   cwlVersion: v1.0
2   class: Workflow
3
4   inputs:
5     srcImpW0: File
6     srcImpW1: File
7     srcViewW0: File
8
9   outputs:
10    outViewW:
11      type: File[]
12      outputSource: viewer/outView
13
14  steps:
15    importer:
16      run: docker_VisIVOImporter.cwl
17      in:
18        srcImp0: srcImpW0
19        srcImp1: srcImpW1
20      out: [outImp0, outImp1]
21    viewer:
22      run: docker_VisIVOViewer.cwl
23      in:
24        srcView0: srcViewW0
25        srcView1: importer/outImp0
26        srcView2: importer/outImp1
27      out: [outView]
```

COMMON
WORKFLOW
LANGUAGE



Interactive Computing Service

CINECA

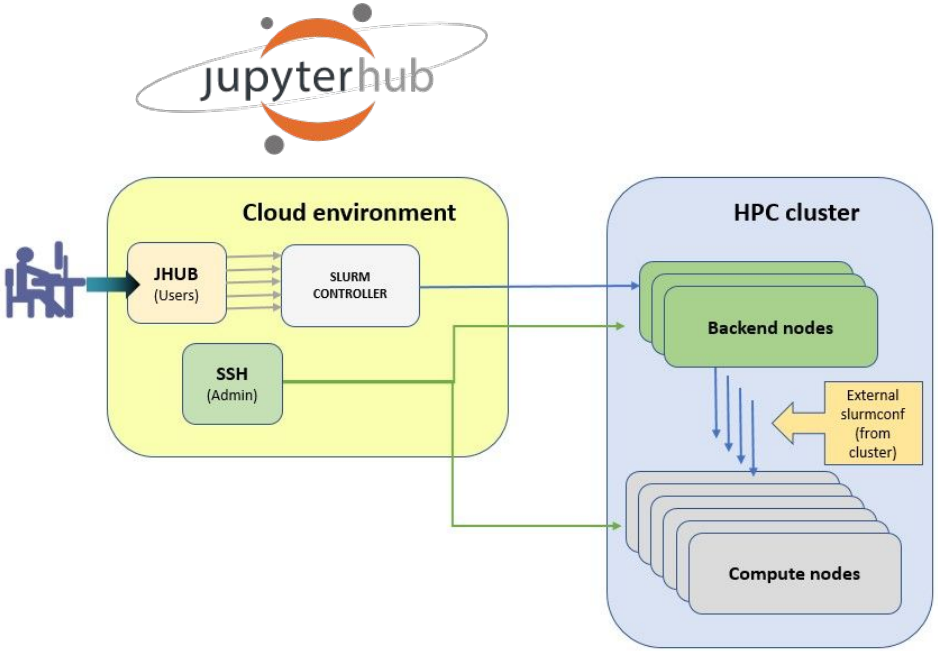
Interactive Computing Service is an alternative approach to the traditional access to HPC resources.

Main idea: interaction on the fly with the workflow during its execution

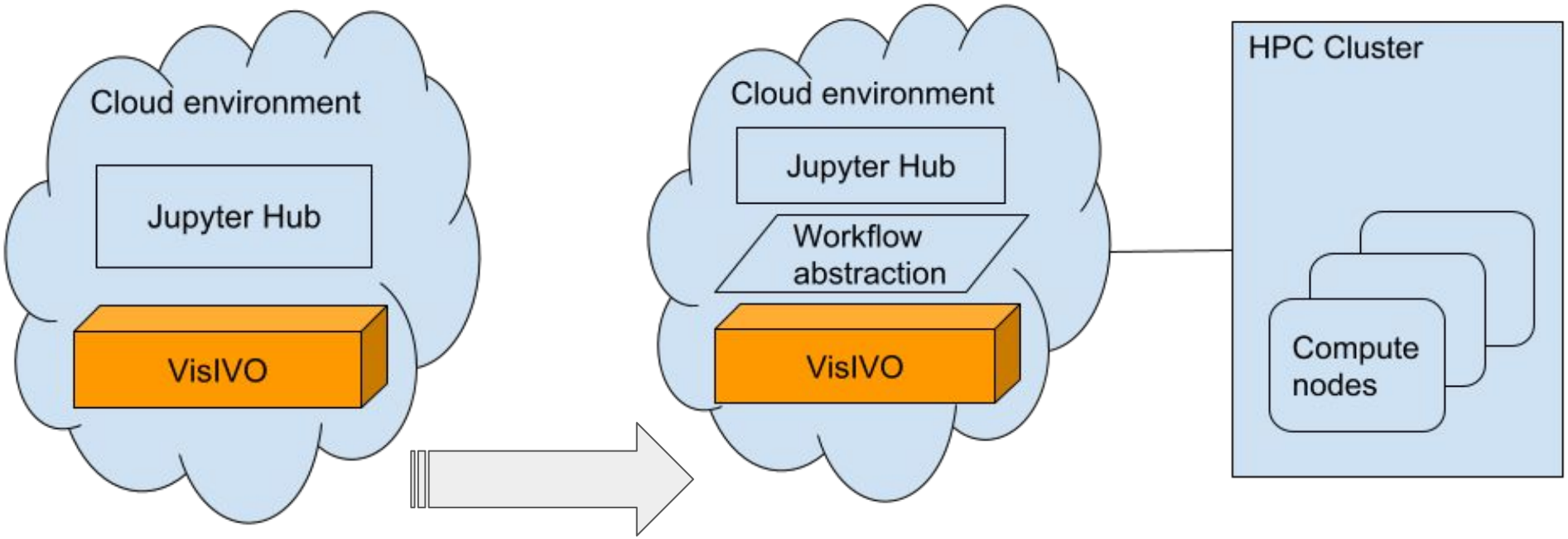
- Resources are allocated to the user despite the code is running or not (near-immediate access in some implementations)
- User can employ those resources on the fly (and not queueing them)

Advantages:

- User can easily check intermediate results and change the workflow accordingly



HPC-Cloud integration



Evolution of the current Cloud deployment prototype (left diagram) toward allowing workflow abstractions and integrating HPC clusters to better exploit the parallelization of VisIVO modules.

CAPIO

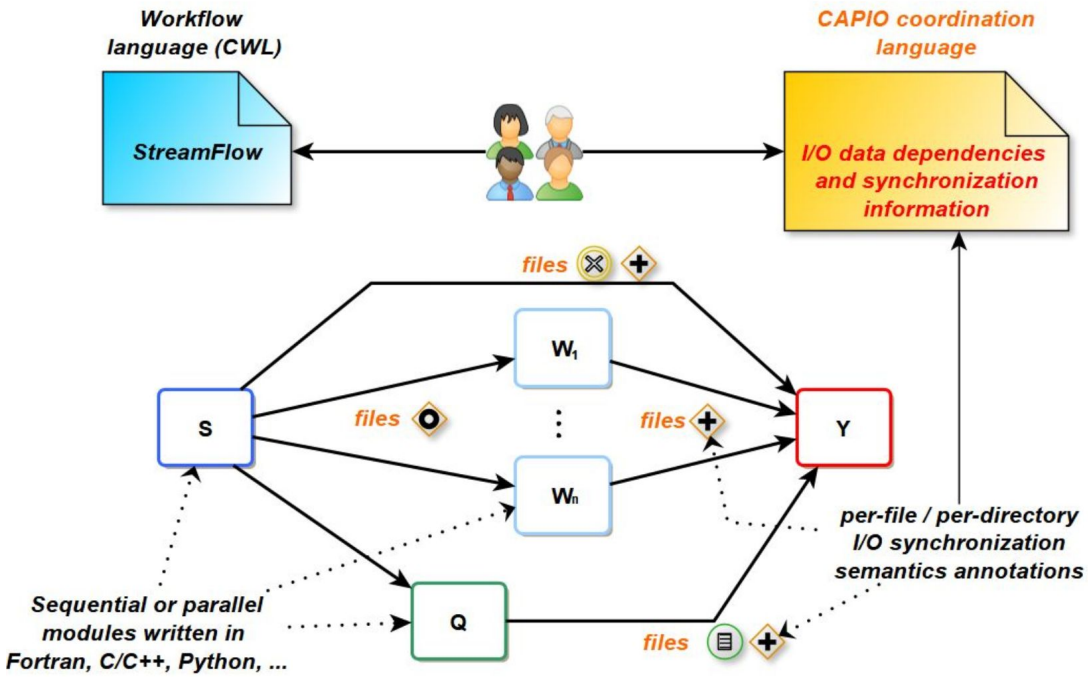
UNIVERSITÀ DI PISA

Cross-Application Programmable I/O

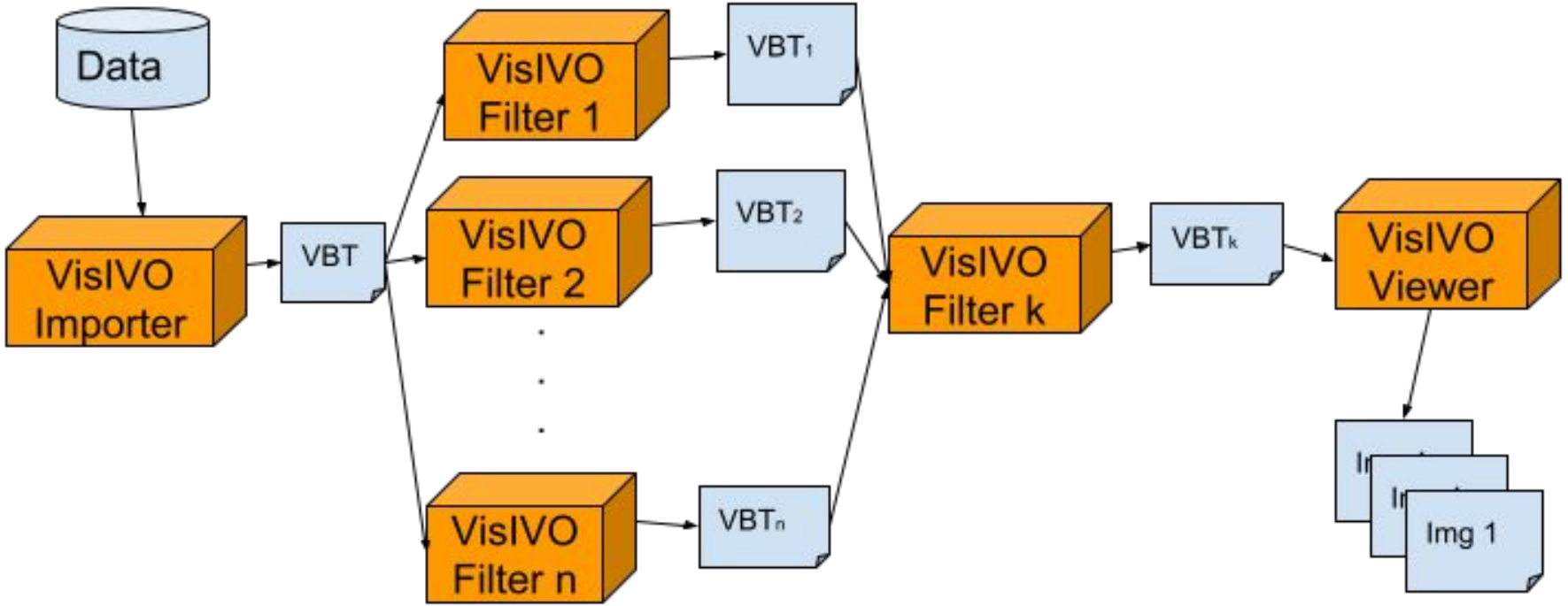
User-space middleware enabling I/O coordination in Scientific Workflows through a declarative coordination language (JSON-based)

- No need to modify the Workflow software components (steps)
- POSIX I/O SCs (open,read,write,lseek,stat,.....) are transparently intercepted using the dynamic linker features

CAPIO couples with existing WMSs to inject data streaming capabilities for file data movements



I/O Performances Improvement



Computing Request @ HPC4AI

VM instances: 2

VCPU: 8 each

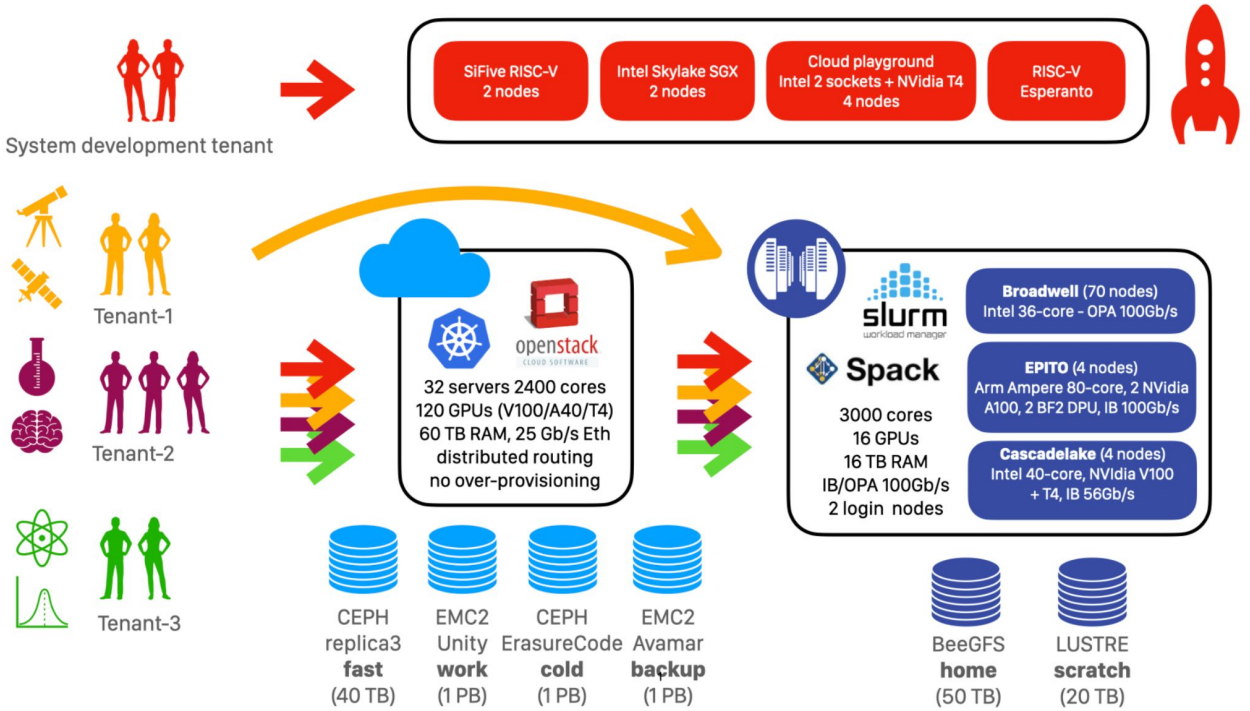
RAM: 128

HD space: 500GB

GPU: 4

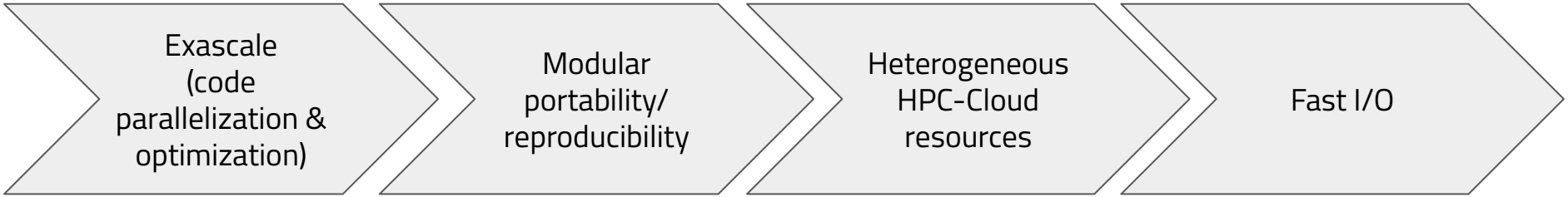
Access by Openstack

Access to HPC resources

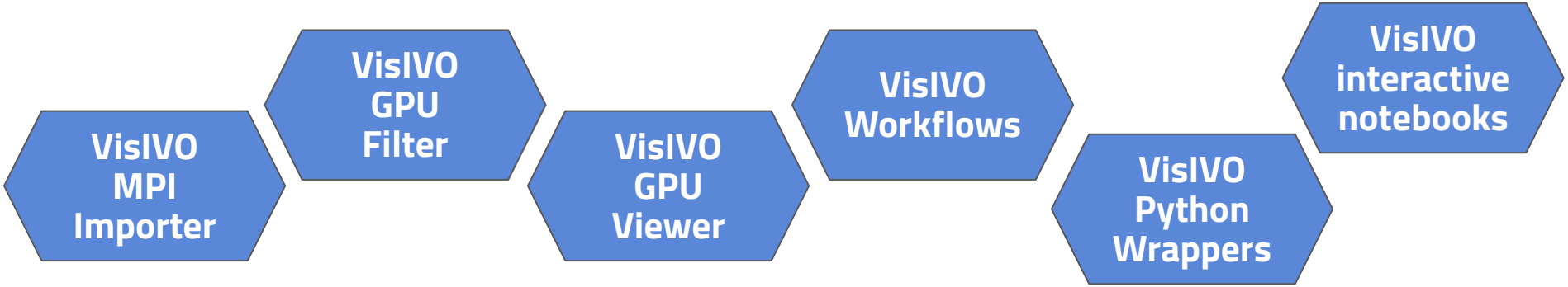


Conclusions

In this project we expect to further evolve VisIVO toward the following directions:



Including the following key results:



Acknowledgements

This work is

funded by the European High Performance Computing Joint Undertaking (JU) and Belgium, Czech Republic, France, Germany, Greece, Italy, Norway, and Spain under grant agreement No 101093441

supported by the spoke "FutureHPC & BigData" of the ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing – and hosting entity, funded by European Union – NextGenerationEU"