A Data Lake prototype for Open Science

Xavier Espinal (CERN) - ESCAPE WP2 leader

PaN ESCAPE Data Management Workshop, 12 January 2020

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

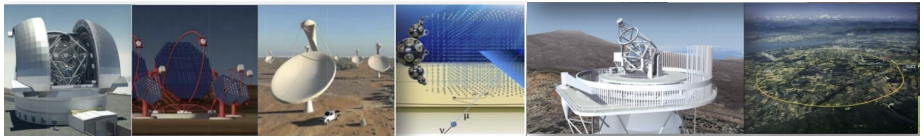
Science Projects

HiLumi HL-LHC PROJECT
SKA SQUARE KILOMETRE ARRAY
VIRGO
FAIR
KM3NeT
JIVE Joint Institute for VLBI ERIC
cta cherenkov telescope array
LSST
EST european solar telescope
ESO
CERN

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

EUROPEAN OPEN SCIENCE CLOUD

- Prototype an infrastructure adapted to **Exabyte-scale** needs of large science projects
- **Common** data infrastructure for Astro-particle, Radio-astronomy, Gravitational Waves, Cosmology and Particle Physics
- Ensure the **sciences** drive the development of the EOSC
- Address **FAIR** data management principles



Data centres

CERN
INFN
GSI
DESY
Nikhef
SURF SARA
rijksuniversiteit groningen
CCIN2P3
cnrs
PIC port d'informació científica
LAPP Laboratoire d'Annecy de Physique des Particules
INAF ISTITUTO NAZIONALE DI ASTROFISICA NATIONAL INSTITUTE FOR ASTROPHYSICS

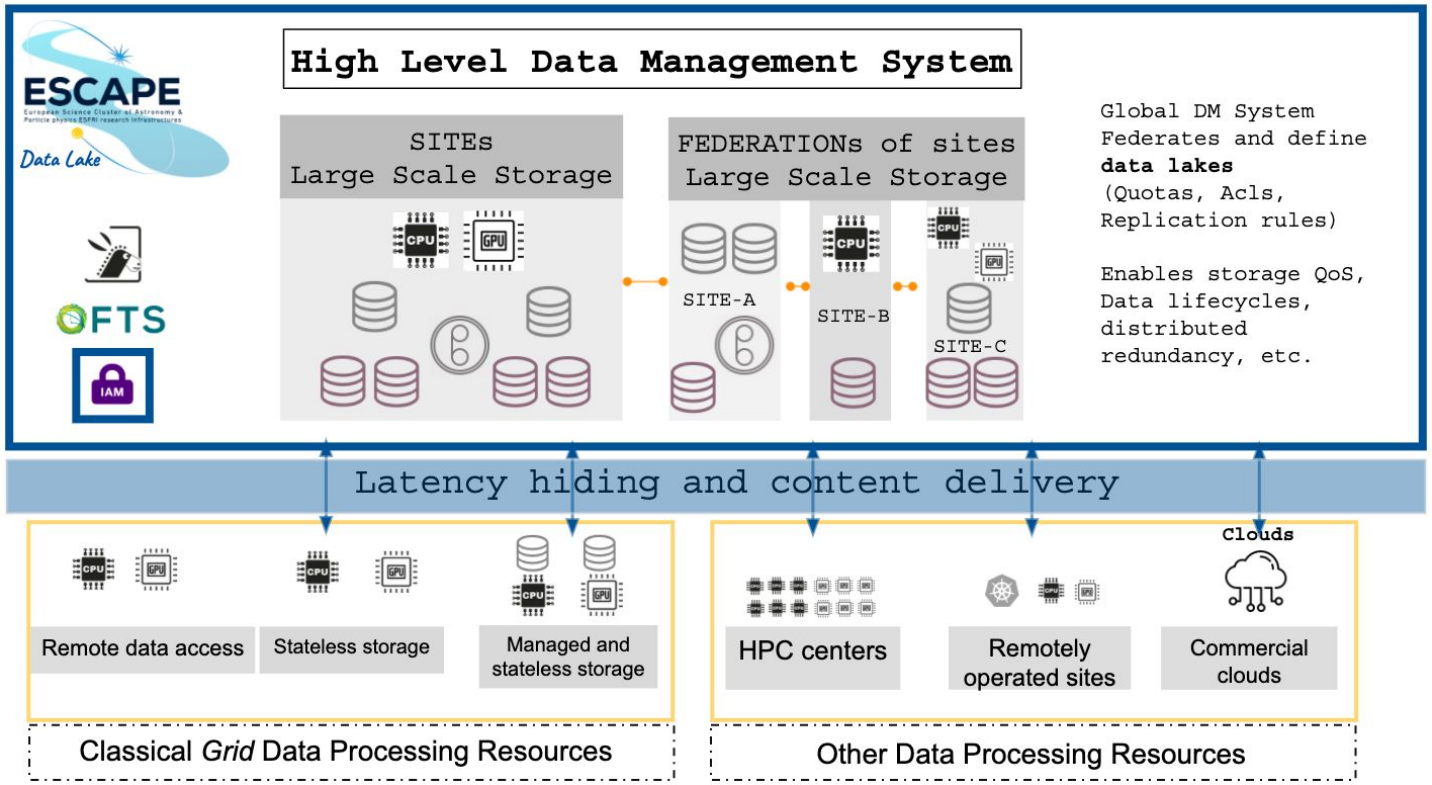
Funded by the European Union's Horizon 2020 - Grant N° 824064

The ESCAPE Data Infrastructure for Open Science

- **Define, integrate and commission** an ecosystem of tools and services to build a data lake

- Contributes to deliver **Open Access and FAIR data services**: trustable data repositories; enable data management policies; transparent data access layer

- Science **projects to drive** the services requirements most suitable to their needs

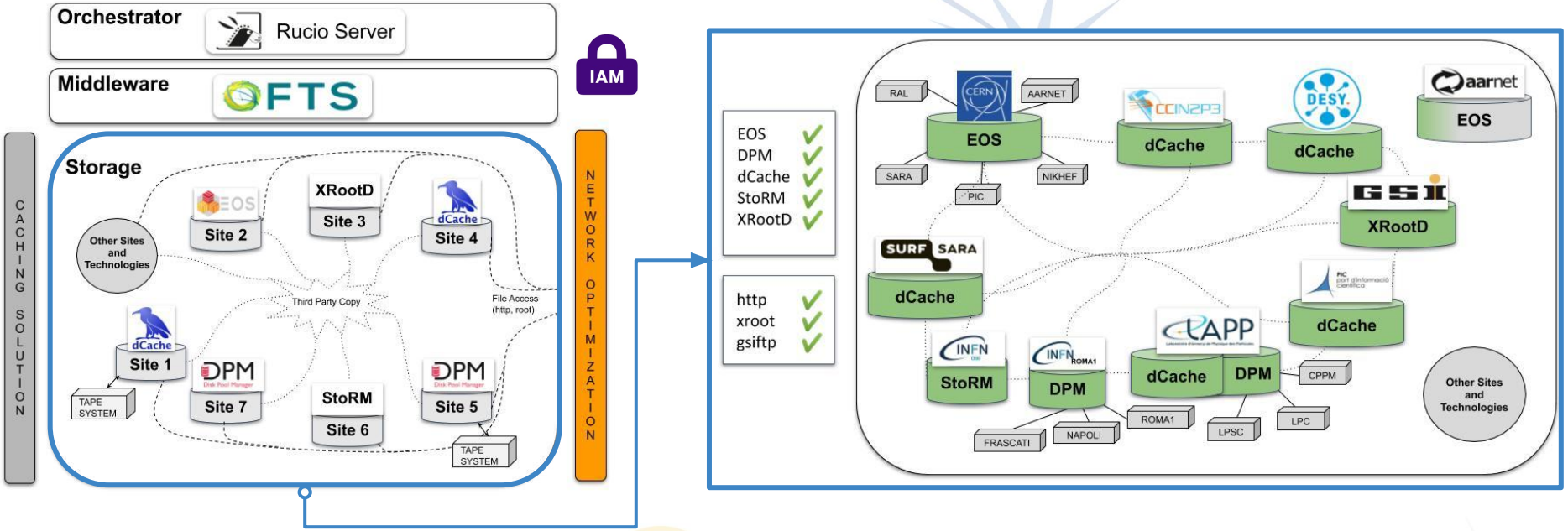


Funded by the European Union's
Horizon 2020 - Grant N° 824064

- The DIOS Work Package aims at **delivering a prototype of the Data Lake concept**
- The backbone of the prototype consists of services operated by the partner institutes and connected through reliable networks:
  - Data management and orchestration: **Rucio**
  - File transfer and data movement: **FTS**
  - Content delivery and latency hiding: **XCache**
  - Data Lake Information System: **CRIC**
  - AAI: Indigo **IAM** (tokens and legacy x509 support)
- The Data Lake harness heterogeneous facilities, with different storage systems:
  - EOS, dCache, DPM, STORM, xrootd, and any type of http-enabled storage
- Widening the access to several access protocols: **http**, **xrootd** and *gridftp*.
  - Allowing to serve the data to heterogeneous facilities, from conventional **Grid** sites to **HPC** centres and **Cloud** providers

Funded by the European Union's Horizon 2020 - Grant N° 824064

The ESCAPE Data Lake (1/4)



- **Hiding complexity** and providing transparent access to data
- Heterogeneous federated storage and operations model
- Some centers joining even if not funded by ESCAPE

Further info: https://wiki.escape2020.de/index.php/WP2_-_DIOS#Datalake_Status

Funded by the European Union's Horizon 2020 - Grant N° 824064

The ESCAPE Data Lake (2/4)

Storage endpoints monitoring

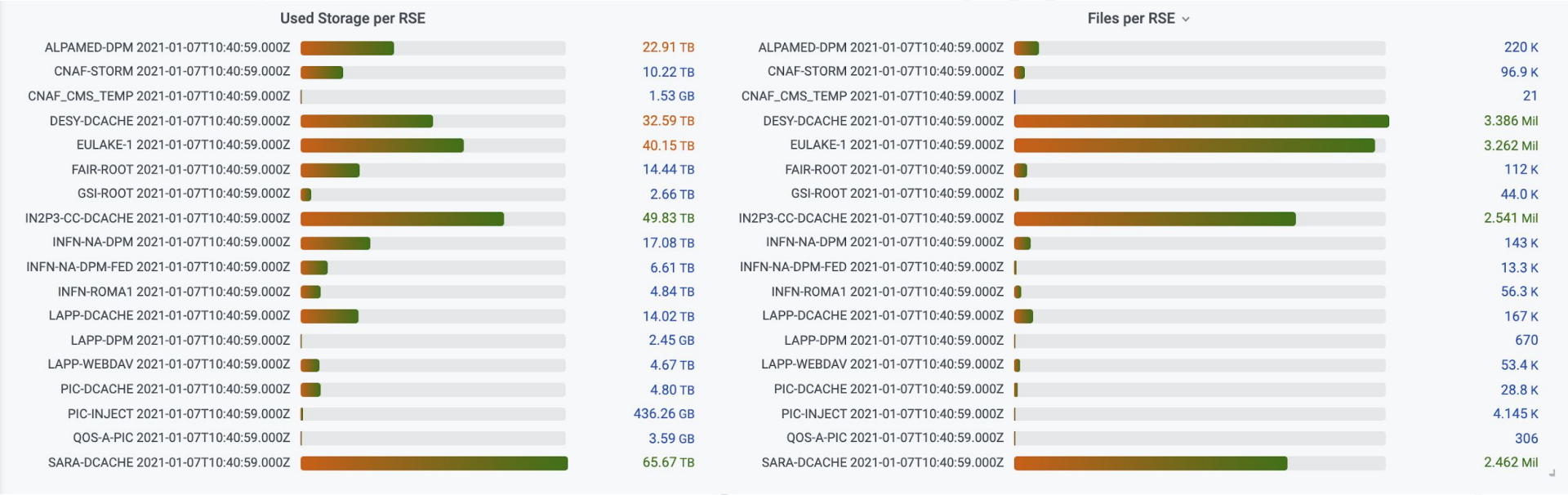
Successful Transfers Percentage (davs)

Source \ Destination	ccdcalitest10.in2p3.fr	dcache-door-doma01.desy.de	dclxwp2dlds1.gsi.de	door05.pic.es	eoseulake.cern.ch	lapp-dcache01.in2p3.fr	lapp-esc02.in2p3.fr	lapp-testse01.in2p3.fr	t2-dpm-dome.na.infn.it	webdav.grid.	xfer.cr.cnaf.infn.it
ccdcalitest10.in2p3.fr	-	100.00%	99.76%	100.00%	100.00%	99.95%	100.00%	99.93%	1.49%	98.69%	99.62%
dcache-door-doma01.desy.de	100.00%	-	99.69%	99.99%	100.00%	99.48%	99.40%	99.74%	0.74%	98.49%	99.33%
dclxwp2dlds1.gsi.de	99.24%	99.55%	-	99.19%	73.65%	98.94%	98.69%	93.85%	0.67%	98.33%	82.23%
door05.pic.es	99.97%	100.00%	99.69%	100.00%	100.00%	99.90%	100.00%	99.65%	1.03%	98.89%	99.69%
eoseulake.cern.ch	100.00%	100.00%	98.97%	100.00%	-	99.87%	100.00%	99.36%	1.77%	98.43%	99.42%
lapp-dcache01.in2p3.fr	99.69%	99.63%	98.89%	99.53%	99.94%	-	99.80%	96.48%	0.65%	98.63%	99.58%
lapp-esc02.in2p3.fr	100.00%	99.82%	99.81%	100.00%	100.00%	100.00%	-	97.30%	-	99.82%	98.98%
lapp-testse01.in2p3.fr	96.16%	96.17%	94.39%	96.35%	94.88%	95.16%	94.50%	-	1.16%	95.72%	95.61%
t2-dpm-dome.na.infn.it	97.44%	100.00%	91.01%	98.18%	92.86%	96.67%	100.00%	97.92%	50.00%	97.92%	100.00%
webdav.grid.sara.nl	100.00%	100.00%	99.85%	100.00%	100.00%	99.79%	99.68%	99.82%	1.15%	-	99.50%
xfer.cr.cnaf.infn.it	100.00%	100.00%	78.27%	100.00%	95.31%	99.96%	100.00%	99.95%	1.37%	99.04%	-

Funded by the European Union's Horizon 2020 - Grant N° 824064

The ESCAPE Data Lake (3/4)

Volume and files monitoring per endpoint

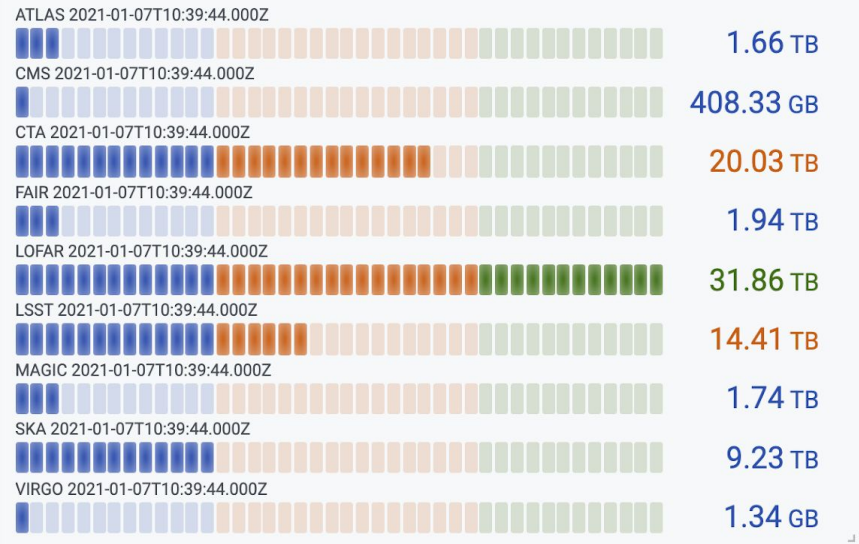


Funded by the European Union's Horizon 2020 - Grant N° 824064

The ESCAPE Data Lake (4/4)

Volume and files monitoring per experiment

experiments



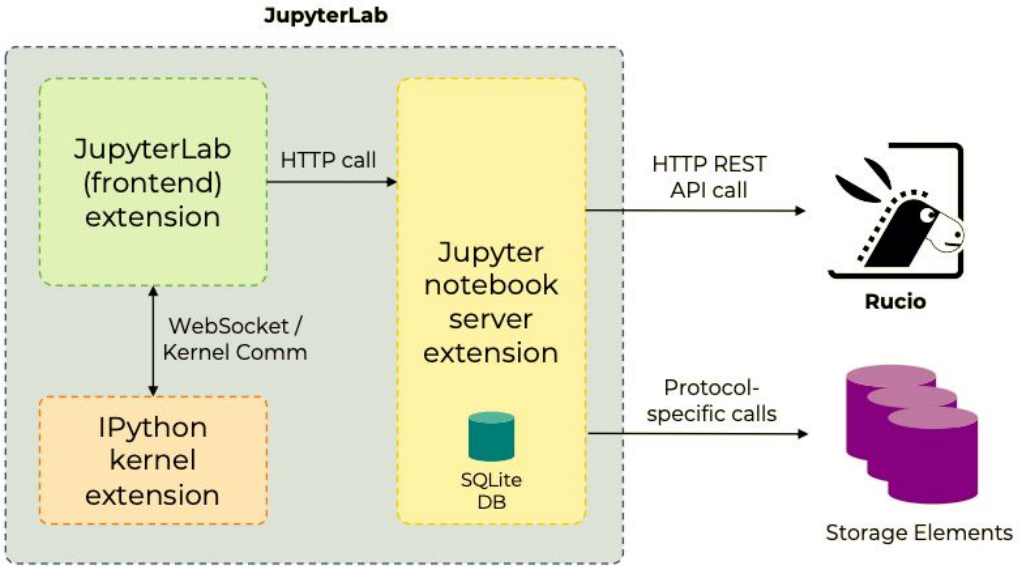
DIDs per Experiment (replica=1)

Experiment	Number of DIDs	Number of files	Number of datasets	Number of containers
LOFAR	25.1 K	25.1 K	4	0
FAIR	1.818 K	1.816 K	2	0
CMS	401	398	3	0
ATLAS	7.441 K	7.051 K	390	0
MAGIC	42.3 K	8.086 K	34.2 K	33
CTA	647 K	644 K	3.074 K	0
LSST	787 K	787 K	10	0
SKA	3.457 Mil	3.417 Mil	39.8 K	40
VIRGO	15.6 K	15.6 K	1	0

Funded by the European Union's Horizon 2020 - Grant N° 824064

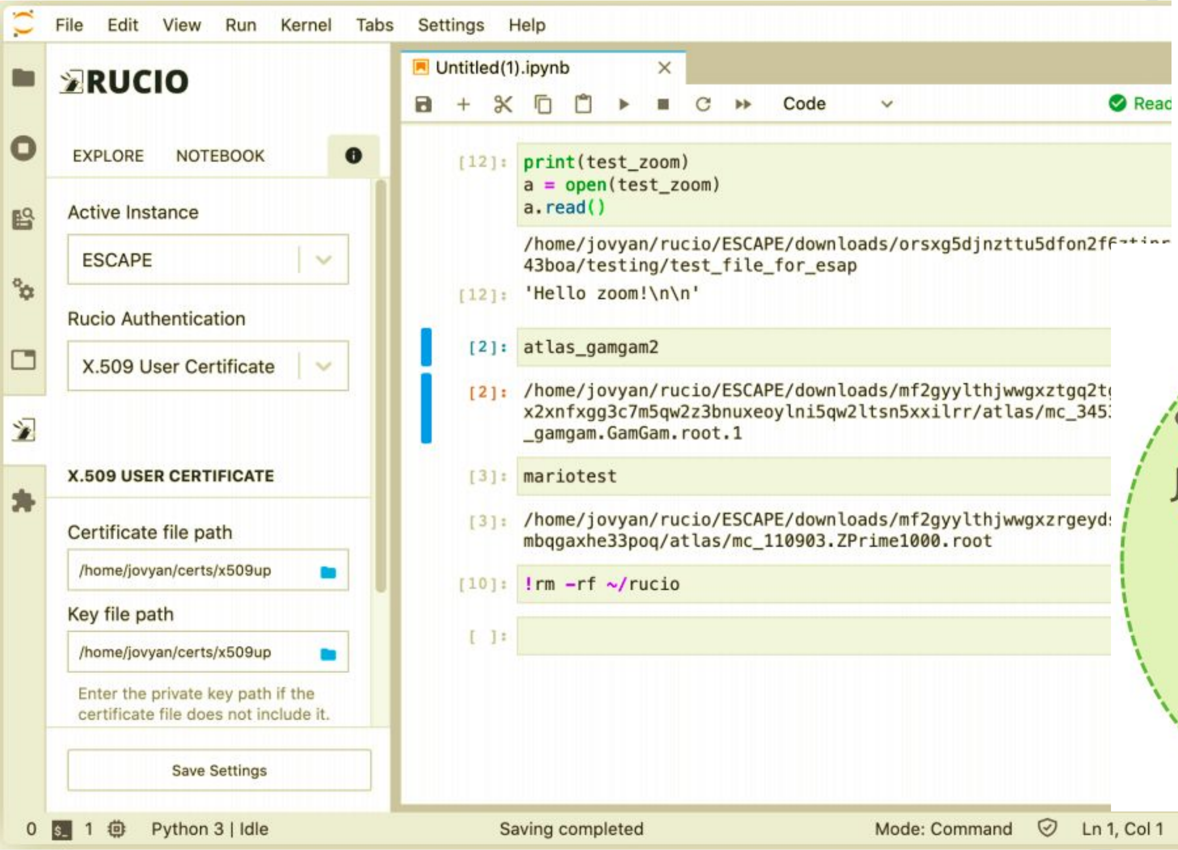
Data Lake integration with notebooks (1/2)

- Developed a JupyterLab extension to **enable data access** from a notebook platform
- Rucio JupyterLab Extension: **notebook datalake integration**
- **Browse/download/replicate** datalake data from the notebook sidebar
- Remote storage fuse mount, XCache integration and multi VO support
- Technology being implemented: fruitful joint work with WP5



Funded by the European Union's Horizon 2020 - Grant N° 824064

Data Lake integration with notebooks (2/2)

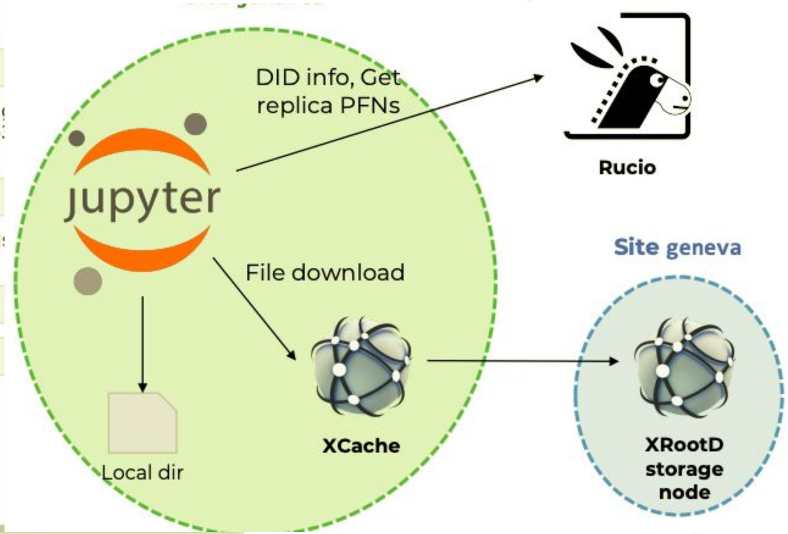


1. Assign a site name to existing RSE
   - rucio-admin rse set-attribute --rse **XRD1** --key site --value **geneva**
2. Register XCache host + port to Rucio
   - rucio-admin config set --section root-proxy-internal --option **jakarta** --value **xcache:1094**

Site name

```
[root@rucio rucio]# SITE_NAME=jakarta rucio list-file-replicas test:file1
+---------+--------+-----------+----------+--------------------------------------------------------------+
| SCOPE   | NAME   | FILESIZE  | ADLER32  | RSE: REPLICA                                                 |
|---------+--------+-----------+----------+--------------------------------------------------------------|
| test    | file1  | 10.486 MB | c9dbba2a | XRD1: root://xcache:1094//root://xrd1:1094//rucio/test/80/25/file1 |
+---------+--------+-----------+----------+--------------------------------------------------------------+
```

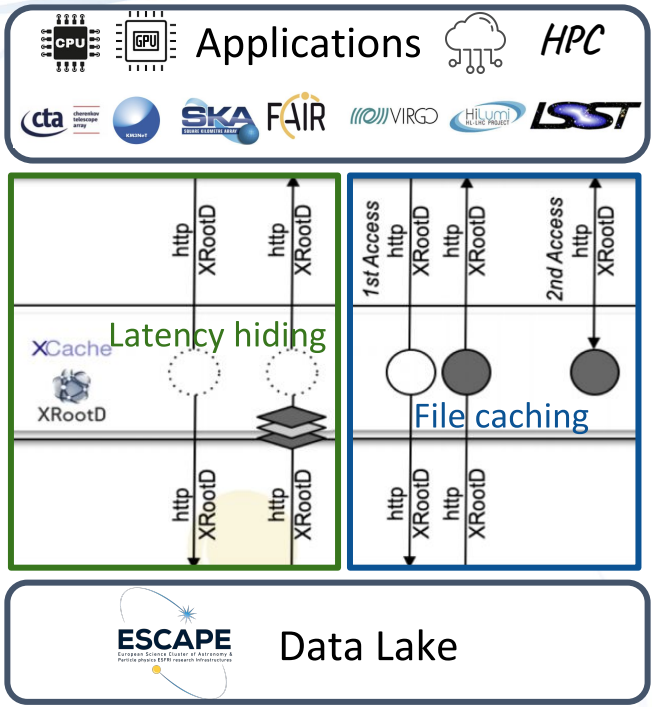
Cache prefix



Funded by the European Union's Horizon 2020 - Grant N° 824064

Content delivery and latency hiding

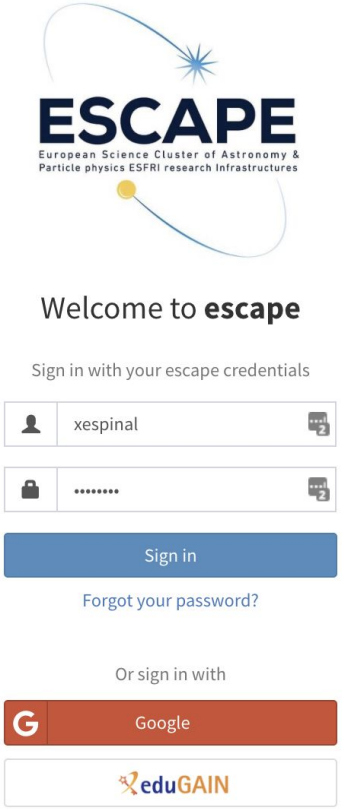
- Streaming caches demonstrate potential on **latency hiding and file re-usability** in Particle Physics workflows
- Understanding whether caching can also help on non-event based formats, e.g. images, data-cubes,...
- Caches can facilitate ingress/egress of data with heterogeneous computing resources: **Commercial Clouds and HPCs**
- Effort made towards a vanilla installation (experiment-unbiased) caching service
  - Based on XCache technology
  - Easy deployable by the partners



Funded by the European Union's Horizon 2020 - Grant N° 824064

Authentication and authorization (AAI)

- **ESCAPE IAM instance deployed and integrated with EduGAIN**, supporting GSI/VOMS and token-based approaches
- Strong user enrollment and X.509/VOMS authN/Z **in place and working reliably**
- Namespace authorization proposal and deployment strategy defined
  - Incremental steps towards finer-grained authz



Funded by the European Union's Horizon 2020 - Grant N° 824064

Summary and next steps

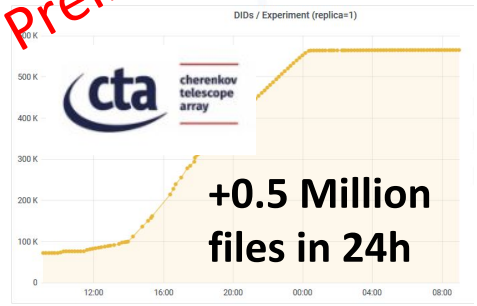
- Pilot datalake with 10 storage endpoints **functional**
- The high level Data Lake orchestration layer is **consolidated**:
  - RUCIO for data management, FTS for file movement and IAM for AAI
  - CRIC information system integrated with RUCIO (endpoints, protocols, and QoS)
- **Strong involvement** from ESFRIs: LOFAR, CTA, FAIR,SKA, LSST, ATLA, CMS and EGO/VIRGO.
  - Currently hosting modest amounts of data but remarkable exercise for this pilot phase.
  - Ready to on-board new use case from new communities
- ESCAPE Data Lake **successfully integrated** with notebook platform/JupyterLab (+info)
  - Browse/download/replicate datalake data from the notebook sidebar
  - Remote storage fuse mount and caching layer integration ability
- 2021: Pilot Data Lake evolution towards the final prototype

Funded by the European Union's Horizon 2020 - Grant N° 824064

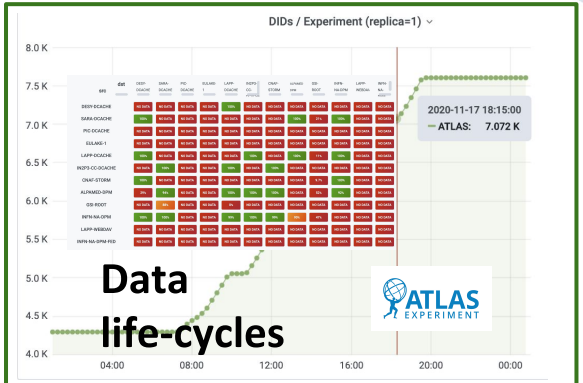
Thanks for listening!

Funded by the European Union's
Horizon 2020 - Grant N° 824064

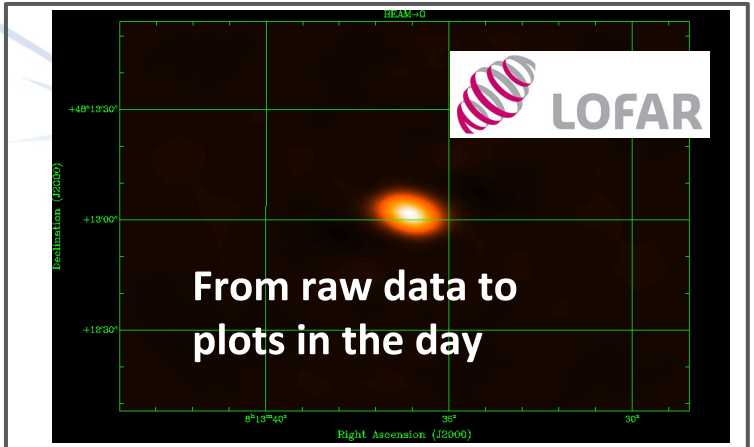
Preliminary Data Lake **24-hour Dress Rehearsal** 17 Nov 2020



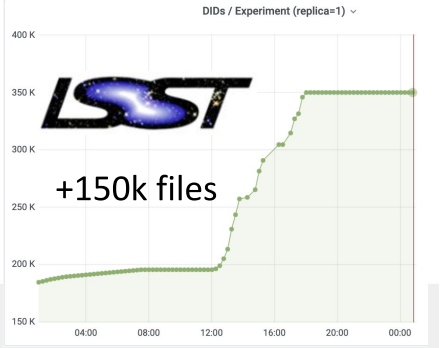
CTA: Simulate a night data captured from telescope in Canary Island for 6 h: ingest 500 Dataset of 10 files.

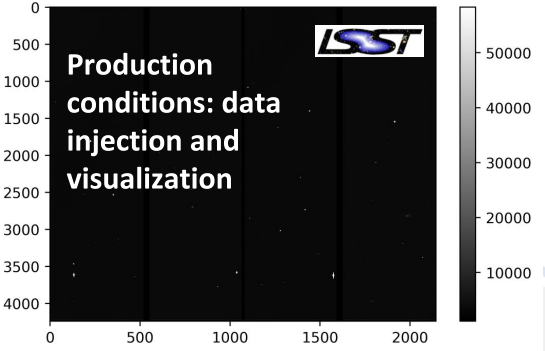


ATLAS: Storage QoS functionality tests: upload files from LAPP cluster to ALPAMED-DPM (FRANCE) and INFN-NA-DPM (ITALY), then request transfer to 1 RSE **QoS=SAFE** and 2 RSEs **QoS=CHEAP-ANALYSIS**



LOFAR: astronomical radio source 3C196 made using LOFAR data. The raw visibility data was downloaded via rucio from the EULAKE-1 and processed on Open Nebula at surfsara using the container based LOFAR software





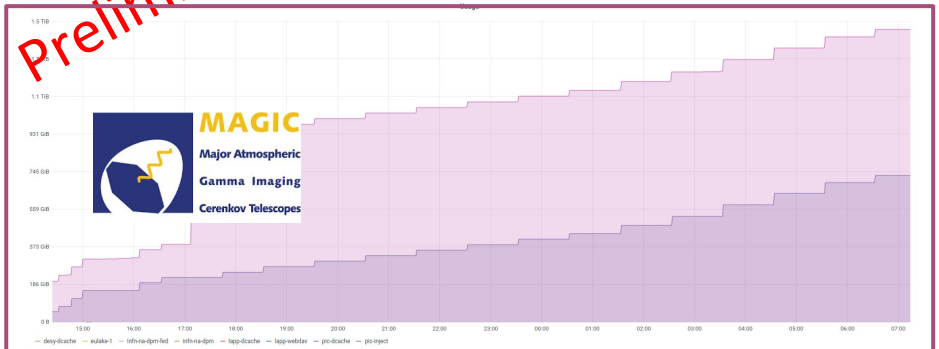
LSST: Simulate production conditions: ingest the HSC RC2 dataset from CC-IN2P3 local storage to the Data Lake, **at a realistic LSST data rate** (20TB/24h). Then **confirm integrity and accessibility of the data via a notebook**.

→ The image is a reconstruction drawn within a Jupyter Notebook accessing the data used in the Full Dress Rehearsal.

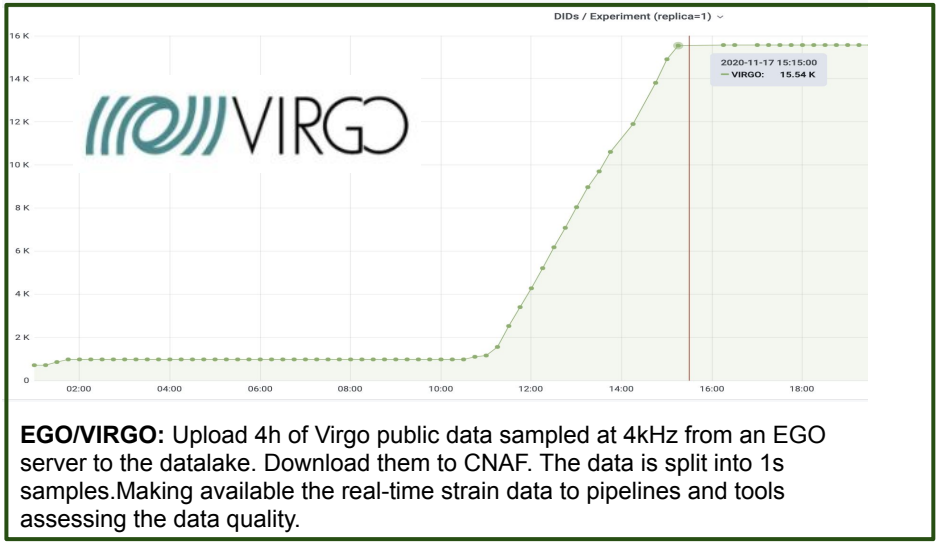
Funded by the European Union's Horizon 2020 - Grant N° 824064

Data Lake **24-hour Dress Rehearsal** 17 Nov 2020

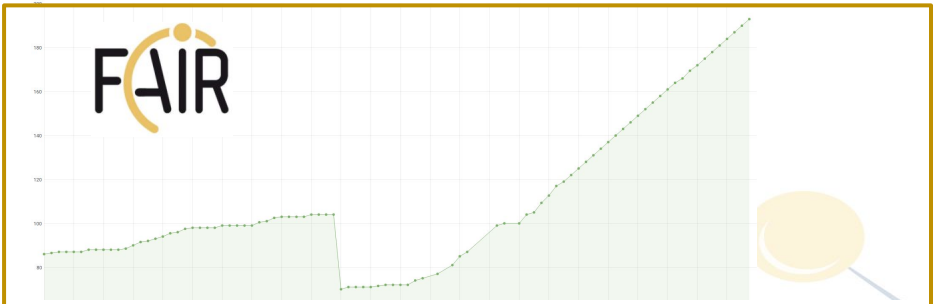
Preliminary



MAGIC: Mimics a real MAGIC observation use case. Remote storage (Data Lake aware) **next to the telescope** acts as a buffer for subsequent data injection to the ESCAPE Data Lake (and local deletion after success)



EGO/VIRGO: Upload 4h of Virgo public data sampled at 4kHz from an EGO server to the datalake. Download them to CNAF. The data is split into 1s samples.Making available the real-time strain data to pipelines and tools assessing the data quality.



FAIR: Upload one 1-GB file every 10 minutes for the whole duration of the rehearsal. Request 2 replicas in QOS=SAFE and 1 replica in QOS=CHEAP-ANALYSIS. File size and QoS tagging approximate data ingestion from CBM (i.e. the FAIR experiment expected to produce the largest volume of raw data)

SKA: Pulsar Observations injection test.For 4 hours at any point during the 24hrs, injecting new group of files in a dataset every ten minutes. Files fall into two containers, representing different SKA Projects. 24-hr test moving data on basis of QoS class.