ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

ESCAPE

Next Generation Management of Exabytes of Cross-Discipline Scientific Data

Riccardo Di Maria on behalf of ESCAPE WP2

CERN

March 24th, 2021 - International Symposium on Grids & Clouds 2021 (ISGC 2021)

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.

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

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

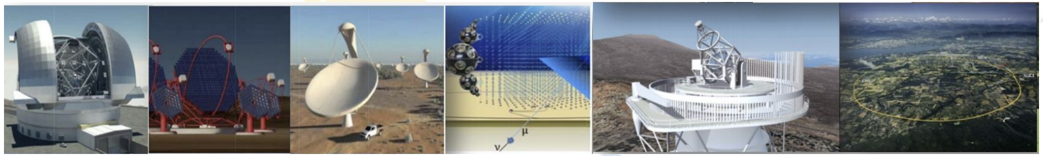
Horizon2020
European Union Funding
for Research & Innovation

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

Project Goals

- Prototype an infrastructure adapted to exabyte-scale needs of large science projects
- Ensure sciences **drive** the development of EOSC
- Address FAIR data management principles



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

The ESCAPE Project Work Packages

- **Management, Innovation, Networking and Dissemination (MIND)**:
  coordination and management
- **Data Infrastructure for Open Science (DIOS)**:
  scalable federated data infrastructure (Data Lake)
- **Open-source scientific Software and Service Repository (OSSR)**:
  repository of scientific software services of the research infrastructures
- **Virtual Observatory - connecting ESFRI projects to EOSC through VO framework (VO)**:
  astronomical high-level products archive and related services
- **ESFRI Science Analysis Platform (ESAP)**:
  flexible science platform for the analysis of open access data
- **Citizen Science - engagement and communication (CS)**:
  open gateway dedicated to the public and communication actions

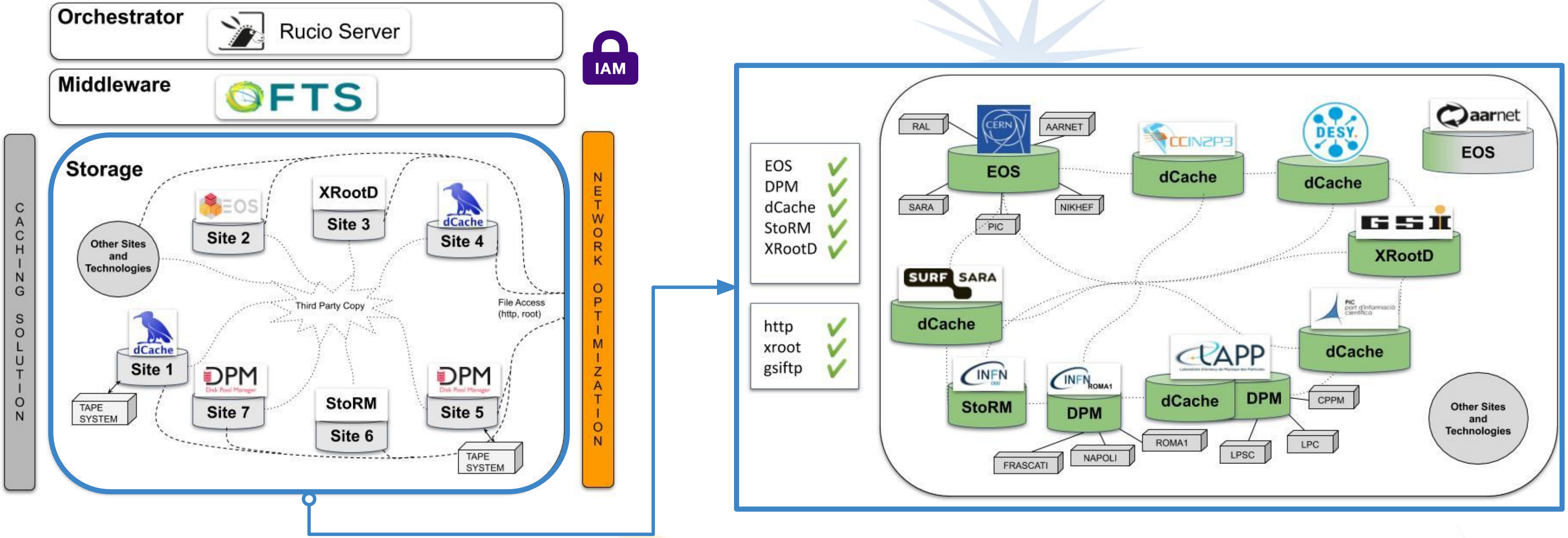
Funded by the European Union's
Horizon 2020 - Grant N° 824064

Data Infrastructure for Open Science

- Deliver a Data Infrastructure for Open Science
    - non HEP-specific implementation of the Data Lake concept
    - modular ecosystem of services and tools shaped around the scientific communities
- ESCAPE sciences at different phases of evolution
    - building up or consolidating computing model
        - needs of science projects drive services requirements
    - interest on data organisation, management and access (DOMA)
- Backbone consists of services operated by the partner institutes
    - leveraging the existing expertise in WLCG
        - e.g. RUCIO, FTS, XRootD-XCache, CRIC, AAI X.509 and Tokens (Indigo IAM), WLCG storage technologies

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

The Data Lake



EOS ✓
DPM ✓
dCache ✓
StoRM ✓
XRootD ✓

http ✓
xroot ✓
gsiftp ✓

- 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

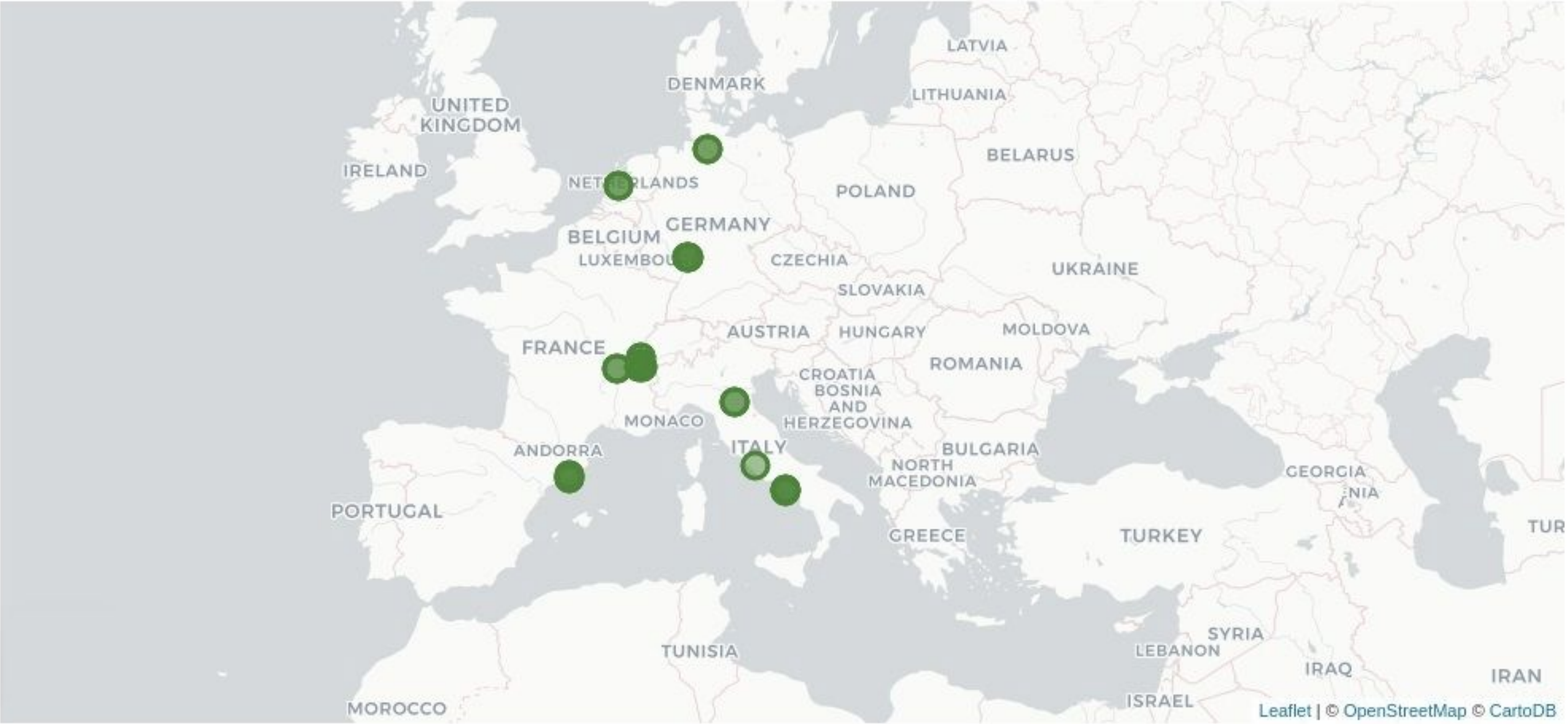
Funded by the European Union's Horizon 2020 - Grant N° 824064

The Data Lake

- Experiments contribute to the total quota at their will/disposal
- 15+ Rucio Storage Elements
  - standard resources from institutes
  - PoC of commercial Clouds integration
  - PoC of HPC integration (through XCache)
- 9 sciences/experiments highly committed
  - 60+ accounts/users as representatives
  - specific needs to be addressed
  - tailored use cases and workflows

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

The Data Lake



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

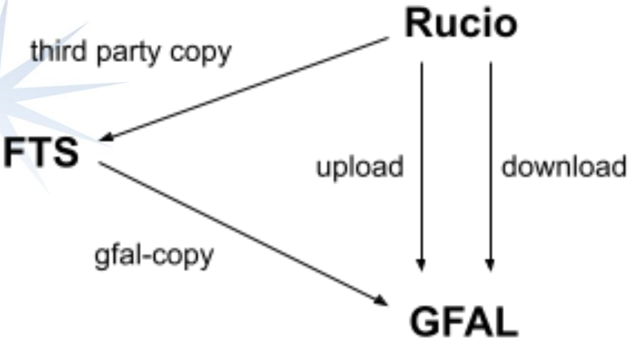
Architecture

- Data Storage Services and Usage
  - facilities diverse in size and expertise
  - diverse backends such as classic RAID systems, Erasure Coding, CEPH, and multi-replica
  - data injection buffering scenarios
  - perfect playground for the implementation of storage Quality of Service intelligence
- Data Orchestration
  - ESCAPE Rucio exploits a specific subset of asynchronous agents
    - redundancy and working-threads designed to fulfil the Data Lake needs
  - testing suite for uploading, downloading, replicating, and deleting a range of different sized files at a different rate for all ESCAPE RSEs
  - dedicated and shared monitoring infrastructure serving ESCAPE community

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

Architecture

- File Transfer Service
  - FTS as asynchronous transfer scheduling service
  - continuous testing for both FTS and GFAL
    - data monitoring and efficiency into ESCAPE dashboards
- Networking
  - exploiting perfSONAR infrastructure
    - network links between sites monitored by point-to-point transfers and latency tests
    - data monitoring and efficiency into ESCAPE dashboards
- Information System
  - ESCAPE Compute Resources Information Catalog (CRIC) instance containing services information and configuration for Rucio

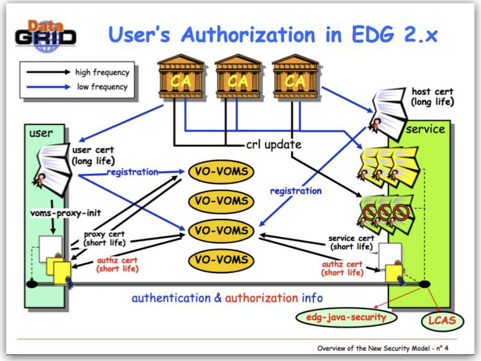


Funded by the European Union's Horizon 2020 - Grant N° 824064

Architecture

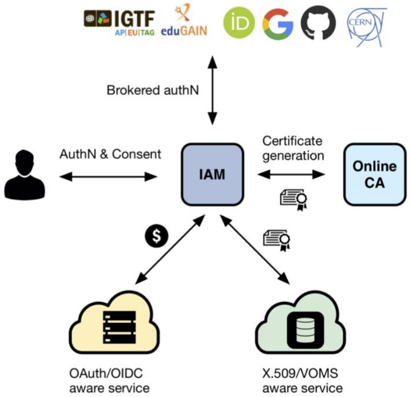
- Authentication and Authorisation

Current, X.509 based AAI



Move beyond X.509

Future, token-based AAI



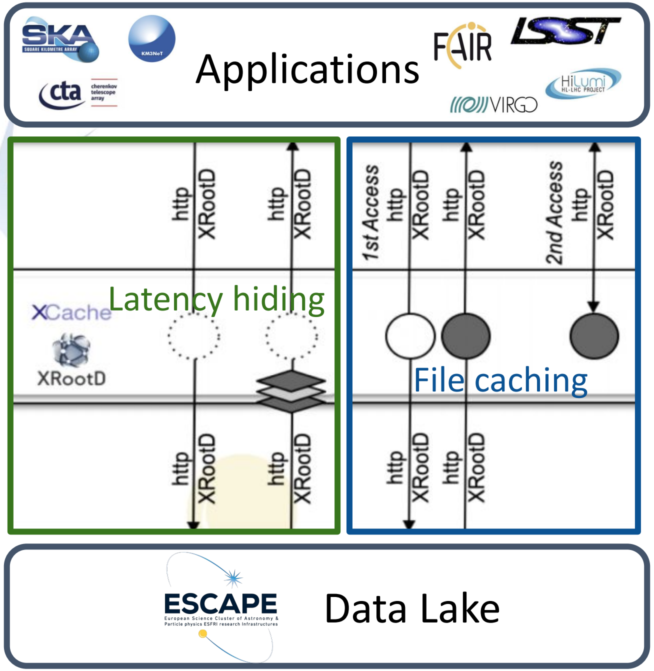
Approach: leverage and build upon the WLCG experience

- Synchronisation Services
  - set of tools designed and deployed to run every 2 hours
  - Rucio pulls necessary information from other ESCAPE services, e.g IAM and CRIC

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

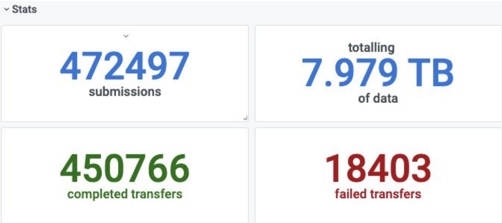
Architecture

- Data Access, Content Delivery and Caching
  - leveraging know-how in DOMA/WLCG
  - vanilla installation (experiment-unbiased) caching service → evaluating multi-VO implementation
  - main use-cases:
    - latency hiding and file re-usability
      - benchmarking multi-caching layers
      - HTTP and Tokens awareness
    - facilitate ingress/egress with Commercial Clouds and HPC
  - main goal to investigate and understand whether caching can help on non-event based files e.g. images, data-cubes, etc.



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

Monitoring - GFAL, FTS, Rucio Events and Stats



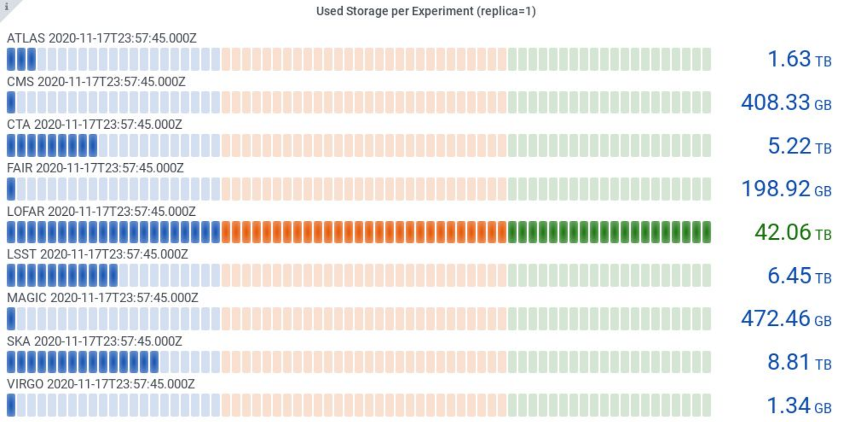
Stats

472497 submissions

totalling 7.979 TB of data

450766 completed transfers

18403 failed transfers



Used Storage per Experiment (replica=1)

ATLAS 2020-11-17T23:57:45.000Z 1.63 TB

CMS 2020-11-17T23:57:45.000Z 408.33 GB

CTA 2020-11-17T23:57:45.000Z 5.22 TB

FAIR 2020-11-17T23:57:45.000Z 198.92 GB

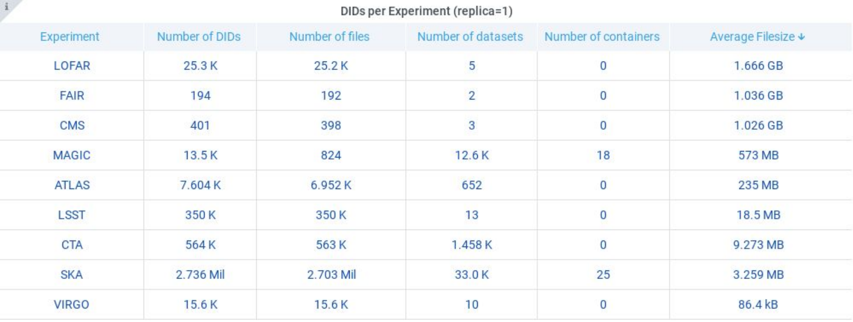
LOFAR 2020-11-17T23:57:45.000Z 42.06 TB

LSST 2020-11-17T23:57:45.000Z 6.45 TB

MAGIC 2020-11-17T23:57:45.000Z 472.46 GB

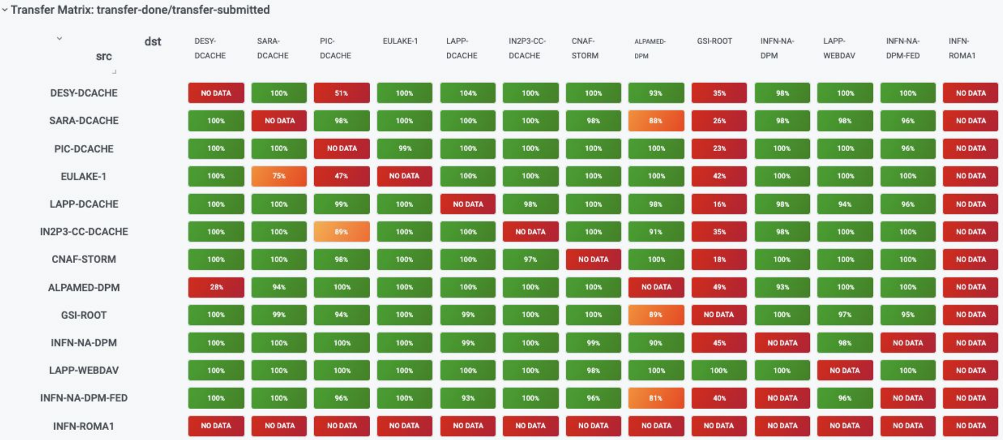
SKA 2020-11-17T23:57:45.000Z 8.81 TB

VIRGO 2020-11-17T23:57:45.000Z 1.34 GB



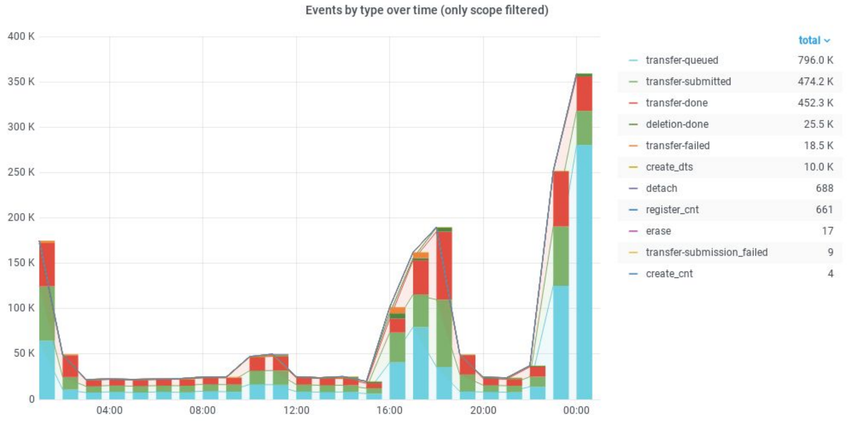
DIDs per Experiment (replica=1)

Experiment	Number of DIDs	Number of files	Number of datasets	Number of containers	Average Filesize ↓
LOFAR	25.3 K	25.2 K	5	0	1.666 GB
FAIR	194	192	2	0	1.036 GB
CMS	401	398	3	0	1.026 GB
MAGIC	13.5 K	824	12.6 K	18	573 MB
ATLAS	7.604 K	6.952 K	652	0	235 MB
LSST	350 K	350 K	13	0	18.5 MB
CTA	564 K	563 K	1.458 K	0	9.273 MB
SKA	2.736 Mil	2.703 Mil	33.0 K	25	3.259 MB
VIRGO	15.6 K	15.6 K	10	0	86.4 kB



Transfer Matrix: transfer-done/transfer-submitted

src \ dst	DESY-DCACHE	SARA-DCACHE	PIC-DCACHE	EULAKE-1	LAPP-DCACHE	IN2P3-CC-DCACHE	CNAF-STORM	ALPAMED-DPM	GSI-ROOT	INFN-NA-DPM	LAPP-WEBDAV	INFN-NA-DPM-FED	INFN-ROMA1
DESY-DCACHE	NO DATA	100%	51%	100%	104%	100%	100%	93%	35%	98%	100%	100%	NO DATA
SARA-DCACHE	100%	NO DATA	98%	100%	100%	100%	98%	88%	26%	98%	98%	96%	NO DATA
PIC-DCACHE	100%	100%	NO DATA	99%	100%	100%	100%	100%	23%	100%	100%	96%	NO DATA
EULAKE-1	100%	75%	47%	NO DATA	100%	100%	100%	100%	42%	100%	100%	100%	NO DATA
LAPP-DCACHE	100%	100%	99%	100%	NO DATA	98%	100%	98%	16%	98%	94%	96%	NO DATA
IN2P3-CC-DCACHE	100%	100%	89%	100%	100%	NO DATA	100%	91%	35%	98%	100%	100%	NO DATA
CNAF-STORM	100%	100%	98%	100%	100%	97%	NO DATA	100%	18%	100%	100%	100%	NO DATA
ALPAMED-DPM	28%	94%	100%	100%	100%	100%	100%	NO DATA	49%	93%	100%	100%	NO DATA
GSI-ROOT	100%	99%	94%	100%	99%	100%	100%	89%	NO DATA	100%	97%	95%	NO DATA
INFN-NA-DPM	100%	100%	100%	100%	99%	100%	99%	90%	45%	NO DATA	98%	NO DATA	NO DATA
LAPP-WEBDAV	100%	100%	100%	100%	100%	100%	98%	100%	100%	100%	NO DATA	100%	NO DATA
INFN-NA-DPM-FED	100%	100%	96%	100%	93%	100%	96%	81%	40%	NO DATA	96%	NO DATA	NO DATA
INFN-ROMA1	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA	NO DATA



Events by type over time (only scope filtered)

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

Deployment Model and Techniques

- Infrastructure should be resource-aware (**minimal environment**) for an improved project sustainability and for the successful exportability adoption by partners and experiments
  - deploy a manifold system utilising a lightweight but complete implementation
  - sciences at different scale and trying to address multiple future use case, including experiments with different data management requirements than ATLAS and CMS
  - Kubernetes cluster on the CERN OpenStack
    - a master of 4 vCPUs, 8 GiB RAM, and 40 GB local storage
    - 6 nodes each of 8 vCPUs, 16 GiB RAM, and 70 GB local storage
- ESCAPE tailors functional services on experiments/sciences needs
- Beyond ESCAPE term, different sciences will be able to deploy and manage the subset of services they will want to run and/or customise at their convenience

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

Deployment Model and Techniques

- Fruitful extended collaboration with teams and experts of the various components within and beyond ESCAPE
  - e.g. MonIT, CERN Cloud, OracleDB, Kubernetes, as well as Rucio, IAM, FTS, CRIC, etc.
  - Rucio/JupyterLab Integration Project within CERN-HSF Google Summer of Code (M. Aditya Hilmy) and used by LOFAR to analyse data during the Data Lake assessment
- Synergy with Rucio team allowed to tailor infrastructure to cope with experiments needs → exploring new Rucio phase space
- Sites, sciences, and experiments strongly involved and committed

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

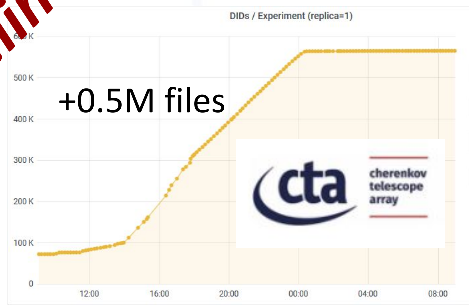
Data Lake **24-hour Full Dress Rehearsal** Preparation

- Data Lake available 24/7 to ESCAPE users even though not-production
  ⟶ FDR as specific testing-focused time-window
  - goal is to cover **experiment data workflow** needs on a single day (from 9 sciences)
  - perspective from **scientists** and from **sites**
  - assessment of robustness of the various Data Lake components, tools, services, etc.
- Fortnightly assemblies for data injector demonstrators towards FDR
  - even out differences in knowledge among partners
  - acquiring know-how on management and utilisation of technology stack - sites deeply involved
  - hot-topics: data injection and access, QoS, data life-cycles
  - deployment&operation team gathered to identify and solve problematic situations
- Improving and deploying features and functionalities for both Kubernetes and Rucio

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

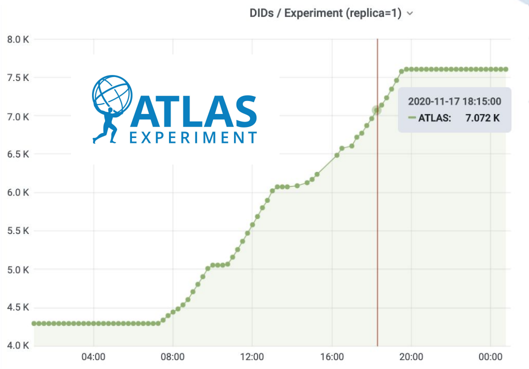
Preliminary

DL **24-hour Full Dress Rehearsal** Takeaway → Workshop



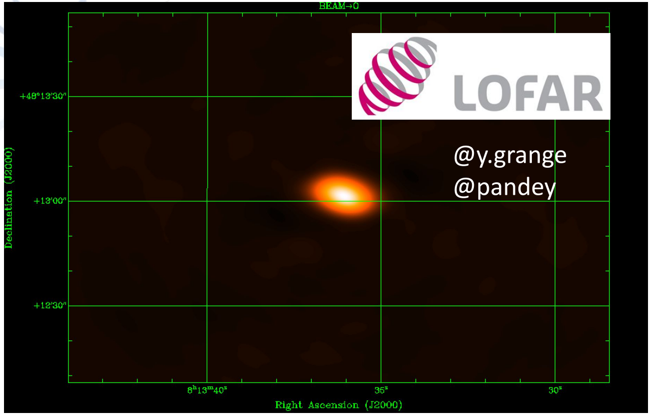
+0.5M files

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



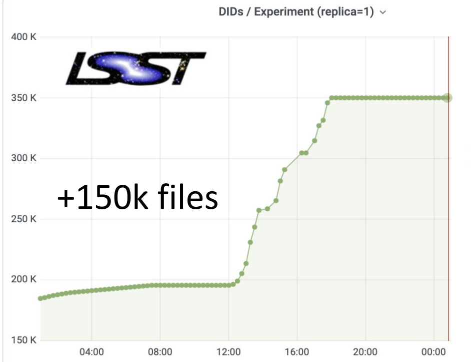
ATLAS - Storage QoS functionality tests: upload files from LAPP cluster to ALPAMED-DPM (FR) and INFN-NA-DPM (IT).

Request transfer to 1 RSE **QoS=SAFE** and 2 RSEs **QoS=CHEAP-ANALYSIS**.

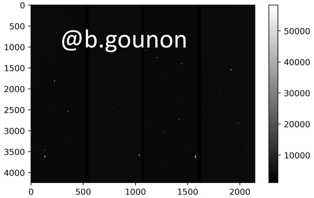


@y.grange
@pandey

LOFAR - Astronomical radio source 3C196 image using LOFAR data. The raw visibility data were downloaded via rucio from the EULAKE-1 and processed on OpenNebula at SURFsara using the container based LOFAR software.



+150k files



@b.gounon

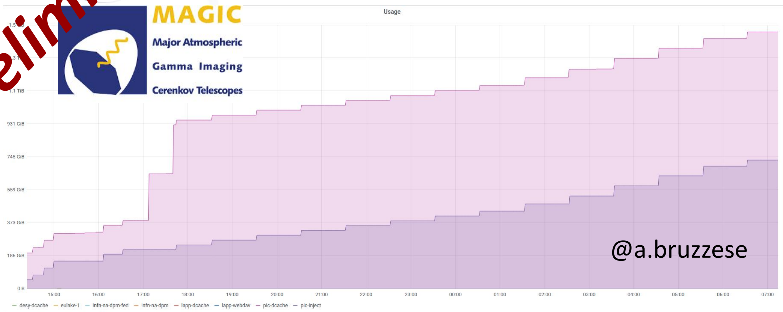
LSST - Simulate production conditions: ingest the HSC RC2 dataset from CC-IN2P3 local storage to the Data Lake, **at a realistic LSST data rate** (20TB/24h); **confirm integrity and accessibility of the data via a notebook**.
The image is a reconstruction drawn within a Jupyter Notebook accessing the data used in the Full Dress Rehearsal.

@x.espinal

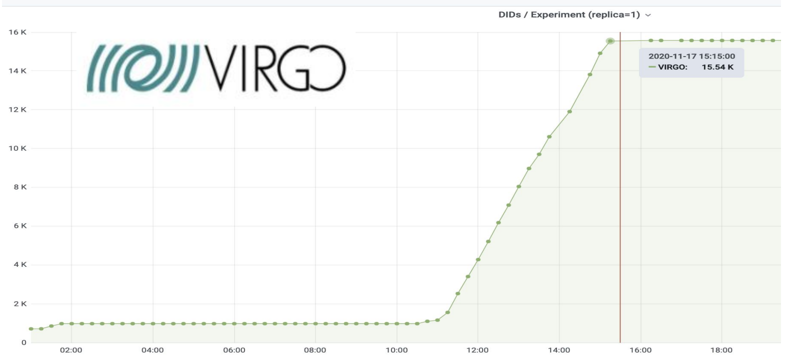
Funded by the European Union's
Horizon 2020 - Grant N° 824064

Preliminary

DL **24-hour Full Dress Rehearsal** Takeaway → Workshop



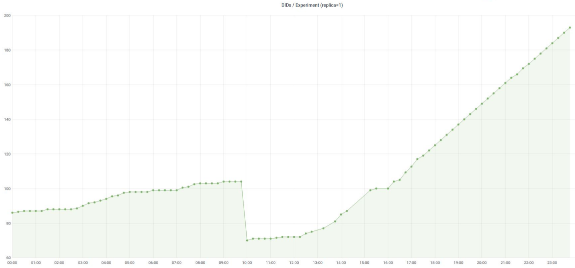
MAGIC - Mimics a real MAGIC observation use case. Remote storage (Data Lake aware) **next to the telescope** acts as a buffer for subsequent data injection to the ESCAPE Data Lake (and local deletion after success).

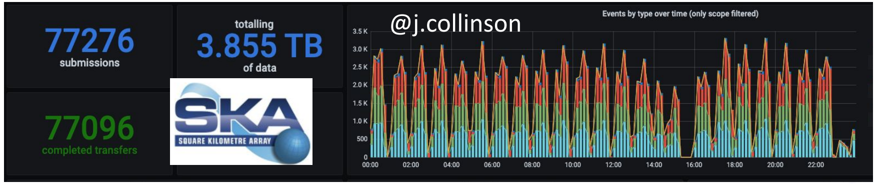


EGO/VIRGO - Upload 4h of VIRGO public data sampled at 4 kHz from an EGO server to the Data Lake. Download data to CNAF-STORM. Data are split into 1s samples. Making available the real-time strain data to pipelines and tools assessing the data quality.

FAIR - Upload 1 file (1 GB) 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).





77276
submissions

totalling
3.855 TB
of data

77096
completed transfers

SKA - Pulsar Observations injection test. For 4 hours at any point during the 24h, injecting new group of files in a dataset every 10 minutes. Files fall into two containers, representing different SKA Projects.
24h test moving data on basis of QoS class.

@x.espinal

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

Conclusion and Next Steps

- ESCAPE DIOS/WP2 managed to pilot a Data Lake infrastructure
  - fulfilling functional data management needs of flagship ESFRIs from several scientific disciplines
  - sensible technologies choice, conceived in WLCG environment and LHC experiments
- Full Dress Rehearsal - Assessment of the Data Lake
  - pivotal role to test model, concepts, and pilot infrastructure
  - chosen technologies offer the right functionality for a broader set of communities
    - Astro-particle Physics, Electromagnetic and Gravitational-Wave Astronomy, Particle Physics, and Nuclear Physics **pursuing together** FAIR and open-access data principles
  - ESCAPE contributing to broaden the scope of some of those technologies according to partners needs (in line/collaboration with providers plans)

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

Conclusion and Next Steps

- ESCAPE mature for prototype phase → full scale exercise for end of the year
  - close-to-production test on data management and data processing
    - experiments needs and interests, FAIR data management vs. embargoed (Open Data policy), fully multi-VO, implementation of token-based AAI
    - complementing existing efforts in WLCG → ESCAPE as perfect environment to test new models/concepts
    - exploring non-HEP-standard scenarios, etc.
  - collaboration with other EU-funded projects on-going
  - ESCAPE partners to explore first-hand technology stack
- ESCAPE end in 2022 → addressing long term sustainability
  - adopting components from established scientific contexts, leveraging services supported by large open source communities, documenting know-how on integration and deployment, ensuring services become part of EOSC-core

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