ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

The ESCAPE Data Lake: The machinery behind testing, monitoring and supporting a unified federated storage infrastructure of the exabyte-scale

Rizart Dona on behalf of the ESCAPE project

CERN

May 20, 2021 - 25th International Conference on Computing in High-Energy and Nuclear Physics (vCHEP 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

EUROPEAN OPEN SCIENCE CLOUD

Horizon2020
European Union Funding
for Research & Innovation

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

Partners

CERN

INFN

GSI

DESY

Nikhef

SURF SARA

rijksuniversiteit groningen

CCIN2P3

cnrs

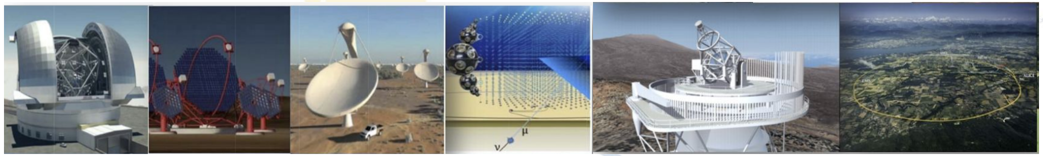
PIC port d'informació científica

LAPP Laboratoire d'Annecy de Physique des Particules

INAF ISTITUTO NAZIONALE DI ASTROFISICA NATIONAL INSTITUTE FOR ASTROPHYSICS

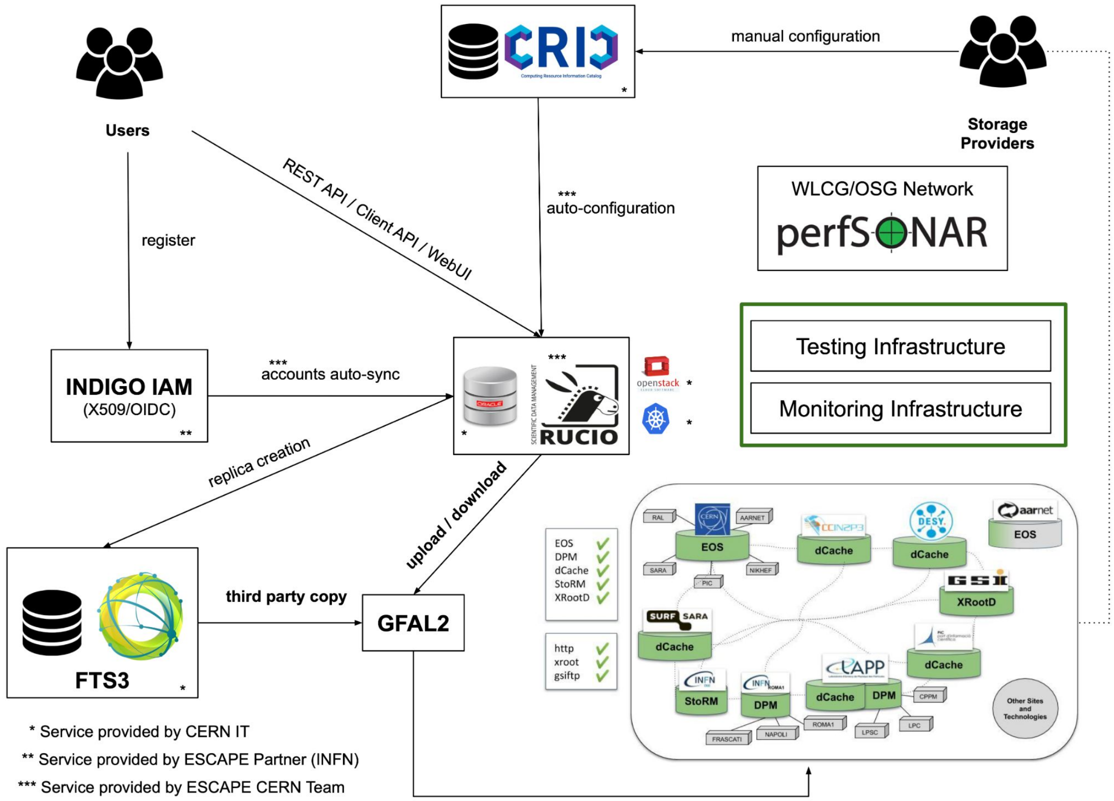
Project Goals

- 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

Data Lake Architecture



* Service provided by CERN IT

** Service provided by ESCAPE Partner (INFN)

*** Service provided by ESCAPE CERN Team

- Federated data infrastructure
  - Multiple **storage/protocol** technologies
  - Combination of HEP-specific services and industry standards
  - **QoS** and file transitions, **distributed redundancy** and **data policies**
- Data Transfer Stack
  - **Rucio** ⟶ Data orchestrator, policy-driven data management
  - **FTS3** ⟶ Middleware, reliable large scale file transfer service
  - **GFAL2** ⟶ Grid file access library, multi-protocol access
- Identity and Access Management
- CRIC Information Catalogue
- perfSONAR boxes deployed in the OSG/WLCG network

- **Testing** infrastructure
- **Monitoring** infrastructure

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

Testing Infrastructure - Continuous Testing

- In order to make sure that all data transfer/access solutions are functioning properly we have **continuous testing** in place

  → Crucial step for consolidating the infrastructure

- Separate tests target each component individually and explore scenarios that involve both **functional testing** as well as **stress testing**

  → Breaking the testing complexity of the stack

- Configurable software has been developed and deployed to make the process **automatic**
- Data from testing is **visualized** in the equivalent Grafana dashboards

  → Part of the monitoring of the Data Lake

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

Testing Infrastructure - GFAL

- All **Rucio Storage Elements (RSE)** consist of one or more endpoints that are associated with a supported protocol
- There are three types of operations that are being done concerning GFAL **functional testing**
  - **Upload** of files to all endpoints of all RSEs
  - **Download** of the files that were uploaded in the previous step
  - **Deletion** of the files that were uploaded in the first step
- This flow examines the **basic data operations** one can perform on the storage level of the Data Lake
  - Results ⟶ per storage element / per endpoint
  - Automatically pushing results in an Elasticsearch datasource
- Integrated with **CRIC**
  - Automatically fetching the configuration of the RSEs before each run

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

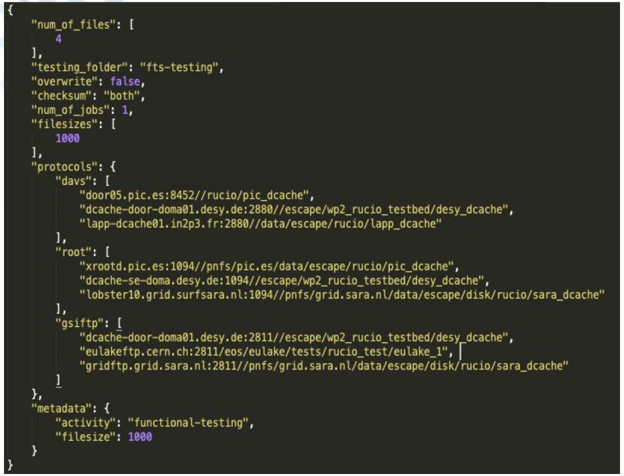
Testing Infrastructure - FTS

- In this case, the same endpoints as in the GFAL tests are examined
- Goal is to trigger **third party copy** transfers between all possible endpoint pairs that participate in the Data Lake

- Configurable software to trigger the transfers asynchronously and do automatic cleanup
- Extensive **error handling**

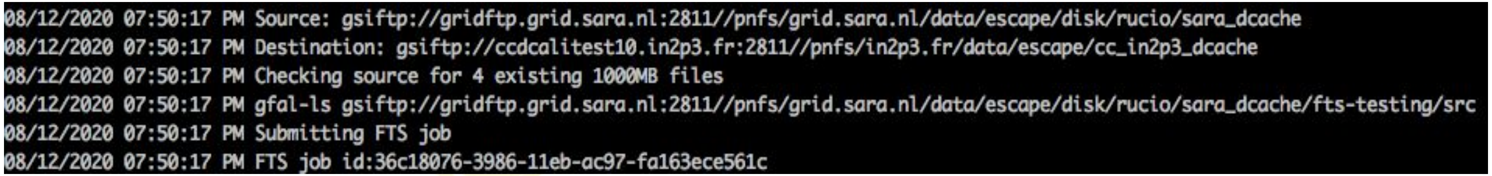
  → testing flow will continue even if endpoints fail mid-test

Config example:

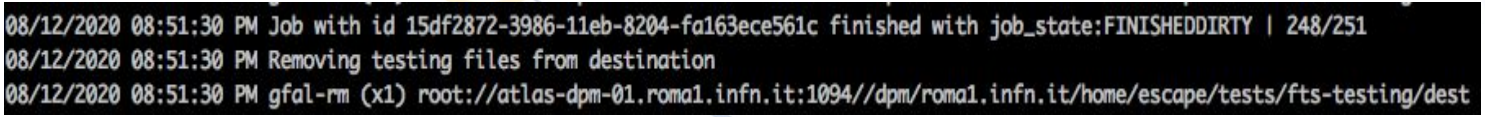


```
{
    "num_of_files": [
        4
    ],
    "testing_folder": "fts-testing",
    "overwrite": false,
    "checksum": "both",
    "num_of_jobs": 1,
    "filesizes": [
        1000
    ],
    "protocols": {
        "davs": [
            "door05.pic.es:8452//rucio/pic_dcache",
            "dcache-door-doma01.desy.de:2880//escape/wp2_rucio_testbed/desy_dcache",
            "lapp-dcache01.in2p3.fr:2880//data/escape/rucio/lapp_dcache"
        ],
        "root": [
            "xrootd.pic.es:1094//pnfs/pic.es/data/escape/rucio/pic_dcache",
            "dcache-se-doma.desy.de:1094//escape/wp2_rucio_testbed/desy_dcache",
            "lobster10.grid.surfsara.nl:1094//pnfs/grid.sara.nl/data/escape/disk/rucio/sara_dcache"
        ],
        "gsiftp": [
            "dcache-door-doma01.desy.de:2811//escape/wp2_rucio_testbed/desy_dcache",
            "eulakeftp.cern.ch:2811/eos/eulake/tests/rucio_test/eulake_1",
            "gridftp.grid.sara.nl:2811//pnfs/grid.sara.nl/data/escape/disk/rucio/sara_dcache"
        ]
    },
    "metadata": {
        "activity": "functional-testing",
        "filesize": 1000
    }
}
```

Run example:



```
08/12/2020 07:50:17 PM Source: gsiftp://gridftp.grid.sara.nl:2811//pnfs/grid.sara.nl/data/escape/disk/rucio/sara_dcache
08/12/2020 07:50:17 PM Destination: gsiftp://ccdcalitest10.in2p3.fr:2811//pnfs/in2p3.fr/data/escape/cc_in2p3_dcache
08/12/2020 07:50:17 PM Checking source for 4 existing 1000MB files
08/12/2020 07:50:17 PM gfal-ls gsiftp://gridftp.grid.sara.nl:2811//pnfs/grid.sara.nl/data/escape/disk/rucio/sara_dcache/fts-testing/src
08/12/2020 07:50:17 PM Submitting FTS job
08/12/2020 07:50:17 PM FTS job id:36c18076-3986-11eb-ac97-fa163ece561c
```



```
08/12/2020 08:51:30 PM Job with id 15df2872-3986-11eb-8204-fa163ece561c finished with job_state:FINISHEDDIRTY | 248/251
08/12/2020 08:51:30 PM Removing testing files from destination
08/12/2020 08:51:30 PM gfal-rm (x1) root://atlas-dpm-01.roma1.infn.it:1094//dpm/roma1.infn.it/home/escape/tests/fts-testing/dest
```

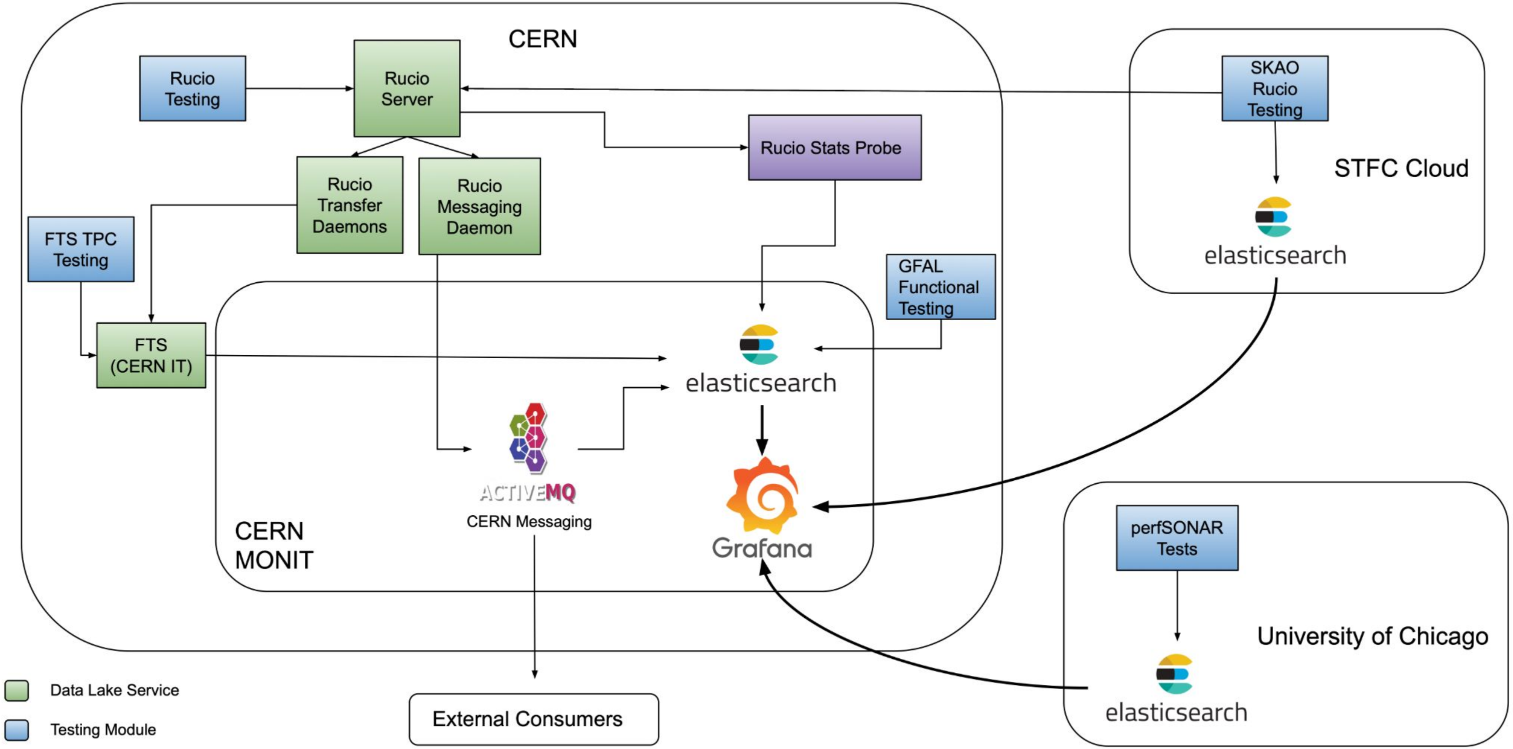
Funded by the European Union's
Horizon 2020 - Grant N° 824064

Testing Infrastructure - Rucio

- An agent is used for fast ad-hoc functional testing
  - Uploads files to all storage elements
  - Performs replica creation (triggers third party copies) between all pairs with so called Rucio rules
- rucio-analysis: extensible python3 framework for yaml-based Rucio tests
  - Developed and deployed by the ESCAPE SKA Team
  - Same type of tests as the testing agent

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

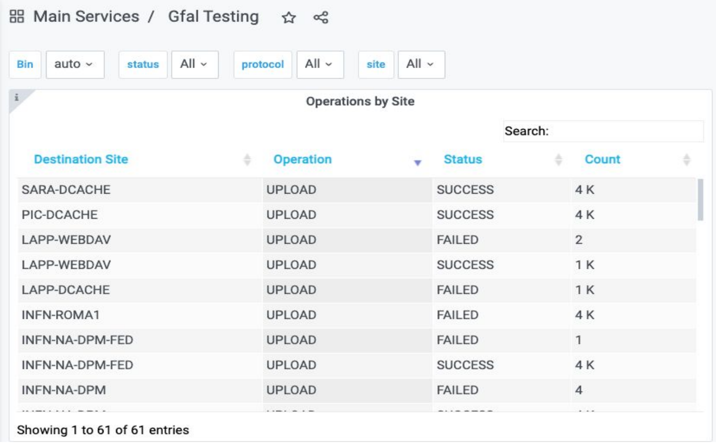
Monitoring Infrastructure - Architecture



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

Monitoring Infrastructure - GFAL Operations

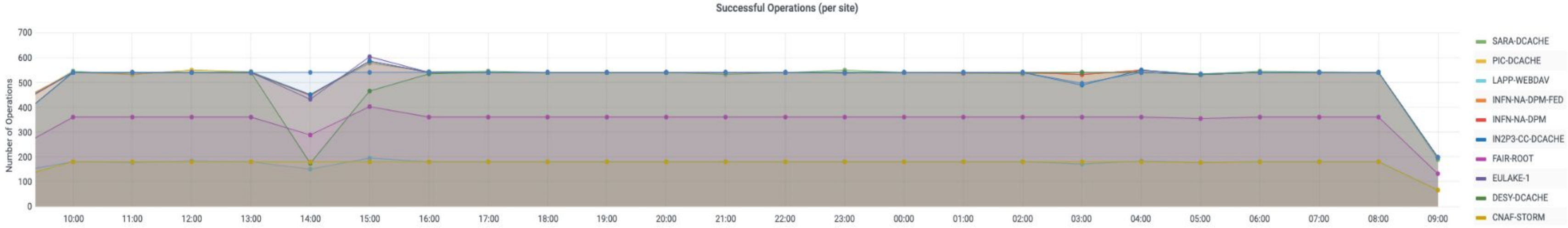
- Monitoring basic **GFAL operations** per storage element
- Users can **filter** plots by
  - Endpoints
  - Operations (upload, download, delete) - SUCCEEDED/FAILED/SKIPPED
  - Protocol (gsiftp, root, http)
- **Error messages** available for failed operations



Operations by Site

Destination Site	Operation	Status	Count
SARA-DCACHE	UPLOAD	SUCCESS	4 K
PIC-DCACHE	UPLOAD	SUCCESS	4 K
LAPP-WEBDAV	UPLOAD	FAILED	2
LAPP-WEBDAV	UPLOAD	SUCCESS	1 K
LAPP-DCACHE	UPLOAD	FAILED	1 K
INFN-ROMA1	UPLOAD	FAILED	4 K
INFN-NA-DPM-FED	UPLOAD	FAILED	1
INFN-NA-DPM-FED	UPLOAD	SUCCESS	4 K
INFN-NA-DPM	UPLOAD	FAILED	4

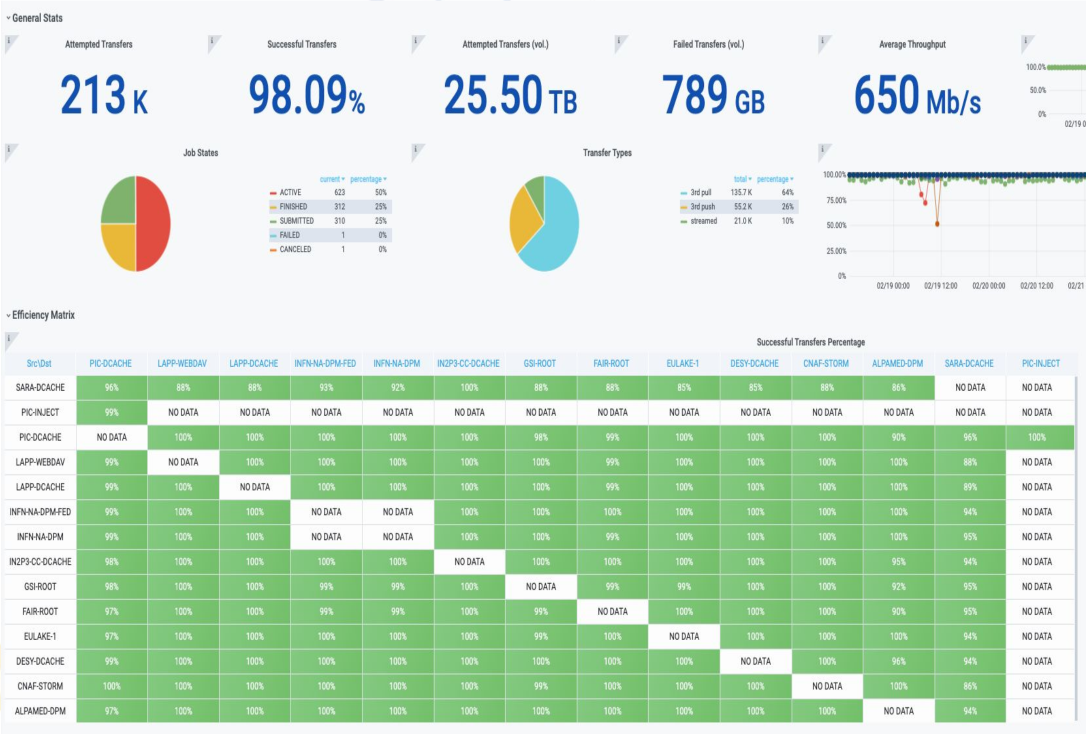
Showing 1 to 61 of 61 entries



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

Monitoring Infrastructure - FTS Transfers

- Monitoring **TPC transfers** between the endpoints
- Main highlights for users
  - Aggregated stats
    - Attempted transfers
    - Percentages
    - Job states
    - Transfer types
  - FTS transfers **efficiency matrix**
  - **Error messages** & links to log files



General Stats

Attempted Transfers	Successful Transfers	Attempted Transfers (vol.)	Failed Transfers (vol.)	Average Throughput
213 K	98.09%	25.50 TB	789 GB	650 Mb/s

Job States

	current	percentage
ACTIVE	623	50%
FINISHED	312	25%
SUBMITTED	310	25%
FAILED	1	0%
CANCELED	1	0%

Transfer Types

	total	percentage
3rd pull	135.7 K	64%
3rd push	55.2 K	26%
streamed	21.0 K	10%

Efficiency Matrix

Successful Transfers Percentage

Src\Dst	PIC-DCACHE	LAPP-WEBDAV	LAPP-DCACHE	INFN-NA-DPM-FED	INFN-NA-DPM	IN2P3-CC-DCACHE	GSI-ROOT	FAIR-ROOT	EULAKE-1	DESY-DCACHE	CNAF-STORM	ALPAMED-DPM	SARA-DCACHE	PIC-INJECT
SARA-DCACHE	96%	88%	88%	93%	92%	100%	88%	88%	85%	85%	88%	86%	NO DATA	NO DATA
PIC-INJECT	99%	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
PIC-DCACHE	NO DATA	100%	100%	100%	100%	100%	98%	99%	100%	100%	100%	90%	96%	100%
LAPP-WEBDAV	99%	NO DATA	100%	100%	100%	100%	100%	99%	100%	100%	100%	100%	88%	NO DATA
LAPP-DCACHE	99%	100%	NO DATA	100%	100%	100%	100%	99%	100%	100%	100%	100%	89%	NO DATA
INFN-NA-DPM-FED	99%	100%	100%	NO DATA	NO DATA	100%	100%	100%	100%	100%	100%	100%	94%	NO DATA
INFN-NA-DPM	99%	100%	100%	NO DATA	NO DATA	100%	100%	99%	100%	100%	100%	100%	95%	NO DATA
IN2P3-CC-DCACHE	98%	100%	100%	100%	100%	NO DATA	100%	100%	100%	100%	100%	95%	94%	NO DATA
GSI-ROOT	98%	100%	100%	99%	99%	100%	NO DATA	99%	99%	100%	100%	92%	95%	NO DATA
FAIR-ROOT	97%	100%	100%	99%	99%	100%	99%	NO DATA	100%	100%	100%	90%	95%	NO DATA
EULAKE-1	97%	100%	100%	100%	100%	100%	99%	100%	NO DATA	100%	100%	100%	94%	NO DATA
DESY-DCACHE	99%	100%	100%	100%	100%	100%	100%	100%	100%	NO DATA	100%	96%	94%	NO DATA
CNAF-STORM	100%	100%	100%	100%	100%	100%	99%	100%	100%	100%	NO DATA	100%	86%	NO DATA
ALPAMED-DPM	97%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	NO DATA	94%	NO DATA

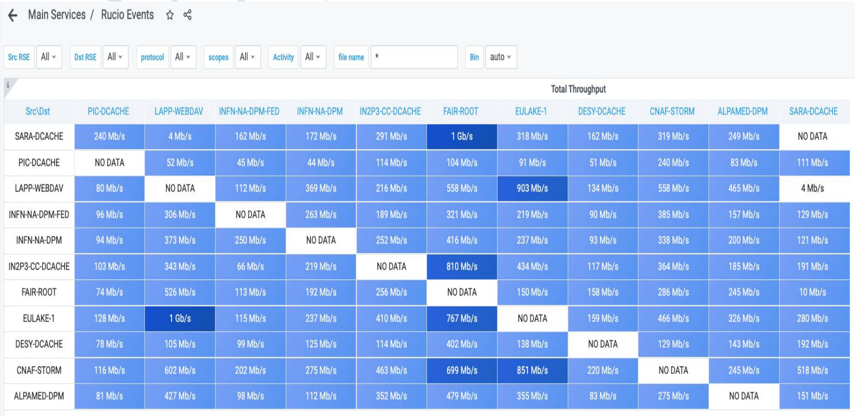
Funded by the European Union's
Horizon 2020 - Grant N° 824064

Monitoring Infrastructure - Rucio Activities

- Monitoring **Rucio** specific activities

  → **replica** creation/deletion

- **Matrices** help users understand the connectivity between storage elements (similar to the FTS one)
  - ○ Transfer successes / Throughput / Transferred volume
- Table with **error messages** and links to the actual FTS transfer logs

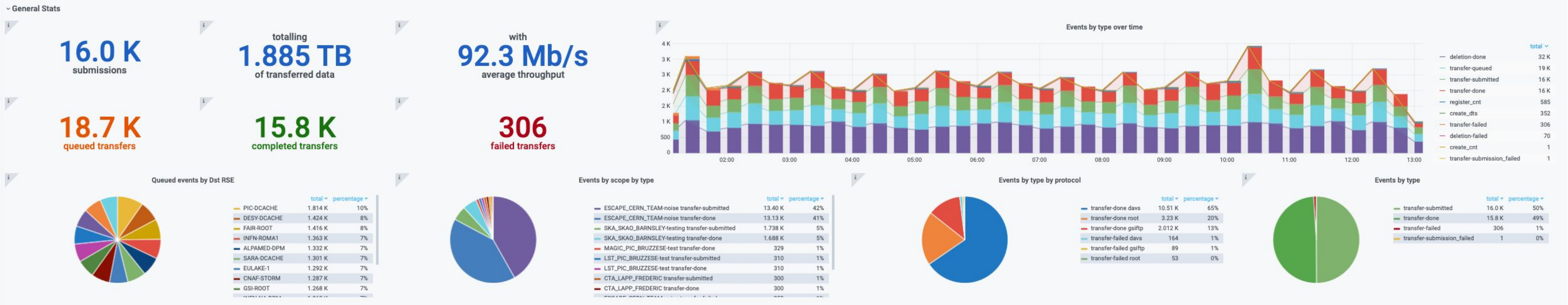


Main Services / Rucio Events

Src RSE: All | Dst RSE: All | protocol: All | scopes: All | Activity: All | file name: * | Bin: auto

Total Throughput

Src\Dst	PIC-DCACHE	LAPP-WEBDAV	INFN-NA-DPM-FED	INFN-NA-DPM	IN2P3-CC-DCACHE	FAIR-ROOT	EULAKE-1	DESY-DCACHE	CNAF-STORM	ALPAMED-DPM	SARA-DCACHE
SARA-DCACHE	240 Mb/s	4 Mb/s	162 Mb/s	172 Mb/s	291 Mb/s	1 Gb/s	318 Mb/s	162 Mb/s	319 Mb/s	249 Mb/s	NO DATA
PIC-DCACHE	NO DATA	52 Mb/s	45 Mb/s	44 Mb/s	114 Mb/s	104 Mb/s	91 Mb/s	51 Mb/s	240 Mb/s	83 Mb/s	111 Mb/s
LAPP-WEBDAV	80 Mb/s	NO DATA	112 Mb/s	369 Mb/s	216 Mb/s	558 Mb/s	903 Mb/s	134 Mb/s	558 Mb/s	465 Mb/s	4 Mb/s
INFN-NA-DPM-FED	96 Mb/s	306 Mb/s	NO DATA	263 Mb/s	189 Mb/s	321 Mb/s	219 Mb/s	90 Mb/s	385 Mb/s	157 Mb/s	129 Mb/s
INFN-NA-DPM	94 Mb/s	373 Mb/s	250 Mb/s	NO DATA	252 Mb/s	416 Mb/s	237 Mb/s	93 Mb/s	338 Mb/s	200 Mb/s	121 Mb/s
IN2P3-CC-DCACHE	103 Mb/s	343 Mb/s	66 Mb/s	219 Mb/s	NO DATA	810 Mb/s	434 Mb/s	117 Mb/s	364 Mb/s	185 Mb/s	191 Mb/s
FAIR-ROOT	74 Mb/s	526 Mb/s	113 Mb/s	192 Mb/s	256 Mb/s	NO DATA	150 Mb/s	158 Mb/s	286 Mb/s	245 Mb/s	10 Mb/s
EULAKE-1	128 Mb/s	1 Gb/s	115 Mb/s	237 Mb/s	410 Mb/s	767 Mb/s	NO DATA	159 Mb/s	466 Mb/s	326 Mb/s	280 Mb/s
DESY-DCACHE	78 Mb/s	105 Mb/s	99 Mb/s	125 Mb/s	114 Mb/s	402 Mb/s	138 Mb/s	NO DATA	129 Mb/s	143 Mb/s	192 Mb/s
CNAF-STORM	116 Mb/s	602 Mb/s	202 Mb/s	275 Mb/s	463 Mb/s	699 Mb/s	851 Mb/s	220 Mb/s	NO DATA	245 Mb/s	518 Mb/s
ALPAMED-DPM	81 Mb/s	427 Mb/s	98 Mb/s	112 Mb/s	352 Mb/s	479 Mb/s	355 Mb/s	83 Mb/s	275 Mb/s	NO DATA	151 Mb/s



General Stats

16.0 K submissions

totalling **1.885 TB** of transferred data

with **92.3 Mb/s** average throughput

18.7 K queued transfers

15.8 K completed transfers

306 failed transfers

Events by type over time

	total
deletion-done	32 K
transfer-queued	19 K
transfer-submitted	16 K
transfer-done	16 K
register_cnt	585
create_dts	352
transfer-failed	306
deletion-failed	70
create_cnt	1
transfer-submission_failed	1

Queued events by Dst RSE

	total	percentage
PIC-DCACHE	1.814 K	10%
DESY-DCACHE	1.424 K	8%
FAIR-ROOT	1.416 K	8%
INFN-ROMA1	1.363 K	7%
ALPAMED-DPM	1.332 K	7%
SARA-DCACHE	1.301 K	7%
EULAKE-1	1.292 K	7%
CNAF-STORM	1.287 K	7%
GSI-ROOT	1.268 K	7%

Events by scope by type

	total	percentage
ESCAPE_CERN_TEAM-noise transfer-submitted	13.40 K	42%
ESCAPE_CERN_TEAM-noise transfer-done	13.13 K	41%
SKA_SKAO_BARNSLEY-testing transfer-submitted	1.738 K	5%
SKA_SKAO_BARNSLEY-testing transfer-done	1.688 K	5%
MAGIC_PIC_BRUZZESE-test transfer-done	329	1%
LST_PIC_BRUZZESE-test transfer-submitted	310	1%
LST_PIC_BRUZZESE-test transfer-done	310	1%
CTA_LAPP_FREDERIC transfer-submitted	300	1%
CTA_LAPP_FREDERIC transfer-done	300	1%

Events by type by protocol

	total	percentage
transfer-done davs	10.51 K	65%
transfer-done root	3.23 K	20%
transfer-done gsiftp	2.012 K	13%
transfer-failed davs	164	1%
transfer-failed gsiftp	89	1%
transfer-failed root	53	0%

Events by type

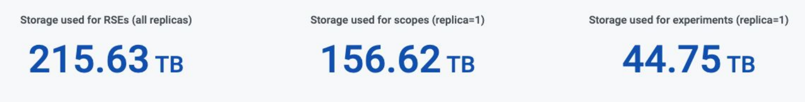
	total	percentage
transfer-submitted	16.0 K	50%
transfer-done	15.8 K	49%
transfer-failed	306	1%
transfer-submission_failed	1	0%

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

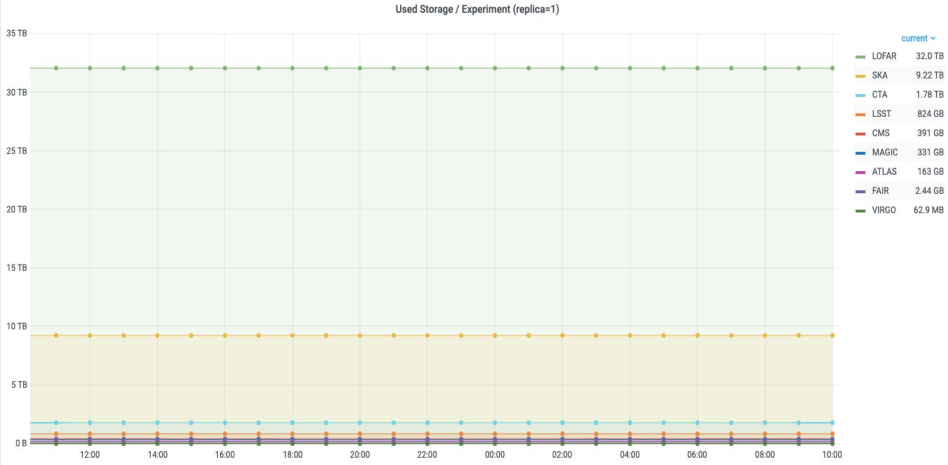
Monitoring Infrastructure - Rucio Accounting

- Rucio **accounting data**
- Storage used / Number of files
  - Per **storage element**
  - Per **experiment**
- Captures trends over time

  → Easy to observe changes on the storage level

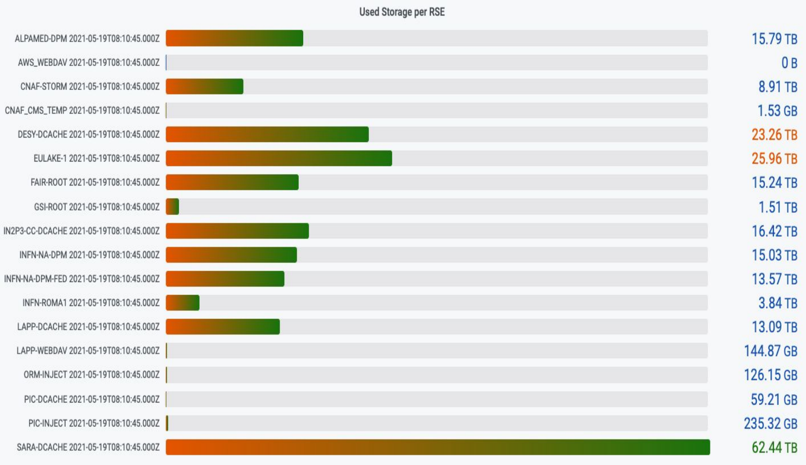


Storage used for RSEs (all replicas)	Storage used for scopes (replica=1)	Storage used for experiments (replica=1)
215.63 TB	156.62 TB	44.75 TB



Used Storage / Experiment (replica=1)

	current ⌄
LOFAR	32.0 TB
SKA	9.22 TB
CTA	1.78 TB
LSST	824 GB
CMS	391 GB
MAGIC	331 GB
ATLAS	163 GB
FAIR	2.44 GB
VIRGO	62.9 MB



Used Storage per RSE

ALPAMED-DPM 2021-05-19T08:10:45.000Z	15.79 TB
AWS_WEBDAV 2021-05-19T08:10:45.000Z	0 B
CNAF-STORM 2021-05-19T08:10:45.000Z	8.91 TB
CNAF_CMS_TEMP 2021-05-19T08:10:45.000Z	1.53 GB
DESY-DCACHE 2021-05-19T08:10:45.000Z	23.26 TB
EULAKE-1 2021-05-19T08:10:45.000Z	25.96 TB
FAIR-ROOT 2021-05-19T08:10:45.000Z	15.24 TB
GSI-ROOT 2021-05-19T08:10:45.000Z	1.51 TB
IN2P3-CC-DCACHE 2021-05-19T08:10:45.000Z	16.42 TB
INFN-NA-DPM 2021-05-19T08:10:45.000Z	15.03 TB
INFN-NA-DPM-FED 2021-05-19T08:10:45.000Z	13.57 TB
INFN-ROMA1 2021-05-19T08:10:45.000Z	3.84 TB
LAPP-DCACHE 2021-05-19T08:10:45.000Z	13.09 TB
LAPP-WEBDAV 2021-05-19T08:10:45.000Z	144.87 GB
ORM-INJECT 2021-05-19T08:10:45.000Z	126.15 GB
PIC-DCACHE 2021-05-19T08:10:45.000Z	59.21 GB
PIC-INJECT 2021-05-19T08:10:45.000Z	235.32 GB
SARA-DCACHE 2021-05-19T08:10:45.000Z	62.44 TB

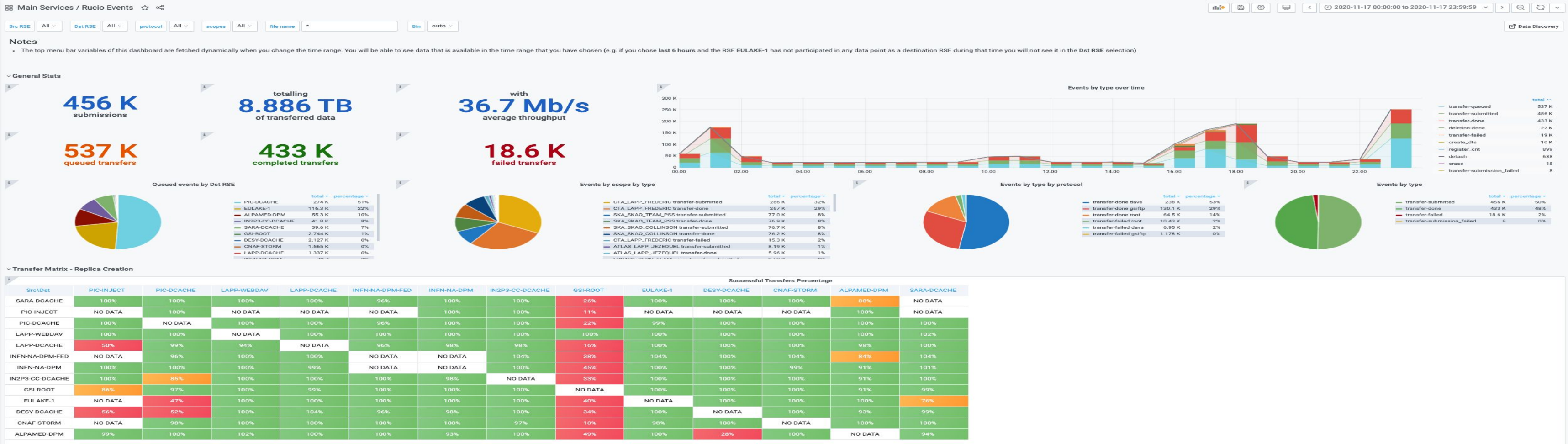
Funded by the European Union's
Horizon 2020 - Grant N° 824064

Monitoring Infrastructure - The FDR experience

- Full Dress Rehearsal → 24h **full scale commissioning exercise**
  - Experiments were called to run their workflows on the Data Lake

Please find more details in Riccardo's presentation (May 19th, Storage session)

- The monitoring infrastructure proved to be a **critical component** during this exercise
  - Instrumental for **identifying issues** and speeding up **debugging**
  - Users could follow up their data injection activity and **analyse** on the fly how this reflects on the Data Lake



Notes

- The top menu bar variables of this dashboard are fetched dynamically when you change the time range. You will be able to see data that is available in the time range that you have chosen (e.g. if you chose **last 6 hours** and the RSE **EULAKE-1** has not participated in any data point as a destination RSE during that time you will not see it in the **Dst RSE** selection)

General Stats

456 K submissions

totalling **8.886 TB** of transferred data

with **36.7 Mb/s** average throughput

537 K queued transfers

433 K completed transfers

18.6 K failed transfers

	total
transfer-queued	537 K
transfer-submitted	456 K
transfer-done	433 K
deletion-done	22 K
transfer-failed	19 K
create_dts	10 K
register_cnt	899
detach	688
erase	18
transfer-submission_failed	8

Queued events by Dst RSE

	total	percentage
PIC-DCACHE	274 K	51%
EULAKE-1	116.3 K	22%
ALPAMED-DPM	55.3 K	10%
IN2P3-CC-DCACHE	41.8 K	8%
SARA-DCACHE	39.6 K	7%
GSI-ROOT	2.744 K	1%
DESY-DCACHE	2.127 K	0%
CNAF-STORM	1.565 K	0%
LAPP-DCACHE	1.337 K	0%

Events by scope by type

	total	percentage
CTA_LAPP_FREDERIC transfer-submitted	286 K	32%
CTA_LAPP_FREDERIC transfer-done	267 K	29%
SKA_SKAO_TEAM_PSS transfer-submitted	77.0 K	8%
SKA_SKAO_TEAM_PSS transfer-done	76.9 K	8%
SKA_SKAO_COLLINSON transfer-submitted	76.7 K	8%
SKA_SKAO_COLLINSON transfer-done	76.2 K	8%
CTA_LAPP_FREDERIC transfer-failed	15.3 K	2%
ATLAS_LAPP_JEZEQUEL transfer-submitted	8.19 K	1%
ATLAS_LAPP_JEZEQUEL transfer-done	5.96 K	1%

Events by type by protocol

	total	percentage
transfer-done davs	238 K	53%
transfer-done gsiftp	130.1 K	29%
transfer-done root	64.5 K	14%
transfer-failed root	10.43 K	2%
transfer-failed davs	6.95 K	2%
transfer-failed gsiftp	1.178 K	0%

Events by type

	total	percentage
transfer-submitted	456 K	50%
transfer-done	433 K	48%
transfer-failed	18.6 K	2%
transfer-submission_failed	8	0%

Transfer Matrix - Replica Creation

Successful Transfers Percentage

Src\Dst	PIC-INJECT	PIC-DCACHE	LAPP-WEBDAV	LAPP-DCACHE	INFN-NA-DPM-FED	INFN-NA-DPM	IN2P3-CC-DCACHE	GSI-ROOT	EULAKE-1	DESY-DCACHE	CNAF-STORM	ALPAMED-DPM	SARA-DCACHE
SARA-DCACHE	100%	100%	100%	100%	96%	100%	100%	26%	100%	100%	100%	88%	NO DATA
PIC-INJECT	NO DATA	100%	NO DATA	NO DATA	NO DATA	100%	100%	11%	NO DATA	NO DATA	NO DATA	100%	NO DATA
PIC-DCACHE	100%	NO DATA	100%	100%	96%	100%	100%	22%	99%	100%	100%	100%	100%
LAPP-WEBDAV	100%	100%	NO DATA	100%	100%	100%	100%	100%	100%	100%	100%	100%	102%
LAPP-DCACHE	50%	99%	94%	NO DATA	96%	98%	98%	16%	100%	100%	100%	98%	100%
INFN-NA-DPM-FED	NO DATA	96%	100%	100%	NO DATA	NO DATA	104%	38%	104%	100%	104%	84%	104%
INFN-NA-DPM	100%	100%	100%	99%	NO DATA	NO DATA	100%	45%	100%	100%	99%	91%	101%
IN2P3-CC-DCACHE	100%	85%	100%	100%	100%	98%	NO DATA	33%	100%	100%	100%	91%	100%
GSI-ROOT	86%	97%	100%	99%	100%	100%	100%	NO DATA	100%	100%	100%	91%	99%
EULAKE-1	NO DATA	47%	100%	100%	100%	100%	100%	40%	NO DATA	100%	100%	100%	76%
DESY-DCACHE	56%	52%	100%	104%	96%	98%	100%	34%	100%	NO DATA	100%	93%	99%
CNAF-STORM	NO DATA	98%	100%	100%	100%	100%	97%	18%	98%	100%	NO DATA	100%	100%
ALPAMED-DPM	99%	100%	102%	100%	100%	93%	100%	49%	100%	28%	100%	NO DATA	94%

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

Conclusions & Future work

- Successfully reached the **Pilot Phase**
  - Robust architecture that **serves** the needs of the sciences
  - **Testing** infrastructure and **monitoring** capabilities
    - Essential for validating the Data Lake status
    - Allow users and **experiments** to efficiently track their activities
- Moving towards the **Prototype Phase**
  - Consolidation of topics
    - **Token** based authentication/authorization
    - **Quality of Service** mechanisms
    - **Network optimized** transfers
- The ESCAPE Data Lake is constantly **evolving**

→ In sync with the **WLCG** activities and the **DOMA** project data challenges

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

References

- ESCAPE Project, https://projectescape.eu/
- ESCAPE Data Lake Wiki, https://wiki.escape2020.de/index.php/WP2_-_DIOS
- FTS3, https://fts.web.cern.ch/fts/
- GFAL2, https://dmc-docs.web.cern.ch/dmc-docs/gfal2/gfal2.html
- Rucio, https://rucio.cern.ch/
- GFAL testing software, https://github.com/ESCAPE-WP2/Utilities-and-Operations-Scripts/tree/master/gfal-sam-testing
- FTS testing software, https://github.com/ESCAPE-WP2/fts-analysis-datalake
- Rucio testing software, https://github.com/ESCAPE-WP2/rucio-analysis
- Elasticsearch, https://www.elastic.co/elasticsearch
- Grafana, https://grafana.com
- Apache ActiveMQ, http://activemq.apache.org
- ESCAPE Grafana Org, https://monit-grafana.cern.ch/d/cHBQ2NjWz/escape-home?orgId=51

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

Thank you!

Questions/Comments?