



# ESCAPE

European Science Cluster of Astronomy &  
Particle physics ESFRI research Infrastructures

## Leveraging ESCAPE monitoring & testing tools for the WLCG Data Challenges

Rizart Dona

CERN

March 17, 2021 - DOMA / TPC Meeting



## Science Projects



# ESCAPE

European Science Cluster of Astronomy &  
Particle physics ESFRI research Infrastructures



## EUROPEAN OPEN SCIENCE CLOUD



Horizon2020  
European Union Funding  
for Research & Innovation

## Data Centres

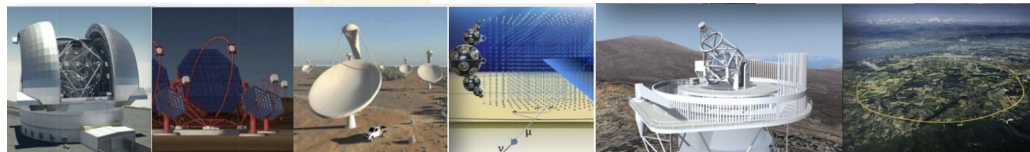


rijksuniversiteit  
 groningen



## 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

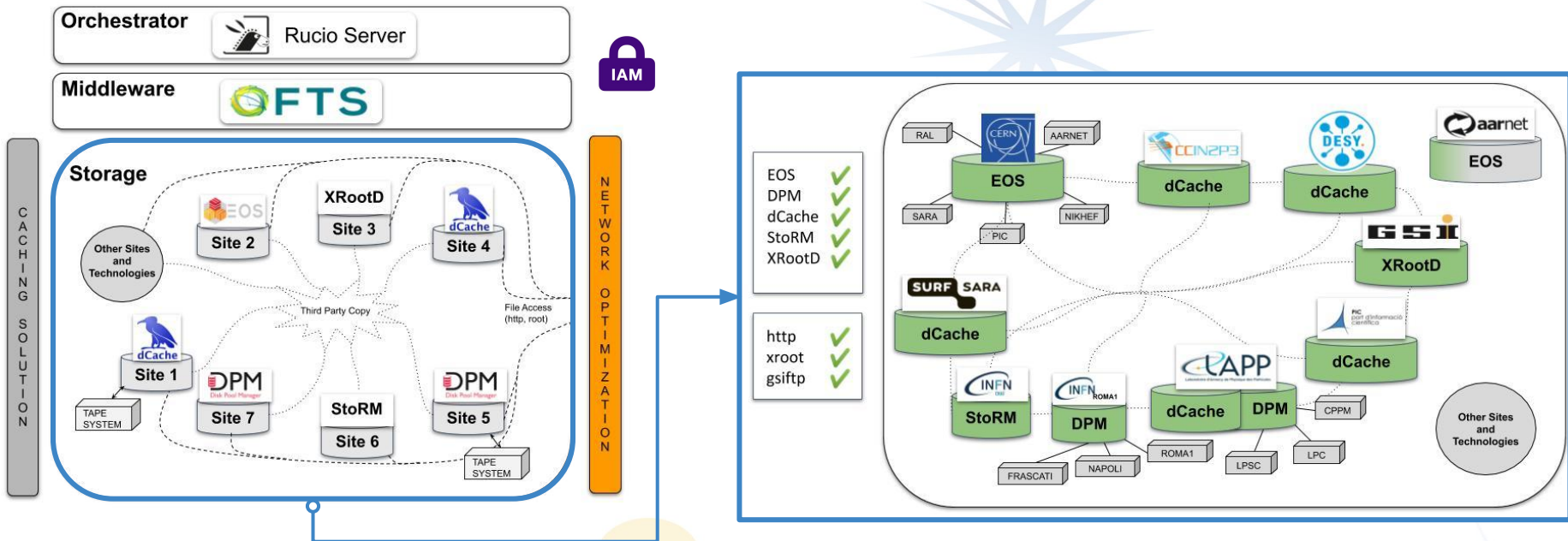


# Overview

- Data Lake Architecture
- Testing Infrastructure
- Monitoring Infrastructure
- Conclusions & Future Work
- References



# Data Lake Architecture - Storages



- Federated data infrastructure

- Multiple storage/protocol technologies

- Diverse participation of sites



# Data Lake Architecture - Services

From CERN Team



**Rucio**

**XCache**



From CERN-IT

**FTS**

**OpenStack**

**perfSONAR**

**CRIC**

**Grafana**

**OracleDB**

**k8s**

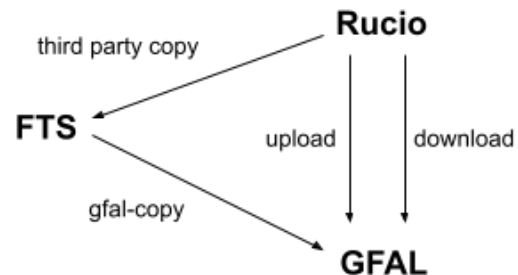
From ESCAPE Partners

**IAM**



# Data Lake Architecture - Transfer Stack

- The Datalake currently employs three major tools/systems that deal with data transfer/access
- **GFAL** (Grid File Access Library), a multi-protocol data management library providing an abstraction layer of the grid storage system complexity (supports protocols like GridFTP, Http & Root)
- **FTS** (File Transfer Service), open source software to transfer data reliably and at large scale between storage systems
- **Rucio**, the data orchestration service, a scalable policy-driven scientific data management system that can work with large amounts of data, this is the service that users basically interact with



# Testing Infrastructure - Continuous Testing

- In order to make sure that all three 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** → Breaks testing complexity
- Configurable software has been developed and deployed to make the process **automatic**
- Data from testing is **visualized** in the equivalent Grafana dashboards that consist the monitoring of the Datalake





# Testing Infrastructure - GFAL

- All RSEs (Rucio) 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 a file that is a few bytes long to all the endpoints of all RSEs
  - **Download** of the file that was uploaded in the previous step
  - **Deletion** of the file that was uploaded in the first step
- This flow examines the **basic data operations** one can perform on the storage level of the Datalake
  - Per RSE per endpoint results
  - Automatically pushing results on an Elasticsearch datasource
- Integrated with **CRIC**
  - Automatically fetching the RSEs configuration before each run
- [Python script](#), deployed inside a container in a Kubernetes cluster @ CERN





# Testing Infrastructure - FTS (1/2)

- In this case, the same endpoints as in the GFAL testing are examined
- Goal is to trigger **TPC** transfers between all possible endpoint pairs that participate in the Datalake
- The flow is the following:
  - The toolkit reads from a **configuration file** all endpoint pairs that are to be tested (config example in the picture)
  - Testing folders are being setup, automatic detection of **problematic** endpoints that will be excluded from the testing if this cannot be achieved
  - For each pair of compatible endpoints an FTS transfer will be triggered
    - 1) Check source for existing files, generate & upload ad-hoc if needed
    - 2) Trigger FTS job, **asynchronous** action
  - Wait for all jobs to finish, when done **delete** the files that were transferred on the destination endpoint (ensures no quota is exceeded)
- Extensive **error handling**, flow will continue even if endpoints fail mid-test
- [fts-analysis-datalake](#) repo, python code (major refactoring is in progress)

```
{
  "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
  }
}
```



## Testing Infrastructure - FTS (2/2)

- Testing results are automatically pushed from the FTS server to an **Elasticsearch** datasource (all systems maintained by the relevant teams @ CERN)
- Deployed inside a container in a Kubernetes cluster @ CERN

```
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://ccdcaltest10.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
```



# Testing Infrastructure - Rucio

- *Reminder:* All RSEs (Rucio) consist of one or more endpoints that are associated with a supported protocol
- A [bash script](#) is used for fast ad-hoc functional testing
  - Uploads files to all RSEs
  - Triggers replica creation between all pairs with so called Rucio rules
  - Deployed inside a container in a Kubernetes cluster @ CERN
- [rucio-analysis](#): extensible python3 framework for yaml-based Rucio tests
  - Developed and deployed by SKA
  - Same type of tests as the bash script

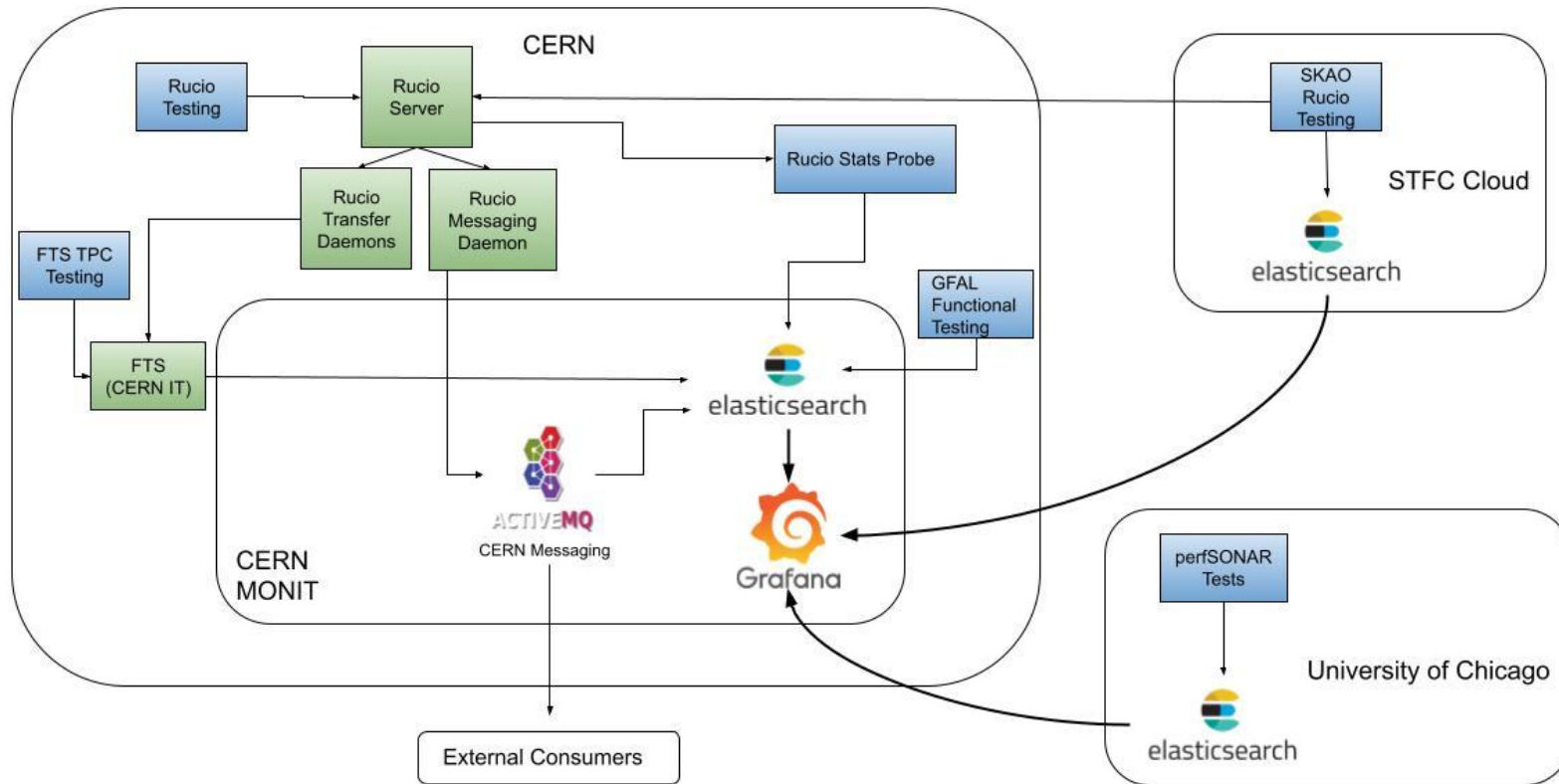


# Monitoring Infrastructure - Software Stack

- Backend storage/messaging technologies used
  - **Elasticsearch**, a distributed multitenant-capable full-text search engine
  - **ActiveMQ**, an open source, multi-protocol, Java-based messaging broker
- Visualization platform → **Grafana**
  - A multi-platform open source **analytics** and interactive **visualization** web application
  - Hosted @ CERN by Monit, separate ESCAPE organization
  - Supports multiple data sources (ES, InfluxDB, Graphite, MySQL, etc.)
- Our main data sources
  - **Elasticsearch** @ CERN (hosted by Monit) → Retention policy of 12 months
  - **Elasticsearch** @ SKAO → No retention policy in place, limited by storage given



# Monitoring Infrastructure - Architecture



- Basic GFAL upload/download/delete operations per endpoint
- A user can filter plots by
  - Endpoints
  - Operations (upload, download, delete) - SUCCEEDED/FAILED/SKIPPED
  - Protocol (gsiftp, root, http)

Main Services / Gfal Testing ☆ 🔗

Bin auto ▾ status All ▾ protocol All ▾ site All ▾

1 Operations by Site

Search:

Destination Site	Operation	Status	Count
SARA-DCACHE	UPLOAD	SUCCESS	180
PIC-DCACHE	UPLOAD	SUCCESS	118
LAPP-WEBDAV	UPLOAD	SUCCESS	59
LAPP-DCACHE	UPLOAD	SUCCESS	59
INFN-ROMA1	UPLOAD	SUCCESS	178
INFN-NA-DPM-FED	UPLOAD	SUCCESS	180
INFN-NA-DPM	UPLOAD	SUCCESS	180
IN2P3-CC-DCACHE	UPLOAD	SUCCESS	180
GSI-ROOT	UPLOAD	SUCCESS	120

Showing 1 to 42 of 42 entries



# Monitoring Infrastructure - FTS (1/2)

- Monitoring TPC transfers between the endpoints

- Main highlights for user

- Aggregated stats

- Attempted transfers

- Percentages

- Job states

- Transfer types

- FTS transfers **efficiency matrix**

- Error codes & logfile links

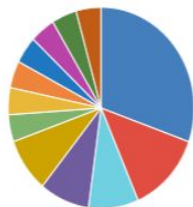




# Monitoring Infrastructure - FTS (2/2)

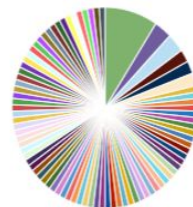
## ▼ Error Messages & Failures

Failures By Error Category



	current	percentage
COMMUNICATION_ERROR_ON_SEND	7	30%
INVALID_EXCHANGE	3	13%
NO_SUCH_DEVICE	2	9%
HOST_IS_DOWN	2	9%
CONNECTION_TIMED_OUT	2	9%
PERMISSION_DENIED	1	4%
NO_SUCH_FILE_OR_DIRECTORY	1	4%
NO_MESSAGE_OF_DESIRED_TYPE	1	4%
INVALID_ARGUMENT	1	4%
INPUT/OUTPUT_ERROR	1	4%
FUNCTION_NOT_IMPLEMENTED	1	4%

Failures By Source RSE



	current	percentage
SARA-DCACHE PERMISSION_DENIED	9	9%
SARA-DCACHE COMMUNICATION_ERROR_ON_SEND	3	3%
FAIR-ROOT NO_SUCH_DEVICE	3	3%
GSI-ROOT NO_SUCH_FILE_OR_DIRECTORY	2	2%
GSI-ROOT NO_SUCH_DEVICE	2	2%
GSI-ROOT CONNECTION_TIMED_OUT	2	2%
SARA-DCACHE NO_MESSAGE_OF_DESIRED_TYPE	1	1%
SARA-DCACHE INVALID_EXCHANGE	1	1%
SARA-DCACHE INPUT/OUTPUT_ERROR	1	1%
SARA-DCACHE HOST_IS_DOWN	1	1%
SARA-DCACHE FUNCTION_NOT_IMPLEMENTED	1	1%

Latest Failed Logfiles

Source RSE	Dest RSE	Logfile Link	Error Message
SARA-DCACHE	PIC-DCACHE	<a href="https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/49320e82-76fe-11eb-965f-fa163e9eca79">https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/49320e82-76fe-11eb-965f-fa163e9eca79</a>	Result (Neon): SSL handshake failed: sslv3 alert certificate unkn
SARA-DCACHE	LAPP-WEBDAV	<a href="https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/47e649ba-76fb-11eb-ad70-fa163ed6e1b4">https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/47e649ba-76fb-11eb-ad70-fa163ed6e1b4</a>	TRANSFER ERROR: Copy failed with mode streamed, with error: I
SARA-DCACHE	LAPP-DCACHE	<a href="https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/42919d0a-7702-11eb-98d2-fa163e297a88">https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/42919d0a-7702-11eb-98d2-fa163e297a88</a>	Result (Neon): SSL handshake failed: sslv3 alert certificate unkn
SARA-DCACHE	INFN-NA-DPM-FED	<a href="https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/546d9018-7704-11eb-8adf-fa163ec6a1a1">https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/546d9018-7704-11eb-8adf-fa163ec6a1a1</a>	globus_ftp_client: the server responded with an error 535 Auther
SARA-DCACHE	INFN-NA-DPM	<a href="https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/aba106ce-7706-11eb-8d5b-fa163e14af38">https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/aba106ce-7706-11eb-8d5b-fa163e14af38</a>	globus_ftp_client: the server responded with an error 535 Auther
SARA-DCACHE	GSI-ROOT	<a href="https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/34297508-7702-11eb-bb57-fa163e006cf5">https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/34297508-7702-11eb-bb57-fa163e006cf5</a>	Result (Neon): SSL handshake failed: sslv3 alert certificate unkn
SARA-DCACHE	FAIR-ROOT	<a href="https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/784a9534-7282-11eb-8526-fa163ed6e1b4">https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/784a9534-7282-11eb-8526-fa163ed6e1b4</a>	TRANSFER ERROR: Copy failed with mode 3rd pull, with error: Re
SARA-DCACHE	EULAKE-1	<a href="https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/87828ba4-7707-11eb-aedc-fa163ece561c">https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/87828ba4-7707-11eb-aedc-fa163ece561c</a>	globus_ftp_client: the server responded with an error 535 Auther
SARA-DCACHE	RESV-DCACHE	<a href="https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/776b80fa-7706-11eb-854b-fa163ec6a1a1">https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/job/776b80fa-7706-11eb-854b-fa163ec6a1a1</a>	globus_ftp_client: the server responded with an error 535 Auther



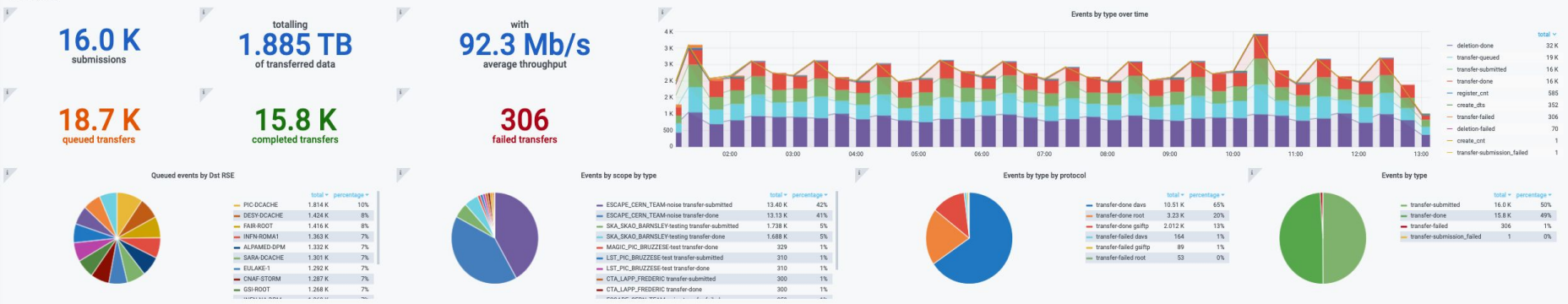
# Monitoring Infrastructure - Rucio

- Monitoring replica creation/deletion
- Transfer matrix helps user understand the connectivity between RSES
  - Percentages of successful replication
- Table with error codes and logfile links to the actual FST transfers

Transfer Matrix - Replica Creation

SrcRSE	Successful Transfers Percentage													
	PIC-DCACHE	LAPP-WEBAV	LAPP-DCACHE	INFN-NA-OPM-FED	INFN-NA-OPM	INDP3-CCDCACHE	OS-ROOT	FAIR-ROOT	ELLAKE1	DESY-DCACHE	CNAF-STORM	ALPAMED-OPM	SARA-DCACHE	
SARA-DCACHE	10%	91%	99%	10%	99%	100%	98%	92%	7%	99%	93%	8%	NO DATA	
PIC-INJECT	11%	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%	100%	100%	100%	100%	100%	9%	94%	
LAPP-WEBAV	100%	NO DATA	100%	100%	100%	100%	100%	100%	100%	100%	10%	100%	8%	
LAPP-DCACHE	100%	100%	NO DATA	100%	100%	100%	100%	100%	100%	100%	100%	100%	97%	
INFN-NA-OPM-FED	100%	100%	100%	NO DATA	NO DATA		100%	100%	100%	100%	100%	10%	91%	
INFN-NA-OPM	100%	100%	10%	NO DATA	NO DATA		100%	9%	9%	100%	9%	100%	9%	
INDP3-CCDCACHE	100%	10%	100%	10%	9%	NO DATA	9%	100%	100%	100%	9%	9%	9%	
OS-ROOT	9%	100%	100%	100%	9%	NO DATA	9%	9%	9%	9%	100%	9%	9%	
FAIR-ROOT	9%	9%	9%	9%	100%	9%	9%	NO DATA	9%	9%	9%	9%	100%	
ELLAKE1	100%	9%	10%	100%	9%	9%	9%	100%	NO DATA	10%	100%	100%	94%	
DESY-DCACHE	100%	9%	100%	9%	9%	9%	9%	100%	100%	NO DATA	100%	9%	9%	
CNAF-STORM	100%	100%	100%	100%	100%	100%	100%	100%	100%	10%	NO DATA	100%	9%	
ALPAMED-OPM	100%	9%	100%	100%	10%	100%	9%	100%	100%	100%	100%	NO DATA	9%	

General Stats



# Monitoring Infrastructure - Rucio Stats

- Rucio Stats Probe → fetches periodically the relevant stats from Rucio, using the client API to query the DB
- Displays usage per RSE/experiment, number of files and trends over time



**general stats**

**NOTE**  
From this point downwards, this dashboard displays always the latest values of the following statistics. A cron job runs every 10 minutes and registers those values. Timestamp values are UTC.

**RSE DATA (all replicas)**

## 70.53 TB

**EXPERIMENTS DATA (replica=1)**

## 46.28 TB

**SCOPES DATA (replica=1)**

## 63.93 TB

**rses**

Used Storage per RSE

SARA-DCACHE 2021-02-25T12:10:24.000Z	23.83 TB
EULAKE1 2021-02-25T12:10:23.000Z	5.28 TB
DESY-DCACHE 2021-02-25T12:10:23.000Z	7.88 TB
PIE-DCACHE 2021-02-25T12:10:24.000Z	2.94 TB
IN2P3-DCACHE 2021-02-25T12:10:24.000Z	517.23 GB
PIE-INJECT 2021-02-25T12:10:24.000Z	130.80 GB
INFN-NDPM 2021-02-25T12:10:23.000Z	6.86 GB
GSI-ROOT 2021-02-25T12:10:23.000Z	17.39 GB
ALPAMED-OPM 2021-02-25T12:10:22.000Z	12.95 GB
FAIR-ROOT 2021-02-25T12:10:23.000Z	26.67 GB
CNAF-STORM 2021-02-25T12:10:23.000Z	8.77 GB
INFN-NA-OPM-FED 2021-02-25T12:10:23.000Z	4.89 GB
LAPP-DCACHE 2021-02-25T12:10:23.000Z	4.97 GB
LAPP-WEBDAV 2021-02-25T12:10:23.000Z	4.91 GB
INFN-NDPM-FED 2021-02-25T12:10:23.000Z	4.90 GB
LAPP-OPM 2021-02-25T12:10:23.000Z	1.23 GB
CNAF_OPM_TEMP 2021-02-25T12:10:23.000Z	765.48 MB
ODS-A-PIE 2021-02-25T12:10:24.000Z	19.00 B

Files per RSE

DESY-DCACHE 2021-02-25T12:10:23.000Z	3,139 M
EULAKE1 2021-02-25T12:10:23.000Z	501 K
SARA-DCACHE 2021-02-25T12:10:24.000Z	119 K
PIE-DCACHE 2021-02-25T12:10:24.000Z	48.9 K
IN2P3-DCACHE 2021-02-25T12:10:23.000Z	22.5 K
PIE-INJECT 2021-02-25T12:10:24.000Z	617
LAPP-OPM 2021-02-25T12:10:23.000Z	336
ALPAMED-OPM 2021-02-25T12:10:22.000Z	192
GSI-ROOT 2021-02-25T12:10:23.000Z	110
LAPP-DCACHE 2021-02-25T12:10:23.000Z	108
INFN-NA-OPM 2021-02-25T12:10:23.000Z	107
FAIR-ROOT 2021-02-25T12:10:23.000Z	180
LAPP-WEBDAV 2021-02-25T12:10:23.000Z	110
INFN-NDPM-FED 2021-02-25T12:10:23.000Z	99
INFN-ROMA1 2021-02-25T12:10:23.000Z	98
CNAF-STORM 2021-02-25T12:10:22.000Z	95
CNAF_OPM_TEMP 2021-02-25T12:10:23.000Z	11
ODS-A-PIE 2021-02-25T12:10:24.000Z	1



# Conclusions & Future work

- Testing tools for transfers
  - GFAL
  - FTS
  - Rucio
  
- Unified Monitoring
  - Views on each component of the transfer stack
  - Network metrics (connected to specific transfers) included → some more accurate than others (Rucio > FTS)
  - Lots of consolidated dashboards for reference
  
- Future Work
  - Identify which tools/methodologies will be useful for the WLCG Data Challenges → modifications might be needed
  - Code/functionality enhancements and better error handling
  - Effort to make the tools more generic



# References

- FTS, <https://fts.web.cern.ch/fts/>
- GFAL, <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>



Thank you!  
Questions?

