

DataLake Status and Plan

Riccardo Di Maria

CERN

September 9th, 2020 - WP2 Fortnightly Meeting, CERN

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 Monitoring and Information System (1/2) perfS NAR

③FTS

🌀 Grafana 🖻



Efficiency Matrix										
Source \ Destination	ccdcalitest10.in2p3.fr	dcache-door-doma01.desy.de	dcbxwp2dlds1.gsi.de	door05.pic.es	eoseulakehttp.cem.ch	lapp-esc02.in2p3.fr	lapp-testse01.in2p3.fr	lapp-xdc01.in2p3.fr	webdav.grid.sara.nl	xfer.cr.cnaf.infn.it
ccdcalitest10.in2p3.fr	*	100.00%	100.00%	100.00%	-	÷	100.00%	2	100.00%	0%
dcache-door-doma01.desy.de	100.00%		100.00%	100.00%	ż	ż	100.00%		100.00%	0%
dcbxwp2dlds1.gsi.de	100.00%	100.00%	-	100.00%	4		100.00%	1	100.00%	×
door05.pic.es	100.00%	100.00%	100.00%	100.00%	0%		33.33%	6	100.00%	0%
eoseulakehttp.cern.ch		-	-	-	-	-	0%	0%	-	
lapp-testse01.in2p3.fr	100.00%	100.00%	100.00%	100.00%	ā.	ž.	с. – В	6	100.00%	0%
lapp-xdc01.in2p3.fr		2	-	-	2	0%	0%	1	-	
webdav.grid.sara.nl	100.00%	100.00%	100.00%	100.00%	÷.	ż	100.00%	<i>5</i>	2	0%



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

Monitoring and Information System (2/2) ESCAPE



filter by Experiment site	filter by Storage Unit	ACTIVE	
Experiment site	Storage Unit	State	
CNAF-STORM-ES	CNAF-STORM	ACTIVE	
CNAF_CMS_TEMP	CNAF_CMS_TEMP	ACTIVE	
DESY-DCACHE	DESY-DCACHE	ACTIVE	
EULAKE-1	EULAKE-1	ACTIVE	
EULAKE-2	EULAKE-2	ACTIVE	
N2P3-CC-DCACHE	IN2P3-CC-DCACHE	ACTIVE	
NFN-NAPOLI-DPM	INFN-NAPOLI-DPM	ACTIVE	
APP-DCACHE	LAPP-DCACHE	ACTIVE	
APP-DPM	LAPP-DPM	ACTIVE	
PIC-DCACHE	PIC-DCACHE	ACTIVE	
PIC-DCACHE	PIC-DCACHE	ACTIVE	
SARA-DCACHE	SARA-DCACHE	ACTIVE	

DOMA RucioStorageElement PIC-DCACHE

General Information RSE Name PIC-DCACHE Storage Unit PIC-DCACHE Last modification date 2019-11-20 16:51:25.774627 State Object state ACTIVE State comment Attributes Space Usage URL Deterministic True Volatile False LFN to PFN Algorithm hash Credentials

RSE Type	
lation to FTS	https://fts3-pilot.cern.ch:8446

Re

Edit

DELETE LAN	DELETE WAN	READ LAN	READ WAN	Third party copy	WRITE LAN	WRITE WAN
Priority	Rucio Priority	Endpoi	int	Protocol	Activity	Edit protoco
1	2	door05	.pic.es:8452	davs	delete_lan	C
2	2	xrootd.	pic.es:1094	root	delete_lan	G

Re-arrange protocols Use new Protocol

Create new Protocol



4

Riccardo.Di.Maria@cern.ch

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

ESCAPE Caching Status and WP2-WP5 Cross-Activities

- <u>XCache@IN2P3</u> + updates
- INFN activity (report needed in following WP2 Fortnightly Meetings)
- GSI Activity <u>Disk-Caching-On-The-Fly</u>
- <u>XCache@CERN</u>

- <u>Scalable JupyterHub deployment on bare-metal Kubernetes</u> by Rohini Joshi (SKAO) presented at <u>WP5 Tech Talk</u>
- <u>Rucio/JupyterLab Integration Project</u> by Muhammad Aditya Hilmy (GSOC Student) presented at <u>August WP2 Fortnightly Meeting</u>

5



ESCAPE On-Going and To-Do-ASAP

- Stable infrastructure as a primary goal → Rucio+k8s upgrade
- Notices (breaking changes, scheduled downtime, etc.) and site status → DepOps Team
- Improve tests (incl. HammerCloud) and monitoring
- Transfer matrix green as main goal → all ESCAPE partners
- Collaboration for a second/test rucio instance (DB at CERN for the time being)
- Enable or Develop more features, e.g. Rucio multi-VO, tokens, etc.







Funded by the



Riccardo.Di.Maria@cern.ch

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



- <u>RocketChat Channel DataLake Status and Support</u>
- <u>Rucio Documentation</u>
- WP2 Wiki DataLake Status
- ESCAPE AAI Documentation
- FTS Monitoring

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



8



Riccardo.Di.Maria@cern.ch



9



- Heterogeneous federated storage and operations model
- Orchestrated by Rucio
- Using FTS for storage to storage transfers
- Caching and Network optimization solutions are investigated in parallel
- Hiding complexity and providing transparent access to data





ESCAPE ESCAPE Data Infrastructure for Open Science (DIOS)

- Data Lake Infrastructure and Federation Services Xavier Espinal/Riccardo Di Maria (CERN)
- Data Lake Orchestration Service Patrick Fuhrmann/Paul Millar (DESY)
- Integration with Compute Services Yan Grange (ASTRON-NWO-I)
- Networking Rosie Bolton (SKAO)
- Authentication and Authorization Andrea Ceccanti (INFN)

Xavier Espinal (CERN) as WP leader and Rosie Bolton (SKAO) as deputy





- Implementing Science Analysis Platforms for EOSC researchers to stage data collections, analyse them, access ESFRIs' software tools, bring their own custom workflows.
- Contributing to the EOSC global resources federation through a Data-Lake concept implementation to manage extremely large data volumes at the multi-Exabyte level.
- Supporting "scientific software" as a major component of ESFRI data to be preserved and exposed in EOSC through dedicated catalogues.
- Implementing a community foundation approach for continuous software shared development and training new generation researchers.
- Extending the Virtual Observatory standards and methods according to FAIR principles to a larger scientific context; demonstrating EOSC capacity to include existing frameworks.
- Further involving SMEs and society in knowledge discovery.

