



Caching Activities for the DataLake within WP2 - DIOS

Riccardo Di Maria
CERN

February 26-27th, 2020 - H2020 ESCAPE Progress Meeting, Royal Library of Belgium, Brussels



Disclaimer: *The on-going caching investigation is based on XCache technology.*



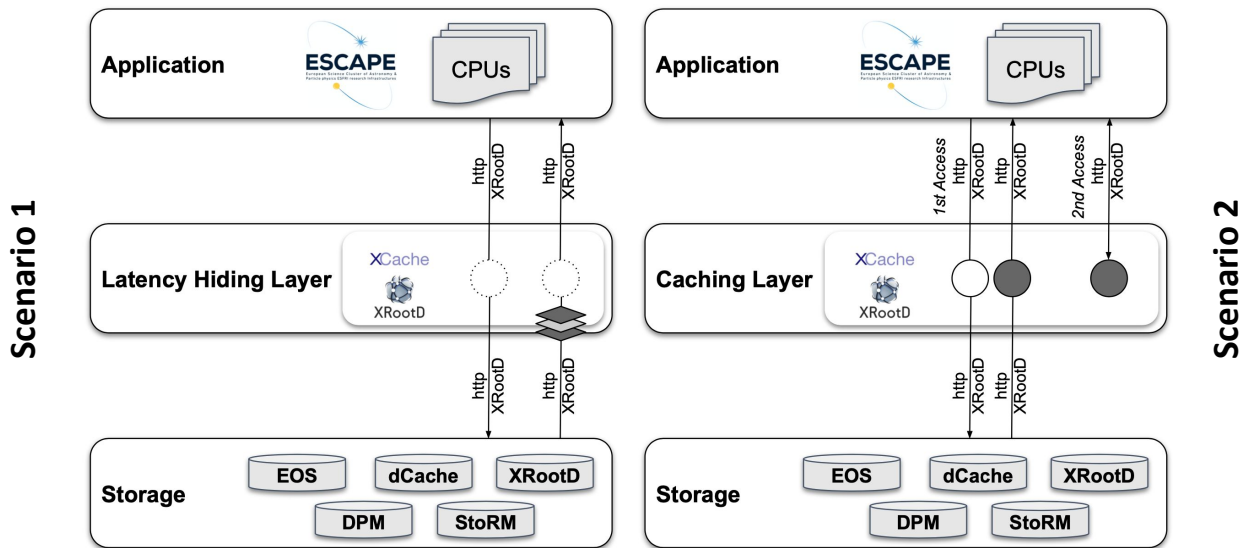
Overview

- What is XCache?
- XCache @ CERN
- MockData
- AuthN/Z
- On-Going Activities
- Next Steps



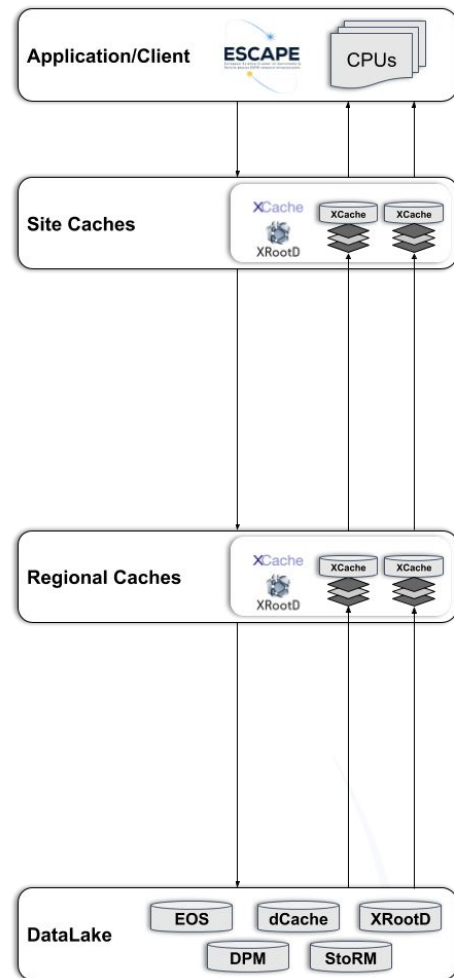
What is XCache?

- Deploy a caching layer flexible enough to serve all ESCAPE experiments.
- XCache natively based on [XRootD protocol](#) and support HTTP.
 - Caching and prefetching full files or only blocks already requested.



XCache @ CERN

- The CERN XCache (Disk Caching Proxy cluster) has two layers/levels.
 - A Site Cache service towards which the client requests are sent.
 - A Regional Cache service towards which the Site Cache requests are sent.
 - Regional Caches points towards the DataLake[/Any] storage.



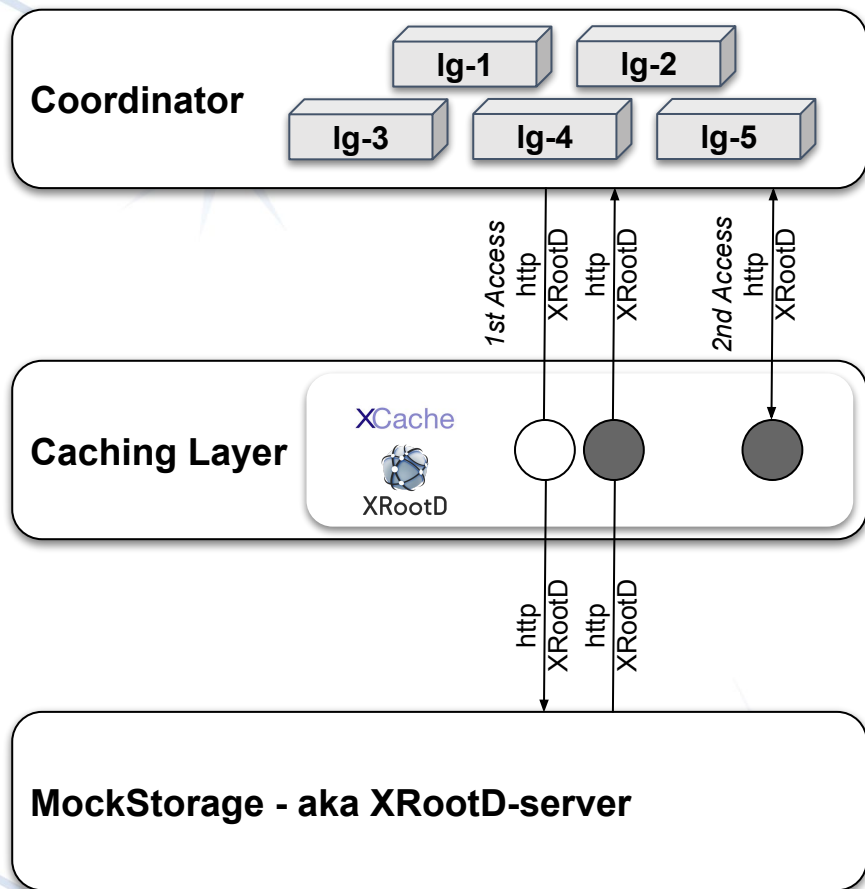
XCache @ CERN

- Each level has one redirector and two caches.
 - 16 VCPUs, 29.3 GB RAM, 160 GB disk (local).
 - 2.5 TB disk (120MB/s - 500 IO ops r/w) for the Site Caches.
 - 1.5 TB disk (120MB/s - 500 IO ops r/w) for the Regional Caches.
 - **XCache/Direct Mode Proxies** - for Site Caches.
`xrdcp -f -v xroot://escape-cache.cern.ch/escape/file.root /dev/null`
 - **XCache/Forwarding/Combination Mode Proxies** - for Regional Caches.
`xrdcp -f -v xroot://escape-cache.cern.ch//xroot://datalake.cern.ch//escape/file.root /dev/null`



XCache & MockData

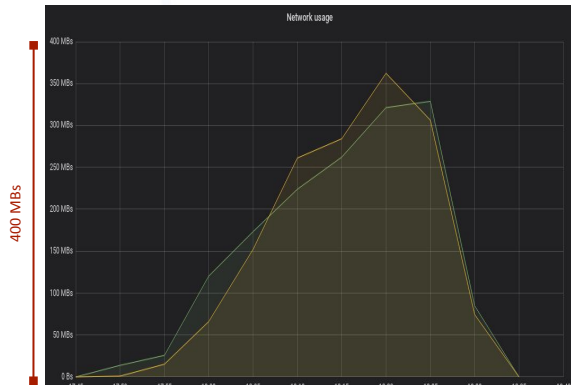
- The goal is to acquire knowledge on the behaviour of an XCache using MockData (tool developed by David Smith, CERN).
 - XRootD-server, load-generator(s), and coordinator (all Puppet-managed) with 8 VCPUs, 14.6 GB RAM, 80 GB disk.
- CERN Puppet setup provides a basic monitoring for the machines managed by it.
- A [python script](#) collect and push useful XCache data to an ElasticSearch cluster by looking at CINFO files (*metadata*).
- Visualisation at monit-kibana/grafana.cern.ch.



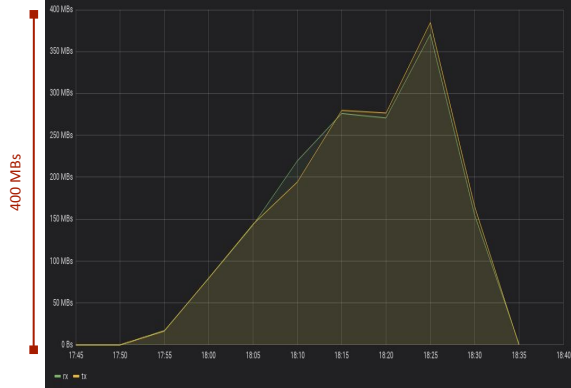
MockData

1st Access

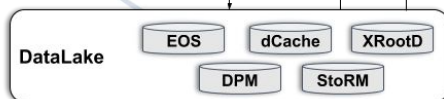
Site Caches



Regional Caches

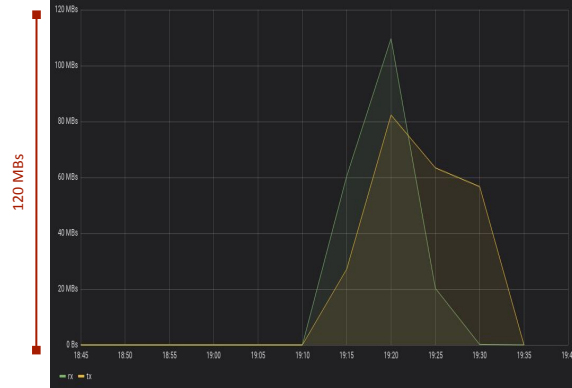
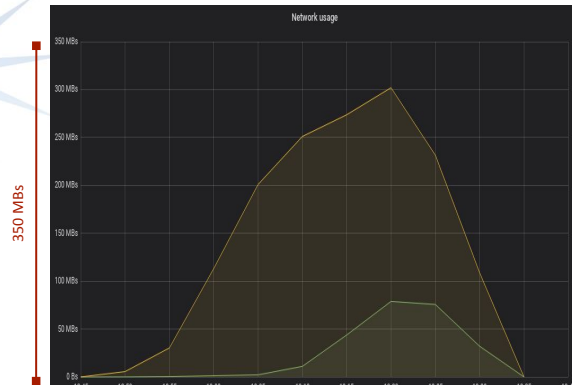


55 min



Green (rx): receives to the caches (misses)
Yellow (tx): transmits from the caches (hits)

2nd Access



55 min



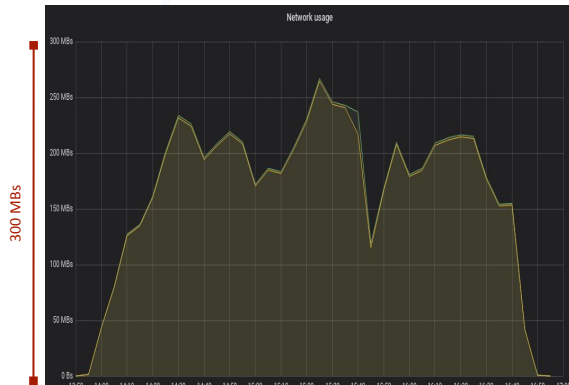
LOFAR (Real) Data



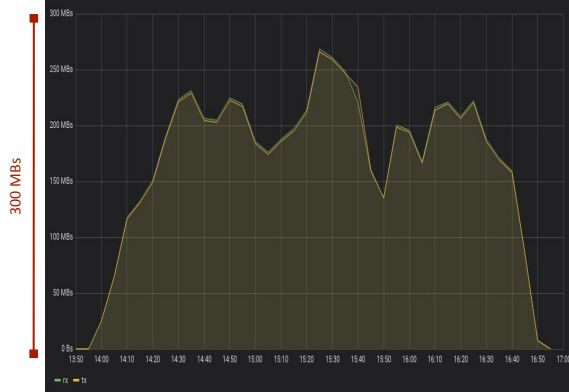
Green (rx): receives to the caches (misses)
Yellow (tx): transmits from the caches (hits)

1st Access

Site Caches

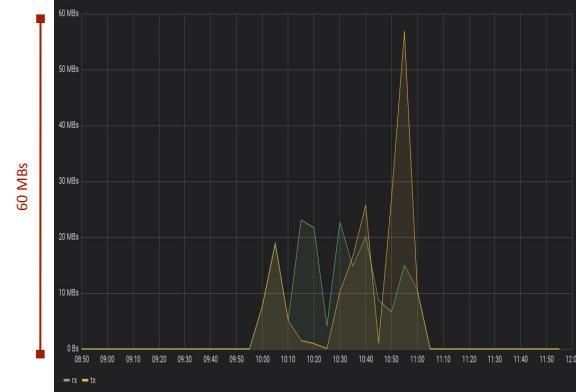
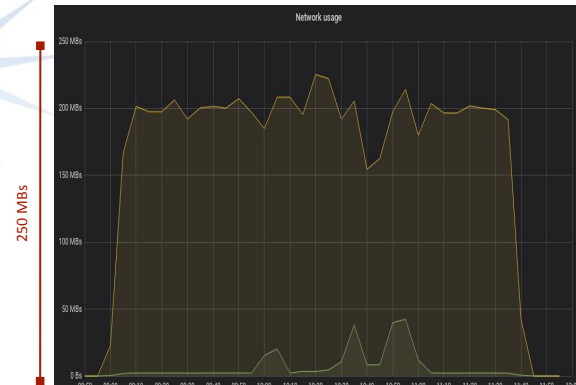


Regional Caches

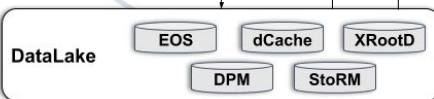


190 min

2nd Access



190 min



Authentication & Authorisation

- The client-cache AuthN/Z is via X.509 certificate and GSI protocol.
 - In each cache, host certificate, host key, and trusted CA certificates are present.
 - AuthN is denied if a valid certificate is not provided by the client.
 - The client is AuthZ to access files, both cached and remote, only in the related-organisation path (necessary for embargoed data - using VOMS extension).
 - A path can be opened (r, w, a, etc.) to all users if necessary, e.g. /mockdata.
- The cache-server AuthN/Z is via X.509 certificate, using ESCAPE VOMS extension.
 - A robot certificate is present in all caches and renews itself via cron.
 - In this case, the AuthZ is managed by the remote server.



ESCAPE Community Organisation

- The Site-Regional cache AuthN/Z is via X.509 certificate and GSI protocol.
 - A grid-map file can be used.
 - XCache robot certificate DN is mapped to *ewp2c01* 'user'.
 - User *ewp2c01* is granted all privileges to all files.
 - Organisation=escape, Role=xcache is used to grant all privileges to all caches.
- Restriction and extension of the client's privileges.
 - Groups can be used to manages user's privileges, e.g. of SuperPippo user.
 - ESCAPE SuperPippo has right for /mockdata and /escape.
 - ESCAPE-CMS SuperPippo could see **also** /cms.
 - CMS SuperPippo could see **only** /cms because /escape contains embargoed data.

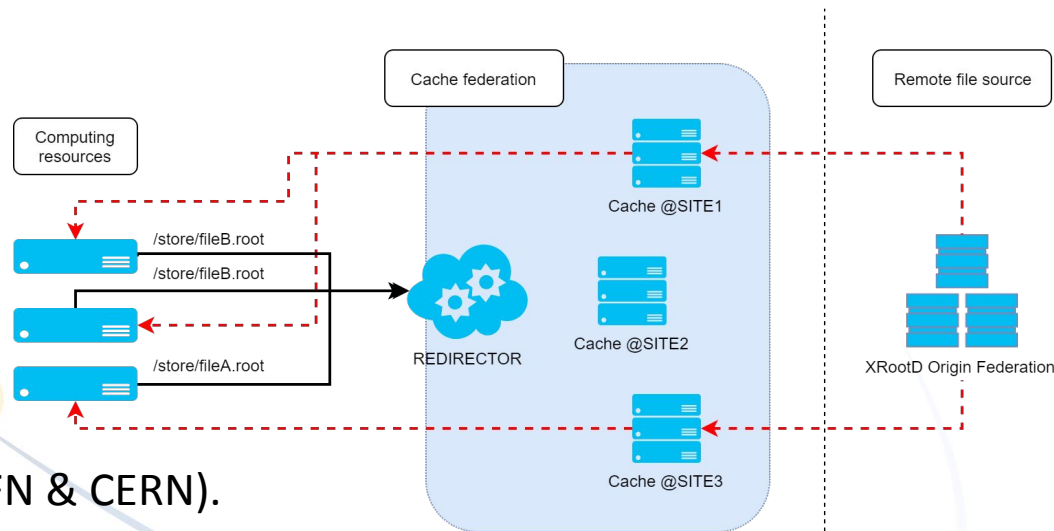


On-Going Activities

- Finalising/implementing dedicated/common monitoring.
- Preparing real analysis workflows from HL-LHC community.
- XCache stress test using MockData.
- Analysing results on the XCache behaviour for overloading.
- Enable the HTTP protocol.
- Simple HTCondor batch jobs requesting files through caches.
- Full integration with DataLake.
 - Investigate horizontal scaling with several stages, load balancing, etc.



- INFN is in the process of integrating the distributed XCache setup into the ESCAPE-DataLake testbed at CNAF.
 - A caching layer to access data in the DataLake with XRootD or HTTP.
 - **AuthN/Z: both X509 and full token flow will be integrated.**
- Importing CMS data into the DataLake and performing tests with:
 - real skimming workflows on medium input format (MINIAOD);
 - real end-to-end analysis workflows on small input format (NANOAOD).
- Next step will be the integration into the multi-lake scenario (testbed @ INFN & CERN).



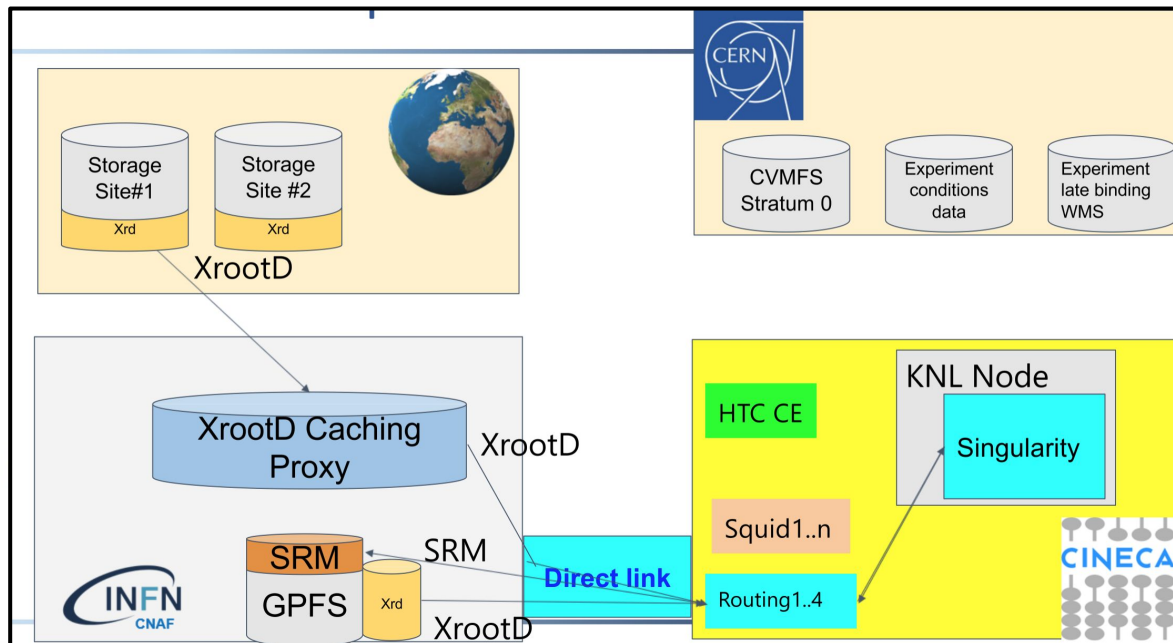
Compute Resources Integration: The Plan


- INFN will test the DataLake testbed following three scenarios:

- regular WLCG Tier resources;
- HPC;
- cloud provisioned resources (**see next slide**).

- The INFN strategy is to make integration of heterogeneous compute resources fully transparent to the end user.

Example of HPC integration @ CINECA





INDIGO - DataCloud

Welcome to **dodas**

Sign in with your dodas credentials

[Sign in](#)

[Forgot your password?](#)

Or sign in with

[Google](#)

[eduGAIN](#)

[egi](#)

Not a member?

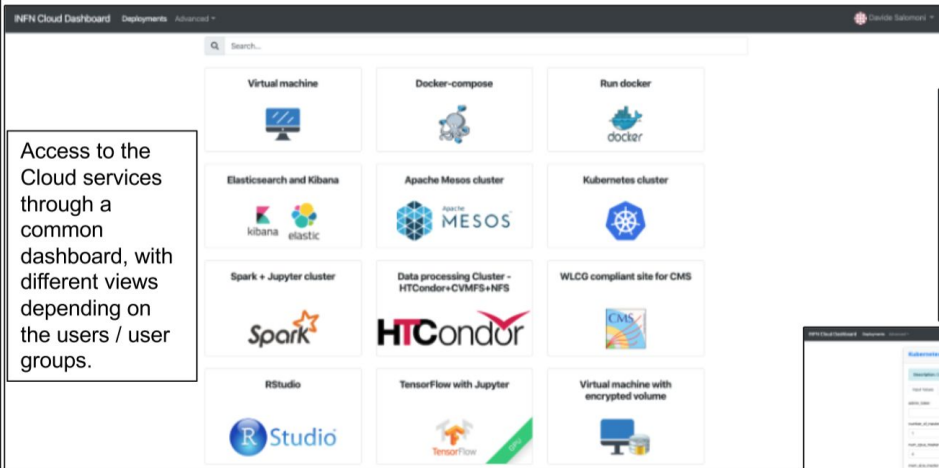
[Register a new account](#)

[Privacy policy](#)

Authentication *can* be enabled for::

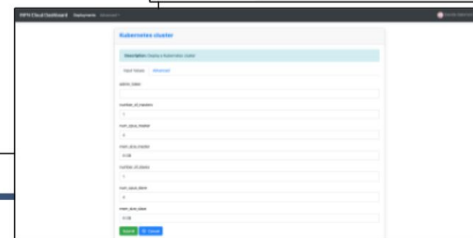
- Local username/password
- Google accounts
- EduGAIN (e.g. University, research centers, etc.)
- Other OIDC providers

Access to the Cloud services through a common dashboard, with different views depending on the users / user groups.



Transparent, multi-site federation for users of Cloud resources belonging to INFN and/or to other Cloud providers (private or public)

Composed, high-level services easily customizable a configurable directly by users



Courtesy of D. Salomoni



Next Steps

- Real data and analysis workflows from ESCAPE community.
 - Measure data access w/ and w/o XCache, evaluating pros/cons.
- XCache stress test using HammerCloud (stability, reliability, etc.).
- Further integration with the DataLake orchestrator: RUCIO.

Progress towards the implementation of the content delivery service.

- Investigate federated European XCache.
- Benchmark multiple-layers caching.

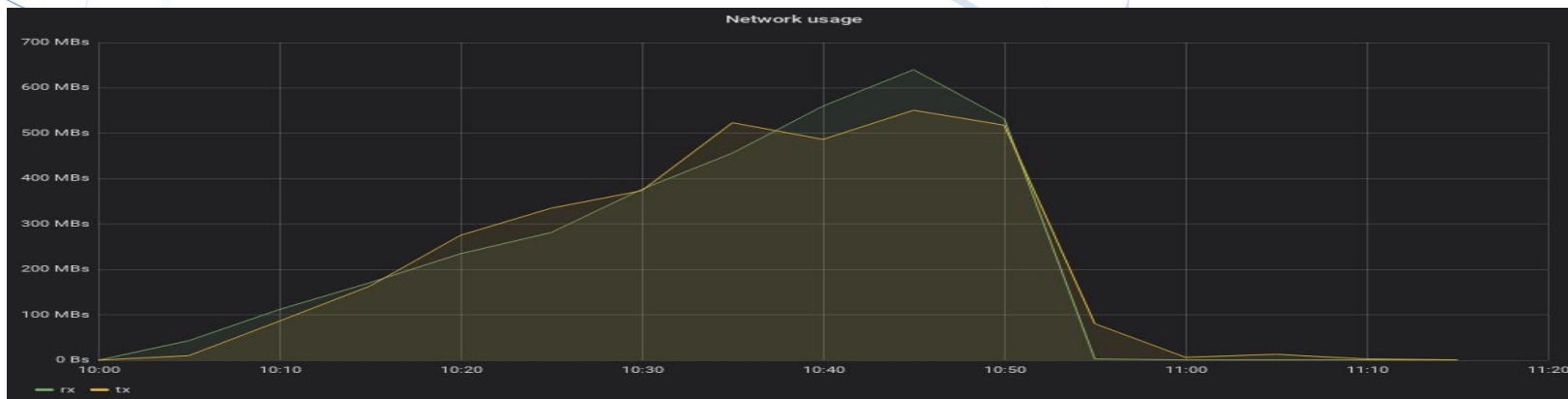


Backup

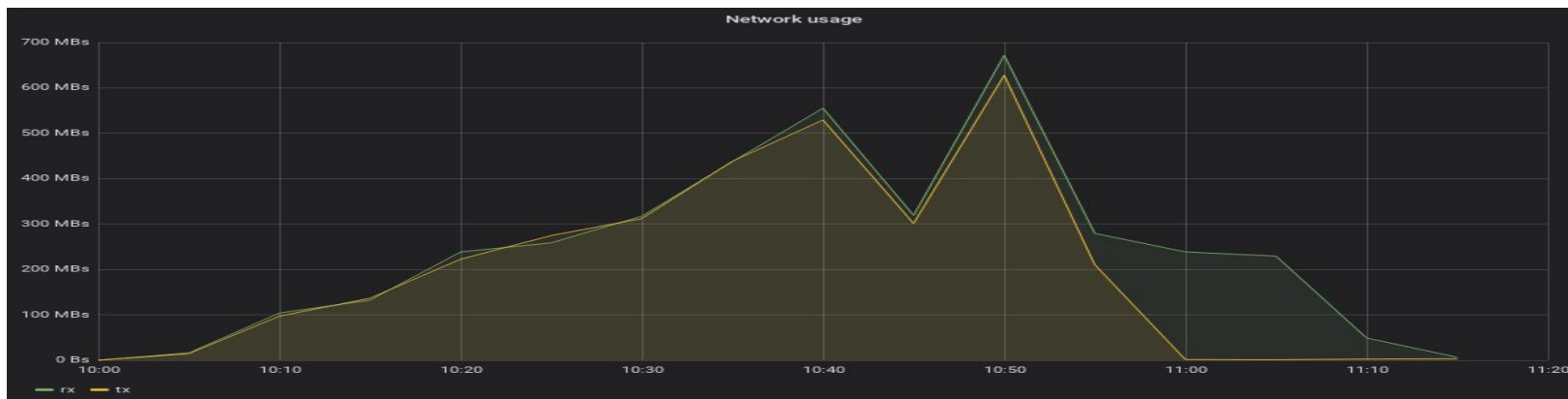


MockData 1st Access (hot cache)

Site Caches

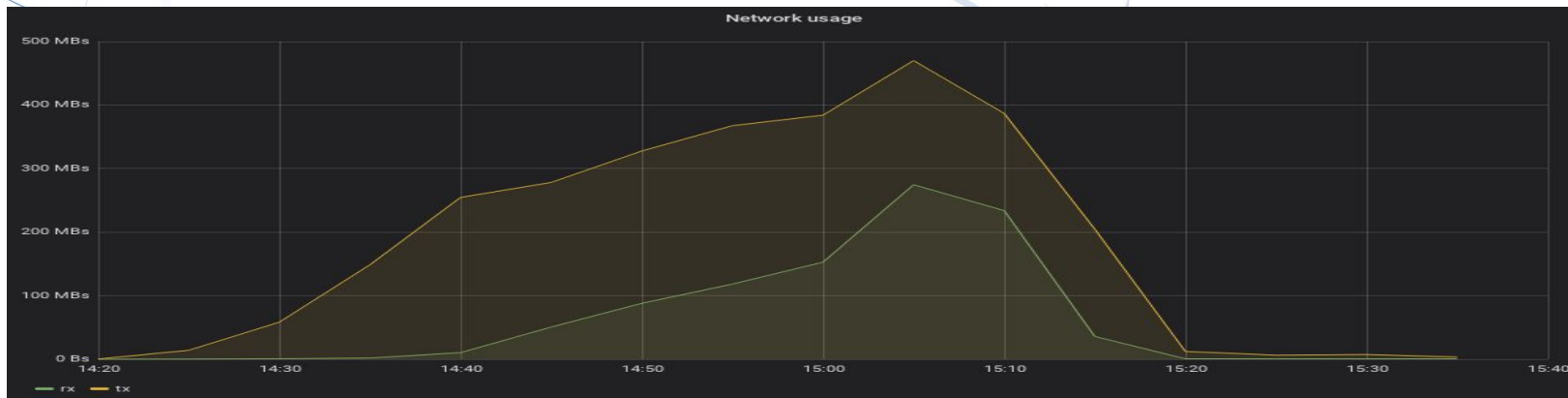


Regional Caches

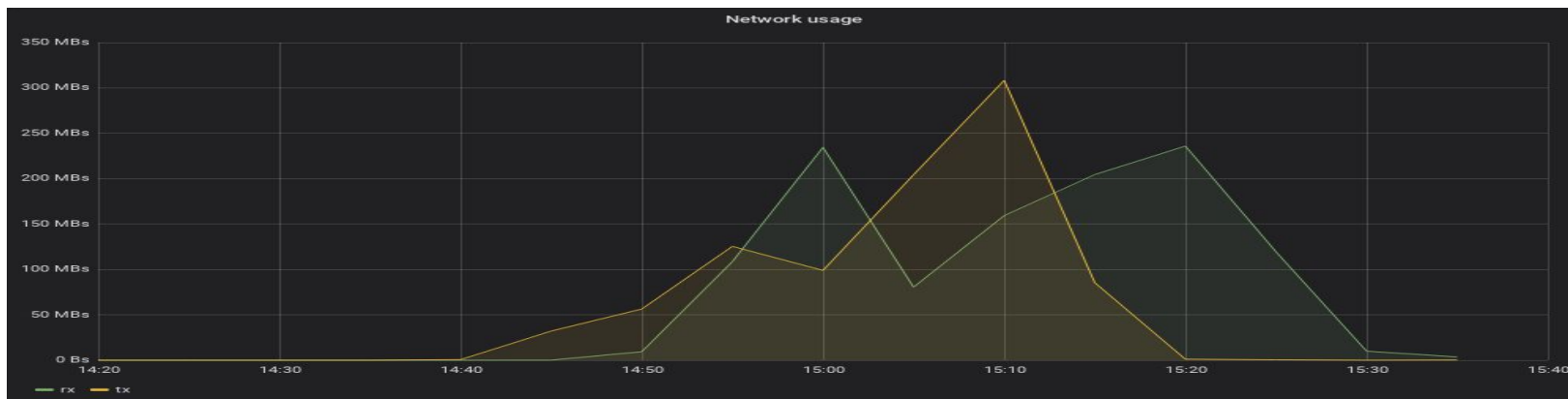


MockData 2nd Access

Site Caches



Regional Caches



ESCAPE Data Infrastructure for Open Science (DIOS)

- Data Lake Infrastructure and Federation Services - Xavier Espinal, CERN
- Data Lake Orchestration Service - Patrick Fuhrmann, DESY
- Integration with Compute Services - Yan Grange, ASTRON-NWO
- Networking - Rosie Bolton, SKAO
- Authentication and Authorization - Andrea Ceccanti, INFN

Simone Campana, CERN as WP leader and Rosie Bolton, SKAO as deputy



ESCAPE Goals

- 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.

