CERN XCache Activities for the ESCAPE DataLake

Riccardo Di Maria

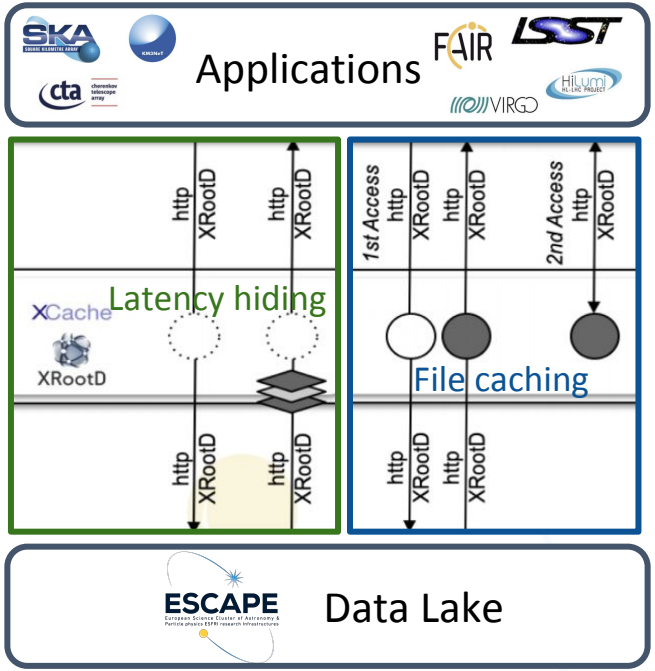
CERN

July 15th, 2020 - WP2 Fortnightly Meeting

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.

Content Delivery and Caching

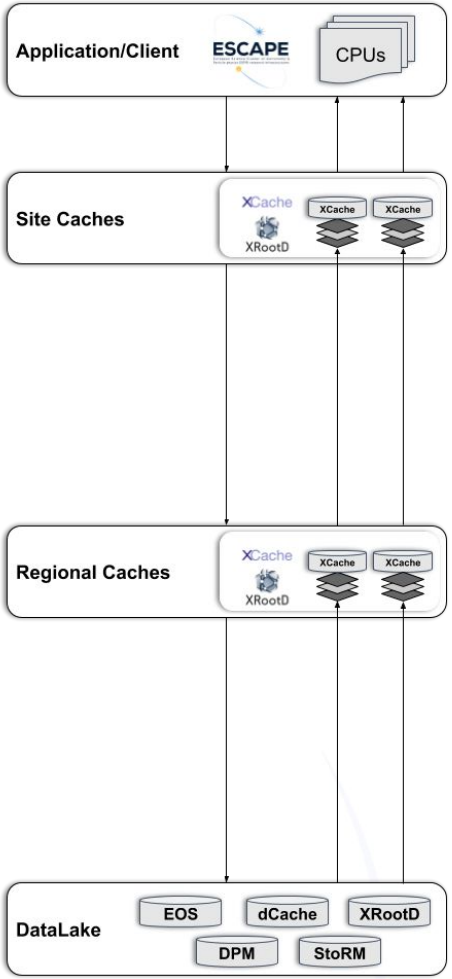
- Leveraging the know-how being acquired in DOMA/WLCG with XCache investigations.
- Effort made towards a vanilla installation (experiment-unbiased) caching service.
  - Installations at CERN, INFN/CNAF, and CC-IN2P3.
- Main use-cases and goal.
  - Latency hiding and file re-usability.
    - Benchmarking multi-caching layers between client and origin.
    - HTTP and Tokens awareness.
  - Facilitate ingress/egress with Commercial Clouds and HPC.
  - 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

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.



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

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
```


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

Recycle of XCache @ CERN

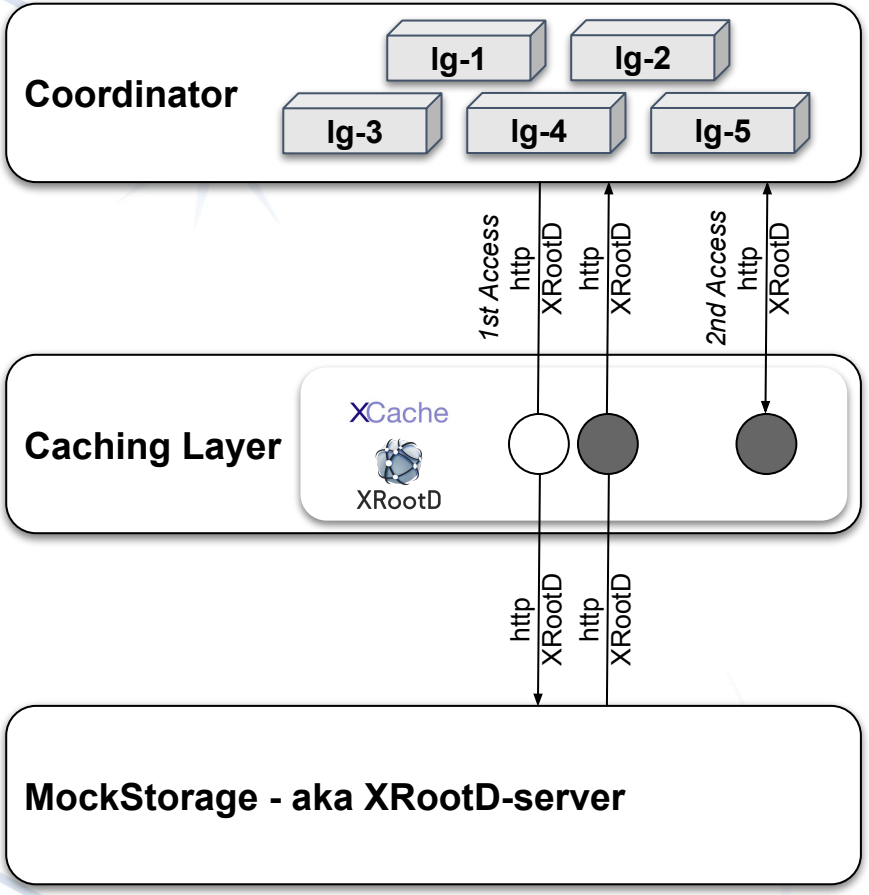
- The two layers/levels have been re-assigned to carry out different tests.
  - One level hosts XRootD HEAD, i.e. version > R5-rc5.
  - The other level has latest EPEL/OSG XRootD stable version.

- Tests consist in using xrootd and HTTP(s) protocols in various combination (4).
  - `xrdcp` for xroot://-xroot:// & xroot://-http(s)://.
  - `davix-get -P grid` for http(s)://-xroot:// & http(s)://-http(s)://.
  - `curl -L -H "Authorization: Bearer $TOKEN" -k -XGET` for http(s)://-xroot:// & http(s)://-http(s)://.
    - These tests have led to tickets in XRootD.

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

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.



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

Docker-isation of XCache @ CERN

- XRootD containers summarise one year of work, meaning that they are tailored to work straight away @ CERN.
- **This is temporary, as a generalisation of the images is on-going.**
- Especially if not standalone, they have to be used in a cluster (URL patterns hard-coded).
- However, with very few actions they are exportable worldwide.
- https://github.com/ESCAPE-WP2/containers

- https://wiki.escape2020.de/index.php/WP2_-_DIOS#3.4_Content_Delivery_and_Caching has been updated.

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

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.

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

ESCAPE Community Organisation

- The Site-Regional cache AuthN/Z is via X.509 certificate and GSI protocol.
  - A grid-map file can be used.
  - 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.
- Tokens are integrated and necessitate to be widely tested wrt XRootD and specific configurations.

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

Proposal for common ESCAPE images

- A GitHub repository has been created for ESCAPE WP2: https://github.com/ESCAPE-WP2 .
- Each XCache dev often has his own repo (GitHub/GitLab).
- `git fork` the ESCAPE WP2 GitHub repo.
- Integrate with existing personal repo.
- `git pull request` towards the ESCAPE WP2 GitHub repo.
- To facilitate the maintaining of the repo and the understandability of the loaded images, the README file must always be updated with a dedicated section guiding the user to easily replicate the environment, otherwise the PR will be rejected.
  - Already good work done by Diego and Paul in WP2 wiki and IN2P3 repo.
  - Request to integrate existing material in ESCAPE-WP2/containers/README using dedicated sections.

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

On-Going Activities

- ESCAPE Data Caching Initiative: Minutes .
  - Next meeting TBD - pool soon.

- Finalising/implementing common monitoring.
- Real analysis workflows from HL-LHC community are ready.
- XCache stress test using MockData - 1st round is done.
- Analysing results on the XCache behaviour for overloading.
- HTCondor batch jobs requesting files through caches.
- Full integration with DataLake.
  - Investigate horizontal scaling with several stages, load balancing, etc.

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

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.

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

Backup

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