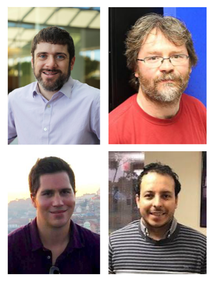
# XCache

Team: Edgar Fajardo[1], Diego Davila[1], Brian Bockelman[2], Frank Wuerthwein[1],
Institutions: [1]University of California San Diego, [2]Morgridge Institute for Research



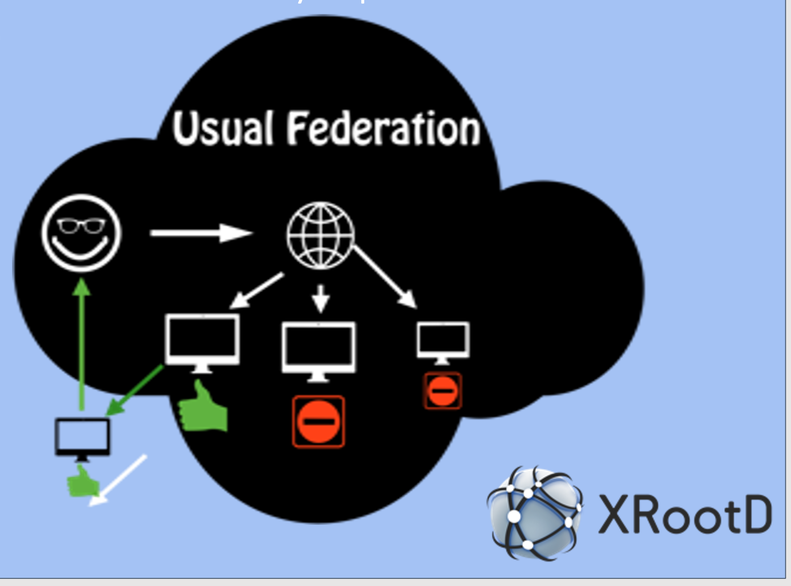
## XCache Basics

### Why to cache?

- When there is significant data reuse, caches may **reduce WAN bandwidth requirements**.
- Caches or buffers **hide latencies** of the network from the client. If a file is read in blocks the cache can prefetch the rest of the blocks before the application needs it.
- Reduce the I/O activity on the data federation. Caches can create additional copies of files, reducing pressure on any single storage service for highly reused files.
- Subfile caches provides **more efficient use of disk**. By only caching what is used, the disk is only used for storing files of the working set rather the entire file. For some datasets, this can save a large percentage of the disk space.

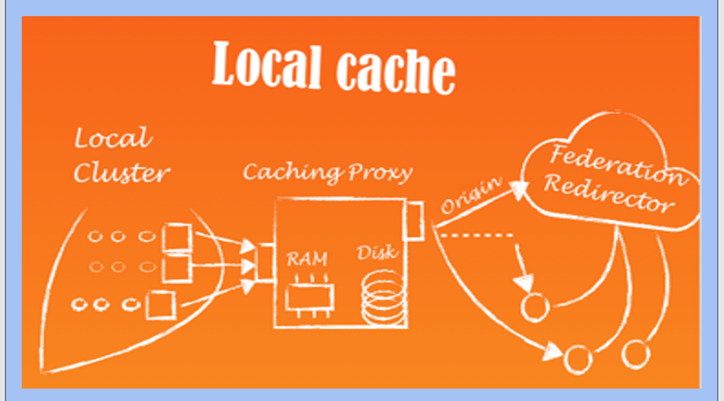
### XRootD Federation

An XrootD data federation provides a uniform namespace out of independently-administered storage services. XRootD federations are arranged in a tree hierarchy consisting of *redirectors* and *data servers*. Redirectors (non-leaf nodes) can broadcast data discovery requests to subscribed data servers.




### XCache in a Nutshell

- An XCache sits in between one (or more) local clusters and a data federation.
- Simple disk systems (RAID or JBOD) and distributed storage file systems can be used to store the cache files.
- When a cache miss occurs, the data is streamed from the data federation to the client and queued to be written into disk.



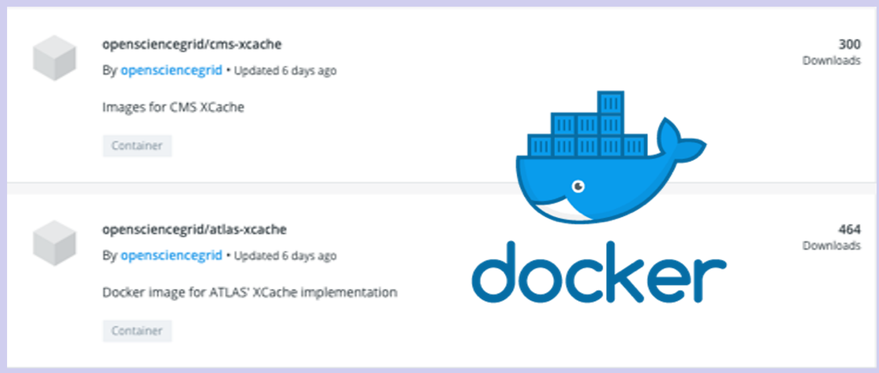

## Packaging and Containers

### Building an XCache Community

- XRootD developers release minor versions of the software several times a year.
- We organize weekly meetings to discuss caching issues among stakeholders from: ATLAS, CMS, XRootD Developers and OSG Software.
- OSG Software packages, tests, and releases new version of XRootD as soon as they are available.
- The OSG Software team distributes a generic XCache RPM along with ATLAS- and CMS-specific versions.

### Adding Docker Support

- To support a **DevOps model**, where newer versions of software are deployed (and rolled back) quickly and easily in production, the OSG-LHC team also supports container images for ATLAS XCache and CMS XCache.




### XCache Features

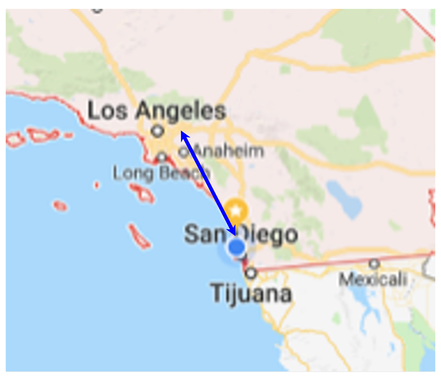
- XCache can cache only a subset of a the entire file, reducing **efficient disk usage**.
- Supports GSI and token-based authentication.
- Provides a **horizontally-scalable** caching system.
- **Automatic data integrity** scans are under development by IRIS-HEP.
- XCaches report to OSG **accounting** (GRACC) and information services.

## SoCal Experience

### SoCal Cache Motivation

Given their proximity, the Caltech and UCSD U.S. CMS sites are able to run a shared "SoCal" cache infrastructure:

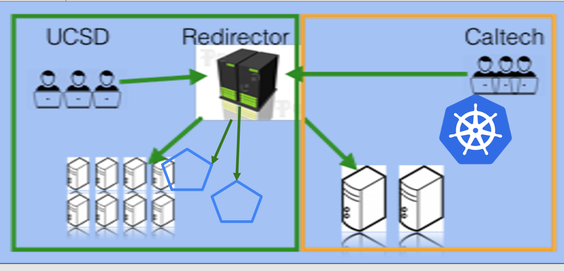
- Separated by **120 miles.**
- **100 Gbit/sec** between datacenters.
- **Less than 3ms network latency**.




### SoCal Setup

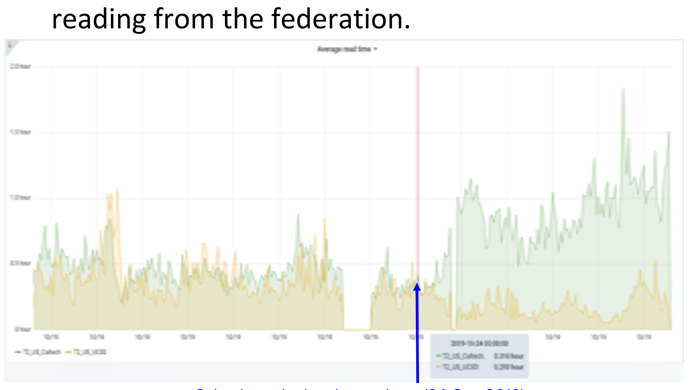
- SoCal setup is made out of 1.2PB of disk deployed among 13 nodes (2 in containers) between the two participating sites.
- Serves as a production service and as testbed for new features (file integrity).

| | UCSD | Caltech |
|---|---|---|
| **Nodes** | 11 (+1 JBOD) | 2* |
| **Disk Capacity per node** | 12 x 2TB = 24TB (+ 48 x 11TB) | 30 x 6TB (HGST Ultrastar 7K6000) |
| **Network Card per node** | 10 Gbps (+ 40 Gbps) | 40 Gbps |
| **Total Disk Capacity** | 264 TB (+ 528 TB) = **792 TB** | **360 TB*** |
| **TOTAL** | 792 TB + 360 TB* = **1,152 TB** | |




### Performance and Data Delivered

- By the end of October 2019, CMS made the switch to have jobs reading from the MINIAOD data tier running at Caltech to read from the federation (WAN) and local storage while UCSD jobs only used the XCache.
- The average read times for MINIAOD jobs reading from the cache was 5 times faster that same mix reading from the federation.




Caltech made the change here (24 Oct. 2019)

This project is supported by National Science Foundation under Cooperative Agreement OAC-1836650. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.