

StashCache

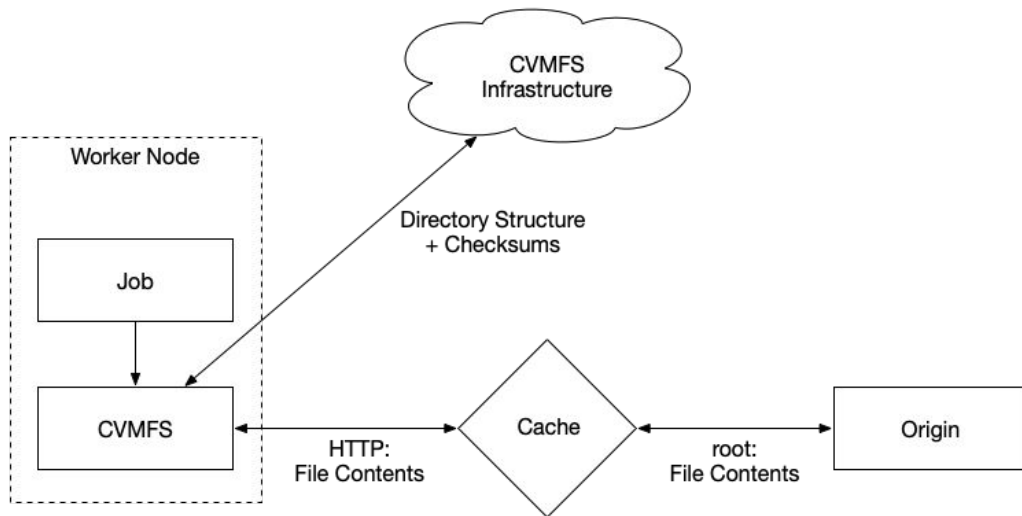
Derek Weitzel, Edgar Fajardo Hernandez

Differences of StashCache from XRootD

- CVMFS Namespace
- GeolP choice of which cache to pick
- Kubernetes deployment
- Caches in the network backbone
- Cache in Amsterdam to support LIGO/MIRGO operations at WLCG sites in EU
- Namespace monitoring

CVMFS

- Export the origin data into CVMFS repos
- CVMFS provides namespace data, while actual data is retrieved from caches
- Uses CVMFS's "External Data" feature
- A central service scans open origins for files, checksums, and publishes them in CVMFS



CVMFS

- Each Origin has a separate CVMFS repository
- An example of an open origin is “stash” hosted by the OSG Connect service
- The user can ‘ls’ through the namespace
- The filesystem is presented as posix (read-only) for all applications to read

```
$ ls /cvmfs/stash.osgstorage.org/user/dweitzel/public -lh
total 4G
drwxr-xr-x 5 cvmfs cvmfs 4.0K Apr 25 2017 blast
-rw-r--r-- 1 cvmfs cvmfs 769M Apr 24 2017 container.img
-rw-r--r-- 1 cvmfs cvmfs 801M Jul 9 20:09 ducks.MOV
-rw-r--r-- 1 cvmfs cvmfs 22M Jul 9 20:08 pdbaa_files.tar.gz
-rw-r--r-- 1 cvmfs cvmfs 22M Jul 9 20:08 pdbaa_files.tar.gz.1
-rw-r--r-- 1 cvmfs cvmfs 1006M Jul 9 20:09 teaching.MOV
-rw-r--r-- 1 cvmfs cvmfs 38M Jul 9 20:08 test_open_terminal.mov
-rw-r--r-- 1 cvmfs cvmfs 187M Feb 21 2018 xrootd.log.gz
```

GeoIP

- The two clients for StashCache, CVMFS and StashCP, use GeoIP to determine a nearby cache to utilize.
- Utilizes CVMFS's existing GeoIP services

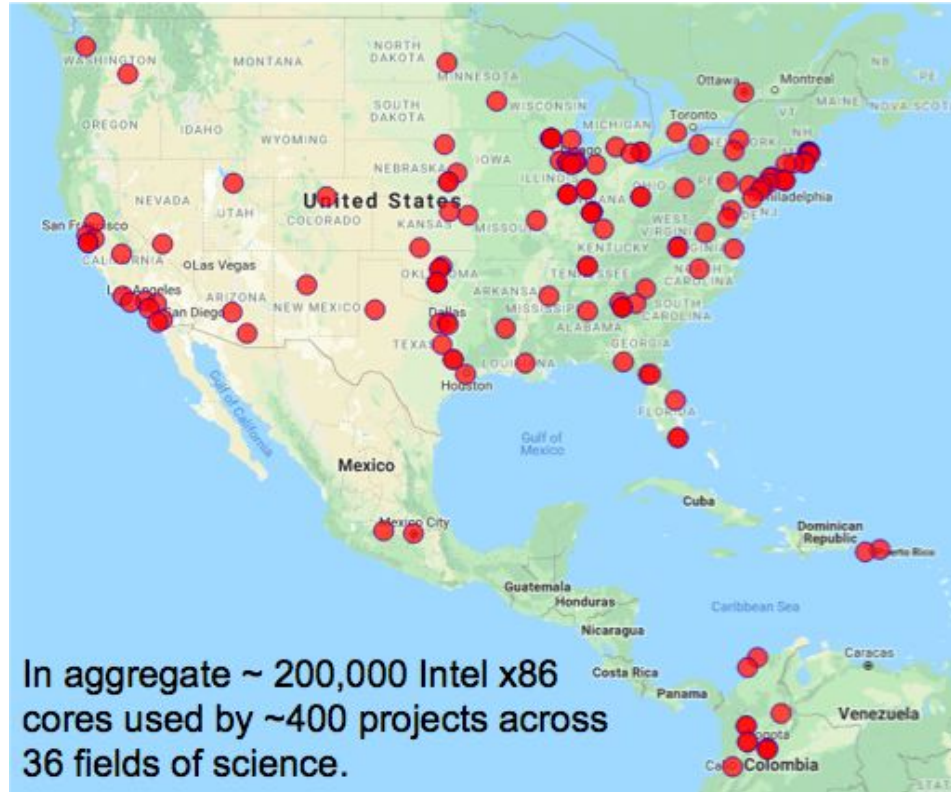


GeoIP

- GeoIP is imperfect
- Nearest != Highest bandwidth or fastest response to client
- But, since network topology is mostly unknown, good approximation



OSG Computing Resources



OSG Data Origins

FNAL: Fermilab based HEP Experiments

U.Chicago: General OSG Community

Caltech: Public LIGO Data Releases

UNL: Authenticated LIGO Data Releases

SDSC: Simons Foundation Planned



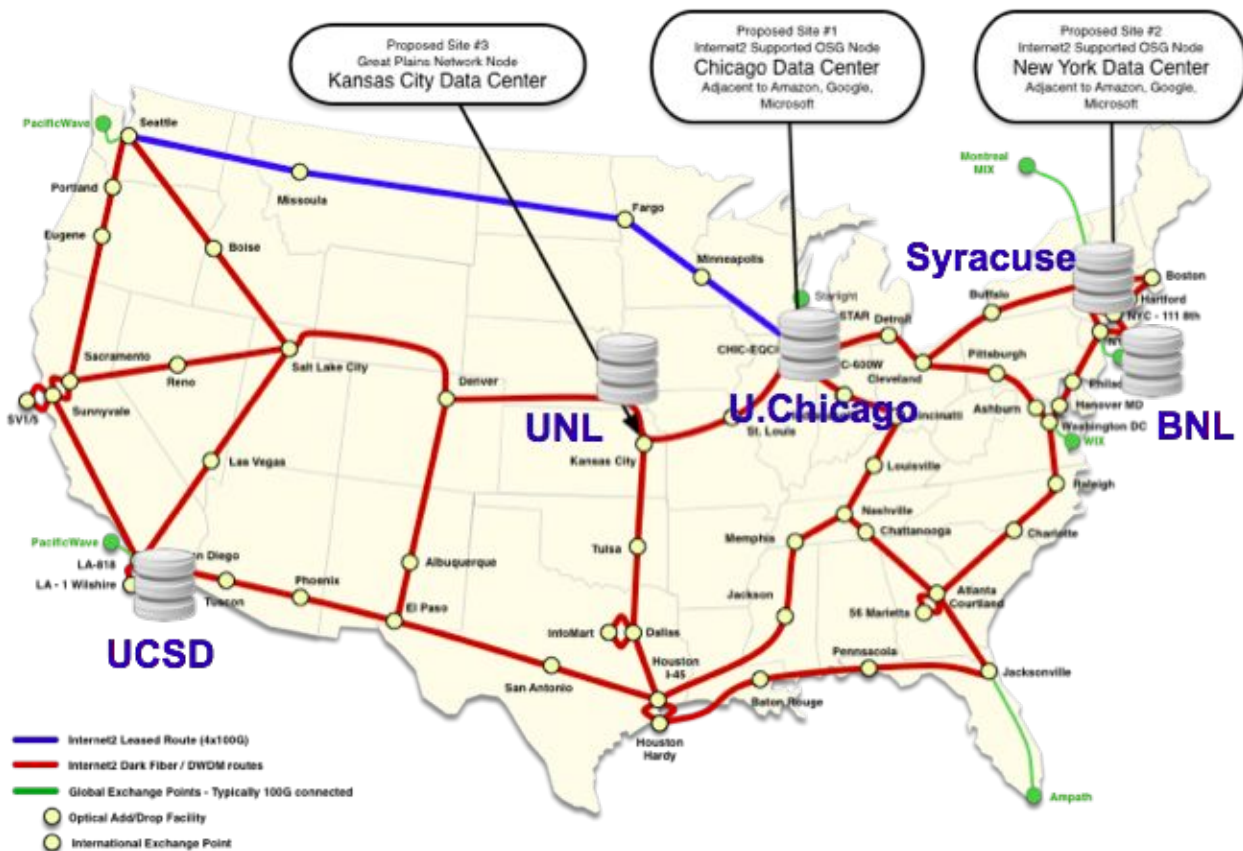
Network backbone Caches



In Collaboration with Internet 2. A pilot model on locating caches in the PoPS of Internet 2.

We are now in talks to place a similar one in GEANT at London.

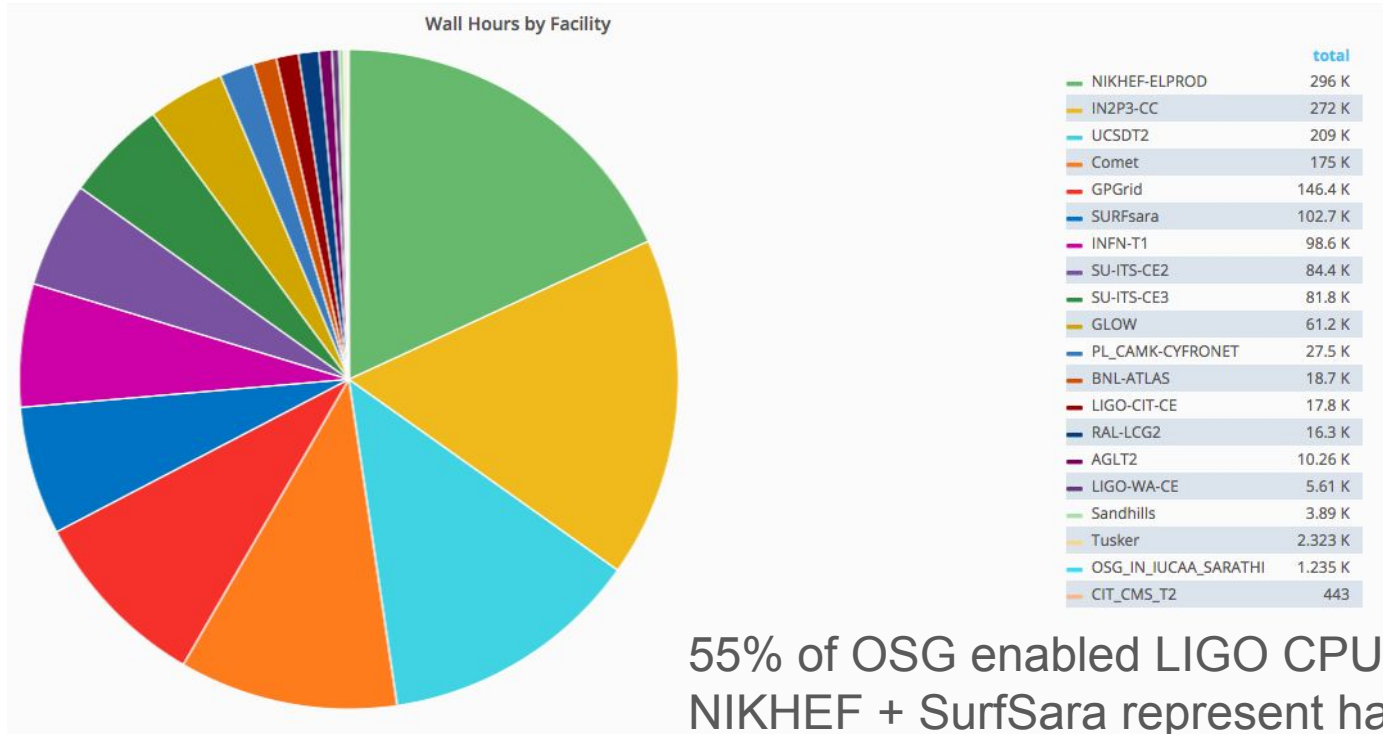
Endpoint Caches



Kubernetes Deployment

- We are leveraging an already existing PRP Kubernetes federation infrastructure maintained by SDSC.
- A shift in traditional grid deployment. The hardware and software responsibilities are split. All software (including the cache) run on docker containers (k8 pods) and it is maintained centrally. The local admins take care of hardware issues (Ex: disk).
- On every node there is a perfsonar pod. This helps deliver a quality of service since now network responsables (PRP) have full access.
- Our current Stashcache Containers can be found at <https://github.com/efajardo/prp-stashcache>

OSG enabled LIGO to seamlessly use VIRGO resources

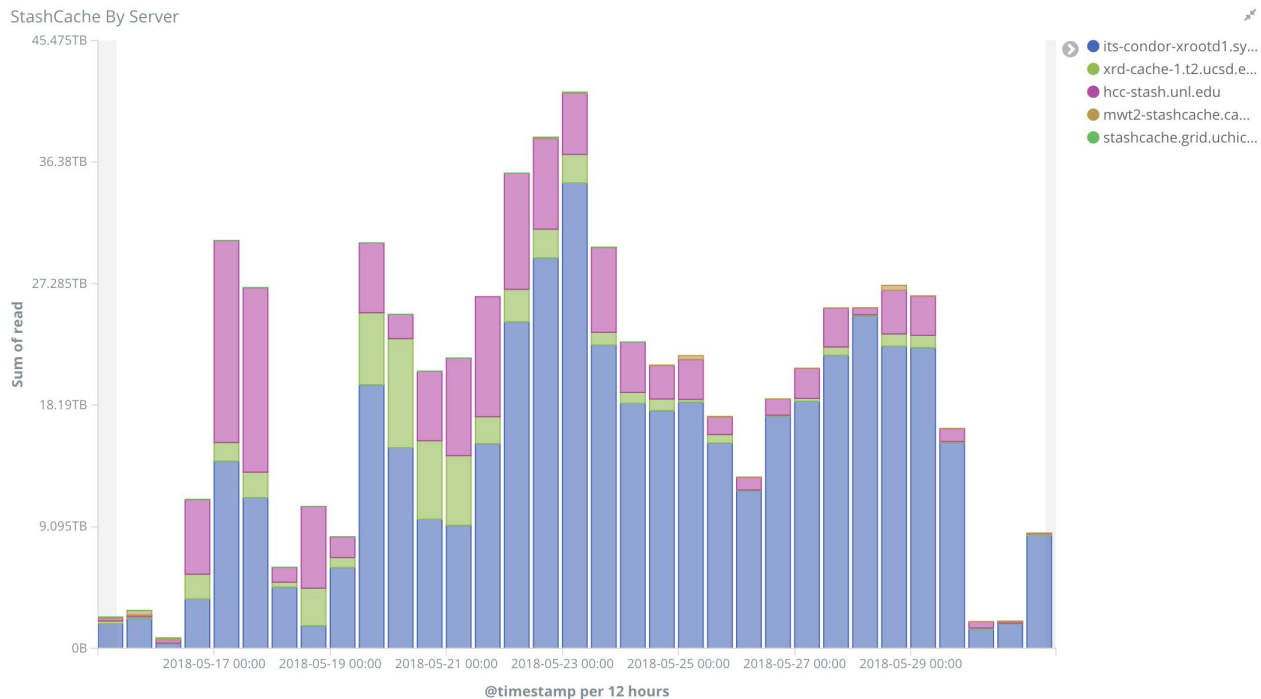


Cache in Amsterdam

- Given the physical location of the Virgo computing resources and the established partnership between PRP and University of Amsterdam it made sense to place a cache there.
- The cache is managed via k8s with same configuration as the I2 backbone ones.
- Right now it is on acceptance testing phase. We ran into some certificate problems that are now fixed.

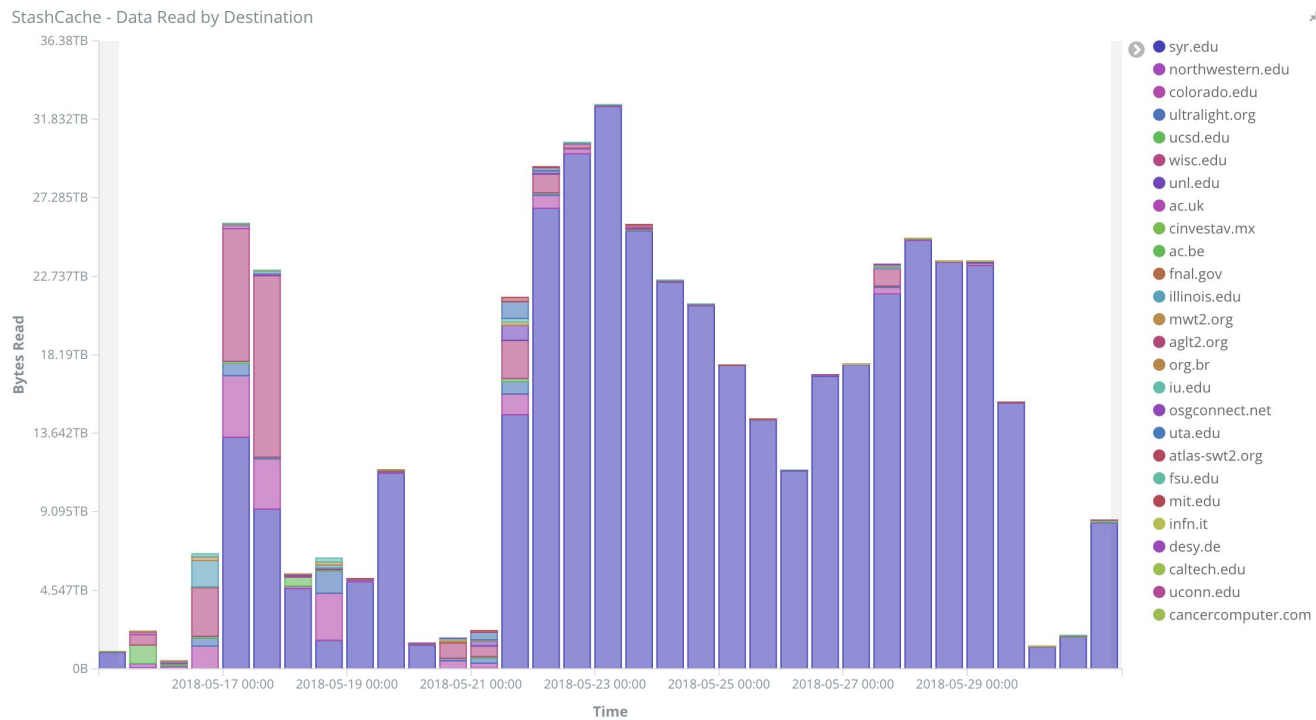
Namespace Monitoring

- Cache Monitoring



Namespace Monitoring

- Destinations



Future Projects

- Cardiff University is on the talks to buy hardware for a kubernetes managed cache for LIGO public and authenticated data.
- Install new origin at NCSA.
- New Caches in Origins and Caches in:
 - Wisconsin (local researchers)
 - Georgia Tech (ligo)

Links for more details

- OSG StashCache Overview and Installation Docs:
<https://opensciencegrid.org/docs/data/stashcache/overview/>
- Presentation with details of StashCache users: (50MB presentation):
<https://unl.box.com/s/l7ej12n6efse0emta8fg0zzvnf56zuq1>
- OSG Presentation on StashCache:
<https://indico.fnal.gov/event/15344/session/7/contribution/32/material/slides/0.pdf>