

A new UK StashCache at Edinburgh for DUNE

Wenlong Yuan

On behalf of DUNE Collaboration



Outline

- StashCache and CVMFS
 - DUNE utilizes CVMFS and CVMFS StashCache to distribute both its software stack and reference files for distributed computing workflows
- Motivations
 - DUNE HTC grid jobs were suffering from low CPU efficiency in the UK, during Dec-'21 to Jan-'22, due to slow access of data from a StashCache instance at Cardiff
- Implemented a new UK StashCache instance at Edinburgh
 - deployment process, from the perspective of the DUNE use case
- StashCache monitoring, Edinburgh site status/performance

StashCache and CVMFS

- DUNE utilizes CVMFS as it provides a read-only POSIX interface to StashCache
- CVMFS client is the first choice for StashCache, as the most redundant features, including
 - built-in GeoIP locating
 - rate monitoring
 - fallback in failures
- DUNE use StashCache to deliver larger payloads such as flux files and shower libraries to grid jobs
- DUNE CVMFS StashCache eliminates need to copy files to every single job all the time, reduced FNAL dCache load

[/cvmfs/dune.osgstorage.org](https://cvmfs/dune.osgstorage.org)

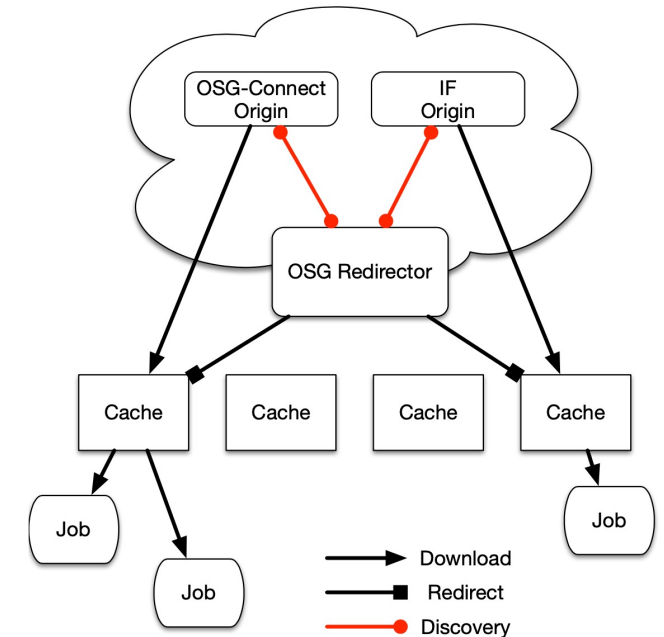
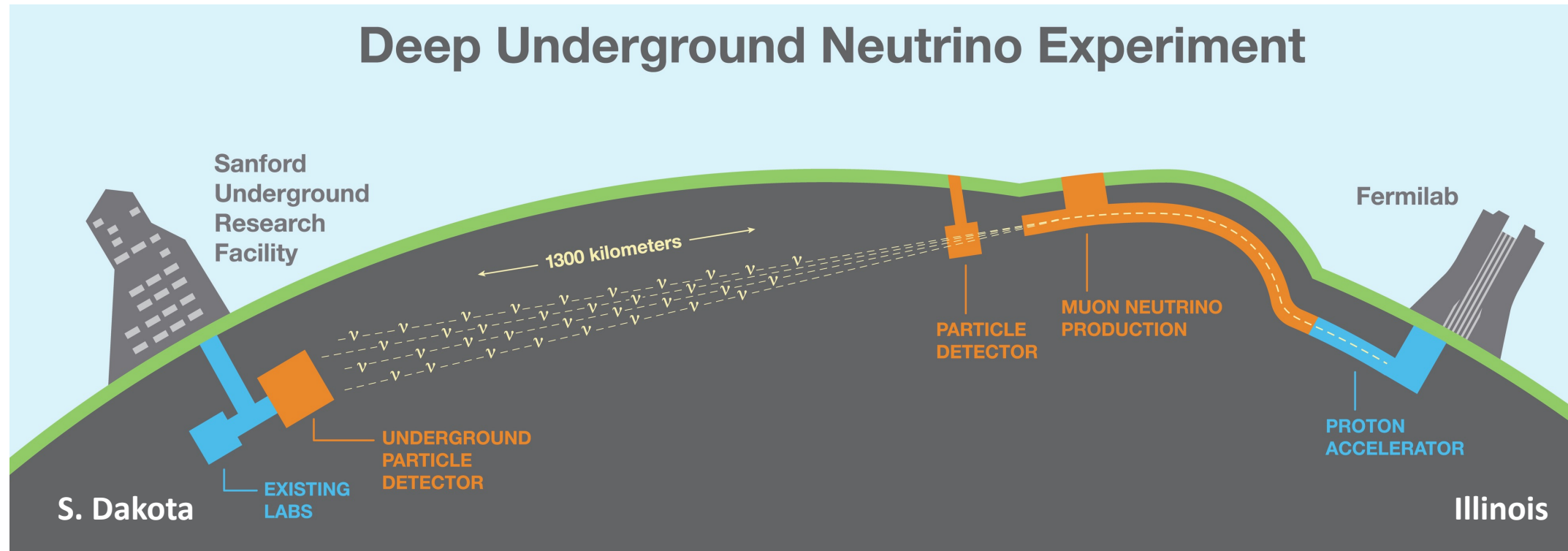


Figure 1: StashCache Architecture: Jobs request data from caches, which in turn query the redirector for the data location. Data is transferred from the origin to the cache, and then to the job.

arXiv:1905.06911

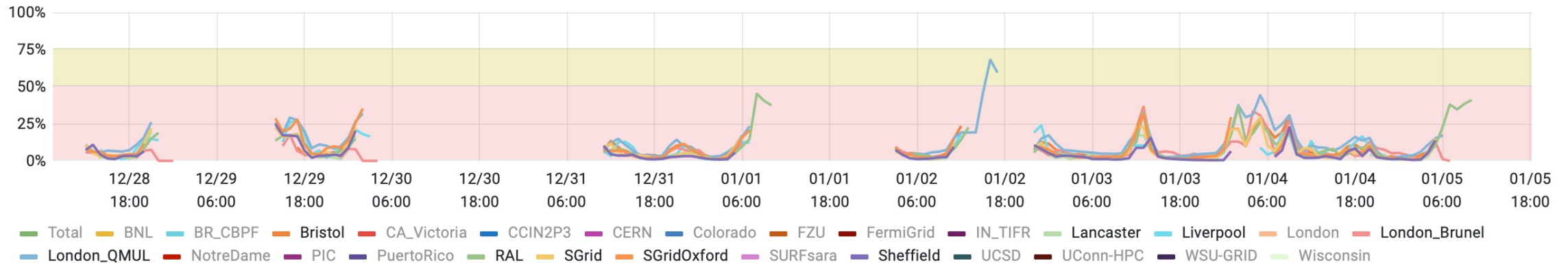
DUNE



Motivation

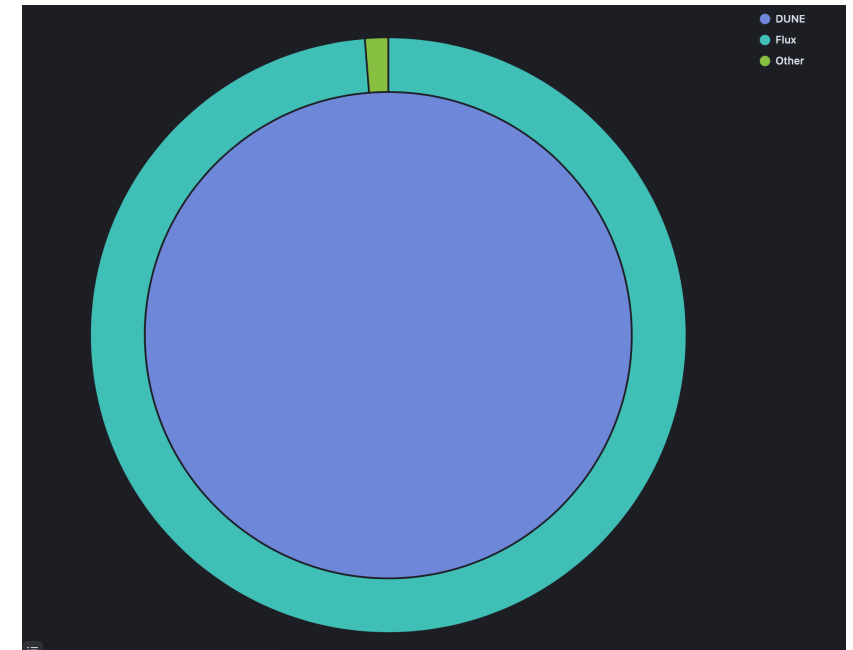
- During Dec-'21 to Jan-'22, DUNE HTC grid jobs were suffering from low CPU efficiency in the UK (<10% eff.), when ran Far Detector Vertical Drift jobs (normally 60% eff.)
- It's not due to the I/O limits, DUNE jobs can get lots of CPU/Mem

Site & Overall Efficiency



Motivation

- Investigations showed long (hours) idle periods in the GENIEgen, when the jobs need many auxiliary files (GENIE flux files) from StashCache
- flux -> event generator -> Geant -> detsim
- DUNE typical flux file size: 50MB-100MB
- The UK sites used Cardiff StashCache, which was intended only for LIGO jobs, DUNE only got limited CPU/Mem
- Edinburgh intended to deploy a new StashCache instance in the UK, to understand the impact of this job efficiency issue



Most DUNE HTC jobs requests from StashCache are Flux files

StashCache implementation at Edinburgh

- Deploying and commissioning since April 2022
 - Installing the Open Science Data Federation Cache (OSDF Cache) from RPM
 - <https://opensciencegrid.org/docs/data/stashcache/install-cache/>
 - straightforward setting up
 - Registering in the OSG
 - <https://osg-htc.org/docs/common/registration/#registering-resources>
 - <https://github.com/opensciencegrid/topology/blob/master/topology/University%20of%20Edinburgh/Scotgrid%20ECDF/UKI-SCOTGRID-CDF.yaml>
- StashCache is OSG Xcache
 - utilizes CVMFS
 - A HTTP(S) based file caching network



Edinburgh registered as an OSG Cache site in the UK with Geo-IP data

StashCache implementation at Edinburgh

- Deploying and commissioning since April 2022
 - Registering in the CVMFS EGI config
 - <https://github.com/cvmfs-contrib/config-repo/pull/92>
 - Edinburgh has been added to
/cvmfs/config-egi.egi.eu/etc/cvmfs/domain.d/osgstorage.org.conf
 - while Cardiff has been removed from that general config file, only for LIGO

```
# extra stashcache servers not used by OSG
CVMFS_EXTERNAL_URL="$CVMFS_EXTERNAL_URL;http://fiona.uvalight.net:8000/;http://xcache.cr.cnaf.infn.it:8000/;
http://xcachevirgo.pic.es:8000/;http://stashcache.edi.scotgrid.ac.uk:8000/"
```

<http://stashcache.edi.scotgrid.ac.uk:8000/>

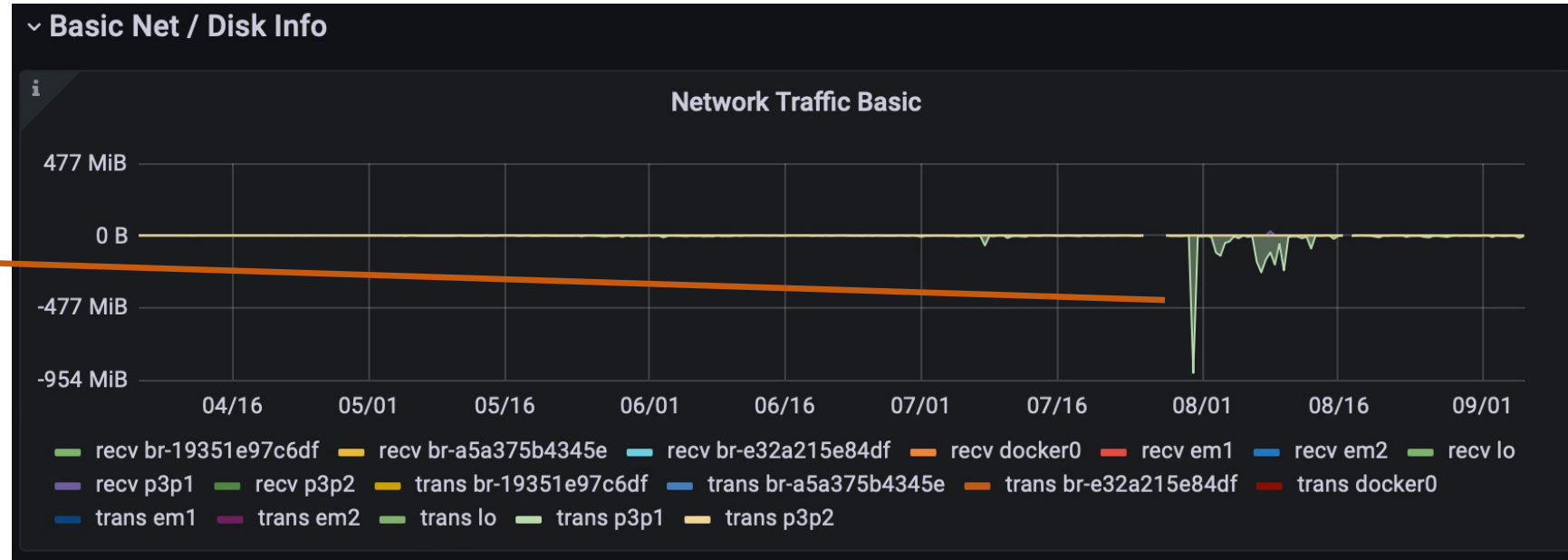


Edinburgh registered as an OSG
Cache site in the UK with Geo-IP
data

StashCache at Edinburgh

- Currently StashCache instances hosted on a physical server 16 Cores 24GB RAM ~10TB of storage with 10Gbps
- Running since April 2022, Ingested ~2TB. Egressed ~120TB
- Very active during the first half August 2022, when DUNE jobs requested “Flux files”

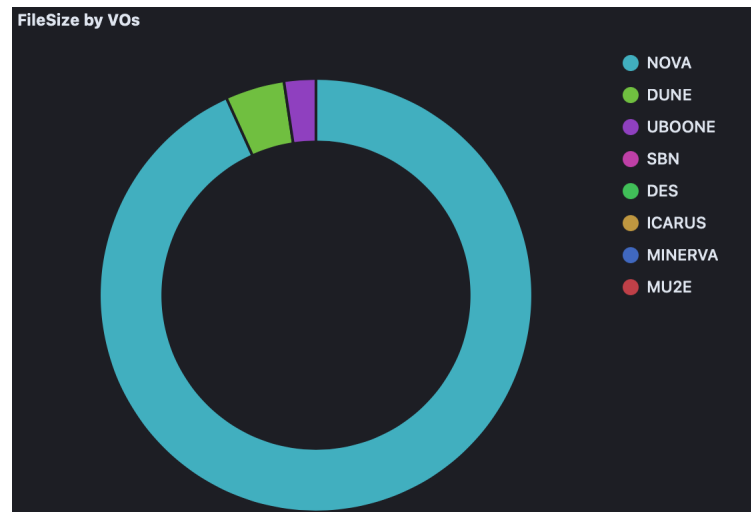
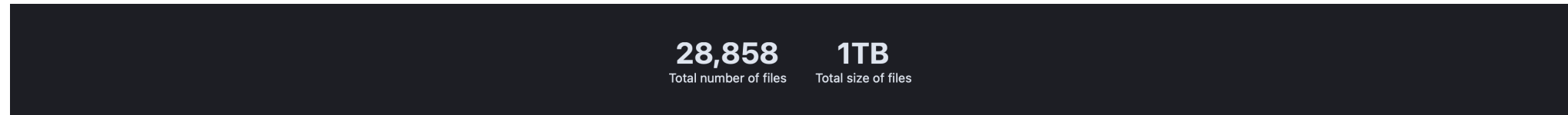
**DUNE activities
in August**



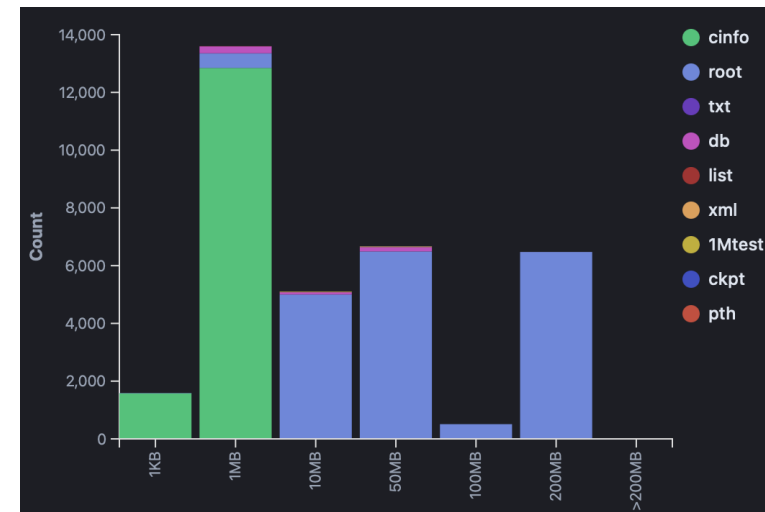
Local StashCache Monitoring

- Node Exporter/prometheus
 - Reliable system and network monitoring
- StashCache logging
 - Fluentd analyzing StashCache logs to OpenSearch
- StashCache metadata
 - Internal state of the StashCache for each file is stored on disk in metadata
 - Developed custom metadata tracker to record/report details to OpenSearch

Local StashCache Monitoring

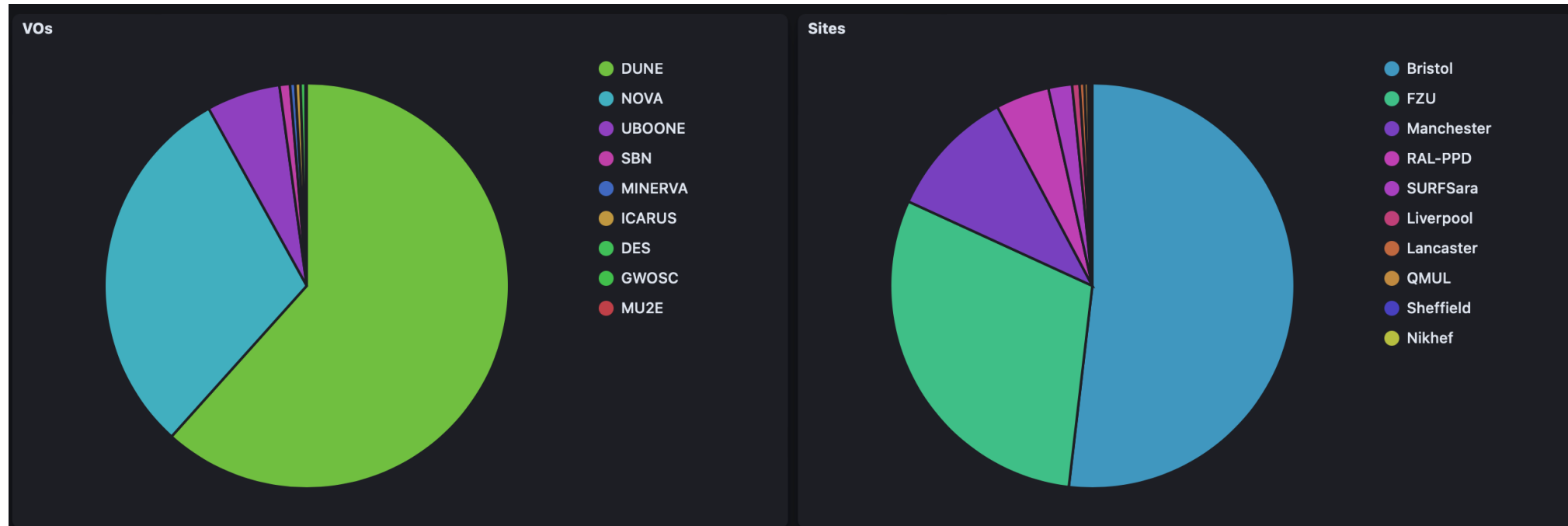


File size by VOs: NOVA occupied most disk, then DUNE, μ BooNE



File size vs type: meta data under 1MB, root files between 10MB – 200MB

Jobs requests in August 2022

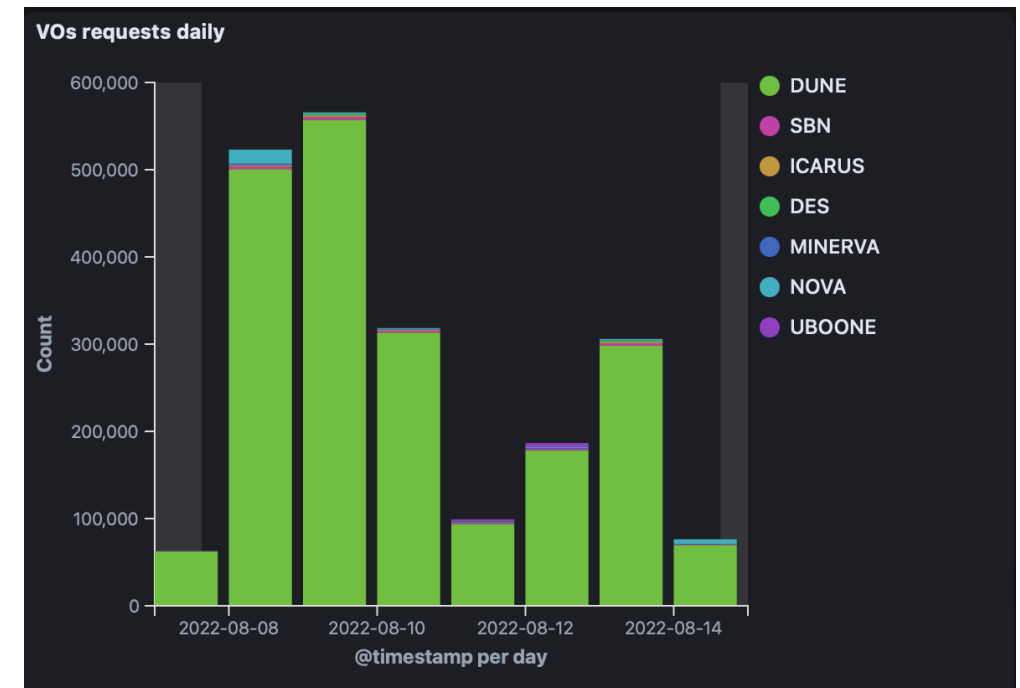
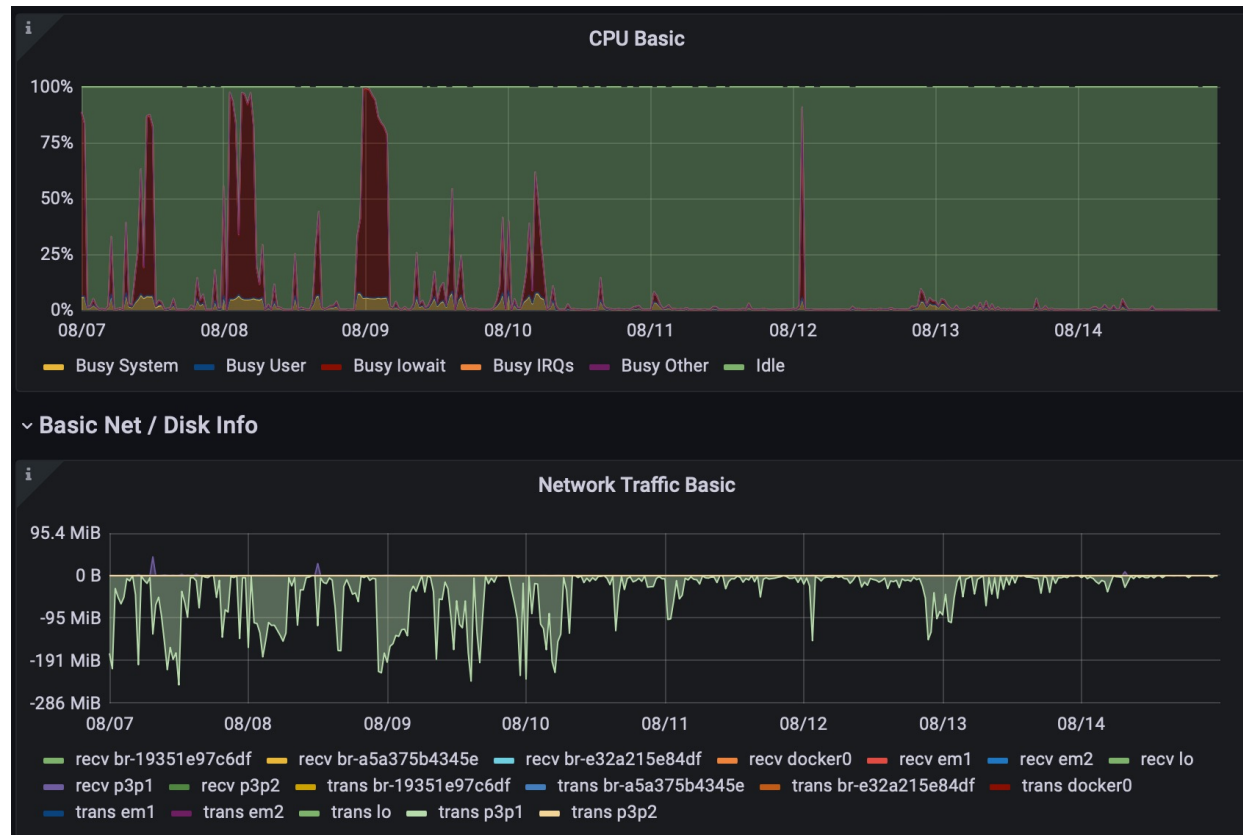


Requests by VOs: most VOs are from FNAL

Requests by Sites: most are from UK, FZU, SURFSara, Nikhef as well

DUNE Jobs in August

- DUNE submitted a large number of HTC Jobs include Flux files during the first half of August
- Heavy loads observed at Edinburgh

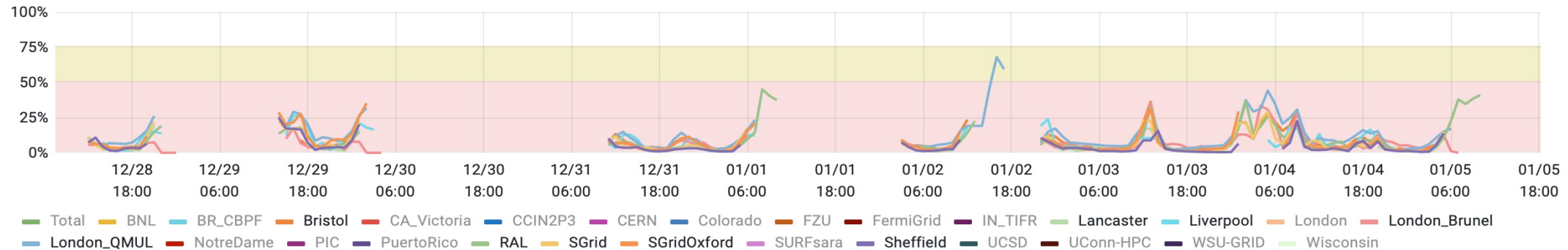


07 Aug. – 14 Aug.

DUNE Jobs in August

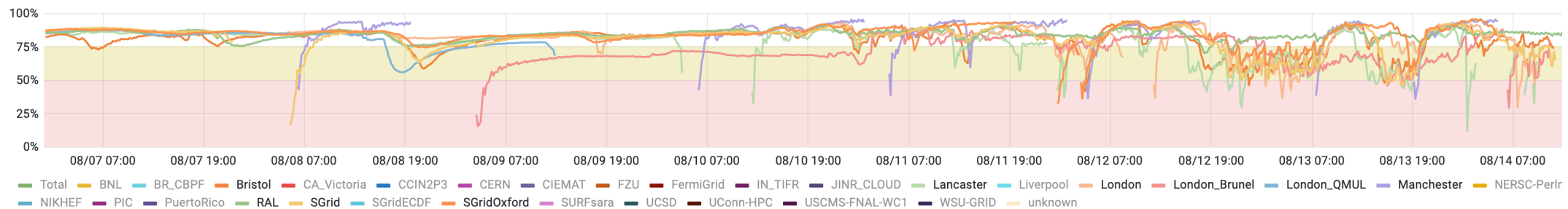
- DUNE UK sites Job Eff. are >60% in average in August
- Edinburgh StashCache solved the flux file low efficiency problem

Site & Overall Efficiency



Dec 21'
- Jan 22'

Site & Overall Efficiency



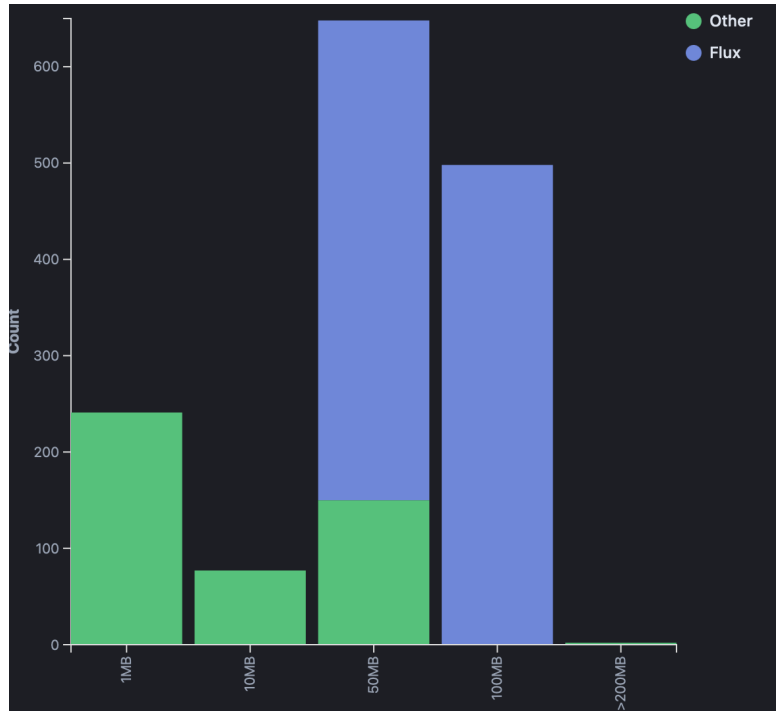
Aug 22'

Summary

- A new UK StashCache instance at Edinburgh for DUNE has deployed since April 2022
- Developed local monitoring to help understanding StashCache usage from various VO's and grid sites, to enhance StashCache operations and maintains.
- Edinburgh StashCache solved the DUNE flux file low efficiency problem, and runs well and smoothly for DUNE as well as other OSG VO's.

*CVMFS has made DUNE computing much easier to grow internationally.
-- Thank you!*

Backups



```
%MSG-i GENIEHelper: GENIEGen:generator@BeginJob 20-Jan-2022 00:20:48 UTC ModuleBeginJob
Convert fiducial volume from master to topvol coords
%MSG
1642638048 INFO GDk2NuFlux : [s] : Attempt to load config "dunevd10kt_v1_1x8x6" from file: /cvmfs/d
1642638048 INFO GDk2NuFlux : [s] : Found config "dunevd10kt_v1_1x8x6"
1642638048 WARN GDk2NuFlux : [s] : start using_param_set: "dunevd10kt_v1_1x6x6"
1642638048 INFO GDk2NuFlux : [s] : Found config "dunevd10kt_v1_1x6x6"
1642638048 INFO GDk2NuFlux : [s] : set user length units to "m"
1642638048 INFO GDk2NuFlux : [s] : set entry reuse = 10
1642638048 WARN GDk2NuFlux : [s] : done using_param_set: "dunevd10kt_v1_1x6x6"
Real time 0:01:18.337647, CP time 10.510
1642652701 INFO Pythia6Decay : [n] : Switching ON all PYTHIA decay channels for particle = -15
```

Site & Overall Efficiency