StashCache for flux files

Robert Illingworth

24 August 2015

Open Science Grid

The problem

- OSG provides resources and tools for distributed computing, but not so much for dealing with distributed data
  - Large VOs, like CMS and ATLAS have implemented their own systems, but these are not exportable to other users
  - They rely a lot on site managed storage elements which are not easily available to opportunistic users

Open Science Grid

Current options

- Smaller files can be distributed via HTCondor, or HTTP with Squid caching, or through CVMFS
  - This is only suitable for smaller datasets
- Otherwise you're left transferring everything from the original source (ie FNAL dCache)
  - Bottlenecks and latency can make this inefficient

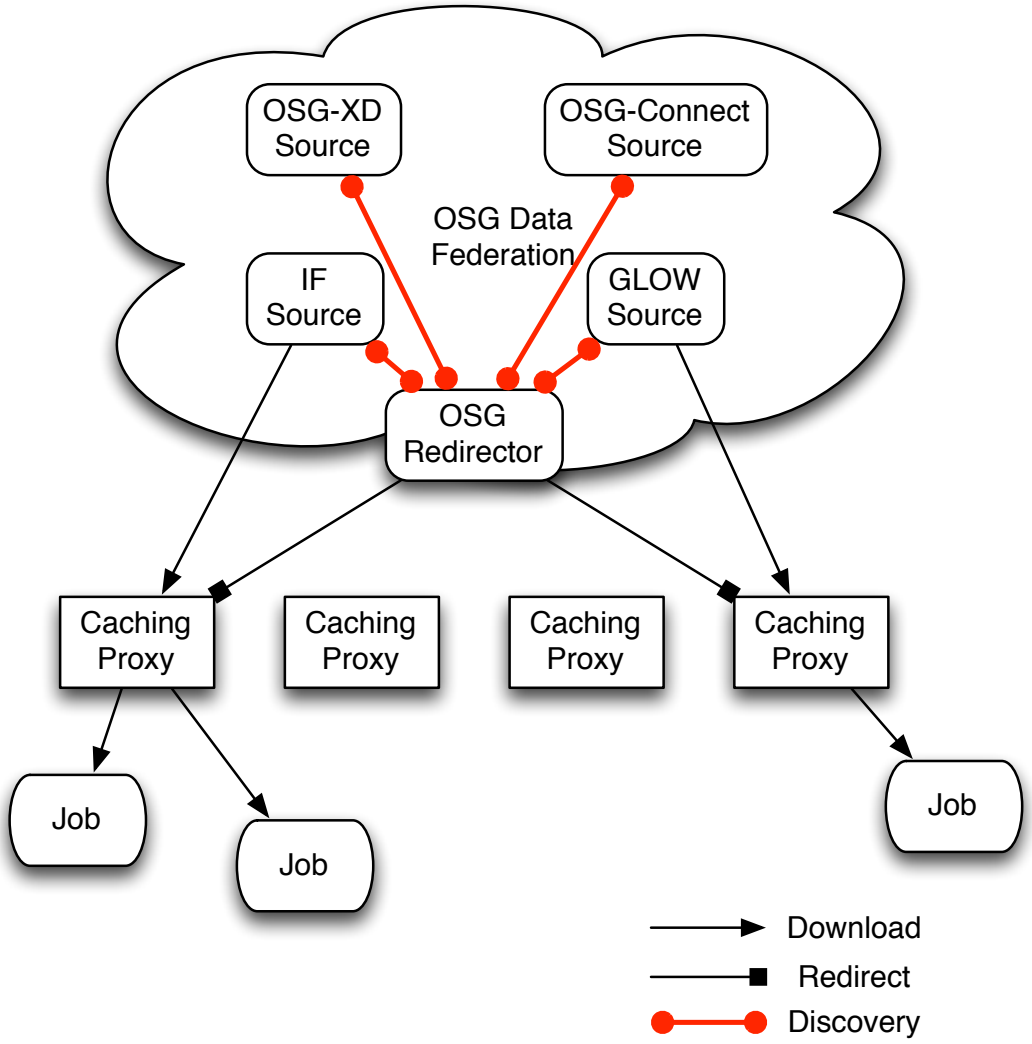
Open Science Grid

StashCache

- StashCache is an OSG project intended to improve certain data access patterns across the grid
  - The initial target is shared input datasets up to a scale of ~1TB
  - “Shared input” meaning that each file should be accessed more than once from the cache
- The caching is transparent and requires no active management by the VO

Open Science Grid

Architecture



Open Science Grid

Architecture

- A source is where the input files come from
  - Managed by the VO
- The redirector points requests to the appropriate source
  - Managed by OSG
- The caches serve out files if they're already there; if not the cache asks the redirector where to get them and adds them to the cache
  - Managed by sites/OSG

Open Science Grid

Architecture

- Implemented using xrootd
- Data access can be either via native xrootd, or using a preload library, through a (mostly) POSIX filesystem interface
- If using xrootd directly you do need to modify your access URLs to point to the appropriate cache server

Open Science Grid

Current status

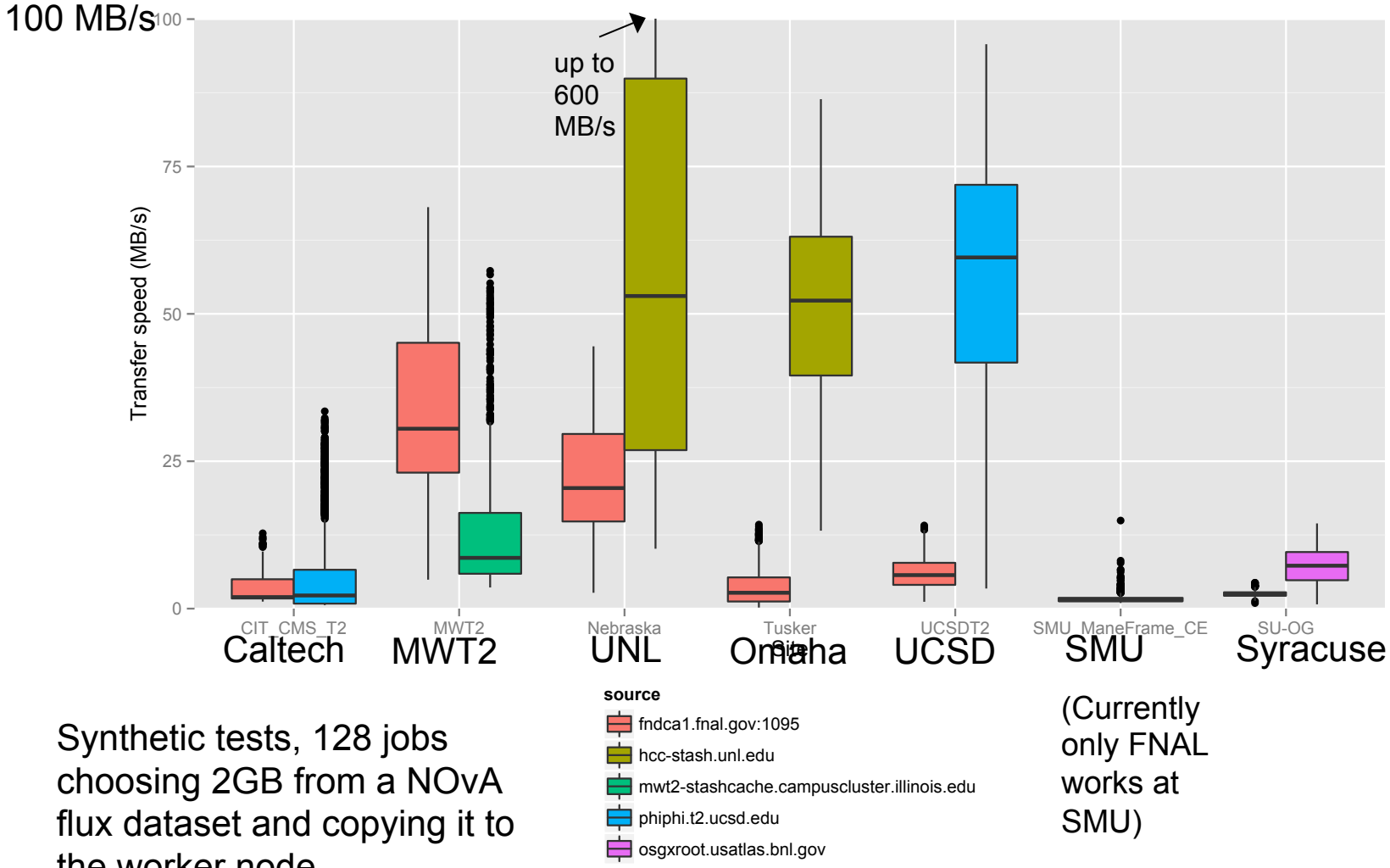
- This is not yet a production service
- Currently “by invitation only”
- We want to ease in rather than promise and not deliver
- We think the flux files for Monte-Carlo generation are a reasonable place to start

Open Science Grid

Uses

- The intended usage seems to match well with the needs of NOvA flux files
  - 10-100s GB datasets, each job randomly selects a small portion of this, but the entire set is going to be accessed multiple times during large scale production
- NOvA data is likely to be a bit big for this
  - But subsets for certain purposes may be possible
- One caveat for FNAL dCache as a source – you must allow unauthenticated read access to your files
  - Opt in at the directory level

Example testing with NOvA flux files



Synthetic tests, 128 jobs choosing 2GB from a NOvA flux dataset and copying it to the worker node

(Currently only FNAL works at SMU)

What's needed to use this

- We think ifdh already provides most of what is necessary
  - The only change is allowing you to override the source host; currently it always uses fndca1.fnal.gov
- Other than that the jobs shouldn't care where the data comes from
- But as the previous page shows, some sites appear anomalous

Open Science Grid

Summary

- StashCache provides a fully automated data distribution mechanism for opportunistic grid jobs
- It looks to be a good fit for flux files
- We have evidence that it speeds up some sites considerably compared to reading direct from FNAL dCache
- Adapting NOvA MC generation to use StashCache shouldn't be difficult