Lightweight Sites

Cedric Serfon on behalf of DDM team
University of Oslo

Introduction

- Some sites have very limited storage/manpower
- Operating a storage element is non-trivial
- From operational experience, the small sites (i.e. <400 TB) and very small sites (<100 TB) generate more problems/operational load :
  - Lost files/Dark Data, and often the site doesn't regular provide dumps
  - Sometimes not very responsive, some sites in downtime for months
  - These sites are not used to store primary data
- The situation is even worse now that we keep the logs where they have been produced :
  - Space is getting filled by logs with time

Solutions for lightweight sites

- Different possibilities for lightweight sites :
  - Storage-less site
  - Site using cache (arc-cache or xrootd cache)
  - Distributed Storage
  - Federated Storage

Storage-less sites

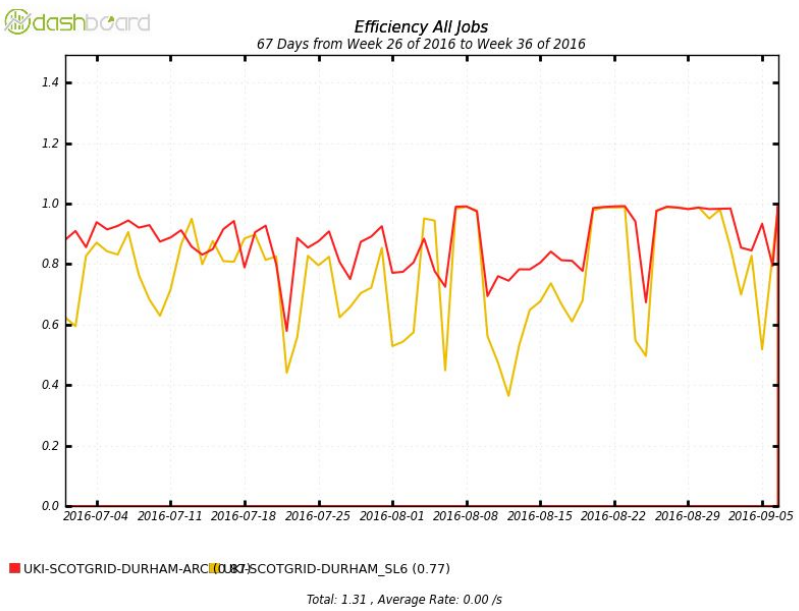
- Having a local Storage Element is not a requirement to run production jobs :
  - E.g. : Uni Dortmund had production queues writing to Wuppertal for years
- Even easier now with new site mover
- 2 sites already in the process of migrating to storage-less : RO-14 and RO-16
- More candidates would be good. DDM ops will help in decommissioning the endpoints.

Arc cache

- Rucio supports the concept of caches which are controlled outside itself and may not be consistent.
- Arc CE cache can publish its content to Rucio through add/delete messages.
- The cache service can create dumps of cache content, and a separate script runs periodically to calculate the differences and send messages to Rucio.
- The cache RSEs are associated to the CE’s PanDA queue and so PanDA can broker jobs to queues where the data is cached.

Arc cache performances

- Queue using Arc cache running for many months in Durham
- Efficiency of queues using Arc-cache and local storage very similar (arc cache queue even a bit better)



xrootd cache

What is it and Why?

Squid like cache proxy on surface

- Use disk to cache data
- Work around firewall
- Easy to use : http_proxy, In the future: **xroot_proxy**=root://mycompany.edu:port/

Different under the hook:

- For static, large files
- Multi-thread to handle data intensive load
- Capable of both whole file caching and file block caching (focus on the later)
- Protocols to client: xroot, http, and (add your protocol plugin here)

Can be a transparent, (mostly) configuration free layer

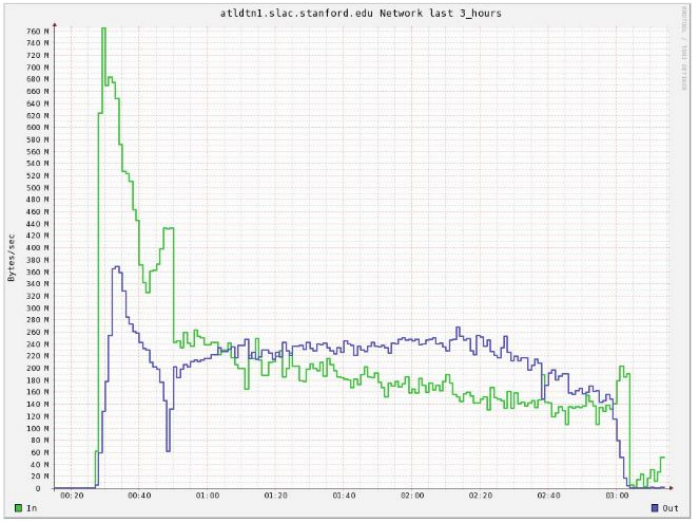
At large scale, it is an unmanaged storage.

Goal: Improve data access efficiency, reduce data manage overhead

xrootd cache test

- The green line shows the data coming from the remote data source.
- The blue line shows the data sent to the batch jobs.

Warm cache



Cold cache

R. Gardner CHEP 2016

Distributed Storage

- A la Nordugrid. Works transparently for Nordic sites (dCache) for years
- Never tested for DPM but should be doable on well connected sites :
  - One site runs the headnode + some disk nodes
  - Other sites only run disk nodes
  - Probably needs some work on the firewall rules
  - If one wants to consolidate existing sites, non trivial to merge the DB from the different sites (remember LFC consolidation)

Federated storage

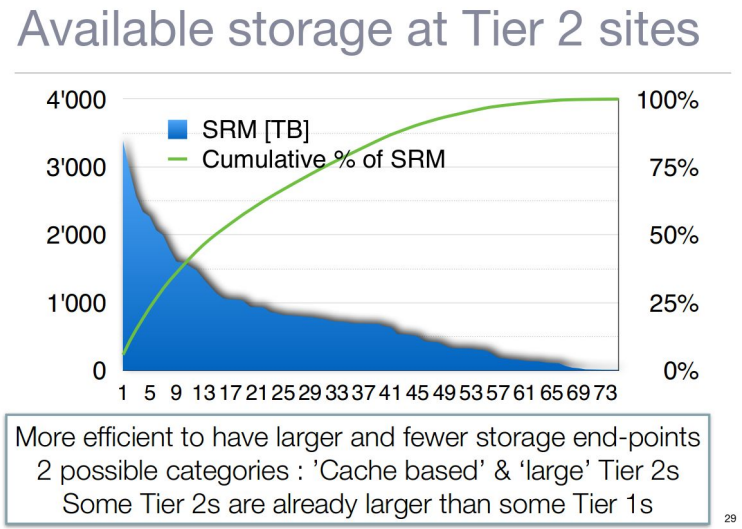
- WebDAV federation — DynaFed
  - This is not distributed storage, but a hosted service is providing the federation
  - WebDAV is now supported both in ROOT (TDavixFile) and DDM level (Rucio + Metalink)
- Provide single **volatile** RSE as "entry point" to the Federation
- Federated storage systems **must not** be registered as RSEs separately
  - Double accounting of data
  - Unavoidable deletion/transfer race conditions
- Allows the transparent use of Microsoft Azure and Amazon-style S3 cloud storages
  - Private keys under control of site, does not need to be published to ATLAS

SRMless storage

- Upload/download in the pilot via rucio upload allows use of non-SRM protocols (gsiftp, xroot, WebDAV)
- 3rd party transfer with gsiftp, xrootd validated. Still need to validate WebDAV

Proposal

- Very small sites (<100 TB) :
  - Keep the panda queues
  - Decommission the Storage element
  - Possibility to setup a cache (arc/xrootd)
- Small sites (>100 TB and <400 TB) :
  - Try to consolidate with one close site
  - Or go to federated storage



From Eric one year ago