SLAC NATIONAL ACCELERATOR LABORATORY

Distributed Computing and SLAC

Wei Yang

After more than a decade of Grid computing ...

Grid computing was sold to LHC with the idea that it would

- Save manpower? → a dashed dream
  - We still need site administrators, and we still have site/local security policies
  - Cross site problem solving is hard - need good communication and working relationship
  - Thick stacks of software developed for Grid computing - cost of maintenance is coming to us
- Share resource? → some success
  - LHC computing runs at a scale that no single site can host
  - Resource (and expertise) are distributed anyway
    - But DOE is pushing for (not so easy to use) HPCs (HEP is not a priority). NSF is not.
  - Remote resources can be used if experiments invest manpower (anther no-easy-to-use example)
    - This does increase resource utilization.

Distributed computing

Evolution of the Grid computing:

- Components to depart, or departed, or not frequently used:
  - Information system (BDII, etc.), SRM, GridFTP, X509 (?)
    - GridFTP was one of the few successful stories. But the Globus team discontinued it
    - WLCG wants to move away from unfriendly X509 - Json Web Token is the future, and is user friendly (really?)
- Components to stay, or are coming
  - HTCondor-CE and ARC-CE, CVMFS, Containers, RUCIO, FTS, Xrootd, Globus, JWT
  - Workflow Management System (WMS).
  - DOE and NSF HPCs - we may need to use them
  - No commercial clouds ← they are too expensive, especially on the e-gress cost

Distributed computing at SLAC

Several SLAC HEP groups work with distributed computing in some ways

- DESC: data flow between NERSC and IN2P3, also use CVMFS
- CDMS: resources at SLAC, Texas A&M, Compute Canada (and FermiLab)?
- EXO: resource at SLAC, use CVMFS from IHEP.ac.cn, working on data movement
- LDMX: Asked for an ACE-CE, interested in RUCIO
- ILC: Globus data transfer, encountered DIRAC system several years ago.
- ATLAS: …
- Fermi: mostly local at SLAC, using Xrootd.
  - **What about Fermi pipeline?**
- KIPAC: Globus

CVMFS and Containers

Both are mainly use to distribute software

- CVMFS client
  - Popular and available at SLAC.
  - Also used to distribute non-software: container images, conditions data.
  - How about CVMFS server/service? Not sure whether SLAC is interested
- Containers
  - Singularity is the most popular one in HEP, supported by SCS
  - LSST DM uses Docker and Kubernetes.
  - NERSC use Shifters (compatible with Docker and Singularity)
  - In some cases, containerization is critical to scale up on HPCs.

Workflow management systems (WMS)

- Most HEP experiments have several main workflows:
  - (Event Generation/Simulation+pile up)/Reconstruction/Derivation+Slimming
  - Group production of simulated or real data (some steps of the above)
  - Adding to the above are calibration, and chaotic user analysis.
- A WMS handle these steps
  - In predefined order, with dashing board and monitoring
  - Distribute workload, handle job failure, storage failure, transfer failure, etc.
- **Fermi pipeline** → Developed at SLAC. What is the current status?
- ATLAS Panda → Powerful but complicate, likely provide service instead of software
- Pegasus: I heard some SLAC experiments mentioned it…
- DIRAC:

Moving jobs and data

- HTCondor-CE and ARC-CE
  - CE (Grid Computing Element) is an interface to submit batch jobs to SLAC from outside
  - ATLAS uses both: HTCondor-CE with LSF, ARC-CE with SLURM
  - CDMS and LDMX expressed some interests
  - CE is hard to use by hand. Usually used by WMS
- Managed data transfer / Third Party Copy (TPC) with Globus or FTS
  - Many SLAC groups use GridFTP with Globus for free. ATLAS uses GridFTP with FTS.
    - Free Globus / GridFTP won't last forever -- Will we pay them license?
    - last time I heard from BNL and Lund Univ. - $90K /yr
  - LHC/WLCG is moving away from GridFTP to use xrootd and dCache for TPC

RUCIO

- RUCIO - a data management system
  - Group data files in data sets.
  - Keep track of access mechanism at sites (RUCIO Storage Element, RSE)
  - Keeps track of data locations and replication rules / policy. Support info query
  - Drive data movement (mostly through FTS)
  - Quota, data and rule lifetime, popularity, etc. Not much about access control (except quota)
- Developed by ATLAS, picking up many HEP and Astrophysics experiments,
  - CMS, Belle II, DUNE, SKA
- RUCIO at SLAC
  - LCLS is closely looking into it and will likely use it
    - rucio-dev.slac.stanford.edu is setup to evaluate RUCIO and FTS
  - LDMX maybe checking it out.
  - If SLAC will adapte RUCIO, it is best to have a team of experts to operate the service

Xrootd

Xrootd is one of the key components in HEP offline data systems

1. As a storage system
    - Used by SLAC and other sites in US and EU. A variation, DPM is used widely in EU
2. Xroot protocol is HEP's de facto standard for remote access (LAN & WAN)
    - http(s) protocol is later added to the Xrootd software stack
3. Xrootd and HTTP TPCs on track to replace GridFTP in WLCG in 1-2 years.
    - SLAC is a leader in this effort
    - This function is integrated with RUCIO and FTS
    - LCLS is also using Xrootd TPC between SLAC and NERSC (not Globus)
        i. Likely will integrate with RUCIO and FTS as well, but without X509
    - WLCG DOMA TPC Working Group is also a testing ground for Json Web Token

Xrootd / Xcache

1. Xcache is gaining very wide and strong interests in WLCG
   - Jointly developed by SLAC and UCSD (since 2016)
2. a cache based on Xrootd software stack
   - using xroot protocol but http protocol is coming
   - Working at either file block level or whole files (work with both large and small files)
   - Multi-thread
3. Several studying show that Xcache is effective in hiding network latency for HEP remote data access.
4. Combine with data management system such as RUCIO, we can simplify and speed up user data access

Xrootd R&D at SLAC

Xrootd gains its ground in LHC and HEP in general because

- Babar harden the core software stack
  - Otherwise Xrootd won't survive rapid changes in the last 7-8 years.
- Open software and plugin architecture enable others to contribute, e.g.
  - DPM, dCache Xrootd, GridFTP-Posix, xrootd-CEPH, VOMS, HTTP, JWS / SciToken, etc.
  - Andy H. embraces and enables those contributions in Xrootd core software stack, and provides leadership for the Xrootd collaboration.

Xrootd is one of SLAC's core competence in HEP distributed computing software

- Likely to continue playing a critical role in HEP in the next decade or two
- Xrootd R&D and support activities at SLAC is largely funded by ATLAS
- SLAC / HEP need to put on it agenda to address long term personal issue
  - Knowledge stack is deep and wide. Not a 12 month job to take over.

Summary

- Some software components and solutions developed for distributed computing will last for a long time.
- SLAC HEP is using several of them. What is our model to support these usage within SLAC? individual group and work together?
- Xrootd is expanding its role in HEP computing - It is a key SLAC competence that need a long term plan.
- HEP groups will likely need some kind of WMS - What is our solution? Will we have a common solution?

Data analysis model (not production) for LHC Run 3 and Run 4:

- Chaotic, but many new ideas, especially on using industry data science tools