DPM and xrootd

Greig A. Cowan

University of Edinburgh

2nd July 2009 / Storage workshop

Disclaimer

- I have been out of the storage-game for a few months now.
- Things may have moved on from what I say here.
- I will give quick overview and links to more detailed information.

What is xrootd? (1)

In my experience, this is a source of much confusion!

Scalla/xrootd

- Scalable Cluster Architecture for Low Latency Access using independent **xrootd** and **olbd** servers.
- Byte level access to any kind of data.
- Organized as a hierarchical filesystem-like namespace.
- Fault tolerance.
- POSIX-like file access using the **xrootd** protocol.

`http://xrootd.slac.stanford.edu`

`http://xrootd.slac.stanford.edu/papers/Scalla-Intro.htm`

What is xrootd? (2)

- Supervisor/Redirector issues file requests to lower nodes in tree and collects response for passing to client.
- oldb servers federates multiple xrootd servers.

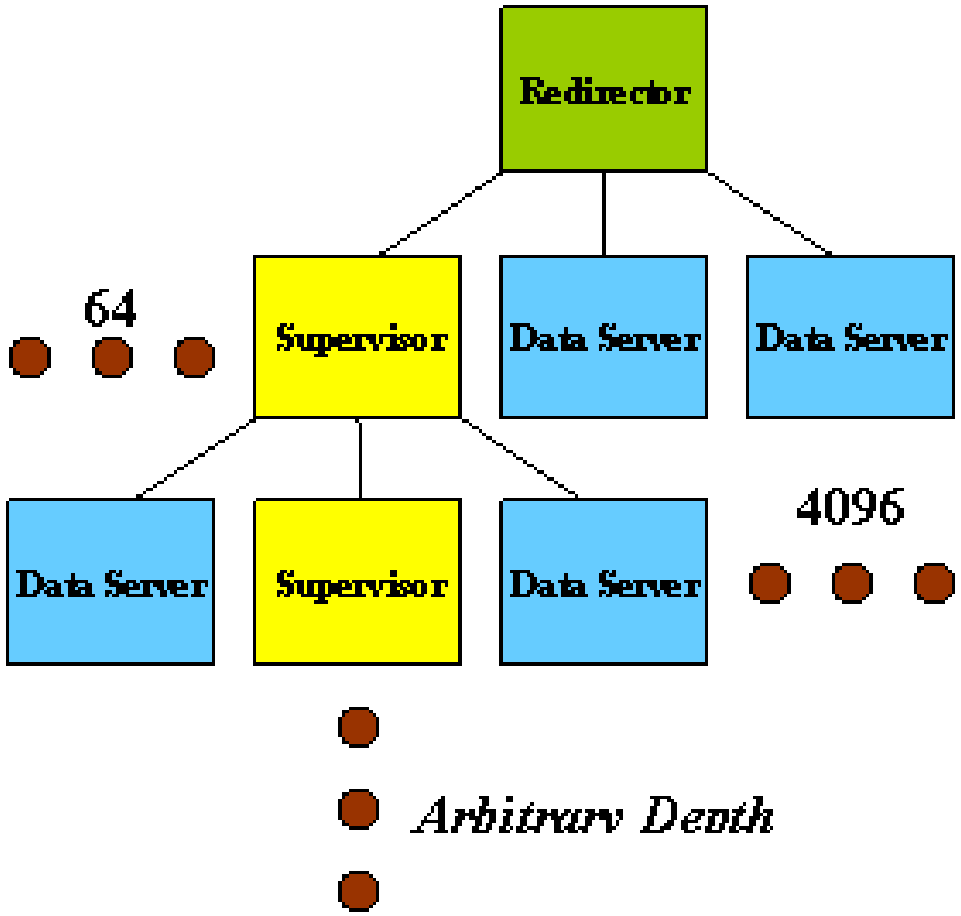


Figure 3: Olbd B64-Tree Architecture

What is xrootd? (3)

DPM/dCache/CASTOR + xrootd

- D/d/C native file transfer protocols: rfio/dcap.
- D/d/C also support the **xrootd protocol**.
- So, if you have a DPM system and a client that wants to use xroot (ALICE), then your site can support them without installing Scalla.
- Access to the DPM namespace to get relevant information.

`https://twiki.cern.ch/twiki/bin/view/LCG/DpmXrootAccess`
`http://trac.dcache.org/projects/dcache/wiki/xrootd`

xrootd protocol features

- Spec available online.
- Fault tolerance (adding or removing servers, failover).
- Performance (TCP connection multiplexing, load balancing) .
- Smart client supports server by understanding redirects and doing several retries in case of server failures.

DPM + xrootd: configuration

Use YAIM or done manually.

`/etc/shift.conf` on head node

```
DPNS TRUST <headnode> <diskserverA> <diskserverB>
DPM TRUST <headnode> <diskserverA> <diskserverB>
DPM PROTOCOLS ..... xroot
```

`/etc/sysconfig/dpm-xrd` on pool nodes

```
export DPNS_HOST=dpm.host.node.name
export DPM_HOST=dpm.host.node.name
```

DPM+xroot appears to have some space token support.

DPM + xrootd: services

On all nodes:

- `service dpm-xrd <start|status|stop...>`
- `service dpm-olb <start|status|stop...>`
- Think of `dpm-xrd` as the equivalent of `rfiod`.

and on the head node:

- `service dpm-manager-xrd ...`
- `service dpm-manager-olb ...`

DPM + xrootd: (lack of) security

- Necessary to turn off GSI due to problems with some certificates!
- **By default the basic config exports the complete `/dpm` namespace with public read-write access to all DPM files!**
- Restrict the public namespace in `/etc/xrd.dpm.config`:
  `export /dpm/myhost/public-dir/`

DPM + xrootd: firewall

- Manager xrootd listens on port **1094**.
- Diskserver xrootd port listens on port **1095**.
- Allow incoming access towards this port from expected client machines.
- All transfers run via port **1095**.
  - No need for an open high port range.

DPM + xrootd: old problems

- Files >2GB were not supported. **Now fixed**.
- Had to replace `/lib`→`/lib64` in config files.
- Performance (in terms of transfer rates during analysis jobs) at Edinburgh appeared similar to rfio.

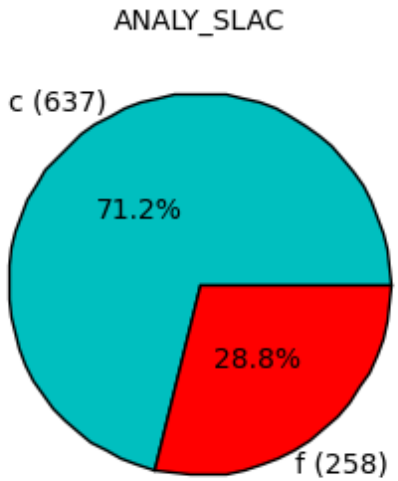
Clients

1. Put:
   ```
   /opt/lcg/bin/xrdcp /etc/groups
   root://<mydpm>//dpm/<mydomain>/xroot/test1
   ```

2. Get within ROOT:
   ```
   TXNetFile::Open("root://<mydpm>//dpm/<mydomain>/xroot/t
   ```

   - Always good idea to test opening files from within ROOT - this is what HEP users do!

Do we care?



- STEP'09 results from ATLAS were not favourable for Scalla/xrootd.
  - Caveat: It looks like there were problems due to lack of support within ATLAS software.
- ATLAS are seeing much higher ($O(10)$?) higher CPU efficiency by copying data to the WN before processing.
  - i.e., rfio/dcap/xrootd not being used anyway.
  - What when we have 8 jobs on a single WN using a single disk?
- Only ALICE that has a concrete requirement for xroot.

Conclusions

- I probably wouldn't invest time in looking at DPM/dCache+xrootd.
- Unless the experiments really say that they want it.