GridPP
UK Computing for Particle Physics

Tier-2 experiences of dCache

Greig A Cowan

University of Edinburgh

GridPP
UK Computing for Particle Physics

Outline

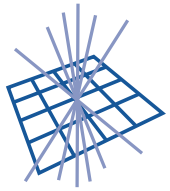
1. Storage at Tier-2's
2. What (GridPP) Tier-2's would like to see from dCache
3. Conclusions

GridPP
UK Computing for Particle Physics

A *typical* Tier-2

No such thing as **typical**, but there are some similarities.

- Limited hardware resources:
  - One or two nodes attached to a few TB of RAID'ed disk.
  - Some storage NFS mounted from another disk server.
  - No tape storage.
- Limited manpower to spend on administering/configuring an SRM.
- Choice of SRM applications (dCache, DPM, StoRM . . .)
- Require SRM to be optimised in order to handle the data flows from the LHC.
  - GridPP service challenge set target for T1$\rightarrow$T2 transfer rate of $\boxed{\geq 300\text{Mb/s}}$.



We would like to...

Storage accounting

Find out how much storage is used/available per VO

- dCache information provider is integrated in YAIM and publishes via GIP.
- Requires that pools are assigned to VO specific pool groups.
- If pools shared between VO pool groups then a single VO can use up all available storage.
  - T2s are not always able to give VO specific pools.
  - Share partitions with other non-Grid users.

Storage accounting

Find out how much storage is used/available per VO

- dCache information provider is integrated in YAIM and publishes via GIP.
- Requires that pools are assigned to VO specific pool groups.
- If pools shared between VO pool groups then a single VO can use up all available storage.
  - T2s are not always able to give VO specific pools.
  - Share partitions with other non-Grid users.
- Alternatively, can get used space per VO:

```
[root:/pnfs/domain.ac.uk/data/]$ du
```

  - This is an expensive operation - takes 50 mins on the PNFS server at RAL Tier-1.

VO quotas

Set a quota on how much storage each VO (each user?) can use

- dCache cannot provide quotas using PNFS.
- Can set limits on VO usage only at the pool level.
- Sys-admins would like finer grained control of pool management. i.e.
  - Allocate different size based on the VOMs role within a VO.
  - Quotas within pools/pool groups would mean T2s do not have to set up more pools for each new VO $\rightarrow$ improved service.
  - Would allow dynamic changing of VO allocations.
  - Performance improvement since data could be spread around multiple pools.

VO quotas

Set a quota on how much storage each VO (each user?) can use

- dCache cannot provide quotas using PNFS.
- Can set limits on VO usage only at the pool level.
- Sys-admins would like finer grained control of pool management. i.e.
  - Allocate different size based on the VOMs role within a VO.
  - Quotas within pools/pool groups would mean T2s do not have to set up more pools for each new VO $\rightarrow$ improved service.
  - Would allow dynamic changing of VO allocations.
  - Performance improvement since data could be spread around multiple pools.
- Theoretically possible with **Chimera**, but time required to implement and test.

Administration scripts

Utilise scripts to interface with and control dCache components

- Tier-2 sites do not typically have the resources for 1FTE to spend administering dCache.
- Would like pool management scripts for easy day-to-day running.

Administration scripts

Utilise scripts to interface with and control dCache components

- Tier-2 sites do not typically have the resources for 1FTE to spend administering dCache.
- Would like pool management scripts for easy day-to-day running.
  - **Draining a pool** so that it can be taken offline for maintenance or reconfiguration.
    - dCache `CopyManager` is available.
    - Could multiple destination pools be used?

Administration scripts

Utilise scripts to interface with and control dCache components

- Tier-2 sites do not typically have the resources for 1FTE to spend administering dCache.
- Would like pool management scripts for easy day-to-day running.
  - **Draining a pool** so that it can be taken offline for maintenance or reconfiguration.
    - dCache `CopyManager` is available.
    - Could multiple destination pools be used?
  - DB consistency checker.
    - Are all my PNFSid files in PNFS?

      ```
      [root:/pnfs/domain.ac.uk/data/]$ pathfinder list-of-ids.txt
      ```

    - The list of files should be PNFSid's or paths.
    - Output is list of PNFSid files that can be deleted.

Administration scripts (2)

- Finding the pool that erroring files are meant to be on.
- Checking files have gone to and from HSM after they have moved.
- Tracking down partial file transfers and deleting the leftover file.

Feature request:

- File permissions manager as PNFS does not support the sticky bit.
  - CMS Phedex makes this important.

GridPP
UK Computing for Particle Physics

Hierarchical Storage Management

Be able to easily interface dCache with an HSM backend

Use case:

- Edinburgh T2 will use SAN storage mounted over NFS as a HSM backend.
  - NFS mounted disk pools did not perform well when writing.
- Custom scripts have to be written by the site to interface with their own HSM backend.
- Lack of manpower at T2s to setup such an HSM interface may impede use of this functionality.
  - Would be good if HSM scripts were made available for study.

Logging

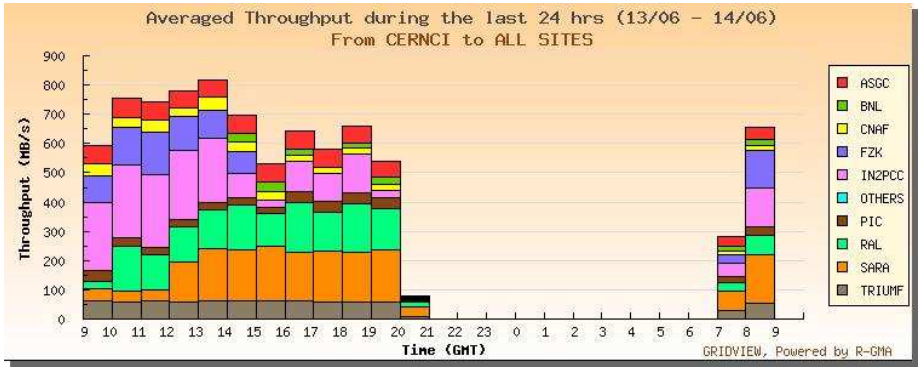
Have improved logging for accounting/monitoring/security

- DNs recorded for srmPuts, srmGets and srmCopies when your dCache is the source SRM, not when your dCache is the destination.
  - Security issue.

Logging

Have improved logging for accounting/monitoring/security

- DNs recorded for srmPuts, srmGets and srmCopies when your dCache is the source SRM, not when your dCache is the destination.
  - Security issue.
- dCache GridFTP logs cannot be used to publish into R-GMA. Need to query the `billing` database.
  - Cannot use GridView to monitor data transfers.



Error messages

See an improvement in the user-friendliness of dCache error/log messages

- Very scary for those new to dCache.
- Multiple (sometimes repeating) messages at same time stamp.
  - Difficult to `grep` logs to discover the result of an admin action.
- Log bloat.
- `strace` has been used to debug dCache behaviour.

Error messages

See an improvement in the user-friendliness of dCache error/log messages

- Very scary for those new to dCache.
- Multiple (sometimes repeating) messages at same time stamp.
  - Difficult to `grep` logs to discover the result of an admin action.
- Log bloat.
- `strace` has been used to debug dCache behaviour.

Suggestions:

- Creation of a list of common messages?
- Integration of dCache logging with syslog?

GridPP
UK Computing for Particle Physics

Optimisation

Admins would appreciate tools to be able to identify bottlenecks

- Optimal mover queue size for each access method.
- Tuning of configuration parameters (*.batch files).
  - Unclear what these parameters do. Should we be changing them?
- There is a tool in the GUI admin interface...
  - could this be extended to show information about these parameters?
  - could it show the effect of making changes to the configuration?

GridPP
UK Computing for Particle Physics

Bulk requests optimisation

- There are reports that when allocating pools to a bulk file upload, dCache uses the same pool for all the files.
- This not a load balanced utilisation of resources.
- dCache.org know about this and has been resolved.

SRM v2.X

GridPP would like to perform some interoperability testing between SRM v2 servers

- Testing of existing SRM v2 servers, even if they just support srmPut's and srmGet's.
- This could help with the debugging of the forthcoming SRM v2.2 release.
- Would also like to test additional functionality like pinning and reservations.

GridPP
UK Computing for Particle Physics

Conclusions

- dCache a very good disk pool management system for Tier-2 sites.
- Additional functionality very useful for many Tier-2 sites.
  - Trying to take advantage of NFS mounted storage.
  - Resilient dCache across WNs.
- GridPP happy with the responsiveness of the dCache team to feature requests and bug reports.
- Deployment of a basic system is well integrated with YAIM via work done by GridPP.
  - Development continuing to allow for more further flexibility in setup.
  - Admins have to get hands dirty to take advantage of additional functionality.
- We would like to see additional admin tools made available.