dCache and Storage Evolution

or

An update from the workshop

Paul Millar

on behalf of the dCache team.

GDB 2015-06-10
35 registrants from 13 countries world-wide. Workshop covered 2 ½ days.

Contributions from: development team, our collaborators at CERN, dCache admins, storage hardware companies. Several interactive sessions: hands-on tutorials, fish-bowl, genius bar.
Funding

2010

NFS 4.1 / pNFS
HTTP / WebDAV
Contributing to the Dynamic Federation

Standardization

2013

Deploying new technologies into Production and exploring new communities

LSDMA

2015

AAI

Data Life Cycle

2018

INDIGO DataCloud
Multi Tier Storage
Quality of Service
Migration Archiving
Indico DataCloud: Cheat Sheet

- H2020 project, started April 2015 for 30 months.
- INFN-Italy project lead; DESY WP4-lead & involved with WP5.
- Budget of 11.1 M€ (0.8 M€ for dCache).
- 26 partners from 11 European countries
- The project aims for an Open-Source Data and Computing platform targeted at scientific communities, deployable on multiple hardware, and provisioned over private and public e infrastructures.

**dCache focus is Data Quality of Service and data-lifecycle management.**
SRM

• Very much **supported** by dCache team:
  e.g., recent GFAL/FTS bug

• *If* SRM goes away, **functionality remains**:
  • SRM is an interface, not the functionality itself
  • Will provide equivalent, RESTful interface (e.g., CDMI), through Indigo DataCloud.
  • **Take input** from experiments in shaping such an interface.
WLCG HTTP deployment WG

- dCache **active participant**, involved from beginning.
- Oliver is doing a **great job**! Strong agreement with the direction of WG
- dCache **supports** the HTTP dynamic federation
  ... not in coding, but deployment and finding user-cases.
- Plan to use within Indigo DataCloud:
  Deploying FTS & WebFTS at DESY so we can further investigate
CEPH integration

Each dCache pool still only ‘sees’ his own private repository.

dCache pools can use shared repositories. Requires new pool semantics. (Focus on protocol engine)
NFS v4.1/pNFS usage by dCache sites

- **6** sites: No immediate plans
- **4** sites: Plan to start testing
- **3** sites: Actively testing
- **2** sites: Beta-like service
- **6** sites: In production

Notes: “In production” services currently do not include direct worker-node access. Data source: dCache user-forum survey 2015-06-08.
KIT evaluation of NFS v4.1/pNFS

Slides courtesy of Preslav Konstantinov, (SCC, KIT) presentation at GridKa TAB meeting 2015-05-06

Conclusions

- In addition to convenience of usage NFSv4.1 in dCache shows some performance improvements over dcap.
- I think it is worth pursuing further testing with production workflows and growing workload in order to achieve stability.

Eff = (User + System) / Wall
Roughly 35% better with NFS
Federated AAI

AARC → INDIGO - DataCloud → egi

LSDMA
Other highlights from the workshop

Future directions in storage tech.
  Manfred Berger, HGST

Embedding dCache in DDN storage controllers
  Simon Liu, TRIUMF

dCache and ELK
  Johan Guldmyr, CSC-IT

The HTTP ecosystem
  Fabrizio Furano, CERN

... plus many other presentations
Thanks for listening … any questions?
Backup slides