dCache: dependable storage for new distributed communities

Paul Millar
(on behalf of the dCache team)

EGI-CF-2013, Manchester, 2013.04.10
Ideal storage system?

Unlimited storage

(but who pays for it?)

for

Unlimited communities

(but who is authorised?)

to

Unlock data's potential

(how long is data stored?)
dCache: advanced storage solution

- Stable, scalable, open-source storage:
  - Project has been going for more than 10 years,
  - Supported by three research institutes (which rely on the software),
  - Easily handle many (many) Petabytes of data
- Strong on supporting standards:
  - HTTP, WebDAV, NFS, FTP, (GLUE, SRM, StAR, ...)
- Many advanced features, including:
  - Transparent tape archiving and retrieval,
  - Easy to increase capacity (no interruption to service),
  - Advanced load management: overload protection, load balancing,
  - Pluggable user identity system
Rolling out the Scientific Cloud
Security: X.509 is (probably) too difficult
Security: identifying your users
LSDMA: working with Big Data

• Solving how Big Data is handled in Germany

• dCache is the major storage system involved in the project

• WP-1 looking at harmonising Federated Identity:

dCache.org is driving this work
LSDMA & dCache: SAML Web profile

• Initial support will be for “SAML Web profile”
  • Access dCache with a web-browser
• Viewing this as a proof of principle
  • Requires web-browser for it to work
  • Upload might be tricky
• Might also investigate non-SAML methods:
LSDMA & dCache: real data usage

- **SAML-Web profile isn't the complete solution:**
  - It's designed for a web-browser
  - Bulk data transfers aren't normally transferred via a web-browser

- **Non-web SAML** is currently an area of R&D
  - Two main contenders are:
    - Project Moonshot / ABFAB vs SAML ECP + GSS binding
  - Currently **Project Moonshot** seems the most promising
Security: Don't forget authorisation!
Bringing our experience

Technology provides increasing amounts of data:

dCache.org has experience in handling large amounts of data
Use standards!

• Communities often buy into **proprietary systems:**
  
  Look at the boundaries – can you replace a component?
  Done for pragmatic reasons, but may hurt in the long run.

• Is **CDMI** a solution?
  The ISO standard for storing data in the cloud
  Powerful enough to support many (most?) use-cases

• dCache is implementing support for CDMI
  Early days, but work has started
The Future

• Storing and accessing data is a very common requirement.

• Research communities are faced with ever increasing storage requirements.

• By engaging with new communities, dCache can grow while helping those communities maximise their research output.
Thanks!