

Cloud Interfaces and optimised workflows

2nd EUDAT User Forum Martin Hellmich (CERN) EUDAT WP7







Warning

I'm part of WP7

EUDAT R & D

The following does not display the main direction of EUDAT



Data Archive Interface: we want

- High performance transfers
- Always-On Data
- Single Sign On
- Hierarchical namespaces
- Interdisciplinary research
- Move Millions of objects at once





Cloud Interfaces



S3 and Swift

or

Hide them

Two solutions: Hide everything or hide a little





Why hide it?

- Established Authentication/Authorization
- Hierarchical namespace
- Massive Writes
- Needs special clients





Case 1: iRODS-OpenStack

Use of iRODS interfaces

iRODS as Cache

Cost-effective with commercial clouds





Case 2: DPM with S3 attached /arch







Case 2: DPM with S3 attached







The case I'd like to make

- HTTP for users
 - Various clients
 - Mostly reading (from anywhere)
 - List objects
 - Change permissions
- S3/Swift/CDMI for administration
 - Few clients
 - Automated processing
 - Move data in federations
- Both: high performance



HTTP gives Users

111010010**1**

- Browser access
- In-browser processing (CORS headers)
- Browser upload with POST
- Mount with WebDAV → POSIX-like
- Multi-stream GET
- Partial reads \rightarrow random I/O
- Multiple redirection
- Magnet links \rightarrow

client-side replica choice (clients actually understand this)





S3/Swift/CDMI gives admins

- ACLs
- In-storage COPY
- Multi-stream PUT
- Storage classes (S3 Glacier)
- High-performance object store
- "fake" hierarchical listing
- Easy 'translatability' into HTTP





Optimised Workflows

- Christian already showed the main ideas:
 - Workflows
 - Move code, not data
- Two building blocks
 - MapReduce + iRODS
 - Third Party Transfers





MapReduce + iRODS



Developed by Jedrzej Rybicki (JSC) Prototype ready





Direct Transfer to HPC



Developed at CERN. In construction. You are invited to contact me, Martin Hellmich

