Kindura: Cloud Storage
OGF 31 / Taipei

Jens Jensen1
with Mark Hedges2, Simon Waddington2, Jun Zhang2, Roger Downing3
1=STFC RAL; 2=King’s College, London
3=STFC Daresbury
Kindura Project

- 12 month **pilot** project in **JISC** Flexible Service Delivery Programme
- Centre for e-Research at King’s College London
- Science and Technology Facilities Council (STFC), RAL+DL
- DuraSpace
  - Not for profit organisation
  - Specialises in open source technologies in the fields of digital repositories and clouds
  - Created in 2009 from a merger of Fedora Commons and the DSpace Foundation
Objectives

- **Hybrid** cloud data management
  - iRODS (via a cloud interface)
  - NGS cloud
  - Commercial cloud providers.
  - Private cloud (e.g. Based on Eucalyptus)
- Real data from science and humanities
- Cost effective archiving
  - “Non-elastic cloud”
  - **Migrator**: knows when to move data
  - Based on **cost models**
Kindura Services

- Researcher
- Storage services
- Compute services
- Common services

Federated iRODS dataservers

NGS

iRODS

Integrated Rule-Oriented Data System

Eucalyptus Systems

Amazon web services

Rackspace
Specific scenarios

• *Environmental Science*
  • monitoring and modelling of environmental change
  • e.g. remote sensing, real-time data feeds, and large scale simulations of environmental models.

• *Financial Mathematics*
  • execute complex mathematical models, for scenarios such as contract pricings or risk computation.

• *Biomedical sciences*
  • crystallography
  • nanoimaging (e.g. microscopy techniques for cell biology),
  • molecular neurobiology (e.g. determining factors influencing gene expression).

• *Humanities – textual scholarship*
Architecture
Technical Approach

- **DuraCloud**, developed by the DuraSpace organisation.
  - Available as a service or software download
  - Java open source software
  - Provide a common REST API to multiple cloud providers.
  - Pilot programme – Kindura is the only UK participant.
  - Common services (runnable in the cloud)
    - Replication
    - Bit integrity checking
    - Format conversion
- **iRODS**, with a cloud interface
- **Fedora Commons** repository
  - Open source repository.
Brokerage

Instance

Value

Sensitivity

Frequency of access

Classification

Lifecycle

Access location

Response time
Emerging standards

- **OGF OCCI**
  - Provides standard CRUD interface

- **SNIA CMDI**
  - SNIA – Storage Networking Industry Association
  - Cloud Data Management Interface (CDMI) enables interoperable cloud storage and data management.
  - Standardises notion of container
  - (cf. Monday’s session: GSM–CG)
Outputs

- Pilot infrastructure
  - “Non-elastic cloud” – based on iRODS
  - Migrator, glue (eg AAA)
- Use cases for targeted areas
- Cost model for clouds (user view)
- Evaluation
Conclusion

• Cloud interface to federated iRODS
• “Less-elastic” storage for cost effective archiving
• Federated access via DuraCloud
• [http://kindura.cerch.kcl.ac.uk/](http://kindura.cerch.kcl.ac.uk/)