Kindura Cloud Computing Pilot
13 Dec 2010

Simon Waddington
Centre for e-Research (CeRch)
King’s College London
Kindura Project

- 12 month project in JISC Flexible Service Delivery Programme
- Centre for e-Research at King’s College London
- Science and Technology Facilities Council (STFC)
- 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

• Develop and pilot a hybrid cloud storage solution for research data at King’s and STFC
• Study user requirements for large scale storage of research data.
• Integrate multiple cloud services into Kindura application including
  • iRODS
  • NGS cloud
  • Commercial cloud providers.
  • Private cloud (e.g. Based on Eucalyptus)
Objectives (ctd)

- Develop brokerage services to provide mediation between storage providers.
- Integrate digital repository to provide archival services for researchers.
- Provide common services for such purposes as data replication, format conversion and digital preservation.
- Provide integration with cloud compute services.
Kindura Services

- Researcher
- Storage services
- Compute services
- Common services

Federated iRODS dataservers

iRODS

NGS

Eucalyptus Systems

Amazon Web Services

Rackspace
Specific scenarios

• *Environmental Science*
  • including the 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 (e.g. determination of protein structures from diffraction images),
  • nanoimaging (e.g. microscopy techniques for cell biology)
  • molecular neurobiology (e.g. determining factors influencing gene expression).

• *Humanities – textual scholarship*
Architecture
Technical Approach

Pilot makes use of:

- **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
    - Replication
    - Bit integrity checking
    - Format conversion
- **Fedora Commons** repository
  - Open source repository.
Brokerage

- Classification
- Value
- Sensitivity
- Lifecycle
- Instance
- Frequency of access
- Access location
- Response time
Emerging standards

**OGF OCCI**
- OGF – Open Grid Forum
- Open Cloud Computing Interface (OCCI) is a free, open, community consensus driven API, targeting cloud infrastructure services.
- 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
Outputs

- Pilot model of ICT provision that uses both cloud and shared services.
- Use cases representing the requirements and activities of targeted researchers.
- Pilot evaluation to demonstrate the potential to increase efficiency and effectiveness.
Further Information

Project blog

http://kindura.cerch.kcl.ac.uk