A national approach for storage scale-out scenarios based on iRODS

Christine Staiger, Ton Smeele, Rob van Schip

Utrecht University

iRODS User Group Meeting, Utrecht, 14/15 June 2017

Context

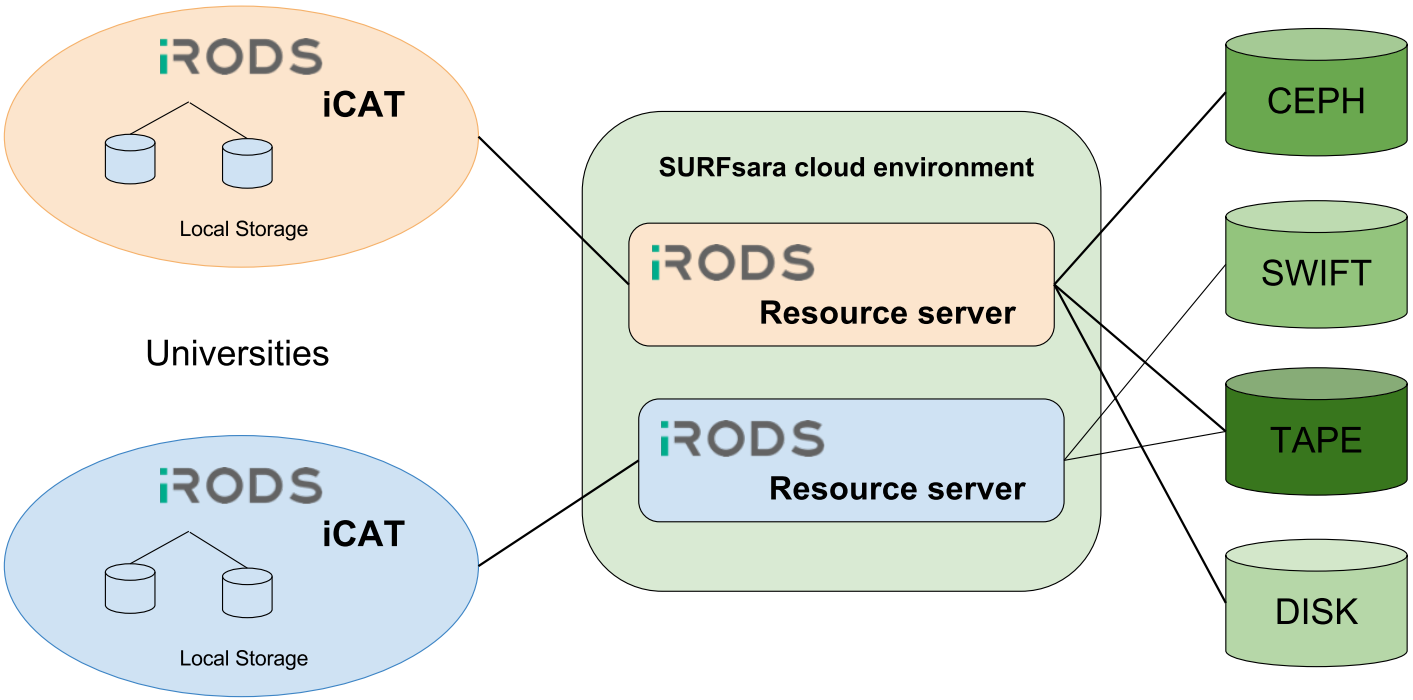
- Dutch universities host research data management platforms based on iRODS
- iRODS instances need to grow with the user base and number of projects
- More storage is needed, different storage systems:
  - Cheap storage for archiving, still managed by iRODS
  - Seamless integration of different storage systems

  → How to scale out storage to a national provider such as SURFsara?

Supported scenarios

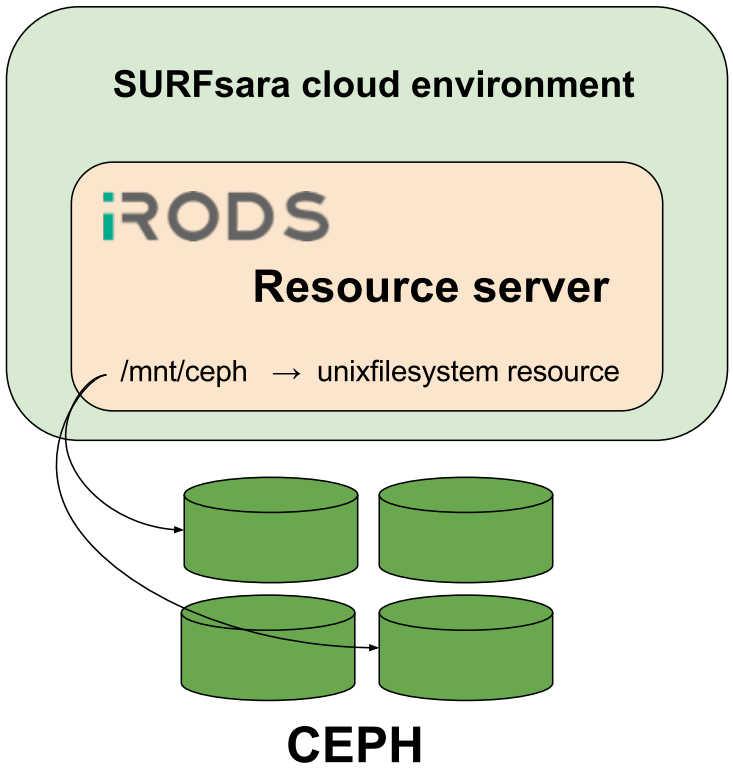
- Data replication
  - Copy of data at SURFsara
  - Permanent copy: Disaster recovery
  - Temporary copy: Bring data closer to compute facilities
- Storage scale-out
  - Data is only located at SURFsara
  - Users work directly on that data through universities' iRODS instances

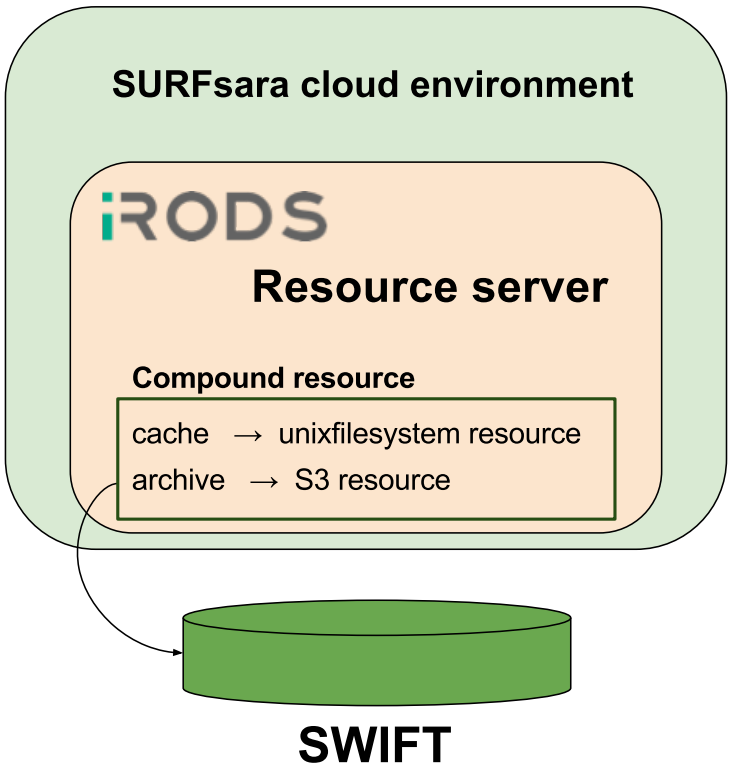
Technical setup



- Proof of concept architecture
- Is it feasible?
- How much effort does it cost the universities and SURfsara, can we provide patterns?

Implementation

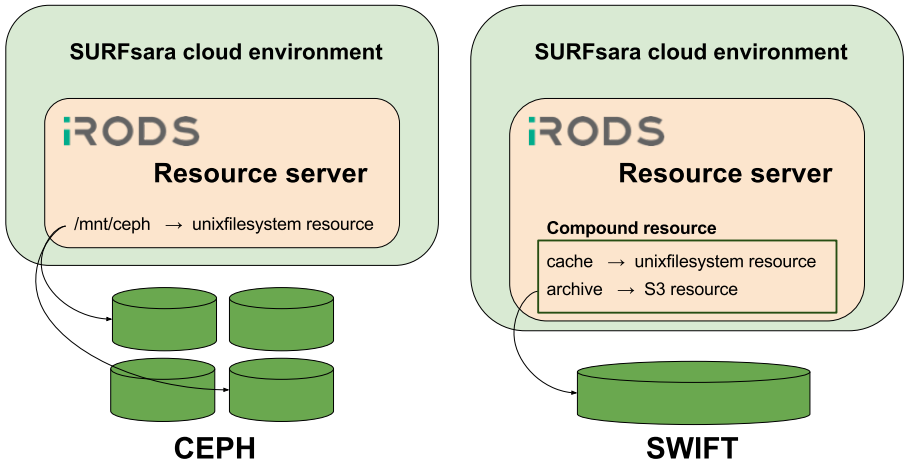




First order resource: File systems attached to VM (POSIX)

Compound resource: Archive environment (tape), SWIFT

Technical setup

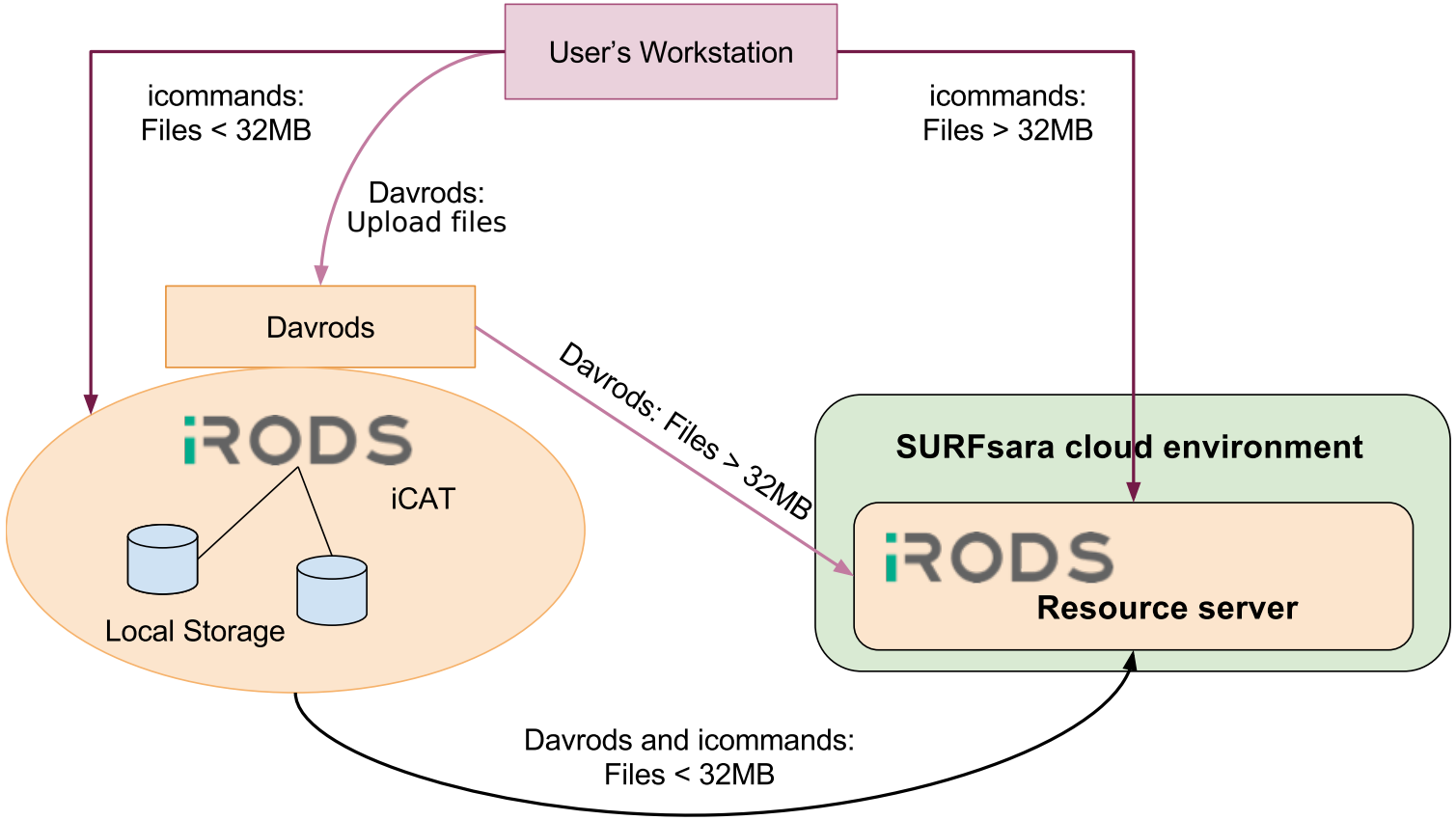


Proof of concept implementation:

- All storage systems can be made available to iRODS
- Effort to setup is moderate → tutorials

- Testing the CEPH resource:
  - Usability from work station with Davrods and icommands
  - Performance of data transfers from HPC
- Out of scope: Testing the throughput from cache to archive resource

Tests from a workstation



Mounting iRODS to workstation

- Running office applications

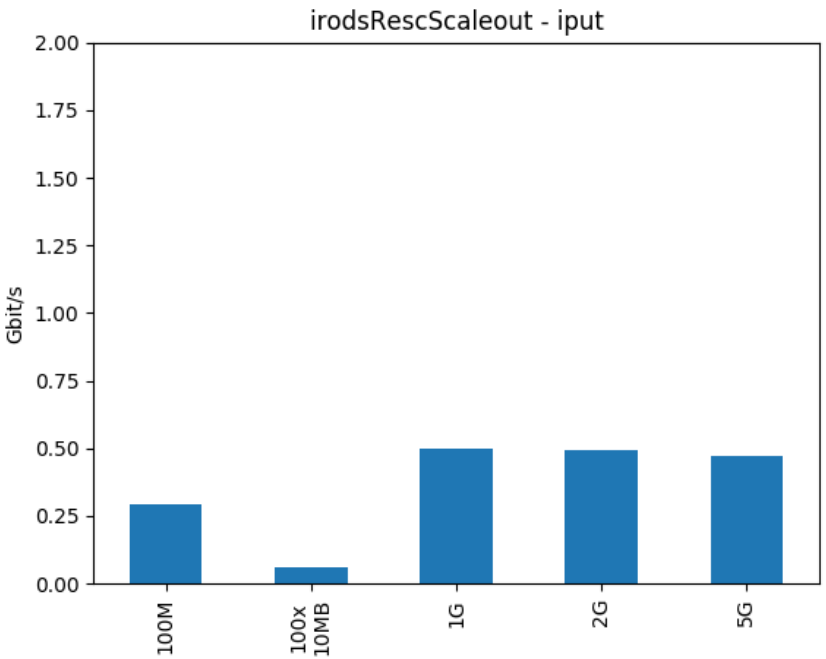
icommands

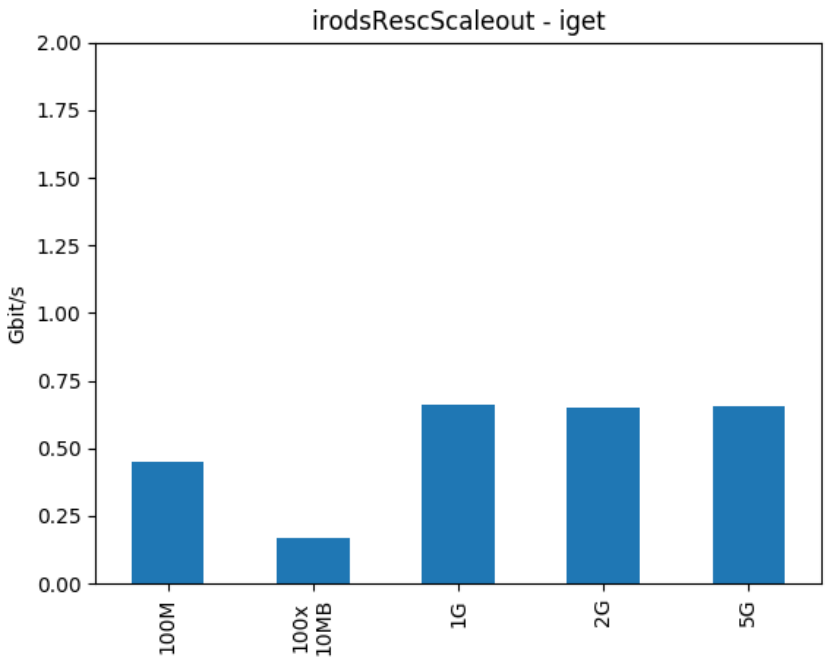
- Programmatic access

Test results: Office workflows

- Usability through Davrods from Windows 7 and Linux client
  - Opening ascii, PDF, spreadsheet files
  - Manipulating and storing ascii and spreadsheet files
- Response time of iRODS resource slightly slower than local access
- No major differences between Windows and Linux

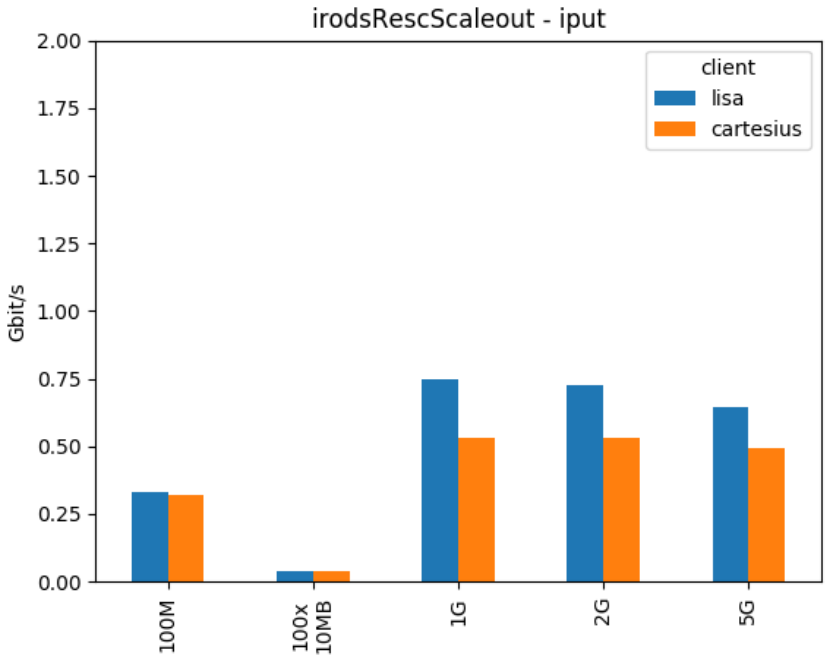
Test results: icommands

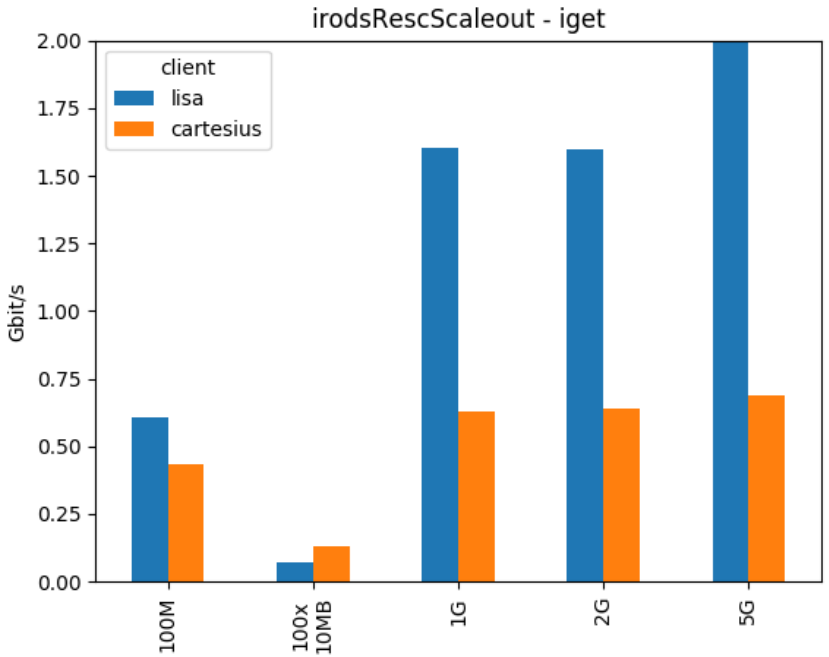




- Tested from workstation
- 100 $\times$ 10MB files in one folder
- Transferred with *iput -r*; no *-b* option
  $\rightarrow$ Causes a lot of overhead: setting up connection

Test results: icommands





- National super computer (orange)
- National compute cluster (blue)
- Difference due to different network environments

Other findings

- Network configuration
  - iCAT and iRODS resource servers need to be addressable with their fully qualified domain name
  - Configurations with load balancer in front of iRODS not fully supported $\rightarrow$ data ports are shielded
- Compound resources
  - Need careful setup
    - Capacity of cache
    - When can data be safely deleted from cache resource
  - Extra monitoring to prevent cache overflows
  - Impact on policies:
    - When can the user assume that data is stored safely on the archive resource?

Summary and Future work

- Performances and user experience looks acceptable
- Limited network configuration
- Need more tests
  - Test throughput from cache resource to different archive resources
  - Test real-life setting
    - Performance when many users access resources on iRODS resource server
  - iRODS federations as a different means to access storage at another site

Thank you! Questions?