Integrating Globus Online with the SDSC Cloud

Choonhan Youn
Chaitan Baru

San Diego Supercomputer Center
University of California, San Diego
Outline

• The SDSC Cloud
• Interfacing to Globus Online (GO)
  – Authentication
  – Swift-DSI module
• Summary
SDSC Cloud

- Announced in October 2011: For storage of academic, research datasets.
- Based on OpenStack Swift Object Store.
- Being used by projects for backup services as well as for object storage.

SDSC Cloud
~5.5 PB Initial Deployment
~2 PB Initial Data (2 and 3 copies)
~20 GB/s Aggregate I/O Throughput
SDSC Cloud Storage Services

- Default: Dual copy on different servers
- Optional: Remote, third copy
- OpenStack Swift Object Storage (Swift)
  - The object-based storage system and multiple interface methods make the system easy to use for the average user, but also provides a flexible, configurable, and expandable solution to meet the needs of more demanding applications.
- Provides academic and research partners with a convenient and affordable way to store, share, and archive data, including extremely large data.
- Linkage to public clouds
  - Fully compatible with the Amazon S3 interfaces
**SDSC Internal Development project to interface GO and SDSC Cloud**

- Develop an OpenStack Object Storage (Swift) Data Storage Interface (Swift-DSI) as an extension to the GridFTP server to enable the use of Globus Online (GO) with the SDSC Cloud.
- GO has not yet prototyped support for the OpenStack Swift Object Store, which powers the SDSC Cloud.
Authentication

• GO and the SDSC Cloud have to reconcile their security methods.
  – GO uses an X509 certificate-based system to identify users.
  – The SDSC Cloud uses Swift’s authentication system from the existing Rackspace architecture.

• Possible solution
  – The authentication/authorization part can be an external system or a subsystem run within Swift as WSGI (Web Server Gateway Interface middleware, and be plugged into the proxy server in Swift.
  – The Swift server will be accessed by the GO GridFTP server via the Swift protocols using GSI authentication.
Data Storage Interface (DSI) module

- Need to interface the GridFTP server to the Swift object store.
- Use GO Data Storage Interface module
  - DSI presents a modular abstraction layer to any storage system, providing functions for reading and writing data to and from the network.
- The Swift-DSI component would be the bridge between GridFTP and Swift
  - All of operation requests and data are routed through this component.
  - The GridFTP server sends requests to this Swift-DSI module, which translates these requests into RackSpace/Swift API calls, which are a subset of Amazon’s S3 API.
  - GO contacts this GridFTP server to access data in the SDSC Cloud.
Summary

- Develop the authentication software framework for the SDSC Cloud and build a Swift-DSI for read/write access to data in the SDSC Cloud.
  - This functionality would allow an existing application that uses a GridFTP compliant client to also utilize SDSC Cloud data resources.
  - More users would more easily be able to use the SDSC Cloud—as an active object store; for archival storage; and for data backups
- Can be used to support the “long tail” of scientific data. Examples:
  - NMR Portal
  - EarthCube
Example: UCSD Cyber NMR Project
Example: EarthCube

Storage Cloud e.g. SDSC Cloud

Internet

GO