



Storage On The Grid



DISTRIBUTED COMPUTING SERVICES

TANYA LEVSHINA

Storage On The Grid

2

- There are various ways you can access Storage on the Grid
 - You have to figure out what is the best fit for your application
- You may encounter a lot of problems and limit your choice of available sites if you are planning to:
 - transfer a large file > 10GB
 - have a large number of input/output files per job
 - expect POSIX access to data on a site

Storage On The Grid – Usage Patterns

3

- **Usage Patterns:**
 - Simulation: small input, big output
 - Searches and Analysis: big input, small output
 - Data processing: about the same input and output
- **Before making any decision you have to determine:**
 - How much total space do you need for your data?
 - How long do you want to keep this data in storage?
 - How much data is read by an individual job?
 - How is the input data for an individual job subdivided into individual files?
 - What kind of output data do you produce, and how much per job?
 - How do you keep track of input and output data?
 - Where do you want to ship your output data?

Storage On The Grid – Job Sandbox

4

- **Condor-G:** Input and output files are transferred from a submission node to a gatekeeper node and then to the worker nodes. Reasonable for file size <10 MB and small number of files. May produce heavy load on CE.
- **GlideinWMS:** Input and output files are transferred directly from a submission node to the worker nodes. Can overload network bandwidth between your host and a worker node. Sends files to WN all at once, possibly incurring in local disk space limitations. Reasonable for file size < 1GB.

Storage on the Grid – Classic SE

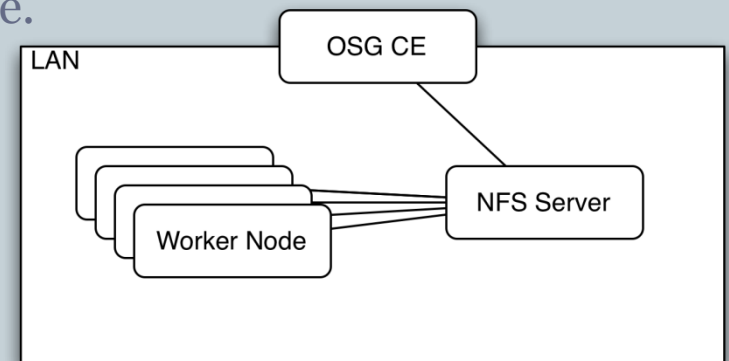
5

- **Classic Storage Element**

- `$OSG_APP` - shared area, used for application installation, read access from a worker node. In the process of being replaced with CVMFS (Dave Dykstra's talk)
- `$OSG_DATA` - shared area, in most cases it is POSIX-mounted storage (usually NFS), for data that has a lifetime > job lifetime. Can be read-only on a worker node. Can be set to UNAVAILABLE. Files could be uploaded by using `globus-url-copy/uberftp`. You need to know CE GridFTP server host, port and end point on a site you want to use.

- **Issues:**

- Scalability
- Difficult to manage space
- Not all the sites support `OSG_DATA`



Slide from B. Bockelman's talk at the OSG Summer School

Storage On The Grid - SE

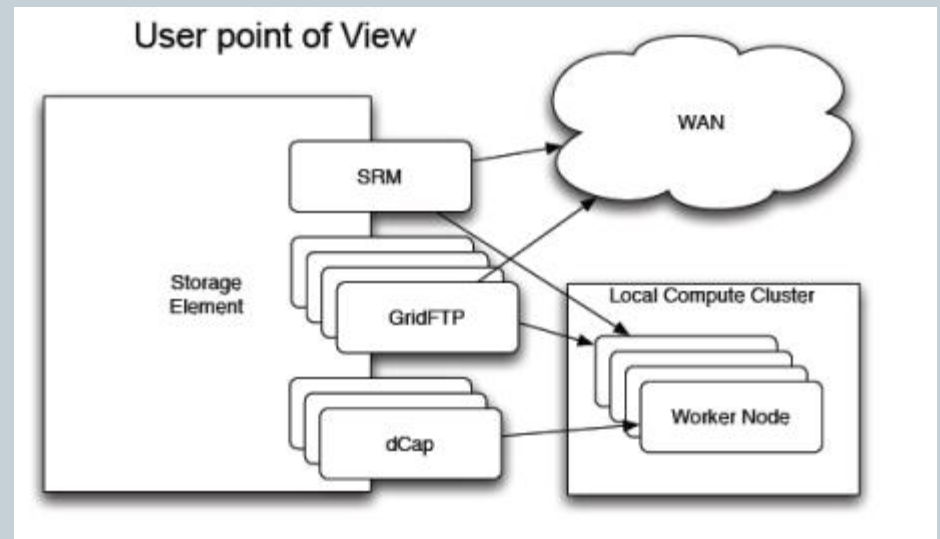
6

- A SE is a cluster of nodes where data is stored and accessed: physical file systems, disk caches, hierarchical mass storage systems.
- Most sites have at least one SE.
- SE software manages storage and enforces authorization policies.
- Scalability and capacity of a SE significantly varies from site to site.
- A user interacts with a SE via a get or put of the file.
- A SE doesn't necessarily provide POSIX access.

Storage On The Grid - SE

7

- A Storage Element (SE) is installed separately from Compute Element
- A typical SE has the following components:
 - ▣ Distributed File System
 - ▣ NFS, GPFS, PVFS, Lustre
 - ▣ HDFS, xrootd
 - ▣ dCache
 - ▣ GridFTP server(s)
 - ▣ Namespace service
 - ▣ SRM endpoint
(BestMan or SRM dCache)



Slide from B. Bockelman's talk at the OSG Summer School

Storage On The Grid – SRM clients

8

- Available from OSG repo
- SRM-Fermi-Client commands (srmcp)
- SRM-LBNL-Client commands (srm-copy)
- LCG-utils is a suite of client tools for data movement written for the LHC Computing Grid. (lcg-cp)
- Each has pros and cons, e.g:
 - Lcg-utils are most efficient and can deal with catalogs and bdi, but don't provide some useful commands like ping or rmdir.
 - Srm-LBNL commands are very verbose but handle errors and exit codes better than fermi client or lcg-utils

Storage On The Grid – Globus Online

9

- GO is a Software-as-a-Service facility that provides file transfer functionality.
- It will do a third-party transfer on your behalf:
 - Performs transfers of files
 - Retries in case of failures
 - Optimizes gridftp parameters for transfer
 - Provides CLI, Rest API, WEB UI
 - Handles certificate via MyProxy servers or uses your proxy certificate
- Issues:
 - SRM is not supported, have to explicitly list all gridftp servers you want to use. GO gridftp-servers load balancing could interfere with SRM load balancing.

Storage On The Grid – Globus Online

10

- To use GO you should:
 - Create public endpoints in GO
 - Prepare the list of files you want to transfer from A to B
 - Get proxy certificate and activate endpoint with your proxy or use myProxy server
 - Initialize the transfer using UI, CLI or Rest API
 - Monitor the progress of transfer
 - Get notification when job succeeded
- Globus Connect allows you to use Globus Online to share and transfer files to/from your local machine (such as a campus server, desktop computer or laptop) — even if it's behind a firewall. The system requires only outbound connections. Globus Connect is available for all major operating systems. It doesn't require root access for installation.

Storage On The Grid – Globus Online

11

The screenshot displays the Globus Online web interface in a Mozilla Firefox browser. The page title is "Transfer Files". The interface includes a navigation bar with "globus online" and "Sign Out" options. Below the navigation bar, there are two panels for managing transfers. The left panel shows a list of files and folders for a specific endpoint, including folders like "iosgvolengagefanya_test_globus" and "iosgvolengagefanya_test_1". The right panel shows a similar list for another endpoint, including folders like "ius/en/engage/1298498152-srmdlent-probe-test-directory.18925" and files like "ius/en/engage/1298332038-gridftp-probe-test-file-remote.30667".

Below the transfer panels, there is a table with columns for "Completion Time" and "Request Time". The table contains several rows of data, including a header row with "View 25 Records".

Completion Time	Request Time
04/29/2011 08:19 PM	04/29/2011 08:19 PM
04/29/2011 08:04 PM	04/29/2011 08:04 PM
04/27/2011 03:10 PM	04/27/2011 12:14 AM
04/27/2011 03:27 AM	04/27/2011 03:22 AM
04/27/2011 03:22 AM	04/27/2011 02:03 AM
04/28/2011 11:45 PM	04/28/2011 11:41 PM
04/28/2011 11:40 PM	04/28/2011 10:13 PM
04/28/2011 08:44 PM	04/28/2011 08:38 PM
04/28/2011 08:38 PM	04/28/2011 06:18 PM
04/28/2011 06:03 PM	04/28/2011 05:52 PM
04/28/2011 04:38 PM	04/28/2011 03:02 PM
04/28/2011 06:40 AM	04/28/2011 06:35 AM
04/28/2011 06:35 AM	04/28/2011 04:00 AM
04/28/2011 01:57 AM	04/25/2011 11:27 PM
04/25/2011 11:27 PM	04/25/2011 11:23 PM
04/25/2011 11:23 PM	04/25/2011 10:45 PM
04/25/2011 08:20 PM	04/25/2011 08:17 PM
04/25/2011 08:16 PM	04/25/2011 08:05 PM
04/25/2011 05:53 PM	04/25/2011 05:48 PM
04/25/2011 05:48 PM	04/25/2011 05:43 PM
04/25/2011 05:00 PM	04/25/2011 05:00 PM
04/25/2011 05:00 PM	04/25/2011 04:59 PM

Storage On The Grid – OSG Public Storage

12

- **Challenges:**

- Most of the OSG sites do not support dynamic storage allocation and do not have tools for automatic management of allocated storage
- the VOs that rely on opportunistic storage have difficulties finding an appropriate storage, verifying its availability and monitoring its utilization
- the involvement of a Production Manager, Site Admins and VO support personnel is required to allocate or rescind storage space.

- **Goals:**

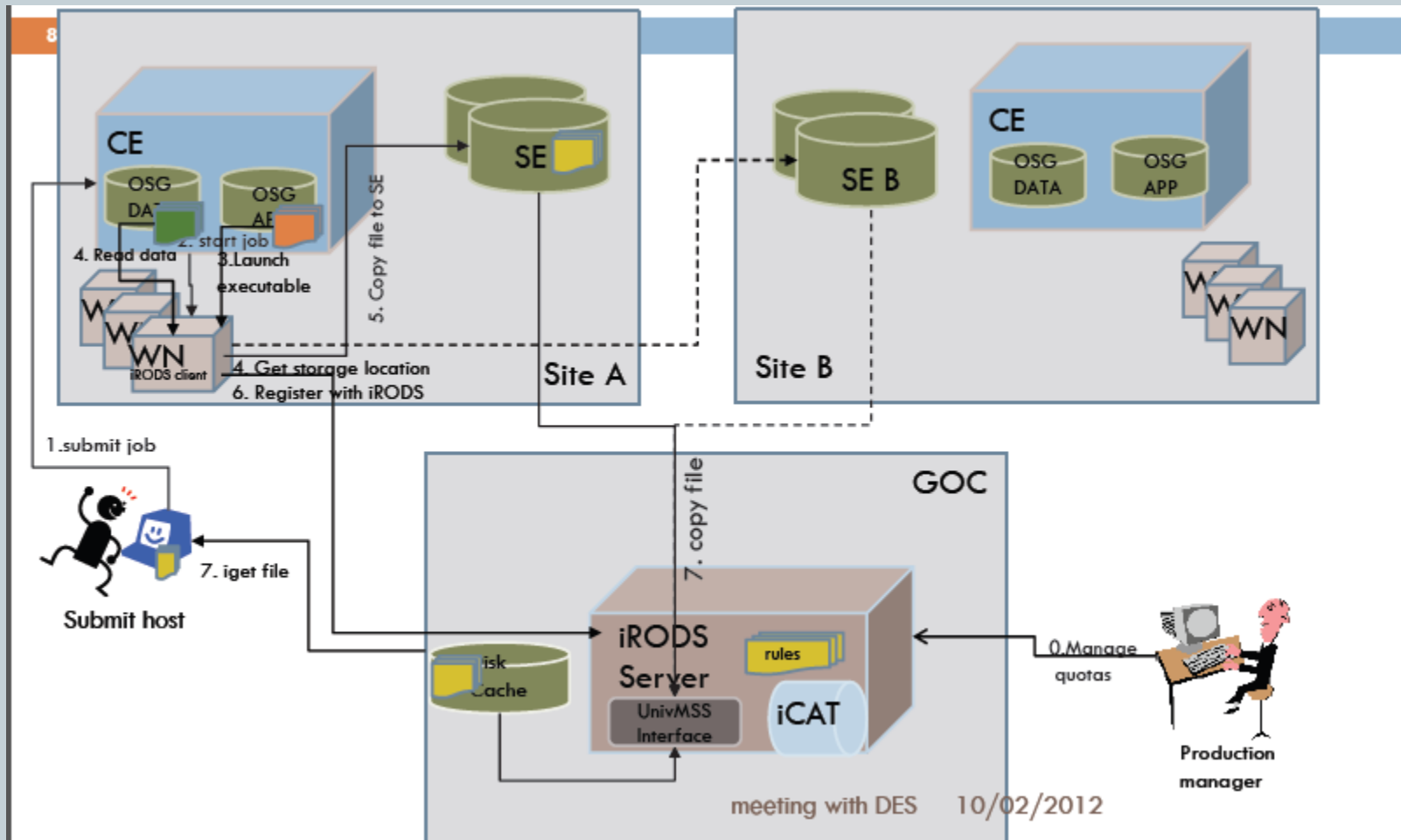
- Ease the task of VO data management

- **Requirements:**

- Allow the OSG Production manager to manage public storage allocation across all the participating sites.
- Impose minimal burden on the participating sites.
- Allow a VO Manager to manage data within VO quota
- Simplify storage selection for data storage.

Storage On The Grid – OSG Public Storage

13



Storage On The Grid – OSG Public Storage

14

- **Data Management:**
 - `iput -R Nebraska my_file`
 - `iput -R osgSrmGroup my_file`
 - `iget my_file`
 - `irepl-osg -R osgSrmGroup my_file`
 - `ils -l my_file`
- **Advantages:**
 - A user can pre-stage data to OSG_DATA, OSG_APP and SE SRMs via iRODS without dealing with sites, gathering scattered information about site resources, worrying about surl and end path
 - Global namespace that have information about files location, and size
 - Quota management
- **Disadvantages:**
 - File pre-staging is happening in two hops. Performance test has shown that irods client -> irods server transfer time is negligible comparing with time consumed by srm copy command. icp-osg command can be used to copy file directly to storage and register file in iRODS.
 - One can not utilize iRODS features fully because of the architecture we are using
- **Status:**
 - Preproduction. Have tested by several experiments including Snowmass

Storage On The Grid – Summary

15

- It's very important to understand your workflow and choose the right data management solution.
- Grid Storage is not always easy to access, be patient while debugging the problems. Usually, after fixing initial problems the data could be successfully moved to/from SE.
- GlobusOnline and OSG-iRODS could be used to simplify data management.

Storage On The Grid – Documentation

16

- Generic Storage documentation

<https://www.opensciencegrid.org/bin/view/Documentation/StorageOverview>

- Storage for End User

<https://twiki.grid.iu.edu/bin/view/Release/StorageEndUser>

- Client Tools

<https://www.opensciencegrid.org/bin/view/Documentation/Release3/LcgUtilities>

<https://www.opensciencegrid.org/bin/view/Documentation/Release3/FermiSrmClientCommands>

<https://www.opensciencegrid.org/bin/view/Documentation/Release3/LbnlSrmClient>

<http://www.globus.org/toolkit/docs/4.0/data/gridftp/rn01re01.html>

- Globus Online

<https://www.globusonline.org/>

- OSG Public Storage

<https://twiki.grid.iu.edu/bin/view/VirtualOrganizations/IRODSOSG>