

Public Storage on the Open Science Grid

Ted Hesselroth

Fermilab





Purpose

• Provide managed storage for non-owner VOs on Storage

Elements

- Public: "Maintained for or used by the people or community"
 - *Community* = OSG VOs
- Plan to develop and/or integrate software
 - Requirements have been reviewed.
 - Backend database and its operations have been defined.





Previous Users

- D0: Tevatron Experiment Monte Carlo production
- Fly's Eye: Highly Energetic Dispersed Radio Transients
- SCEC: Probabilistic Seismic Hazard Assessments
- Human Microbiome Project: Correlate microbiome/health
- LIGO: Einstein @ Home
- LIGO: Binary Inspiral
- LSST: Simulated images.





Current Procedure

- Contact OSG
- Discuss of Storage requirements
- Arrange space with sites
- Do test transfers
- Troubleshoot authorization or BDII info
- Upload files, notify by email
- Track uploaded files
- Run jobs





Current Procedure

- Contact OSG
- Discuss of Storage requirements
- Arrange space with sites
- Do test transfers
- Troubleshoot authorization or BDII info
- Upload files, notify by email
- Track uploaded files
- Run jobs





New Services

- Space Management needed for
 - Coordinate use of space in public storage areas by VOs
 - Track accessibility and use of public storage
- Transfer Management needed for
 - Enforcement of space management
 - Track files, cleanup
 - Storage resource selection





Constraints on the Design

- No alteration of Storage Elements
 - Access continues to be through current clients.
- No centralized OSG service
 - Software to be operated by VO
- Accommodate usage outside the service
 - Use of traditional means will not have an adverse effect





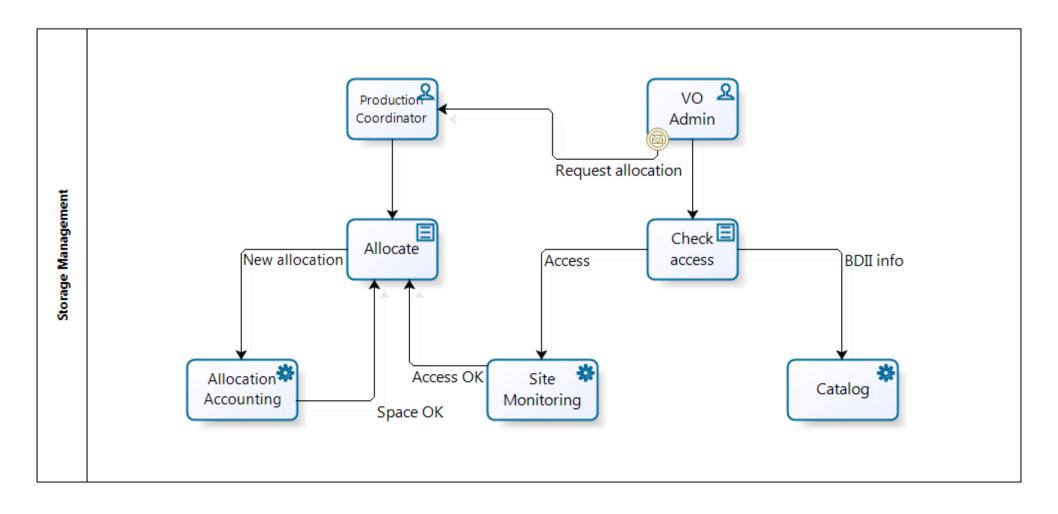
Public Storage Space Management

- Site administrator deploys a storage area and authorizes VOs
 - Puts storage area discovery info on OSG-central BDII.
- VO Administrator discovers and checks access to storage areas.
 - Requests an allocation
- OSG Production Coordinator allocates space on a storage area.
 - Allocation = Storage Area, Size, Max number of files, Expiration.
- User accesses a storage resource through its allocation.
 - Allocation parameters are enforced and updated.
- VO Administrator cleans up expired allocations.

‡ Fermilab



Storage Management Process



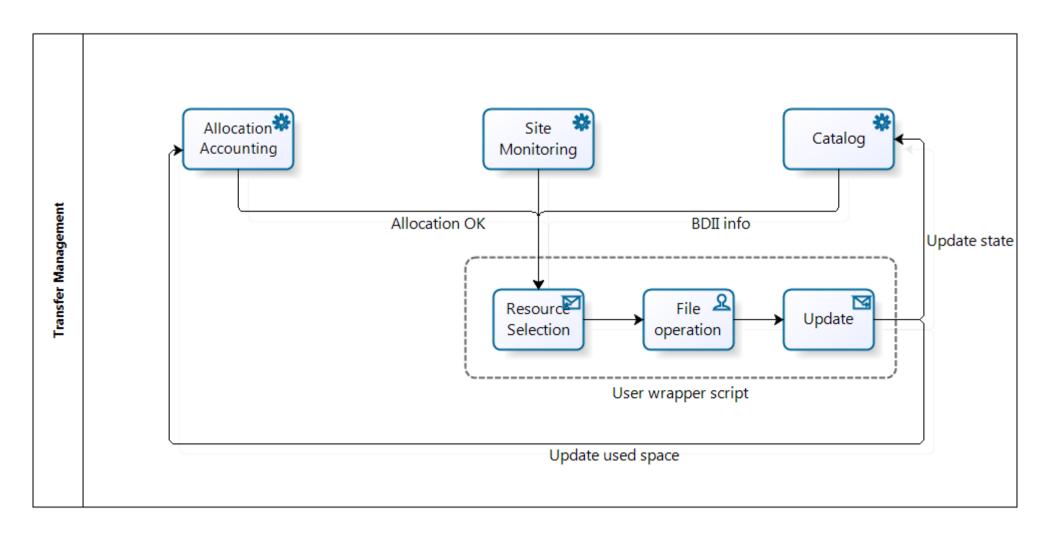




Fermilab



Transfer Management Process







Components of the Public Storage Service

- Allocation Accounting Service
 - Used to grant an allocation to a VO.
 - Tracks space used in allocations.
- Site Monitoring Service
 - Saves records of access check tools.
- Public Storage Catalog
 - Maps logical file name to location on storage resource.
 - Tracks replicas for access and cleanup.
 - Saves dataset definitions.





Support for Dataset Operations

- User defines a dataset.
 - By ancestor directory or metadata entry
- VO Administrator optionally makes suballocations for the dataset.
 - Same parameter types as for allocations.
- User uploads the dataset files.
 - All files to the same destination storage resource.
- User replicates the dataset files to another resource.
- User runs jobs on the dataset.
- User deletes expired datasets.

‡ Fermilab



Cleanup Policy

- Ensure availability of unused space over the long term.
- Expiration times for allocations and datasets.
- In order of preference
 - Users clean up datasets (delete files in them)
 - VO Administrators clean up allocations
 - Site administrators clean up allocations
 - At the request of the OSG Production Coordinator
 - Using a provided tool and input list





The Actual Schema

Public Storage probe_jd : INTEGER name : VARCHAR protocol_jd : INTEGER user_jd : INTEGER passed : BOOLEAN «PKsprobe_PK: probe_jd «FKsprobe_protocol_FK1: prof	Public Storage storage_resource storage_resource_id : INTEGEI storage_access_id : INTEGER storage_resource_handle : V/ default_dataset_replica_id : IN «PK»storage_resource_PK: st «UK»storage_resource_UK1: : «UK»storage_resource_UK2: ; «UK»storage_resource_UK3: : «FK»storage_resource_datas; storage_resource_INDEX1: st	Public Storage storage_area storage_area_id: INTEGER storage_resource_id: INTEGER storage_area_name: VARCH/ total_space: BIGINT used_space: BIGINT non_allocatable_space: BIGIN allowed_num_files: INTEGER used_num_viles: INTEGER 01 allowed_num_files: INTEGER used_num_viles: INTEGER 01 non_allocatable_num_files: IN allowed_connections: INTEGER 01 storage_area_profile: INTEGER 01 storage_area_UK1: storag eVK>storage_area_UK1: storag eUK>storage_area_UK2: storag	01	Public Storage allocation allocation allocation allocation intersection allocation allocation allocation type : VARCHAR acl_id : INTEGER total_space : BIGINT used_space : BIGINT used_space : BIGINT used_reserved_space : BIGINT allowed_num_files : INTEGER used_reserved_num_files : INTEGER update_time : TIMESTAMP VVITI reservation : VARCHAR parent_allocation : INTEGER allocation_INDEX1: allocation_j	01	Public Storage dataset_replica type : VARCHAR acl_id : INTEGER allocation_id : INTEGER expected_size : BIGINT current_size : BIGINT expected_num_files : INTEGER current_num_files : INTEGER state : VARCHAR update_time : TIMESTAMP WITH reserved : BOOLEAN dataset_id : INTEGER ePKsdataset_replica_PK: data: cUKsdataset_replica_UK1: alk eFKsallocation_FK: allocation_ eFKsdataset_replica_acl_FK1:	• 01	Public Storage dataset dataset dataset dataset dataset idataset dataset dataset vance dataset dataset dataset dataset expiration : TIME WITH TIME ZC state : VARCHAR update_time : TIMESTAMP WITI «PK»dataset_PK: dataset_id «FK»dataset_PK: dataset_id	Public Storage namespace file_id : INTEGER name : VARCHAR type : VARCHAR expected_size : BIGINT acl_id : INTEGER parent_id : INTEGER dataset_id : INTEGER update_time : TIMESTAMP VA «PK»namespace_PK: file_id «FK»namespace_acl_FK: ac «FK»namespace_dataset_F «FK»parent_FK: parent_id	 א
	01 01 Public Storage storage_access storage_access_id : INTEGER accessible : BOOLEAN storage_resource_id : INTEGEI «PK»storage_acces_PK: stora «UK»storage_acces_UK1: sto	Public Storage user user_id: INTEGER type: VARCHAR username: VARCHAR DN: VARCHAR FQAN: VARCHAR VO: VARCHAR state: VARCHAR update_time: TIMESTAMP WITI expiration: TIME WITH TIME ZC «PK»user1_PK: user_id 0.1	01	O1 Public Storage acl acl_id : INTEGER action : VARCHAR resource_id : INTEGER user_id : INTEGER user_id : INTEGER ePK-suser_PK: acl_id «FK-sacl_user_FK1: user_id O1 O1	01	* 01		Public Storage file_replica file_replica ifile_id : INTEGER acl_id : INTEGER acl_id : INTEGER dataset_replica_id : INTEGER current_size : VARCHAR state : VARCHAR update_time : TIMESTAMP WITI expected_checksum1 : BIGINT current_checksum2 : BIGINT checksum_type1 : VARCHAR expected_checksum2 : BIGINT checksum_type2 : VARCHAR «PKsfile_replica_PK: file_replica «FKsdataset_replica_FK: data:	01	





The Pigeon Access Checker

- Discover Servers
 - The test "generator" calls Discovery Tools on a daily basis to get a list of servers from BDII that advertise support for the VO.
- Create Tests
 - Using this list, create a custom suite of tests to verify the SRM/gridftp functionality of each site.
- Execute Tests
 - Execute the tests to verify the site's storage functionality.
- Show Results
 - Publish the results on a web page and in a XML.





Pigeon Screenshot

8	Mozilla Firefox	
<u>F</u> ile	<u>E</u> dit <u>V</u> iew Hi <u>s</u> tory <u>B</u> ookmarks <u>T</u> ools <u>H</u> elp	
4	🖙 🔹 🛞 👔 🚺 🔂 👘 ttps://gw014k0.fnal.gov:8443/gip/ 🖓 🕈 Google	Q
	tps://gw014k0I.gov:8443/gip/	

GLOW

lping GLOW	2010-08-27 09:18:01.767522	Success	Command finished.	<u>Archäve</u> <u>Create Länk</u>
2stncp GLOW	2010-05-27 09:23:02.252111	Pailure	Authorization Error	Archive GOC TICKET 8799
5xmkfile GLOW		Not. Run	Dependent, test, 2srmcp_GLOW has failed.	Archive Create Link
4zminkdir 61.0W		Not. Run	Dependent, test, 2srmcp_GLOW has failed.	Archive Create Link
5amlsdir GLOW		Not. Run	Dependent test 4srmmkdir_GLOW has not run yet due to dependencies.	Archive Create Link
özmendir GLOW		Not. Run	Dependent test 4srmmkdir_GLOW has not run yet due to dependencies.	Archive Create Link
7amm GLOW		Not. Run	Dependent, test, 2srmcp_GLOW has failed.	Archive Greate Link
globus url copy GLOW	2010-08-27 09:15:01.485599	Success	Command finished.	Archive Create Link
globus url run GLOW	2010-08-27 09:20:02.205427	Success	Command finished.	Archive Create Link

GRASE-CSE-MAGIC

globus url copy GRASE-CSE-MAGIC	2010-08-27 09:54:01.856264	Success	Commend finished.	Archive Create Link
globus url run GRASE-CSE-MAGIC	2010-08-27 09:39:04.892010	Success	Commend finished.	<u>Archive</u> Greate Link

C.Iduneed Centeri

🗲 Fermilab 🕽



The OSG Discovery Tool

- Reads BDII information
- Can search using XPath and XQuery
 - XML representation more future-proof than LDAP
 - Composition of complex queries is easier than with Idapsearch
- Prewritten Compute Element Searches
 - OS Version, Runtime Version, SE mount point
- Prewritten Storage Element Searches
 - SRM and gridftp URLs. Client command templates.
 - Storage Area available space and authorization info.





get_runtime_versions

• Finds software version with matching string

[discovery-1.0.6]\$ get_runtime_versions --vo engage --match "OSG 1.2."

Site Name	Compute Element ID	Software Version
BNL-ATLAS	gridgk01.racf.bnl.gov:2119/jobmanager-condor-default	OSG 1.2.9
BNL-ATLAS	gridgk02.racf.bnl.gov:2119/jobmanager-condor-default	OSG 1.2.9
CIT_CMS_T2	cit-gatekeeper.ultralight.org:2119/jobmanager-condor-default	OSG 1.2.8
CIT_CMS_T2	cit-gatekeeper.ultralight.org:2119/jobmanager-condor-osg_user	OSG 1.2.8
CIT_CMS_T2	cit-gatekeeper2.ultralight.org:2119/jobmanager-condor-default	OSG 1.2.8
CIT_CMS_T2	cit-gatekeeper2.ultralight.org:2119/jobmanager-condor-osg_user	OSG 1.2.8
GridUNESP_CENTRAL	ce.grid.unesp.br:2119/jobmanager-condor-default	OSG 1.2.11
MIT_CMS	ce01.cmsaf.mit.edu:2119/jobmanager-condor-default	OSG 1.2.3
USCMS-FNAL-WC1	cmsosgce3.fnal.gov:2119/jobmanager-condor-default	OSG 1.2.8
FNAL_FERMIGRID	d0cabosg2.fnal.gov:2119/jobmanager-pbs-osg	OSG 1.2.11
FNAL_FERMIGRID	fermigridosg1.fnal.gov:2119/jobmanager-condor-default	OSG 1.2.11
Firefly	ff-grid.unl.edu:2119/jobmanager-pbs-default	OSG 1.2.4
Firefly	ff-grid2.unl.edu:2119/jobmanager-pbs-default	OSG 1.2.4
Firefly	ff-grid3.unl.edu:2119/jobmanager-pbs-default	OSG 1.2.4





get_mount_path

• Finds the mount point of the Storage Element on worker nodes

[discovery-1.0.6] \$ get_mount_path --vo ligo

COMPUTE ELEMENT ID

osg-gw-2.t2.ucsd.edu:2119/jobmanager-condor-default osg-gw-2.t2.ucsd.edu:2119/jobmanager-condor-group_ligo cit-gatekeeper.ultralight.org:2119/jobmanager-condor-cms_monitor cit-gatekeeper.ultralight.org:2119/jobmanager-condor-default cit-gatekeeper2.ultralight.org:2119/jobmanager-condor-default cit-gatekeeper2.ultralight.org:2119/jobmanager-condor-osg_user ce.grid.unesp.br:2119/jobmanager-condor-default MOUNT POINT

/hadoop/ligo
/hadoop/ligo
/mnt/hadoop/osg/LIG0
/mnt/hadoop/osg/LIG0
/mnt/hadoop/osg/LIG0
/mnt/hadoop/osg/LIG0
/store/ligo





get_storage_area

- Finds total and free space, authorized VOs
- Can be used to find public storage areas

[discovery-1.0.6]\$ get_storage_area

STORAGE ELEMENT ID atlas.bu.edu	STORAGE AREA NAME Default Storage Area	TOTAL 122984	FREE 26264	VOs atlas,cdf,mis,ops,osg
atlas07.cs.wisc.edu	Default Storage Area	0	0	atlas
bsrm-1.t2.ucsd.edu	Default Storage Area	1010185	100463	accelerator,astro,atlas,
charm.ucr.edu	Default Storage Area	6730	1388	atlas,cdf,cms,engage,
cit-se.ultralight.org	Default Storage Area	526869	109617	DOSAR,LIGO,alice,atlas,
cmssrm.fnal.gov	flushPools:replica:online	6984011	1870180	cms,mis
cmssrm.fnal.gov	LFSOnlyPools:replica:online	529440	377902	cms,mis
cmssrm.fnal.gov	ResilientPools:replica:online	428079	195002	cms,mis
cmssrm.fnal.gov	<pre>stagePools:custodial:nearline</pre>	161695	37560	cms,mis
cmssrm.hep.wisc.edu	black-pools:replica:online	20186	11618	LIGO,atlas,cdf,cms,



Fermilab



Further Info

- OSG Twiki
 - OSG Public Storage
 - Opportunistic Storage
 - OSG Storage Discovery Tool
- Pigeon: http://home.fnal.gov/~dstrain/pigeon
- OSG docdb
 - OSG Public Storage Requirements
 - OSG Public Storage Design

