

OSG Storage Overview

Tanya Levshina



Talk Outline

- OSG Storage
 - Storage for VDT
 - Certification
 - Documentation
 - Support
- Storage Software in VDT
 - dCache
 - BeStMan “full-mode”/“gateway”
 - Xrootd
 - Hadoop
- Gratia Probes
 - dCache Probes
 - GridFtp Probe
- Summary

OSG Storage for VTD

OSG Storage for VTD is a well integrated distributed project between Wisconsin and Fermilab. The project is active for 2 years and is carried on by about 2FTE. Among other tasks we are responsible for:

- Helping to package storage software for VTD
 - Srm/dCache
 - BeStMan
 - BeStMan-gateway/Xrootd
 - BeStMan-gateway/Hadoop (NEW)
- Simplify configuration/installation for OSG
- Help VOs to use storage on OSG sites
- Develop and run validation tests
- Develop/maintain/package accounting and monitoring tools
- Provide users and administrators support
- Perform troubleshooting and debugging
- OSG liaison to storage developer groups
- Educate OSG community about storage, provide documentation
- Participate in grid schools organized by OSG

Certification

- Maintain test stands
 - 6 nodes test stand for dCache
 - 5 nodes test stand for BeStMan-gateway/DFS
 - 2U dual Intel Xeon Quad Core X5450 3.0Ghz 12M 1333Mhz Rack Server
 - 20 VMs
- Develop/run validation test suites before software is released to VDT
 - dCache test suite covers:
 - all srm-fermi-client commands
 - data replication
 - space management
 - information provider
 - S2 CERN tests
 - BeStMan testing provided by LBL
 - Site registration, daily test results - <http://datagrid.lbl.gov/osg>
 - Site could run tests with srm-tester-2 – instructions at <https://twiki.grid.iu.edu/twiki/bin/view/Storage/BeStMan>
 - BeStMan-gateway/DFS covers:
 - all supported srm-lbl-client /srm-fermi-client commands
 - Results of the tests are at
 - https://gw014k0:8443/validation_tests

Test Suites Results

OSG Storage Validation

dCache Tests	BeStMan Tests	Ganglia	Test Nodes Assignment
--------------	---------------	---------	-----------------------

OSG dCache tests	S2 tests	dCache Service Info	dCache SRMWatch
------------------	----------	---------------------	-----------------

DATE	RESULT
2009-05-11 17:39	Success
2009-05-05 11:28	Success

OSG Storage Validation

dCache Tests	BeStMan Tests	Ganglia	Test Nodes Assignment
--------------	---------------	---------	-----------------------

OSG dCache tests	S2 tests	dCache Service Info	dCache SRMWatch
------------------	----------	---------------------	-----------------

TEST NAME	START	END	RESULT	LOG
srmkddirtests	Mon May 11 16:58:56 2009	Mon May 11 16:59:18 2009	Success	log
srmcpwithouttokentests	Mon May 11 16:59:18 2009	Mon May 11 17:00:25 2009	Success	log
srmmvtests	Mon May 11 17:00:25 2009	Mon May 11 17:00:29 2009	Success	log
srm1stests	Mon May 11 17:00:29 2009	Mon May 11 17:03:07 2009	Success	log
srmgetpermissiontests	Mon May 11 17:03:07 2009	Mon May 11 17:03:27 2009	Success	log
srmsetpermissiontests	Mon May 11 17:03:27 2009	Mon May 11 17:03:36 2009	Success	log
srmcheckpermissiontests	Mon May 11 17:03:36 2009	Mon May 11 17:03:42 2009	Success	log
srmreservespacetests	Mon May 11 17:03:42 2009	Mon May 11 17:03:51 2009	Success	log
srmcpwithtokentests	Mon May 11 17:03:51 2009	Mon May 11 17:04:10 2009	Success	log

Find: Next Previous Highlight all Match case Phrase not found

OSG Storage Validation

dCache Tests	BeStMan Tests	Ganglia	Test Nodes Assignment
--------------	---------------	---------	-----------------------

OSG BeStMan	LBNL srmtester
-------------	----------------

MODE	FS	STATIC SPACE RESERVATION	SRM CLIENT TYPE	START	END	RESULT	DETAILS
gateway	HDFS	no	lbnl	Thu Jun 11 14:30:05 CDT 2009	Thu Jun 11 14:36:51 CDT 2009	success	details
gateway	HDFS	no	lbnl	Thu Jun 11 14:26:58 CDT 2009	Thu Jun 11 14:27:41 CDT 2009	error	details
gateway	HDFS	no	fermi	Thu May 28	Thu May 28		details

OSG Storage Validation

dCache Tests	BeStMan Tests	Ganglia	Test Nodes Assignment
--------------	---------------	---------	-----------------------

OSG BeStMan	LBNL srmtester
-------------	----------------

```

copy Put - Copies file without Space Token. No options
Executing command ...copy file:///tmp/test srm://gw018k1.fnal.gov:10443/srm/v2/server?SFN=/mnt/hadoop/fnalgrid/test_1244748605.25/test
srm-copy 2.2.1.2.14 Tue Apr 7 10:07:10 EDT 2009
SRM-Client and BeStMan Copyright(c) 2007-2009,
Lawrence Berkeley National Laboratory. All rights reserved.
Support at SRM@LBL.GOV and documents at http://datagrid.lbl.gov/bestman

SRM-CLIENT: Thu Jun 11 14:30:55 CDT 2009 Connecting to http://gw018k1.fnal.gov:10443/srm/v2/server
SRM-CLIENT: Thu Jun 11 14:30:55 CDT 2009 Calling SrmPrepareToPutRequest now ...
request.token=put:0
status=SRM_SUCCESS
explanation=null
SRM-CLIENT: Received URL=gsiftp://gw018k1.fnal.gov:5000//mnt/hadoop/fnalgrid/test_1244748605.25/test
SRM-CLIENT: Thu Jun 11 14:30:57 CDT 2009 start file transfer
SRM-CLIENT:Source=file:///tmp/test
SRM-CLIENT:Target=gsiftp://gw018k1.fnal.gov:5000//mnt/hadoop/fnalgrid/test_1244748605.25/test
SRM-CLIENT: Thu Jun 11 14:31:01 CDT 2009 end file transfer for file:///tmp/test
SRM-CLIENT: Thu Jun 11 14:31:01 CDT 2009 Calling putDone for srm://gw018k1.fnal.gov:10443/srm/v2/server?SFN=/mnt/hadoop/fnalgrid/test_1244748605.25
Result.status=SRM_SUCCESS
Result.Explanation=null
SRM-CLIENT: Request completed with success
SRM-CLIENT: Printing text report now ...

SRM-CLIENT*REQUESTTYPE=put
SRM-CLIENT*TOTALFILES=1
SRM-CLIENT*TOTAL_SUCCESS=1
SRM-CLIENT*TOTAL_FAILED=0
SRM-CLIENT*REQUEST_TOKEN=put:0
SRM-CLIENT*REQUEST_STATUS=SRM_SUCCESS
  
```

Find: Next Previous Highlight all Match case Phrase not found

LBL SRM Tester Results

Subscribe at

<https://hpcrdm.lbl.gov/mailman/listinfo/srmtester>

Test results at

<http://datagrid.lbl.gov/osg>

OSG SE SRM Daily Test Runs - Mozilla Firefox

http://datagrid.lbl.gov/osg/

OSG SE SRM daily test reports

[Monitoring Home](#) [Index@V11](#) / [Index@V22](#) [SDM](#) / [LBNL](#)

Affiliation: OSG_ITB_SRM_v11

Sites	Last Test	Last test runs	Archive
UCSD	06-19-2008 07:38	2, 5, 7, 14, 21	Archive
FNAL_FAPL_ITB_SE	12-04-2007 07:31	2, 5, 7, 14, 21	Archive
LIGO_CIT_ITB	06-19-2008 07:38	2, 5, 7, 14, 21	Archive

Affiliation: OSG_SRM_v22

Sites	Last Test	Last test runs	Archive
TTU_bestman	11-07-2008 09:00	2, 5, 7, 14, 21	Archive
TTU_SIGMORGH	11-07-2008 09:00	2, 5, 7, 14, 21	Archive
UCSD_dcache	11-07-2008 09:11	2, 5, 7, 14, 21	Archive
NERSC_bestman	11-07-2008 09:13	2, 5, 7, 14, 21	Archive
AGLT2	11-07-2008 09:11	2, 5, 7, 14, 21	Archive
UNL	11-06-2008 09:13	2, 5, 7, 14, 21	Archive
NERSC_PDSF_SRM	11-06-2008 09:13	2, 5, 7, 14, 21	Archive
PURDUE	11-06-2008 09:13	2, 5, 7, 14, 21	Archive
BNL_TEST_SE	11-06-2008 09:11	2, 5, 7, 14, 21	Archive
UC_ITB_SE	11-06-2008 09:14	2, 5, 7, 14, 21	Archive
CIT_CMS_T2	11-06-2008 09:18	2, 5, 7, 14, 21	Archive
NWICG_NotreDame	11-06-2008 09:30	2, 5, 7, 14, 21	Archive
FNAL_FERMIGRID_ITB_SE	11-06-2008 09:34	2, 5, 7, 14, 21	Archive
FNAL_GRIDWORKS	05-16-2008 10:31	2, 5, 7, 14, 21	Archive
SBGrid_EAST	11-06-2008 09:39	2, 5, 7, 14, 21	Archive
SBGrid_EXP	11-06-2008 09:39	2, 5, 7, 14, 21	Archive

Done

Start | OSG SE... | WRQ R... | dCache... | root@bg... | fg0x1 | root@f... | xterm (... | Downlo... | https://... | https://... | Present... | dCache... | 11:18 AM

Storage Documentation

- Revised documentation
- Main Page:
<https://twiki.grid.iu.edu/twiki/bin/view/Documentation/WebHome>
- Useful links under Storage Element administrators:
 - Opportunistic Storage/Space Reservation
 - Opportunistic Storage Model for USCMS
 - Gratia Storage Probes
 - Tools, Tips, FAQs
 - dCache Installation/references
 - BeStMan installation guides/references
- Minutes of the OSG Storage weekly meeting (includes list of all current tickets)
 - <https://twiki.grid.iu.edu/bin/view/Storage/MeetingMinutes>

Current Support (I)

- 0.25 FTE is allocated for support
- ~ 9 dcache and ~6 BeStMan/DFS Tier-2 sites
- Work with Tier-3 sites is just starting
- Tickets are submitted via OSG GOC or directly to osg-storage@opensciencegrid.org
- Often help comes from storage administrators and developers subscribed to the list
- The actual bugs are submitted to software developers' bug tracking systems by the group members
- About 10 tickets per week
 - 5/6 dCache related tickets
 - 4/5 BeStMan, Xrootd, etc
 - Expected to grow with inclusion of new software (hadoop, chimera, etc) and influx of Tier-3 sites

Current Support (II)

Types of support:

- Help with installation
- Help with storage configuration (including opportunistic storage configuration)
- Problems troubleshooting and debugging

Support Challenges

- Complicated, highly distributed services
- Huge variety of configuration options (software and hardware)
- Widely diverse utilization patterns
- dCache is known for poor error diagnostic, exception handling and propagation
- We do not have enough experience with Xrootd and Hadoop
- Lack of monitoring/diagnostic tools
- Support team does not have access to the service. Support personnel
 - Often are not authorized to use the service as user
 - Can not access site logs and configuration
 - Often can not access storage monitoring pages on the site

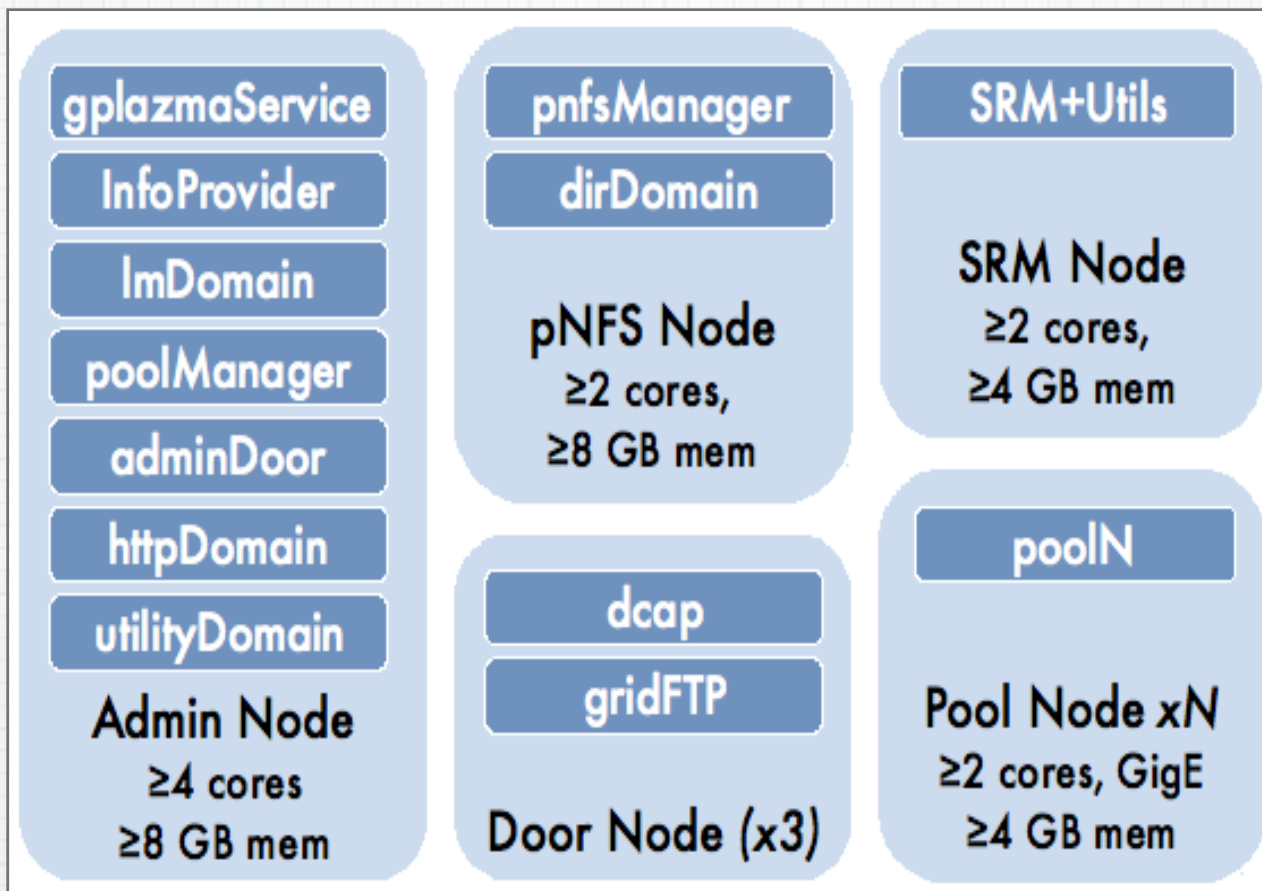
We would like to ask storage administrators for cooperation in:

- Notifying us about the reoccurring problems
- Provide us access to log files, configuration files

dCache in VDT

- dCache could be installed from VDT web page
 - <http://vdt.cs.wisc.edu/components/dcache.html>
- Current version is vdt 2.3.1 (dCache server 1.9.2-5) 2.3.1
- Distribution contains dCache-server, pnfs, postgres, srm-watch, gratia probes rpms and a configuration script tailored to set up dCache for Tier-2/Tier-3
- Configuration script allows to do system setup, enable opportunistic storage, replication etc
- dCache-clients are distributed as a part of VDT client cache
 - Fermi client
 - LBNL client
 - LCG-utils
- Installation documentation at <https://twiki.grid.iu.edu/bin/view/ReleaseDocumentation/DCache>

dCache OSG Tier-2 site Architecture



Slide courtesy of Ted Hesselroth (from presentation: "Installing and Using SRM-dCache")

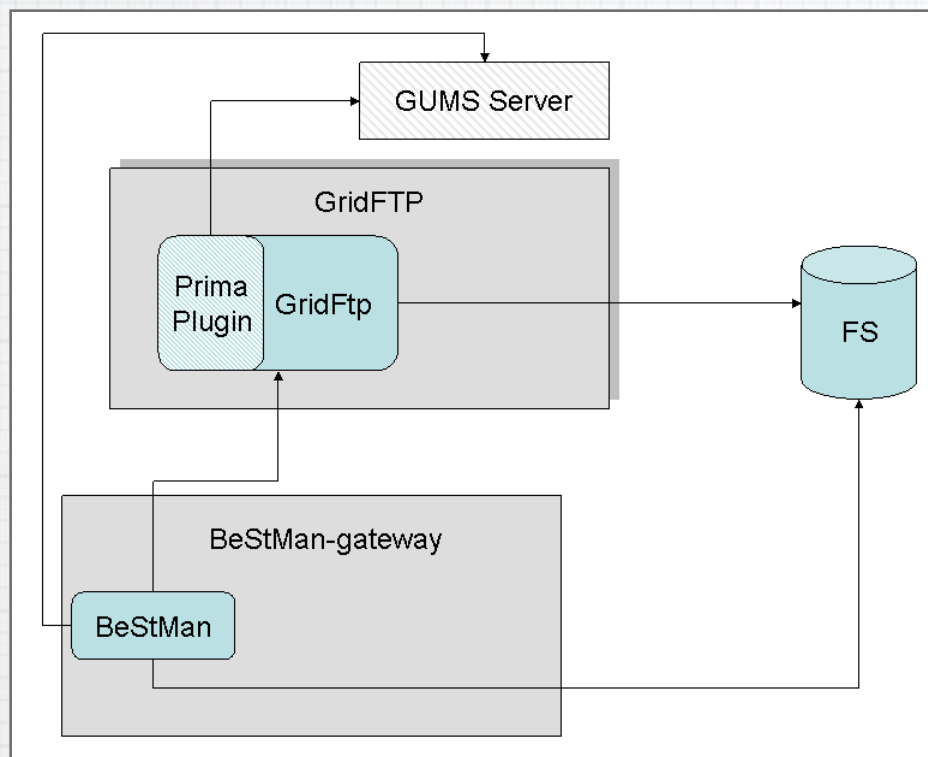
Opportunistic Storage

- Opportunistic Storage in dCache 1.8 with SRM 2.2
 - Provides a capability of specifying a portion of the total storage for opportunistic use
 - Allows particular VOs and Roles a privilege to use space other than that included in opportunistic storage
 - Files created through opportunistic use will not be permanently available in the storage system
 - A storage site administrator may configure the site for opportunistic use through space reservation.
 - Creation of space reservations is controlled by use of link groups
 - The administrator may assign storage pools to link groups
 - Certain pools are designated for opportunistic use.
- Numerous documents describing how to install and operate Opportunistic Storage on Tier-2 sites
 - <https://twiki.grid.iu.edu/twiki/bin/view/Storage/OpportunisticStorageSetup>
 - <https://twiki.grid.iu.edu/bin/view/Storage/OpportunisticStorageModelForCMS>

BeStMan in VDT

- Current version of software
 - BeStMan - 2.2.1.2.i6
 - Prima 0.7.1
- VDT configuration script tailored to set up BeStMan “full mode”/“gateway” for Tier-2/Tier-3
- BeStMan srm-clients are distributed as a part of VDT client cache
 - Fermi client
 - LBNL client
 - LCG-utils
- Installation Guide is at
 - <https://twiki.grid.iu.edu/bin/view/ReleaseDocumentation/Bestman>
 - <https://twiki.grid.iu.edu/bin/view/ReleaseDocumentation/BestmanOnCE>

BeStMan-gateway

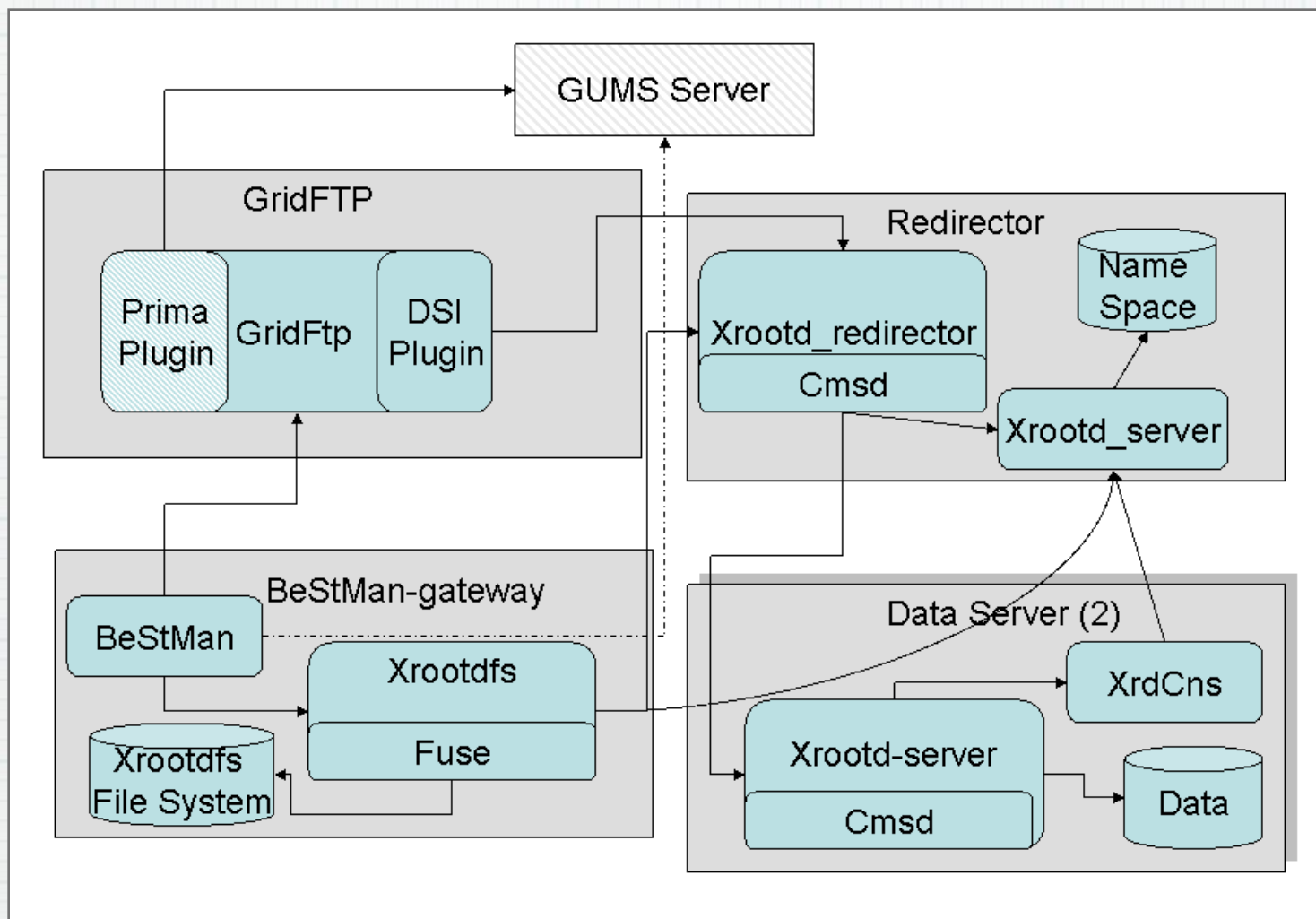


- Generic SRM v2.2 load balancing frontend for GridFTP servers
- Light-weight implementation of SRM v2.2 for POSIX file systems
 - srmPing
 - srmLs
 - srmRm
 - srmMkdir
 - srmRmdir
 - srmPrepareToPut (Status, PutDone)
 - srmPrepareToGet (Status, ReleaseFiles)
- Designed to work with any Posix-like file systems
 - NFS, GPFS, GFS, NGFS, PNFS, HFS+, PVFS, AFS, Lustre, XrootdFS, Hadoop
- Doesn't support queuing or disk space management
- Installation Guide at <https://twiki.grid.iu.edu/bin/view/ReleaseDocumentation/BestmanGateway>

BeStMan/Xrootd in VDT

- Current version of software
 - BeStMan - 2.2.1.2.i6
 - XrootdFS - 2.2
 - GridFTP-Xrootd ,xrootd-dsi-20080828-1632
 - Prima 0.7.1
 - Xrootd - 20080828-1632
- Installation Guide is at
<https://twiki.grid.iu.edu/bin/view/ReleaseDocumentation/BestmanGatewayXrootd>

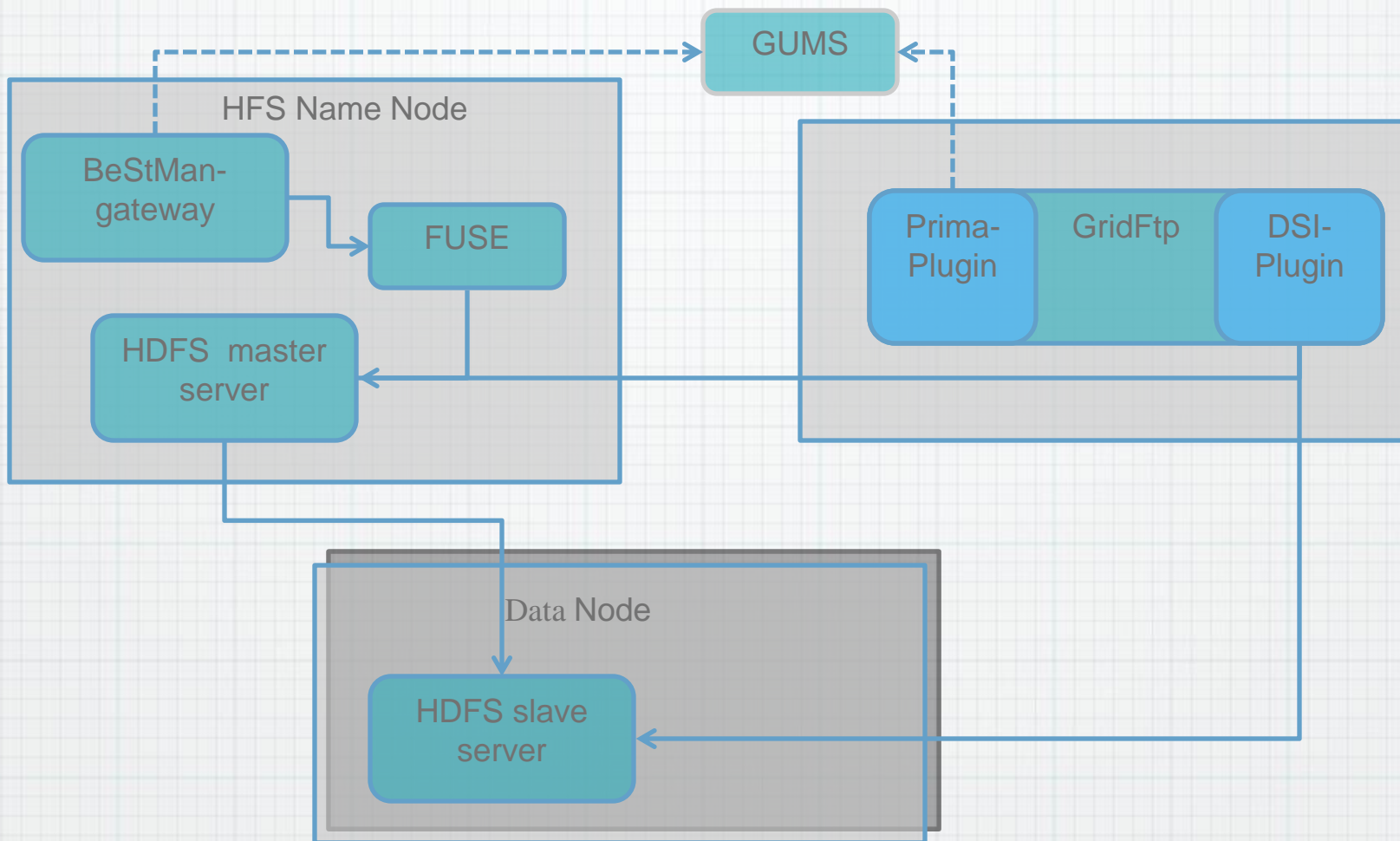
BeStMan-gateway/Xrootd Architecture



BeStMan/HDFS

- Hadoop Distributed FS is one of DFS that is considered by OSG to be accepted into VDT.
- It is currently available in UNL VDT cache
 - <http://t2.unl.edu/store/cache:Hadoop>
 - <http://t2.unl.edu/store/cache:Hadoop-Config>
- Current version of software
 - BeStMan - 2.2.1.2.i6
 - FUSE
 - GridFTP-Hadoop, hadoop-dsi
 - Prima 0.7.1
 - Hadoop
- VDT configuration scripts are not “user friendly” yet – some manual work is needed to change configuration

BeStMan-gateway/Hadoop Architecture



Gratia Service

Gratia is the accounting service for OSG is provided by the Gratia external project.

- Main goal is to provide the stakeholders with a reliable and accurate set of views of the Grid resources usage.
- Job and other accounting information gathered by Gratia probes run on the compute element or other site nodes are reported to a Gratia collectors
 - Fermi collector: <http://gratia-fermi.fnal.gov:8886/gratia-reporting>
 - OSG collector: <http://gratia.opensciencegrid.org:8886/gratia-reporting>
- Accounting records collected by Gratia are forwarded to the EGEE accounting system, APEL:
 - http://www3.egee.cesga.es/gridsite/accounting/CESGA/osg_view.html

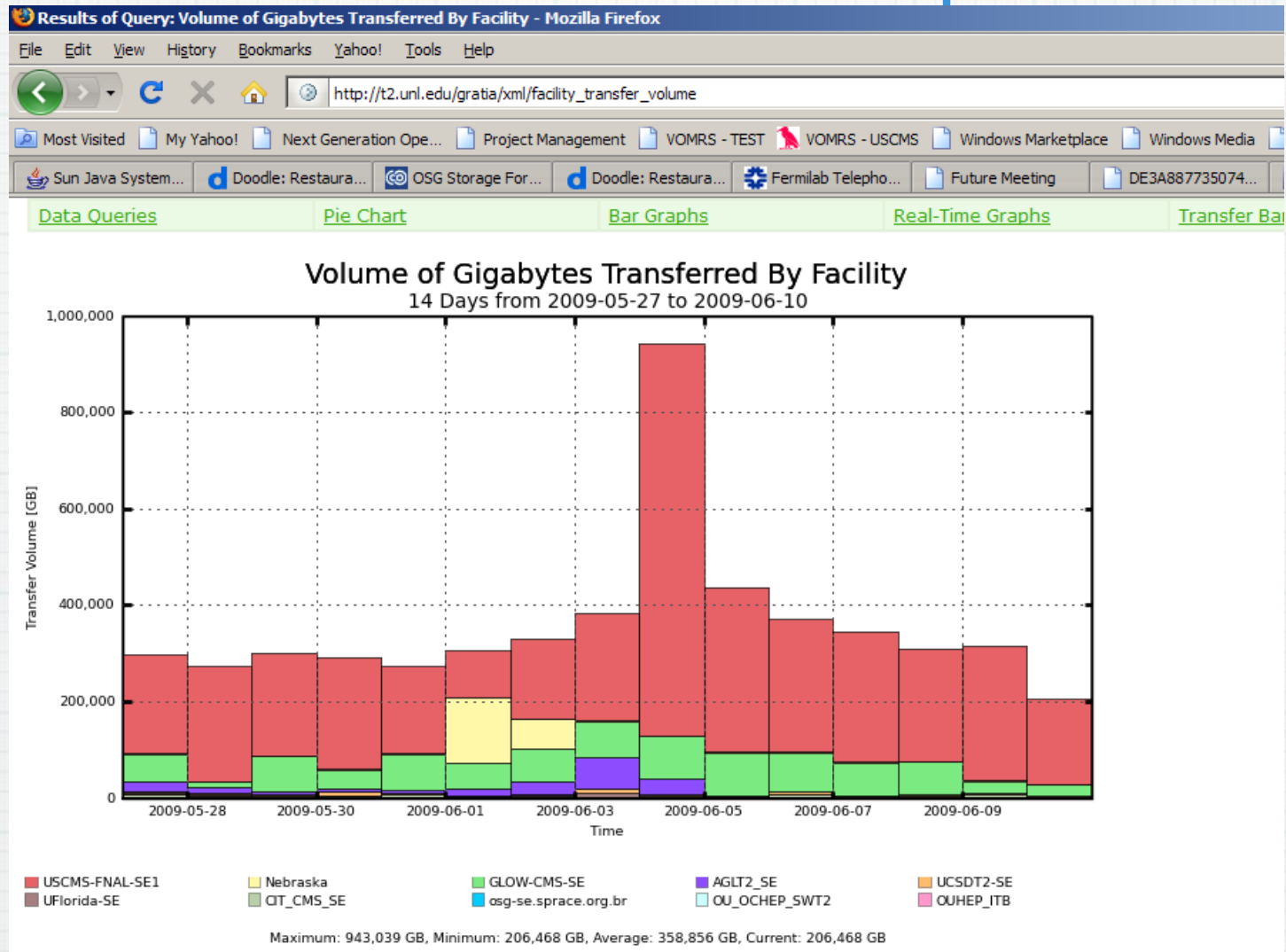
dCache Gratia Probes

- dCache Gratia Probes
 - Storage Probe
 - Transfer Probe
- Storage Probe
 - Is responsible for reporting storage capacity and storage usage
 - Gets the pool information from the dCache admin server
 - Gets the SRM information from the SRM tables in the SRM Postgres database
 - Runs as a cron job on the host running the Postgres database server for SRM
- Transfer Probe
 - Reports the details of each file transfer into or out of a dCache file server
 - Gets this information from the dCache "billing" database.
 - Runs as a daemon process
 - For performance reasons, sites with large dCache billing databases are advised to alter the "billinginfo" table by indexing specific tables in order speed up the search for newly added records
- Installation Guide is at
<https://twiki.grid.iu.edu/bin/view/ReleaseDocumentation/GratiaDcacheProbes>

GridFtp Probe

- Reports the details of each file transfer by GridFTP server
- Gets this information from the gridftp and gridftp-authorization logs
- Runs as a cron job
- Installation Guide is at <https://twiki.grid.iu.edu/bin/view/ReleaseDocumentation/GratiaTransferProbe>

Gratia Transfer Reports



Various report can be found at http://t2.unl.edu/gratia/xml/facility_transfer_volume

Gratia Custom Queries

Custom SQL Report

Enter below a SQL statement and press "Execute Query" to see the results.

```
select * from MasterTransferSummary where Probename like 'gridftp-transfer%unl%'
and StartTime = date(now())
```

Execute Query

Export Data (csv)

#	TransferSummaryID	StartTime	VOcorrid	ProbeName	CommonName	Protocol	RemoteSite	Status	IsNew	Njobs	TransferSize	TransferDuration
1	1,479,001	2009-06-26 00:00:00.0	3	gridftp-transfer:red-gridftp3.unl.edu	/CN=Carl Lundstedt 229191	gsiftp	gpn-husker.unl.edu	0	0	131	16,768	127
2	1,479,002	2009-06-26 00:00:00.0	3	gridftp-transfer:red-gridftp3.unl.edu	/CN=Carl Lundstedt 229191	gsiftp	gpn-husker.unl.edu	0	1	92	11,776	91
3	1,479,337	2009-06-26 00:00:00.0	3	gridftp-transfer:red-gridftp3.unl.edu	/CN=Carl Lundstedt 229191	gsiftp	cithep160.ultralight.org	0	1	1	2,147,483,647	264
4	1,479,338	2009-06-26 00:00:00.0	3	gridftp-transfer:red-gridftp3.unl.edu	/CN=Benedikt Mura	gsiftp	dcache-door-cms05.desy.de	0	0	4	2,147,483,647	5,953
5	1,479,339	2009-06-26 00:00:00.0	3	gridftp-transfer:red-gridftp3.unl.edu	/CN=Carl Lundstedt 229191	gsiftp	f01-120-117-e.gridka.de	0	1	1	2,147,483,647	439
6	1,479,376	2009-06-26 00:00:00.0	3	gridftp-transfer:red-gridftp3.unl.edu	/CN=Iban-Cabrillo-Bartolome	gsiftp	pool04.ifca.es	0	0	10	2,147,483,647	13,583
7	1,479,377	2009-06-26 00:00:00.0	3	gridftp-transfer:red-gridftp3.unl.edu	/CN=Iban-Cabrillo-Bartolome	gsiftp	pool03.ifca.es	0	0	6	2,147,483,647	9,236
8	1,479,378	2009-06-26 00:00:00.0	3	gridftp-transfer:red-gridftp3.unl.edu	/CN=Iban-Cabrillo-Bartolome	gsiftp	pool02.ifca.es	0	0	8	2,147,483,647	10,248
9	1,479,401	2009-06-26 00:00:00.0	3	gridftp-transfer:red-gridftp2.unl.edu	/CN=Jerome Pansanel	gsiftp	sbgse2.in2p3.fr	0	0	1	2,147,483,647	936
10	1,479,402	2009-06-26 00:00:00.0	3	gridftp-transfer:red-gridftp2.unl.edu	/CN=Carl Lundstedt 229191	gsiftp	gpn-husker.unl.edu	0	0	13	1,664	13
11	1,479,403	2009-06-26 00:00:00.0	3	gridftp-transfer:red-gridftp2.unl.edu	/CN=Benedikt Mura	gsiftp	dcache-door-cms01.desy.de	0	0	1	2,147,483,647	1,483
12	1,479,404	2009-06-26 00:00:00.0	3	gridftp-transfer:red-gridftp2.unl.edu	/CN=Carl Lundstedt 229191	gsiftp	gpn-husker.unl.edu	0	1	25	3,200	22

Query Gratia at <http://gratia.opensciencegrid.org:8886/gratia-reporting> to get specific information about file transfers

Summary

- We will continue to work on improving storage packaging in VDT
 - **Feedback is welcome!**
- We are trying to make support more efficient by providing FQA, debugging the most frequently occurred problems, working with developers on improving logging and error diagnostic
 - **The quality of the support depends greatly on Storage Administrators cooperation!!!**
- With addition of new software and influx of Tier-3 we have to figure out how to structure support so it could scale. We hope that Tier-3 will provide additional level of support.
- We are trying to maintain documentation up-to-date, adding new interesting references and “how to do” tips
 - **Please let us know if we are missing some important topics!**
- As a liaison to software developers we will be happy to pass your requests/suggestions