OSG STORAGE OVERVIEW

Doug Strain

OSG Storage Architecture

- ## Classic Storage Element
  - POSIX-mounted storage
  - Mounted and writable on the CE.
  - Readable from the worker nodes
  - Not-scalable under heavy load
  - High-performance FS is not cheap
  - Space management is not trivial
- ## Storage Element
  - SRM endpoint
  - Provides GridFTP Load balancing
  - Transfers via GridFTP servers
  - May provide internal access protocols (xroot, Posix)
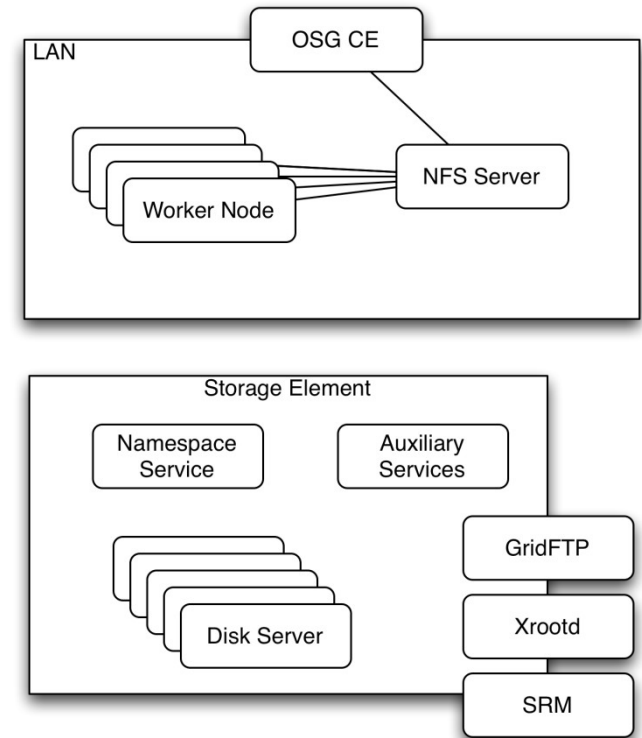
Pictures from B. Bockelman's presentation at OSS2010

VDT – RPM Transition

OSG software used to be provided via packaging manager (Pacman)

Now, transitioning to RPM yum repository

General Procedure

▫ Install epel-release (extra packages for linux)
▫ Install either:
  – vdt-release (VDT yum repo)
  – Or add hadoop repository (will be merged in ~6-12 months)
▫ Install storage software

Now is the time for comments / requests.

Bestman 2

- SRM service:
  - Provides support for the Storage Resource Management Protocol (SRM).
  - Provides access to an underlying storage solution (Hadoop, Xrootd, Lustre, etc)
  - Can access POSIX storage or GridFTP
  - Can load balance multiple (GridFTP) servers
  - Plugins for custom/special file systems
- **Implementation:** Jetty web server container
- **Authentication:** gridmap-file or GUMS

BeStMan Gateway vs Full-mode

- Gateway:
  - Light-weight implementation of SRM v2.2
  - Faster performance
  - Includes basic SRM functionality
  - Designed to work with any Posix system
  - Load balancing frontend for GridFTP servers
  - Static space tokens
- Fullmode:
  - Includes full SRM implementation
  - Includes Dynamic space reservation
 - Request queue management
  - Enhanced plug-in support
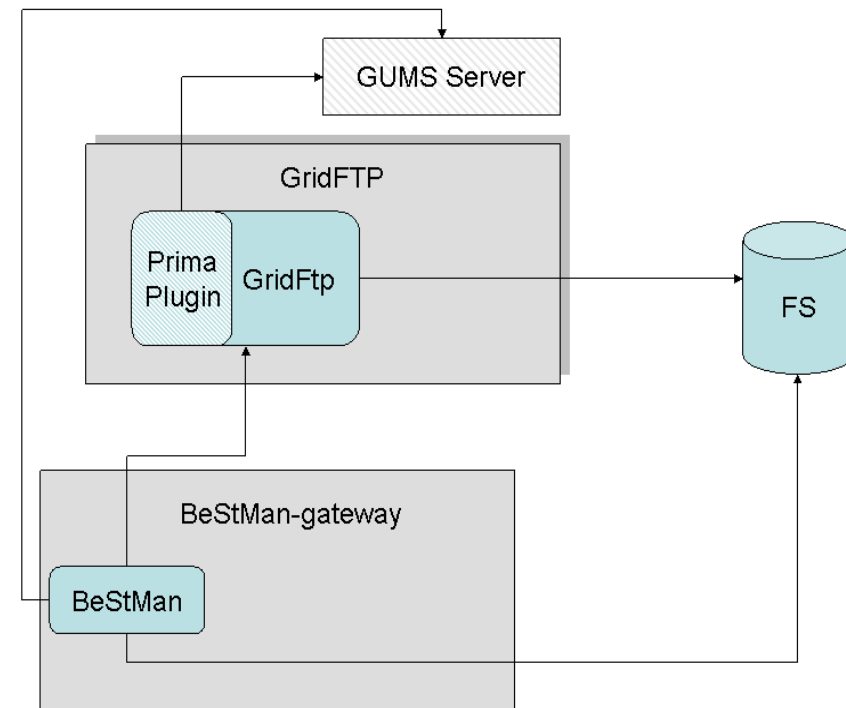
Simple Bestman Gateway Example

BeStMan-gateway/Xrootd

- **Xrootd** (SLAC/ CERN, others)
  - Provide access to storage with:
    - POSIX-like
    - via root framework (root://)
    - Native commands (xrdcp,…)
  - Allows formation of global clusters using a hierarchy of redirectors
  - Allows fallback to redirectors if file is not used.
  - Allows unix-like user/group authorization as well as X509 authentication.
  - Requires FUSE, XrootdFS to hook with BeStMan, GridFTP DSI plugin
- Currently used by most ATLAS and ALICE T2 sites , recommended for Atlas T3
- Can be installed via RPM
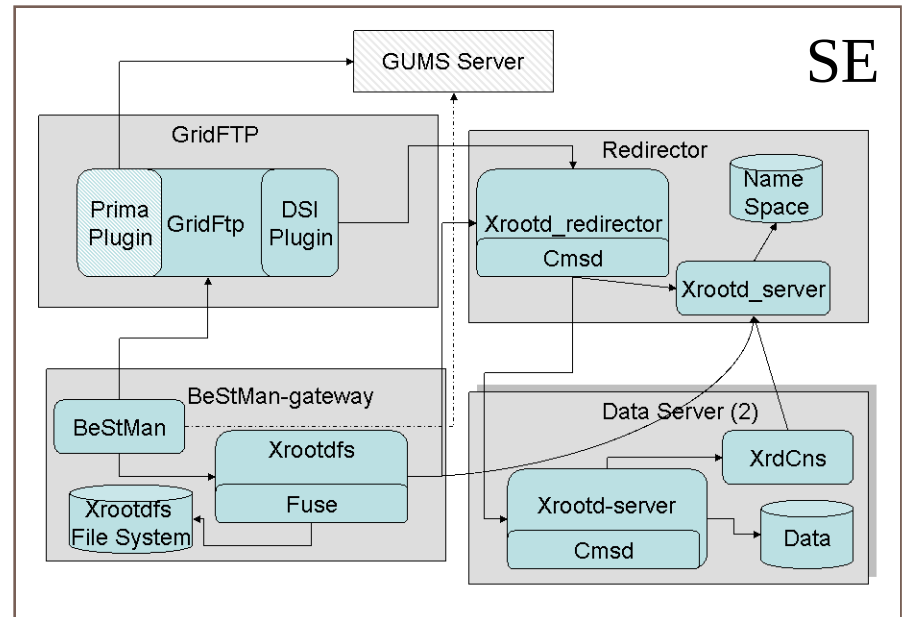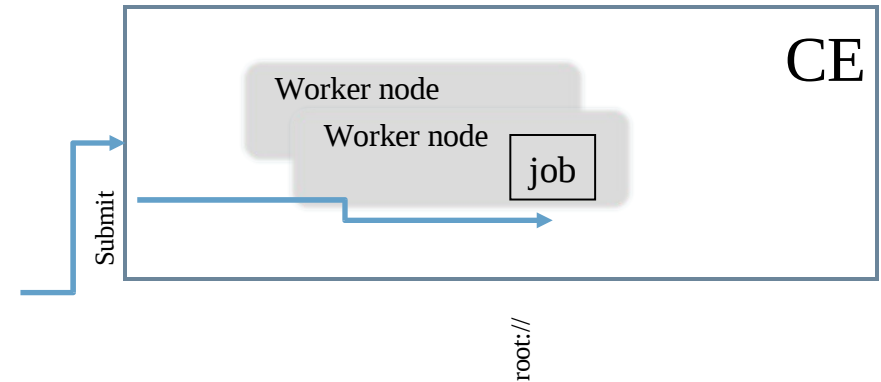
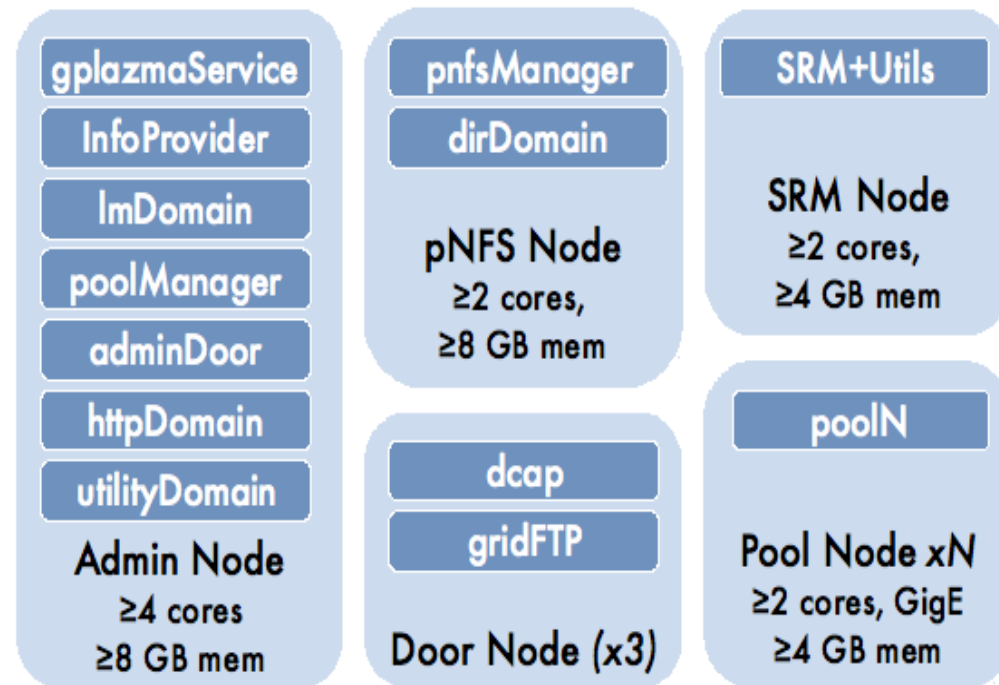# BeStMan-gateway/HDFS

- **Hadoop 0.20 DFS** - Apache project
  - Aimed for commodity hardware, highly distributed storage
  - Replication of data across nodes
  - Unix user/group authorization, For x509 authentication, BeStMan/GridFTP needed
- **Namenode:** central meta-data server contains file/directory information
- **Datanode:** Contains blocks of data distributed.
- **FUSE:** Provides posix access
- **BeStMan, GridFTP:** Requires FUSE to access HDFS w/plugin
- RPMS currently available in special Hadoop Yum repo (tutorial follows)
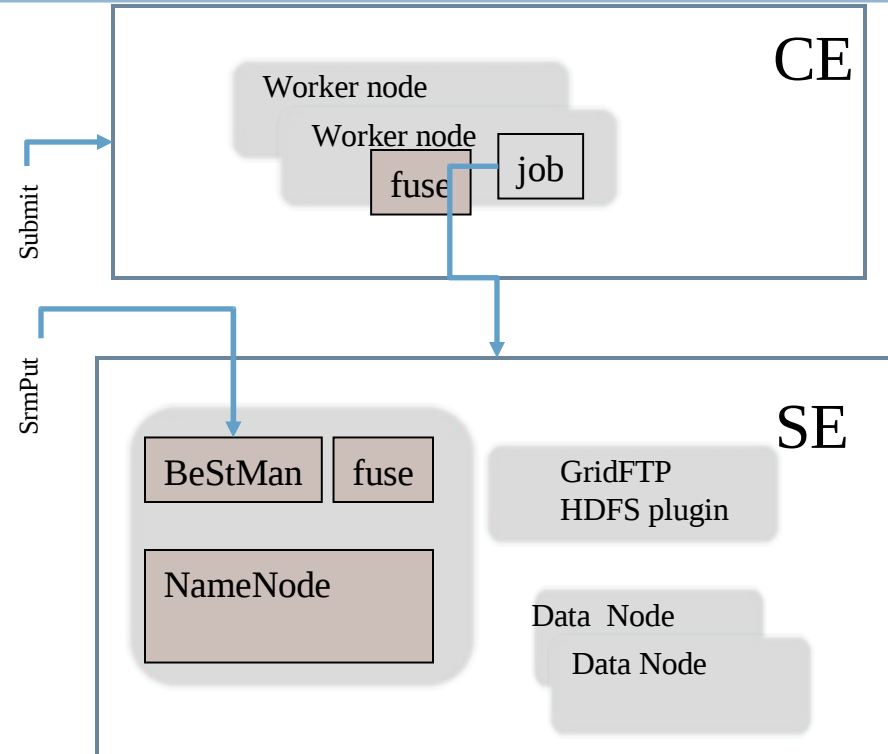
dCache

- **dCache** : distributed storage provides tertiary storage with SRM v2.2, NFS namespace, replication to tape, role-based authorization
  - May require more maintenance than T3 can reasonably support
- **dCache 1.9.5** supported and packaged by OSG-Storage
- dCache 1.9.12 now released but will not packaged by VDT
- dCache now adopted and supported by EMI.
- Going forward, dCache support will be best effort only. Community-supported.
- dcache.org for more information.

Open Science Grid

gplazmaService
InfoProvider
lmDomain
poolManager
adminDoor
httpDomain
utilityDomain

Admin Node
≥4 cores
≥8 GB mem

pnfsManager
dirDomain

pNFS Node
≥2 cores,
≥8 GB mem

dcap
gridFTP

Door Node (x3)

SRM+Utils

SRM Node
≥2 cores,
≥4 GB mem

poolN

Pool Node xN
≥2 cores, GigE
≥4 GB mem

Picture from Ted Hesselroth's
(from presentation: "Installing and Using SRM-dCache"

SRM Clients

- Available from VDT
  - (RPM: **osg-client**, **osg-wn-client**)
- SRM-Fermi-Client (RPM: **srm-client-fermi**)
  - developed and maintained at Fermilab
  - access any Storage Element with the SRM v1 or v2 specification
- SRM-LBL-Client (RPM: **bestman2-client**)
  - developed at LBNL,
  - access any SRM v2.2 based storage components
- LCG-utils client tools(RPM: **lcg-utils**)
  - Suite of tools written for data movement with the LHC
  - based on the Grid File Access Library,
  - access any SRM v2.2 based storage components
  - May use logical file names and require a connection to a BDII-based catalog for some commands file copies and deletions, which take endpoints based on the SRM URL.

Gratia transfer probes

- Included in BeStMan, dCache VDT Cache
- Reports to Gratia Accounting System
- Generates accounting information about file transfers, source, destination, size of the file and owner
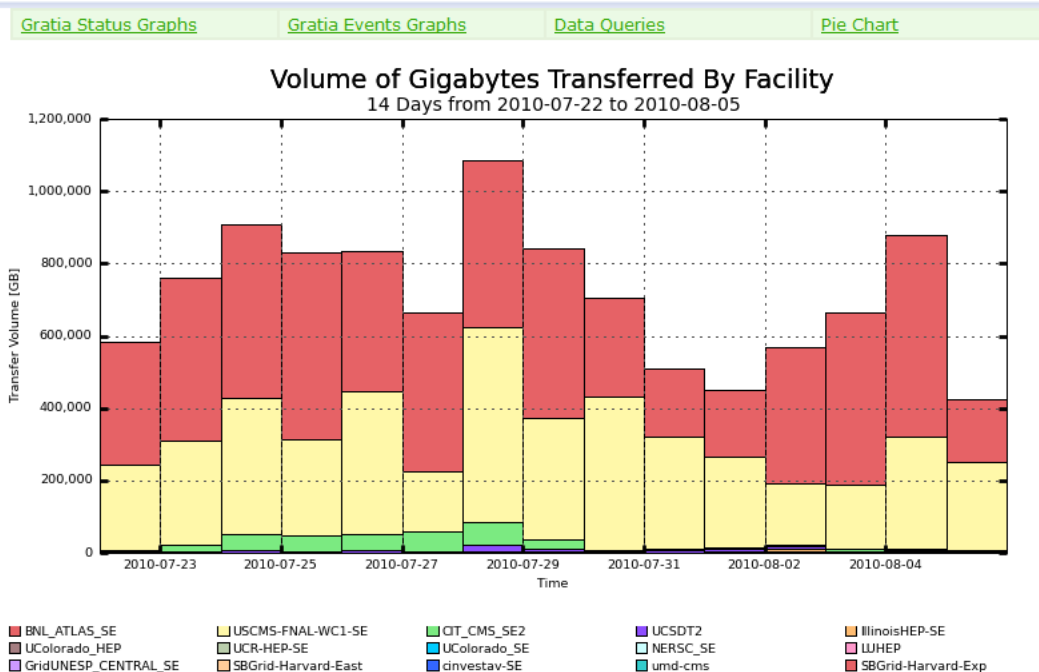
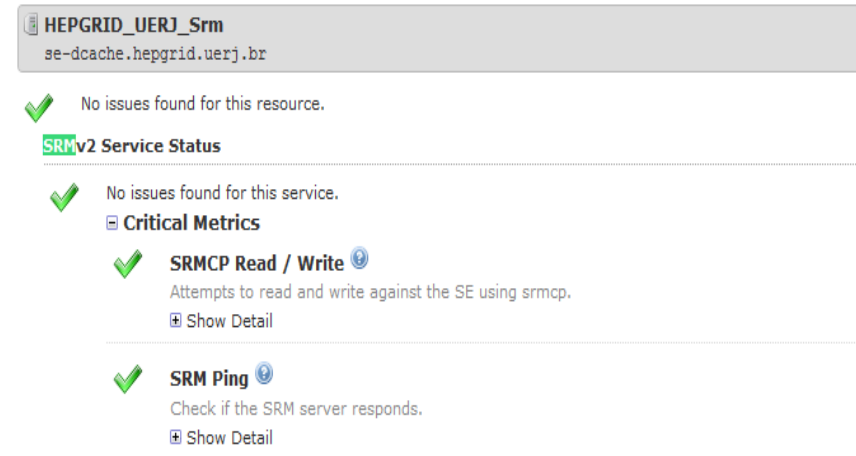http://t2.unl.edu/gratia/xml/facility_transfer_volume

RSV Storage Probes

The Resource and Service Validation (RSV) provides monitoring infrastructure for an OSG site admin.

- included with CE or as standalone

Storage RSV probes:

- **Default probes**: Srmping, srmcp
- **Storage client probes**: srmmkdir, srmrmdir, srmmv, space reservation
- **SRM tester probes**: Tests SRM based on lbnl srmtester application

Discovery and Pigeon Tools

- Discovery Tools
  - Get info about SE's that support a VO
  - Surl, end-path, space
  - Info from BDII

- Pigeon tools
  - Tests endpoints from Discovery Tools.
  - Helps debug site problems
  - Available as RSV Probes

OSG SE Statistics

Unofficial statistics based on BDII:
- Number of sites providing SE's: 48
- Number of sites running dCache: 12
- Number of sites running BeStMan: 36
  - HDFS 8
  - Xrootd 5
  - Lustre 3
- Number of sites reporting Gratia GridFTP Transfer Probes: 15
  - Daily transfer ~170000 files, 800 TB

Storage Documentation On OSG Twiki

- Release Documentation:

https://twiki.grid.iu.edu/bin/view/ReleaseDocumentation

- Main Storage Page:

  https://twiki.grid.iu.edu/bin/view/ReleaseDocumentation/Storage

  - End User Guide
  - Site Admin Guide

- Tier-3 specific documentation:

  https://twiki.grid.iu.edu/bin/view/Tier3/WebHome

- OSG Storage Group Meetings

https://twiki.grid.iu.edu/bin/view/Storage/MeetingMinutes