BeStMan/DFS support in VDT

OSG Site Administrators workshop
Indianapolis
August 6-7 2009

Fermilab

Tanya Levshina
tlevshin@fnal.gov
Fermilab

Storage in VDT

- BeStMan/FS
  - BeStMan-full mode
  - BeStMan-gateway
  - Xrootd
    - XrootdFS
    - GridFtp-Xrootd
  - Hadoop will be distributed via VDT by the end of 2009, could be currently installed via rpms from
    - http://newman.ultralight.org/repos/hadoop/
    - GridFtp-HDFS
  - Gratia transfer probe
- dCache

Why to use VDT?

- Provides means to easily configure and enable services
  - One method to start/stop/enable/disable for all services (vdt-control)
  - Common syntax for all configuration scripts
- Limits configuration options to most commonly used
- Provides support for:
  - handling CA, CRL, logs rotation
- Installs grid software, common libraries
  - Globus, GridFTP, GUMS-client, VOMS-client, SRM-clients
- Relatively easy way to upgrade software
  - Some problem exists with preserving configuration

When VDT is not enough

When you want to create more sophisticated configuration you should use the “native” configuration scripts.

- The following additional options could be specified for BeStMan:
  - Different path for cache, CAs directory
  - Different checksum type (e.g md5)
  - Change event log path or logging level
  - Change the number of concurrent file transfers
  - Change the number of grdiftp parallel streams, gridftp buffer size
- VDT provides very minimal xrootd configuration

BeStMan in OSG 1.2

- BeStMan could be installed from OSG:
  - pacman -get http://software.grid.iu.edu/osg-1.2:Bestman
- Installs the following services:
  - fetch-crl
  - vdt-rotate-logs
  - vdt-update-certs
  - gsiftp
  - gratia-gridftp-transfer
  - gums-host-cron
  - edg-mkgridmap
  - bestman
- Installs libraries (prima, globus, openssl)
- You can configure BeStMan in two modes:
  - Full mode - full implementation of srm protocol
  - Gateway mode – partial implementation of srm protocol (doesn't support dynamic space allocation)

BeStMan full mode in VDT

- Limited amount of options you can specify
  - --server <y,n> default “y” adds bestman start/stop in /etc/init.d
  - --user <bestman user> default “daemon”
  - --cert <bestman service cert> default /etc/grid-security/hostcert.pem
  - --key <bestman service key> default /etc/grid-security/hostkey.pem
  - --http-port <public port number> default 10080
  - --https-port <secure port number> default 10443
  - --globus-tcp-port-range <low_port,high_port> default none
  - --volatile-file-lifetime <lifetime in seconds> default 1800
  - --cache-size <Cache size in MB> default your file system size
  - --gums-host <GUMS hostname> default none
  - --gums-port <GUMS port number> default none
  - --with-transfer-servers <GridFTP server list> default localhost, port 2811
  - --with-allowed-paths <List of accessible paths> default “any”
  - --with-blocked-paths <List of non-accessible paths> default “none”

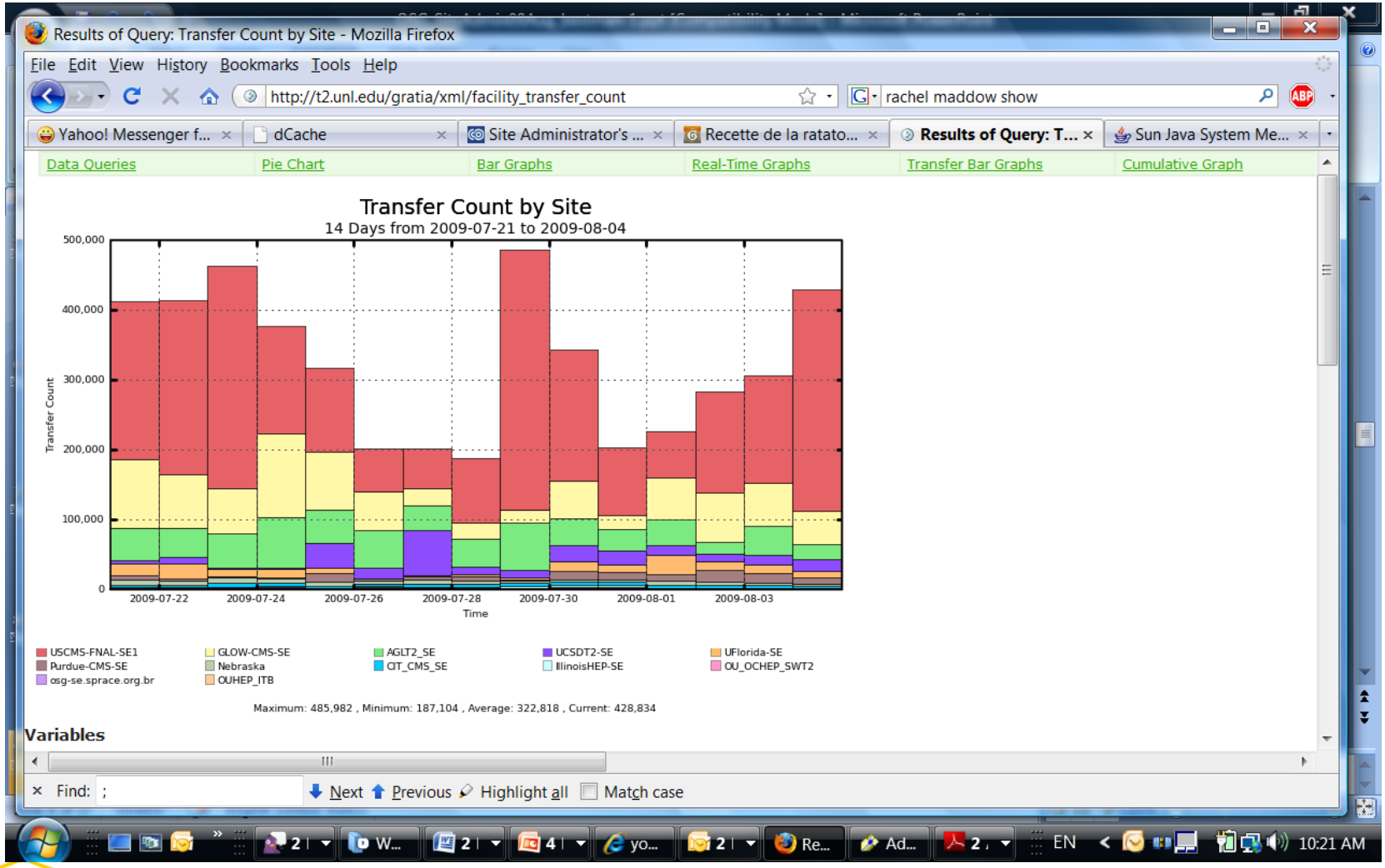
BeStMan-gateway mode in VDT

- Same options as for BeStMan-fullmode
- Additional options:
  - --enable-gateway # mandatory
  - --with-tokens-list <token-list> if you want to use static space reservation
- If you are planning to run BeStMan-gateway/Xrootd:
  - --use-xrootd instead of –enable-gateway

Gratia Transfer Probe

- Gratia is the accounting service for OSG.
  - Provides 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
    - OSG collector: http://gratia.opensciencegrid.org:8886/gratia-reporting
  - Records are forwarded to the EGEE accounting system, APEL:
    - http://www3.egee.cesga.es/gridsite/accounting/CESGA/osg_view.html
- 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

Gratia transfer sites report



Where to find configuration and log files?

- VDT installation and configuration log:
  - $VDT_LOCATION/vdt-install.log
- BeStMan
  - Logs: $VDT_LOCATION/vdt-app-data/bestman/logs/event.srm.log
  - Cache: $VDT_LOCATION/vdt-app-data/bestman/cache
  - Configuration: $VDT_LOCATION/bestman/conf/bestman.rc
- GridFTP
  - Logs: $VDT_LOCATION/globus/var/log
- Gratia probe
  - Logs: $VDT_LOCATION/gratia/var/logs
  - Configuration: $VDT_LOCATION/gratia/probe/gridftp-transfer/ProbeConfig
- Xrootd
  - Logs: $VDT_LOCATION/xrootd/var/log
  - Configuration: $VDT_LOCATION/xrootd/etc
- XrootdFS
  - Configuration: $VDT_LOCATION/xrootdfs/bin/start.sh

BeStMan and FS

BeStMan-gateway has been installed and is successfully working on the following file systems:

- NFS
- Ibrix
- PVFS2
- GPFS
- Xrootd
- Hadoop
- Lustre
- REDDnet

BeStMan-gateway/FS

- BeStMan-gateway/Ibrix (ATLAS T2 – Oklahoma University)
  - 3.2 GHz Xeon, 2 GB, 73 GB, 10 Gbps – + 12 TB Ibrix
- BeStMan-gateway/NFS (ATLAS T2 – Oklahoma University – ITB cluster)
  - 1.4 GHz P4, 512 MB, 20 GB, 100 Mbps
  - Using NFS-mounted ext3 file system as storage location

BeStMan-gateway/Hadoop (I)

- Caltech:
  - BeStman SRM server: 8 cores, 2.33GHz, 12GB RAM, 2x1 GbE ethernet, 4 x 750GB SATA drives
  - 4 GridFtp servers with 2 x 10GbE Name Node 8 cores, 16GB RAM (2GB for Name node jvm)
  - Data nodes: 82 data nodes, 277TB available space, batch worker nodes with fuse-mount
    - (62) 8 cores, 2.5GHz, 16GB RAM, 1GbE ethernet, 2 x 1TB SATA drives
    - (12) 8 cores, 2.33GHz, 12GB RAM, 1GbE ethernet, 4 x 750GB SATA drives
    - (4) 8 cores, 2.33GHz, 8GB RAM, 1GbE ethernet, 13TB hardware raid across 24 drives
    - (4) 8 cores, 2.33 GHz, 8GB RAM, 1GbE ethernet, 6.5TB hardware raid across 12 drives
    - (2) Sun x4500 Thumpers with 36TB zfs with 10GbE ethernet
- Nebraska
  - 3 BeStMan Servers
  - 5 GridFTP servers for HDFS; 1 with 10Gbps card, 4 with 1Gbps cards
  - Name node 8 core (2.2GHz Opteron) 16GB RAM
  - Data nodes: 140 nodes, ~350TB raw online currently, 500+TB soon, most are batch worker nodes
    - 53x 4 core 2.2GHz Opterons 4G RAM
    - 78x 8 core 2.2GHz Opterons with 16GB RAM
    - 2x 4 core 2.8GHz Opterons with 8GB RAM (Sun x4500 'Thumpers' with 48x 1TB disks)
    - 3x 4 core 2.2GHz Opterons with 8GB RAM
    - 2x 4 core 2.0GHz Xeons with 4GB RAM (SCSI attached vaults)

BeStMan-gateway/Hadoop (II)

- USCD
  - BeStMan 8 core, 8GB Mem, dynamic gridftp selector
  - Name Node 8 core, 16 GB Mem
  - Data Nodes 4/8 core, 8/16Gb mem, 1GB up-link: 15 data nodes, 42TB, batch worker nodes with fuse-mount
- LIGO:
  - 1 BeStMan, 1 GridFtp server
  - Data Nodes :13 storage nodes ,23TB, Batch worker nodes with fuse-mount

BeStMan-gateway/Xrootd

- SLAC
  - BeStMan dual 1.8 GHZ, mem 2GB, disk 1GB
  - 2 gridftp servers, they are dual AMD 2218 (2.6GHZ), mem 8GB,3x1GB
  - Redirector and CNS: dual 1.8 GHZ, mem 2GB , disk GB
  - Data servers, total space ~ 268TB usable

    3x Sun x4500 (thumper), dual AMD 285, 16GB, 4x1GB and 17TB usable space on ZFS

    7x Sun x4500 (thumper), dual AMD 290, 16GB, 4x1GB and 31TB usable space on ZFS
- BNL, IN2P3, INFN, FZK, RAL – all have just xrootd installation

BeStMan-gateway/Lustre

- LQCD – Fermilab
  - 65TB
  - 4 servers are used to serve 2 satabeast and the metadata server is c0-hosted on one of these nodes
- TTU

BeStMan-gateway/REDDnet

- Vanderbilt University
  - 2 quad-core Opteron CPU's, 16 GB RAM, 10 Gbit Ethernet
  - 1700 batch slots

Useful links

- BeStMan documentation
  - http://datagrid.lbl.gov/bestman
- Hadoop documentation
  - https://twiki.grid.iu.edu/bin/view/Storage/Hadoop
  - http://indico.fnal.gov/getFile.py/access?contribId=22&sessionId=24&resId=0&materialId=slides&confId=2538
  - http://indico.fnal.gov/getFile.py/access?contribId=3&sessionId=24&resId=0&materialId=slides&confId=2538
  - http://indico.fnal.gov/getFile.py/access?contribId=5&sessionId=24&resId=0&materialId=slides&confId=2538
  - http://indico.fnal.gov/getFile.py/access?contribId=4&sessionId=24&resId=0&materialId=slides&confId=2538
- Xrootd documentation
  - http://xrootd.slac.stanford.edu/
  - http://indico.fnal.gov/getFile.py/access?contribId=15&sessionId=26&resId=0&materialId=slides&confId=2538
- Lustre documentation
  - http://wiki.lustre.org/index.php/Main_Page
  - http://indico.fnal.gov/getFile.py/access?contribId=17&sessionId=27&resId=1&materialId=slides&confId=2538
- REDDnet web page
  - http://www.reddnet.org/mwiki/index.php/Main_Page
  - http://indico.fnal.gov/getFile.py/access?contribId=25&sessionId=25&resId=0&materialId=slides&confId=2538
- OSG Gratia report
  - http://t2.unl.edu/gratia/xml/facility_transfer_volume
- VDT documentation
  - http://vdt.cs.wisc.edu
- OSG Installation Guides
  - https://twiki.grid.iu.edu/bin/view/ReleaseDocumentation/WebHome
- OSG Storage group mailing list: osg-storage@OPENSCIENCEGRID.ORG