Grid Plans @ DESY

Andreas Gellrich

DESY

EGEE ROC DECH Meeting
FZ Karlsruhe, 22./23.01.2009
Contents

• Introduction

• Status of Grid @ DESY

• New Communities and VOs

• Work Load Management (WMS) @ ROC DECH

• ROC SLAs
DESY

- DESY operates a Grid infrastructure as a partner in the German/Swiss federation (DECH) of the EU project Enabling Grids for E-sciencE (EGEE) deploying the middleware gLite.

- DESY provides Grid services and Grid resources to a number of VOs of various disciplines; ONE Grid infrastructure for ALL VOs.

- DESY provides a data repository for CALICE/ILC testbeam and Monte Carlo data accessible via the Grid (Tier-0/1).

- DESY is part of the World-wide LHC Computing Grid (WLCG) as a Tier-2 centre.
Status of Grid @ DESY

“Operating the Grid …”
Grid @ DESY …

- VOs at DESY:
  - Regional: ‘calice’, ‘ghep’, ‘ildg’  [hosted at DESY]
  - Local: ‘desy’, ‘hermes’, ‘icecube’
  - Regional: ‘dech’, ‘xray.vo.egee-eu.org’  [hosted elsewhere]

- Grid Core Services:
  - VOMS, LFC, top-level-BDII, 11 WMS

- Grid Computing Resources at DESY: (CE) [32-bit, SL47]
  - grid-ce3.desy.de  2158 slots @ 373 hosts

- Grid Storage Resources at DESY: (SE) (dCache)
  - dcache-se-atlas.desy.de  O(100 TB) w/ tape backend
  - dcache-se-cms.desy.de  O(100 TB) w/ tape backend
  - dcache-se-desy.desy.de  O(100 TB) w/ tape backend
Federating resources among VOs and their groups
  - Jobs are transient!
  - Grid resources are procured from various sources
    - DESY
    - DESY / Tier-2
    - D-GRID
    - NAF
    - DESY Zeuthen (DESY-ZN) supports mainly ATLAS and LHCb
  - Opportunistic usage of resources
    - guarantee optimal usage of cycle
    - handle peak loads

- keep shares on average (*fair share*)
- limit maximal number of jobs
... Grid @ DESY ...
Fair Share at DESY-HH Dec/Jan 2008/09
Technicalities

- DESY-HH uses Quattor to install the OS
- gLite middleware is installed via ‘yaim’
  - Automatically on the WNs
  - Manually on the core servers
- We currently run Scientific Linux 4.7 in 32-bit
  - All servers and WNs are 64-bit machines!
  - Awaiting standard SL5 and/or 64-bit installation
- ‘sgm’ jobs go to dedicated WNs with NFS r/w access to SW area
- All other WNs mount the NFS SW area r/o
- Move from SUN-cluster based server to NetApp (?)
- Virtualization of services planned …
New Communities and VOs

“The future of scientific computing …”
“DESY conducts basic research in the natural sciences with special emphasis upon accelerators, photon science and particle physics.”

http://www.desy.de/
http://www.xfel.eu/

Accellerators
Petra III
XFEL
ILC

Photon Science
FLASH
Petra III
CFEL
XFEL

HEP
H1,HERMES,ZEUS
ATLAS,CMS
ILC
IceCube
Theory
The VO ‘xfel.eu’ …

- A VO for the future XFEL free electron laser
- Of interest for the FLASH / Petra3 community as well

- Needs of the synchrotron community is fundamentally different:
  - No tradition in big global collaborations
  - Short term experiments (days)
  - Short term users (come-and go)
  - Many/fully independent users
  - Little/no sharing of data
  - Little understanding of (scaling) problems in computing

- No VOs in the sense of HEP
- UIs must be multi-platform or web portal based
- (very) easy to use infrastructure required
... The VO ‘xfel.eu’

- DESY founded the VO ‘xfel.eu’

- All necessary core services are available:
  - VOMS: grid-voms.desy.de (soon)
  - LFC: grid-lfc.desy.de
  - AMGA: grid-amga0.desy.de
  - WMS: xfel-wms.desy.de
  - CE: grid-ce3.desy.de
  - SE: dcache-se-desy.desy.de

- probably Data Grid only
- Investigating use cases
The VO ‘xray’

- ESRF initiated a project to study ‘Grid for Synchrotron’
- ESRF founded the VO ‘xray.vo.eu-egee.org’
- By now 2 sites support ‘xray’: ESRF and DESY-HH

All necessary core services are available:
- **VOMS:** grid-voms.esrf.eu
- **LFC:** grid-lfc.desy.de
- **AMGA:** grid-amga0.desy.de
- **WMS:** xray-wms.desy.de
  wms1.egee.fr.cgg.com
- **CE:** grid-ce3.desy.de
- **SE:** dcache-se-desy.desy.de
‘xray’: Amga …
‘xray’: g-Eclipse
Workload Management Service

“Distributing the load …”
WMS ...

- Running WMS service is difficult ...
- Simple monitoring with (slightly modified) rbwmsmon
- Our Experience: WMProxy is most critical – Condor part is more stable
  - Often growing WMProxy input queue
    /var/glite/workload_manager/input.fl
    /var/glite/workload_manager/input.fl.log
  - If many jobs cannot be matched
    /var/log/glite/workload_manager_events.log
    - Service starts to degrade
    - Not clear if this caused by jobs itself or by malfunctioning services
  - Typical throughput up to 10k jobs on single WMS
    - Advertised number is over 100k!
... WMS

- Published to information system DNS aliases per VO
  - DNS assigned to one (out of 4 WMS) depending on “load metric”
  - Metric evaluated every few minutes
  - Distribute VOgs over WMS servers
  - Avoid usage of “bad” WMS servers

- WMS/LB Hardware
  - Modern multi-core hosts
  - At least 4GB of RAM
  - Some 100GB space for sandboxes (on separate RAID disks)
ROC DECH

“Making the Grid pervasive …”
ROC SLAs

- Create a *reliable, pervasive* Grid infrastructure
- *Mandatory* services at all sites
- Monitoring (SAM, GStat)

- Tier-2 MoUs already signed!
- Motivate sites to join (Tier-3)

- Situation of university sites which depend on local computer centre considered?
• The *local* installation is operated in a *global* environment
  • Some Core Grid services are central (VOMS, LFC)
  • Some Core Grid services are essential (WMS, BDII)

  • Who runs which services?
  • How is the load distributed?

• User *support* is a big issue
  • Not scalable
  • Underestimated
  • Has a huge social factor