Grid Enabling a small Cluster

Doug Olson

Lawrence Berkeley National Laboratory

STAR Collaboration Meeting
13 August 2003

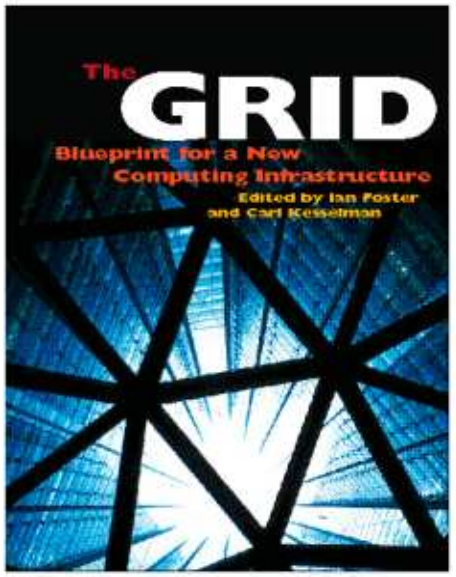
Michigan State University

Contents

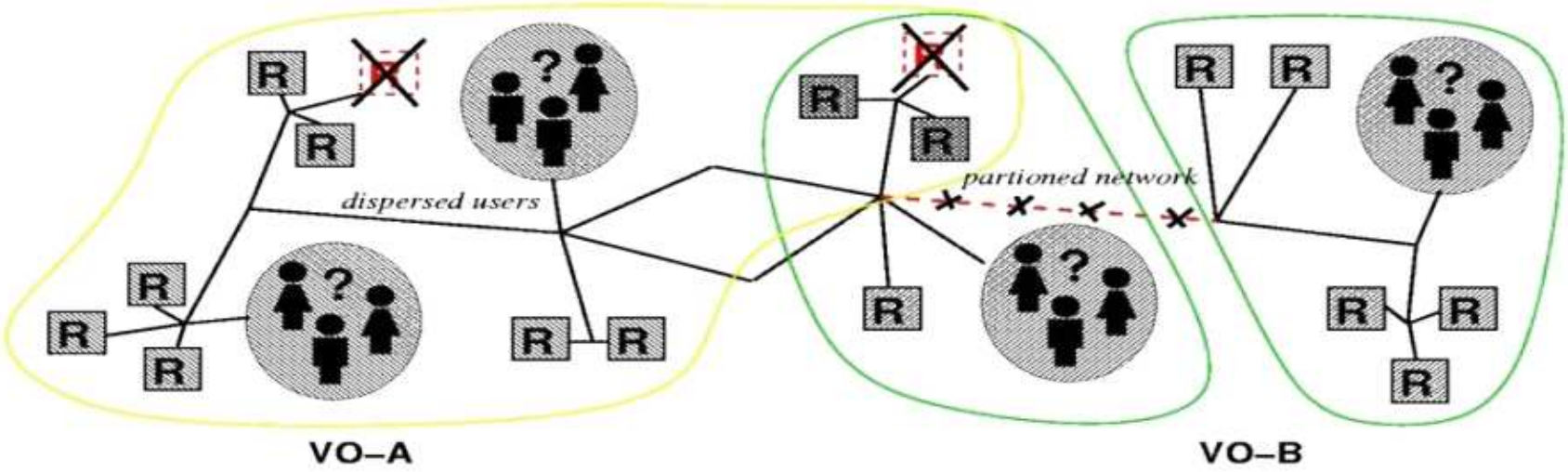
- Overview of multi-site data grid
- Features of a grid-enabled cluster
- How to grid-enable a cluster
- Comments

the globus project™
www.globus.org

And Thus: The Grid



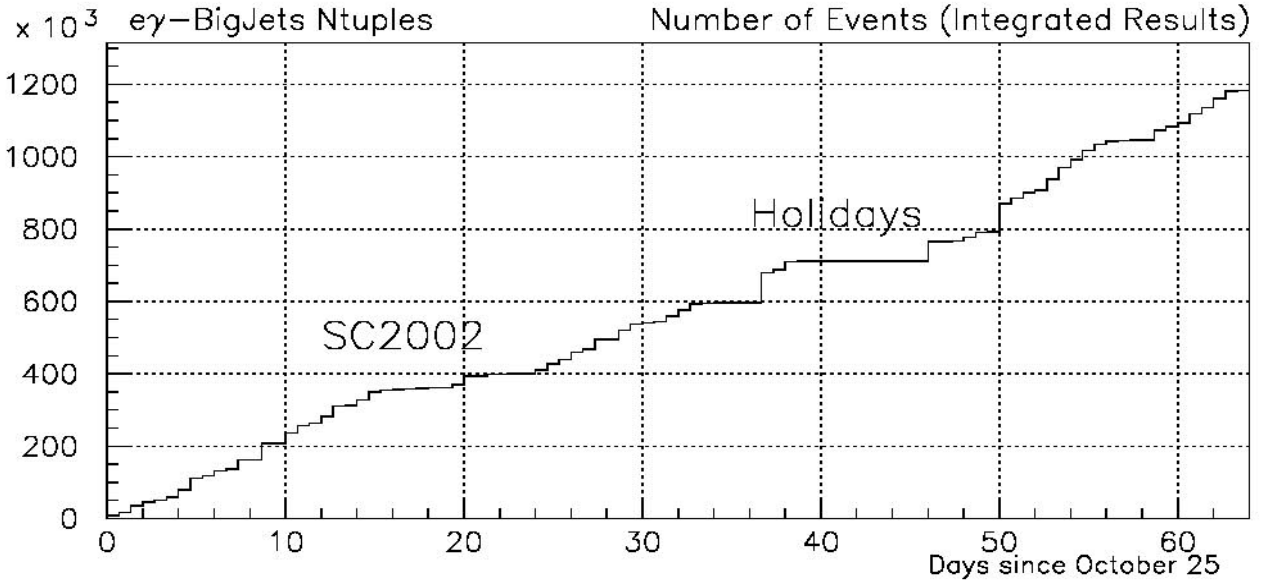
"Resource sharing & coordinated problem solving in dynamic, multi-institutional virtual organizations"

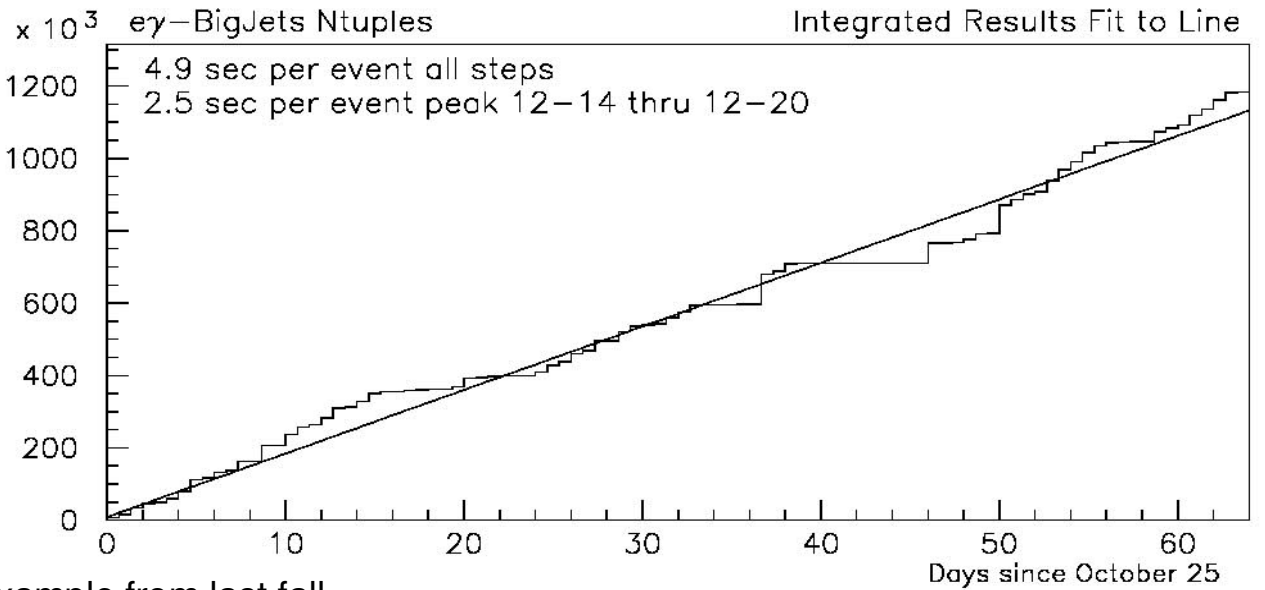


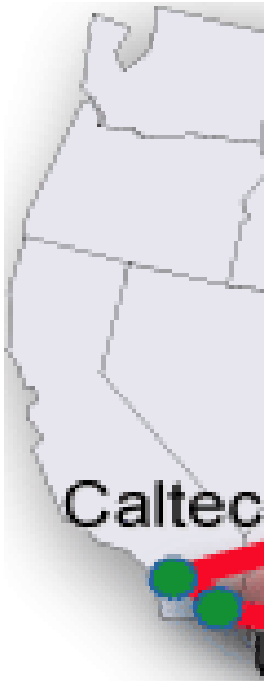
"When the network is as fast as the computer's internal links, the machine disintegrates across the net into a set of special purpose appliances" (George Gilder)

CMS Integration Grid Testbed

Time to process 1 event: 500 sec @ 750 MHz

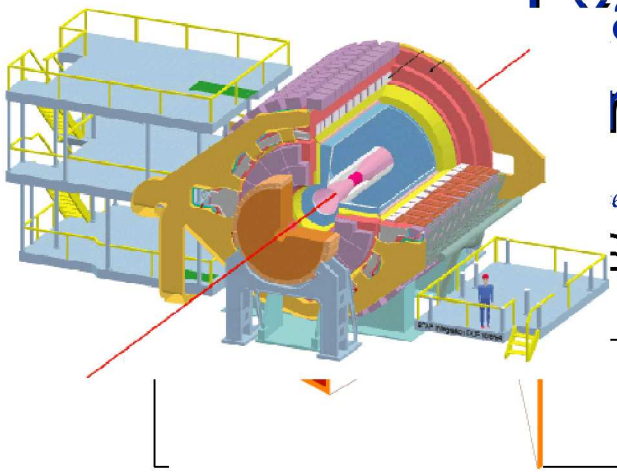






From Miron Livny, example from last fall.

Example Grid Application: Grids for High Energy Physics



es/sec

Online System

~100 MBytes/sec

1 TIPS is approximately 25,000 SpecInt95 equivalents

Offline Processor Farm ~20 TIPS

~100 MBytes/sec

BNL

~622 Mbits/sec or Air Freight (deprecated)

Tier 0

Tier 1

France Regional Centre

Germany Regional Centre

Italy Regional Centre

FermiLab ~4 TIPS

~622 Mbits/sec

Tier 2

Caltech ~1 TIPS

Tier2 Centre ~1 TIPS

Centre ~1 TIPS

Centre ~1 TIPS

Centre ~1 TIPS

~622 Mbits/sec

Institute ~0.25TIPS

Institute

Institute

Institute

Physics data cache

~1 MBytes/sec

Tier 4

Physicist workstations

Physicists work on analysis ?channels?.

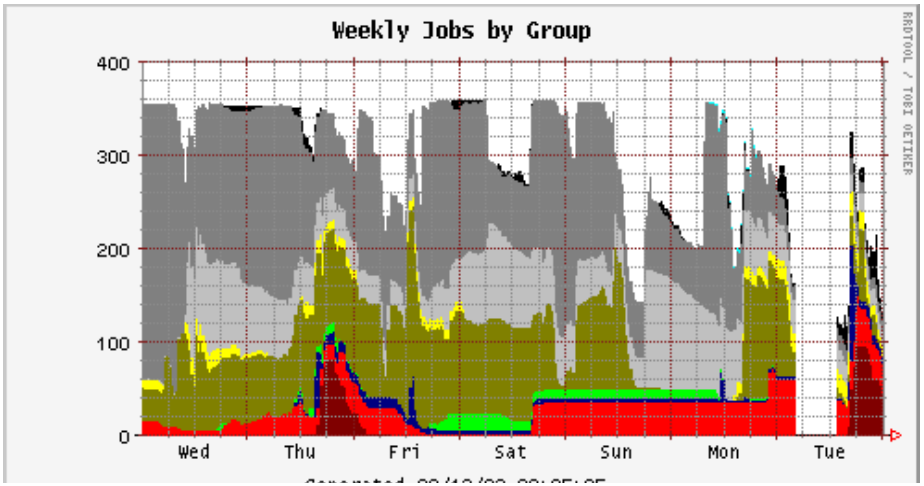
Each institute will have ~10 physicists working on one or more channels; data for these channels should be cached by the institute server

Famous Harvey Newman slide

What do we get?

Distribute load across available resources.

Access to resources shared with other groups/projects.
Eventually sharing across grid will look like sharing within a cluster (see below).



On-demand access to much larger resource than available in dedicated fashion.

(Also spreading costs across more funding sources.)

Features of a grid site (server side services)

- Local compute & storage resources
  - Batch system for cluster (pbs, lsf, condor, …)
  - Disk storage (local, NFS, …)
  - NIS or Kerberos user accounting system
  - Possibly robotic tape (HPSS, OSM, Enstore, …)
- Added grid services
  - Job submission (Globus gatekeeper)
  - Data transport (GridFTP)
  - Grid user to local account mapping (gridmap file, …)
  - Grid security (GSI)
  - Information services (MDS, GRIS, GIIS, Ganglia)
  - Storage management (SRM, HRM/DRM software)
  - Replica management (HRM & FileCatalog for STAR)
  - Grid admin person
- Required STAR services
  - MySQL db for FileCatalog
  - Scheduler provides (will provide) client-side grid interface

How to grid-enable a cluster

- Signup on email lists
- Study globus toolkit administration
- Install and configure
  - VDT (grid)
  - Ganglia (cluster monitoring)
  - HRM/DRM (storage management & file transfer)
- Set up method for grid-mapfile (user) management
- Additionally install/configure MySQL & FileCatalog & STAR software

Background URL's

- stargrid-l mail list
- Globus Toolkit - www.globus.org/toolkit
  - Mail lists, see - http://www-unix.globus.org/toolkit/support.html
  - Documentation - www-unix.globus.org/toolkit/documentation.html
  - Admin guide - http://www.globus.org/gt2.4/admin/index.html
- Condor - www.cs.wisc.edu/condor
  - Mail lists: condor-users and condor-world
- VDT - http://www.lsc-group.phys.uwm.edu/vdt/software.html
- SRM - http://sdm.lbl.gov/projectindividual.php?ProjectID=SRM

VDT grid software distribution

(http://www.lsc-group.phys.uwm.edu/vdt/software.html)

- Virtual Data Toolkit (VDT) is the software distribution packaging for the US Physics Grid Projects (GriPhyN, PPDG, iVDGL).
  - It uses pacman for the distribution tool (developed by Saul Youssef, BU Atlas)
  - VDT contents (1.1.10)
    - Condor/Condor-G 6.5.3, Globus 2.2.4, GSI OpenSSH, Fault Tolerant Shell v2.0, Chimera Virtual Data System 1.1.1, Java JDK1.1.4, KX509 / KCA, MonaLisa, MyProxy, PyGlobus, RLS 2.0.9, ClassAds 0.9.4, Netlogger 2.0.13
    - Client, Server and SDK packages
    - Configuration scripts
  - Support model for VDT
    - The VDT team centered at U. Wisc. performs testing and patching of code included in VDT
    - VDT is the prefered contact for support of the included software packages (Globus, Condor, …)
    - Support effort comes from iVDGL, NMI, other contributors

Additional software

- Ganglia - cluster monitoring
  - http://ganglia.sourceforge.net/
  - Not strictly req'd for grid but STAR uses as input to grid info svcs
- HRM/DRM - storage management & data transfer
  - Contact Eric Hjort & Alex Sim
    - Expected to be in VDT in future
  - Being used for bulk data ransfer between BNL & LBNL
- + STAR software …

VDT installation (globus, condor, …)

(http://www.lsc-group.phys.uwm.edu/vdt/installation.html)

- Steps:
  - Install pacman
  - Prepare to install VDT (directory, accounts)
  - Install VDT software using pacman
  - Prepare to run VDT components
  - Get host & service certificates (www.doegrids.org)
  - Optionally install & run tests (from VDT)
- Where to install VDT
  - VDT-Server on gatekeeper nodes
  - VDT-Client on nodes that initiate grid activities
  - VDT-SDK on nodes for grid-dependent s/w development

Manage users (grid-mapfile, …)

- Users on grid are identified by their X509 certificate.
- Every grid transaction is authenticated with a proxy derived from the user's certificate.
  - Also every grid communicaiton path is authenticated with host & service certificates (SSL).
- Default gatekeep installation uses grid-mapfile to convert X509 id to local user id
  - [stargrid01] ~/> cat /etc/grid-security/grid-mapfile | grep doegrids
  - "/DC=org/DC=doegrids/OU=People/CN=Douglas L Olson" olson
  - "/DC=org/DC=doegrids/OU=People/CN=Alexander Sim 546622" asim
  - "/OU=People/CN=Dantong Yu 254996/DC=doegrids/DC=org" grid_a
  - "/OU=People/CN=Dantong Yu 542086/DC=doegrids/DC=org" grid_a
  - "/OU=People/CN=Mark Sosebee 270653/DC=doegrids/DC=org" grid_a
  - "/OU=People/CN=Shawn McKee 83467/DC=doegrids/DC=org" grid_a
- There are obvious security considerations that need to fit with your site requirements
- There are projects underway to manage this mapping for a collaboration across several sites - a work in progress

Comments

- Figure 6 mo. full time to start, then 0.25 FTE for cluster that is used rather heavily by a number of users
  - Assuming reasonably competent linux cluster administrator who is not yet familiar with grid
- Grid software and STAR distributed data management software is still evolving so there is some work to follow this (in the 0.25 FTE)
- During next year - static data distribution
- In 1+ year should have rather dynamic user-driven data distribution