Grid @ DESY

Andreas Gellrich

DESY

HGF

15.12.2008
Legacy …
... Legacy
History

- Trivially, computing requirements must always be related to the technical abilities at a certain time …

- Until not long ago: (at least in HEP …)
  - Computing was a pure offline task:
    "Let’s first take data and then see how we handle them."
  - Necessary resources could be provided locally
  - In HEP, people have always been used to global approaches

- Nowadays: (LHC, ILC, …)
  - Computing is treated like a detector component
  - Necessary resources cannot be provided locally anymore
  - Larger amounts of resources are not provided locally

- The paradigm changed from local to global
Grid Computing is about *virtualization* of *global* resources.

- It is about *transparent* access to globally distributed resources such as data and compute cycles
- A Grid infrastructure consists of *services* to access resources and (of course) of the *resources* itself
  - Opposite to *distributed computing*, Grid resources are *not centrally controlled*
  - Hence it is mandatory to use *standard, open, general-purpose protocols and interfaces*
  - A Grid must *deliver nontrivial qualities of services*
- In general Grid infrastructures are *generic*; without any dependencies of the applications / experiments
Grid Types

- **Data Grids:**
  - Provisioning of transparent access to data which can be physically distributed within *Virtual Organizations* (VO)

- **Computational Grids:**
  - allow for large-scale compute resource sharing within Virtual Organizations (VO)

- **(Information Grids):**
  - Provisioning of information and data exchange, using well defined standards and web services

**Note!** Jobs are *transient*; data is *persistent.*
A Virtual Organization (VO) is a dynamic collection of individuals, institutions, and resources which is defined by certain sharing rules.

- A VO represents a collaboration
- Users authenticate with personal certificates (*Authentication*)
- Users are members of a certain VO (*Authorization*)
- Certificates are issued by a Certification Authority (CA)

- Grid Infrastructure
  - Core Services (mandatory per VO)
    - VO Membership Services
    - Grid Information Services
    - Workload Management System
  - Resources (brought in by partners (*Grid sites*))
Grid Projects
Objectives:

The EGEE project brings together experts from more than 50 countries with the common aim of building on recent advances in Grid technology and developing a service Grid infrastructure which is available to scientists 24 hours-a-day.

The project provides researchers in academia and business with access to a production level Grid infrastructure, independent of their geographic location. The EGEE project also focuses on attracting a wide range of new users to the Grid.

Because of its needs and its tradition and because of its *simple* use cases, HEP has become the pilot application for the Grid (in EGEE).
... EGEE ...

Statistics:
- Submitted: 16
- Waiting: 6522
- Ready: 442
- Scheduled: 5185
- Running: 4393
- Done: 722
- Aborted: 893
- Cancelled: 229
- Active Sites: 126 - 192300
Highlights from EGEE

- >200 VOs from several scientific domains
  - Astronomy & Astrophysics
  - Civil Protection
  - Computational Chemistry
  - Comp. Fluid Dynamics
  - Computer Science/Tools
  - Condensed Matter Physics
  - Earth Sciences
  - Fusion
  - High Energy Physics
  - Life Sciences
- Further applications under evaluation

Applications have moved from testing to routine and daily usage

~80-90% efficiency

98k jobs/day

Andreas Gellrich, DESY
HGF, 15.12.2008
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 ILC testbeam and *Monte Carlo* data accessible via the Grid

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

- VOs hosted at DESY:
  - Regional: ‘calice’, ‘ghep’, ‘ildg’
  - Local: ‘desy’, ‘hermes’, ‘icecube’

- VOs supported at DESY:
  - Regional: ‘dech’, ‘xray.vo.egee-eu.org’

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

- Grid Computing Resources at DESY: (CE) [2008-12-15] [growing!]
  - grid-ce3.desy.de 1774 slots @ 325 hosts

- Grid Storage Resources at DESY: (SE)
  - 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
... Grid @ DESY ...

- Federating resources among VOs and their groups
  - Jobs are transient
  - Grid resources are procured from various sources
    - DESY
    - DESY / Tier-2
    - D-GRID
    - NAF
    - 1/3 ATLAS, 1/3 CMS, 1/3 others

- Opportunistic usage of resources
  - guaranty optimal usage of cycle
  - handle peak loads

- keep shares on average (*fair share*)
- limit maximal number of jobs
... Grid @ DESY

Jobs at DESY-HH December 2008
Issues

• The local installation is operated in a global environment
  • There is always day light somewhere on the globe
  • Core Grid services are used everywhere (VOMS, LFC)

• One common infrastructure for multiple VOs of multiple disciplines
  • Different groups want different things
  • Computing models differ fundamentally
  • Use and user patterns differ
  • Software requirements differ

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