

# Grid Infrastructure & Services

**Andreas Gellrich**  
 DESY  
 IT Strategy Review  
 DESY, 28.02.2007

## The Grid Dream

**Andreas Gellrich, DESY**      **IT Strategy Review, DESY, 28.02.2007**      **1**



## Grid Computing

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



## 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

Jobs are *transient*; data is *persistent*.



# Grid Building Blocks



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 are members of a certain VO (**Authorization**)
- Users authenticate themselves with a certificate (**Authentication**)
- Certificates are issued by a national **Certification Authority (CA)**
- Grid Infrastructure
  - **Core Services** (mandatory per VO)
    - VO Membership Services
    - Grid Information Services
    - Resources Broker
    - Workload Management System
  - **Resources** (brought in by partners) (*Grid sites*)



# EGEE



- **Enabling Grids for E-Science (EGEE)**
- EU 6th Framework Programme (FP 6)
- EGEE-I started on April 1<sup>st</sup>, 2004
- **EGEE-II** started on April 1<sup>st</sup>, 2006
- 91 partners in 30 countries
- federated in 10 regional Grids
- Headquarter: CERN
- <http://www.eu-egee.org/>



*"EGEE is providing a production quality Grid infrastructure spanning more than 30 countries with over 150 sites to a myriad of applications from various scientific domains, including Earth Sciences, High Energy Physics, Bioinformatics and Astrophysics."*



## EGEE VOs



- **Astrophysics** (MAGIC telescope for air showers, ESA Planck Mission on microwave sky survey, ANTARES large area water Cherenkov detector, NEMO project underwater Cherenkov)
- **Biomedical applications** (medical imaging, bioinformatics gene sequence analysis, drug survey)
- **HEP** (LHC, HERA, ILC, Tevatron, BaBar, NA4)
- **Comp. Chemistry** (CHARON, CompChem)
- **Geophysics** (EGEODE seismic data processing)
- **Earth Science** (ESR/SAR Earth Observation, Solid Earth Physics, Hydrology, Climatology)
- **Fusion** (ITER)
- **Financial** (Egrid project Italy)



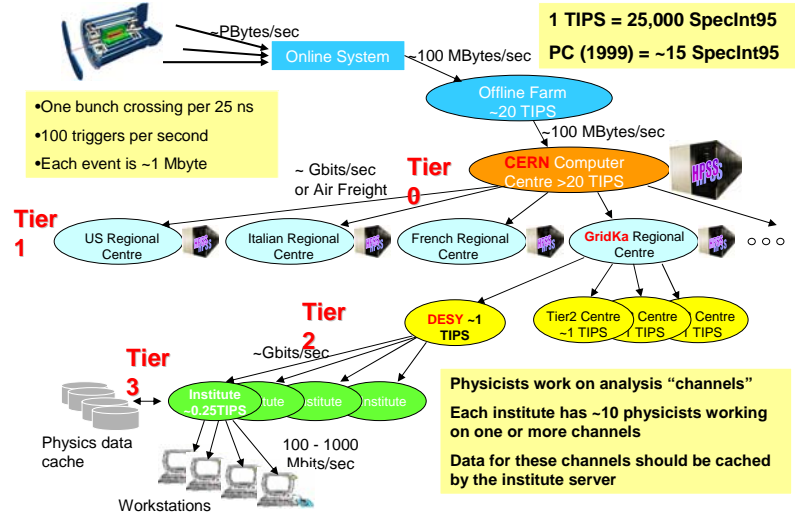
## D-GRID @ DESY



- R&D programme for a **national German e-science infrastructure**
- **DESY** is founding member
- Programme started **end 2005** for 3 year
- Funding volume of 15 MEUR
- Organization in 6 **communities** of community specific projects and 1 integration project
- **DESY** participates in the HEP Community Project (CP)
- **DESY** participates in the Integration project (IP), led by GridKa
- **DESY** brings in **know-how** and **experiences** in **data management**



# The LHC Computing Model



# Grid Reality

- For LHC a combined computing and data Grid is needed
- Jobs are *transient*, data is *persistent*
- The LHC requires computing power which can not be provided by single computer center; for technical and political reasons
  - The traditional model of hosting most of the computing in a single CC close to the accelerator/experiments fails
- Driven by its vast computing demands the LHC Computing Grid (LCG) has been implemented. It incorporates more than 200 resources centers (sites) with 100 thousands of CPUs and PetaBytes of disk space

The tradition of HEP experiments to build global collaborations was successfully mapped to the Grid concept of VOs.



## Grid @ DESY


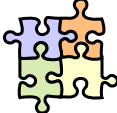


- DESY Grid activities were initially driven by the demand for resources for MC production for HEP by **H1** and **ZEUS**
- The International Lattice Data Grid (**ILDG**) and the International Linear Collider Community (**ILC**) joined the Grid activities
- DESY participates in **EGEE** (since 2004) and **D-GRID** (since 2005)
- DESY runs a registration authority (RA) for GridKa
- DESY is a **Tier-2** centers for **ATLAS** (in federation w/ U Freiburg) and **CMS** (in federation w/ RWTH Aachen)

DESY's Grid infrastructure is *generic* and open for new VOs.  
DESY offers hosting and *support* for new applications and VOs.



## Open Issues

- **Stability** of the middleware components / services
- **Performance** of the job submission; turn-around times
- **Authentication** as a member of more than one VO
- **Security** 
- Global **user support**
- Global **application support**
- Access Control Lists (ACL) 
- Parallel computing (MPI, ...)
- Massive database access
- Data catalogues; catalogue searches; collections; meta data
- Debugging and monitoring of jobs
- Software support; OS; environments



## Future Activities

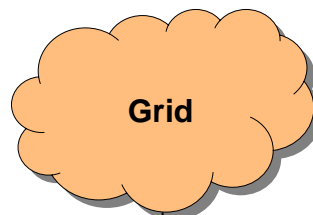
Resources are *virtual* in the Grid. Only *shares* are dedicated!

- Grid computing must become a part of the local computing at sites
- We will see how the relation between of *local* computing and *global* Grid computing will evolve
- More resources are foreseen to fulfill the **ATLAS** and **CMS** requirements for a **Tier-2** center at DESY. They will be added to the generic Grid infrastructure and distributed by shares.
- The DESY Grid infrastructure is ready for new applications and VOs in the context of EGEE and D-GRID
- We are ready for **FLASH** and **XFEL**!

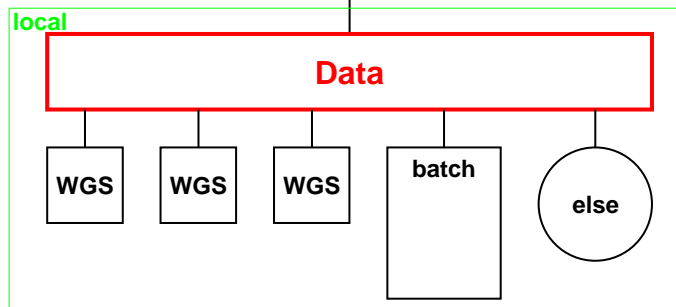


## global vs. local

global



local





## Conclusions

- Grid Computing is a *strategic* technology for the future
- Significant *resources* will be available in the Grid only
- The Grid requires *global* thinking!
- DESY maintains a generic Grid Infrastructure *in production*
- DESY is an LCG *Tier-2* centre for *ATLAS* and *CMS/*
- *H1*, *ILC*, and *ZEUS* heavily use the Grid for *MC* production
- Non-HEP Grid applications gain significance
- We are waiting for *XFEL* to come to the Grid



## Appendix: The Grid on the Web

- The Grid on the web at DESY:
  - ✓ <http://grid.desy.de/>
- German Grid Initiative:
  - <https://www.d-grid.de/>
- LCG and EGEE:
  - <http://cern.ch/lcg/>
  - <http://www.eu-egee.org/>
  - <http://cic.in2p3.fr/>





## Appendix: History of the Grid

„We will probably see the spread of *computer utilities*, which, like present *electric and telephone utilities*, will service individual homes and offices across the country.“ Len Kleinrock (1969)

„A computational grid is a hardware and software *infrastructure* that provides dependable, consistent, pervasive, and *inexpensive access* to high-end computational capabilities.“ I. Foster, C. Kesselmann (1998)

„The sharing that we are concerned with is not primarily file exchange but rather direct access to computers, software, data, and other resources, as is required by a range of *collaborative problem-solving* and resource brokering strategies emerging in industry, science, and engineering. The sharing is, necessarily, highly controlled, with resources providers and consumers defining clearly and carefully just what is shared, who is allowed to share, and the conditions under which sharing occurs. A set of individuals and/or institutions defined by such sharing rules what we call a *virtual organization*.“ I. Foster, C. Kesselmann, S. Tuecke (2000)



## Appendix: The Term Grid

I. Foster: *What is the Grid? A Three Point Checklist* (2002)

“A Grid is a system that

... *coordinates resources which are not subject to centralized controls* ...

integration and coordination of resources and users of different domains vs. local management systems (batch systems)

... *using standard, open, general-purpose protocols and interfaces* ...

standard and open multi-purpose protocols vs. application specific system

... *to deliver nontrivial qualities of services.*”

coordinated use of resources vs. uncoordinated approach (world wide web)



## Appendix: Grid Services @ DESY

- Quattor (SL 3.08 for all nodes; complete installation for WNs)
- GLITE-3\_0, Yaim (for all service nodes)
- Central VO Services: (*unique per VO*)
  - VO Members (VOMS) [grid-voms.desy.de]
  - Catalogue Services (LFC) [grid-lfc.desy.de]
- Distributed VO Services: (*mandatory per VO*)
  - Resource Broker (RB) [grid-rb0/1/2.desy.de]
  - Workload Management (WMS) [grid-wms0.desy.de]
  - Information Index (BDII) [grid-bdii.desy.de]
  - MyProxy (PXY) [grid-pxy.desy.de]



## Appendix: Grid Resources @ DESY

- Site Resources:
  - GIIS: DESY-HH [grid-giis.desy.de]
  - CE: 166\*Opteron/2.4GHz [grid-ce0.desy.de]
  - CE: 324\*Opteron/2.2 Woodcrest/3.0GHz [grid-ce1.desy.de]
  - CE: 90\*XEON/3.0GHz [grid-ce2.desy.de]
  - SE: dCache-based w/ access to DESY data space
- Grid (Tier-2) Resource Planning:
  - 2005: 200 kSpecINT2k 30 TB
  - Now: 850 kSpecINT2k (200) TB
  - 2008: 1600 kSpecINT2k 600 TB
  - 2009: 1800 kSpecINT2k 800 TB



## Appendix: VO Support @ DESY

- VOs hosted at DESY:
  - Global: 'hone', 'ilc', 'zeus'
  - Regional: 'calice', 'dcms', 'ildg'
  - Local: 'baikal', 'desy', 'herab', 'hermes', 'icecube'
- VOs supported at DESY:
  - Global: 'atlas', 'cms', 'dteam'
  - Regional: 'dech'
- H1 Experiment at HERA ('hone') (<http://grid-voms.desy.de:8443/vomses/>)
  - desy, uni-dortmund, cscs, gridpp, bham, ucl, lincs, ox, marseille, cyf-kr, saske
- ILC Community ('ilc', 'calice') (<http://grid-voms.desy.de:8443/vomses/>)
  - desy, ifh, ciemat, lal, polgrid, cclgeli, cam, ic, manchester, ucl
- ZEUS Experiment at HERA ('zeus') (<http://grid-voms.desy.de:8443/vomses/>)
  - desy, uni-dortmund, gridpp, bham, ucl, lincs, ox, marseille, cyf-kr, saske, infn, utoronto, uam, scotgrid, weizmann, scai, bris, tau, ed



## Appendix: Further Reading

### Books:

- Foster, C. Kesselmann: *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann Publisher Inc. (1999)
- F. Berman, G. Fox, T. Hey: *Grid Computing: Making The Global Infrastructure a Reality*, John Wiley & Sons (2003)



### Articles:

- I. Foster, C. Kesselmann, S. Tuecke: *The Anatomy of the Grid* (2000)
- I. Foster, C. Kesselmann, J.M. Nick, S. Tuecke: *The Physiology of the Grid* (2002)
- I. Foster: *What is the Grid? A Three Point Checklist* (2002)