e-Infrastructures for Compute and Data Intensive Applications

Solving the Grand Challenges in Big Science

Wolfgang Gentzsch
The DEISA Project & Board of Directors of OGF
gentzsch at rzg.mpg.de
InfoSys 2009

ICNS 2009: Networking Services
ICAS 2009: Autonomic and Autonomous Systems
INTENSIVE 2009: Intensive Applications and Services

all deal with components important for e-Infrastructures
e.g. Grids, Clouds, Internet, sensor nets, agent networks, etc.
Components of an e-Infrastructure:

HPC Centers, Grids and Clouds
**Terminology**

**Utility Computing**
- Cloud services
- Pay per use
- Grid Techn. & Virtualization

**Distributed Computing**
- Loosely coupled
- Heterogeneous
- Central management

**Cluster**
- Tightly coupled
- Homogeneous
- Cooperative working

**Grid Computing**
- Large scale
- Multi-organizational
- Cross-geography
- Distributed management
HPC Centers

- HPC Centers are service providers, for past 35 years
- Computing, storage, applications, data, etc IT services
- Serve (local) research, education, and industry
- Very professional: to end-users, they look (almost) like Cloud services (Amazon Cloud definition: easy, secure, flexible, on demand, pay per use, self serve)
- But: no virtualization, semi-automatic, not flexible
- HPC centers can become a Cloud customer for dynamic scaling and adopting to changing user demands
1998: The Grid: Blueprint for a New Computing Infrastructure:

“... hardware and software infrastructure ... dependable, consistent, pervasive, inexpensive access to high-end computational capabilities.”

2002: The Anatomy of the Grid:

“... coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations.”

Quotes: Ian Foster, Carl Kesselman, Steve Tuecke
Clouds

- IT resources provisioned outside corporate data center
- Resources accessed over the Internet
- A virtual computing environment (Vmware, Xen,...)
- Abstraction of the hardware from the service
- Service oriented: SaaS, PaaS, IaaS, HaaS
- Variable cost of services (QoS)
- Build and deliver, always-on, pay-per-use IT services
- Scaling up/down: computing, storage, database, services, users
10 Examples of e-Infrastructures

Courtesy: Dennis Gannon
NEESGrid

Realtime access to earthquake
Shake table experiments at remote sites.
BIRN – Biomedical Information

The Biomedical Informatics Research Network (BIRN) Portal provides BIRN members with a single sign on web portal to access data grid files, computation grid resources, and a variety of collaboration tools to facilitate the scientific needs of BIRN researchers. Non-BIRN participants may access the portal through a guest registration.
Geological Information Grid Portal

Select a Subject to Show Resources:
- Biological oceanography
- Chemical oceanography
- Cryology
- Ecology
- Education
- Environmental science
- Forestry
- Geochemistry
- Geologic time
- Geology
- Geophysics
- Human geography
- Hydrology
- Mineralogy or petrology
- Natural hazards
- Other
- Palaeontology
- Physical geography
- Physical oceanography
- Soil science
- Technology
- Structural geology

Resources in Geology
- Title: Arizona Geology Map
- Format: shapefile
- Dataset Id: GEON-250f83d-b71b-11d8-b226-ab22ad7681d0
- Spatial Coverage: North: 37 East: -100.04 South: 31.33 West: -114.82
- Temporal Coverage: any
- Description: This is a geology map of Arizona in USA.
- Semantic Annotations: see details
Mesoscale Meteorology

NSF LEAD project - making the tools that are needed to make accurate predictions of tornados and hurricanes.
- Data exploration and Grid workflow
Renci Bio Portal

Providing access to biotechnology tools running on a back-end Grid.

- leverage state-wide investment in bioinformatics
- undergraduate & graduate education, faculty research
- another portal soon: national evolutionary synthesis center
X-Ray Crystallography

Welcome to the Crystallography Portal

The Purdue Chemistry Crystallography Center

Disable your browser’s cache to get the live stream!

Data from Nonius Kappa CCD detector
(Under development)
Total number of jgs: 10
Frame: 0010010.jpg
All available .jpg images
Remove the 20 latest .jpg images

Streaming video from the lab showing the Nonius instrument

Streaming video from the crystal microscope on the Nonius diffractometer

Local date/time: 2005-09-24 11:36:54
These values are updated approx. every 60 sec.

LabJack U12

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cold, Humid. 43.1 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chill Water in:</td>
<td>15.4 C</td>
<td>2005-09-24 16:36:25</td>
</tr>
<tr>
<td>Chill Water Out:</td>
<td>19.3 C</td>
<td>2005-09-24 16:36:27</td>
</tr>
<tr>
<td>Generator Relay Voltage:</td>
<td>3.42</td>
<td>2005-09-24 16:36:48</td>
</tr>
<tr>
<td>All previous voltages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X-ray Generator is:</td>
<td>OFF</td>
<td></td>
</tr>
</tbody>
</table>
ServoGrid Portal

Welcome to the QuakeSim Computational Portal

Solid Earth Research Virtual Observatory Grid

- QuakeSim home page
- Old GEM General Earthquake Modeling Web Site
- SLIDE Distributed File System for NASA Computational Technology Project
- Report from the Earth Science Enterprise Computational Technology Requirements Workshop April 30-May 1 2002 where SERVO concept first introduced
- Discover the Grid at the Grid Forum or at this collection of papers
- Other collected papers and presentations on SERVOGrid and related topics are available from the Community Grids Lab publications page.

Participating Institutions:
IU CGL | NASA JPL | UC Davis | UC Irvine | USC
Belfast Gene Grid Portal

Welcome to the GeneGrid Prototype - Release 0.6

This is the GeneGrid Test Bed release 0.6 managed by the Belfast e-Science Centre, utilising resources in BeSC, Queen's University of Belfast, Melbourne University, BT and the San Diego Super Computing Centre.

Users are limited to selected staff of both commercial partners - Fusion Antibodies, Amtec Medical - and the Belfast e-Science Centre. To obtain a user account, please contact the appropriate representative - P.V. Jithesh (BeSC), Mark McCurley (Fusion) or Dr. Shane McKee (Amtec). Authorized users will be provided with a username and password by BeSC.

All users are requested to subscribe to the GeneGrid mailing list and to use it for directing queries etc. Mail GeneGrid, and place the word "subscribe" (without the quotes) in the message body.

For more on the GeneGrid project, please click here.

Important Note: Current GeneGrid Users please continue to use the Release 0.5 available here.
myGrid - Bioinformatics

myGrid is a collection of services and components that allows the high level integration of biological applications. The architecture provides the infrastructure necessary, in a web service environment, for an e-science workbench that actively supports the scientific lifecycle. Each component or service contributes to a system that allows the e-scientist to perform complex in-silico experiments across distributed bioinformatics resources.
The DEISA Ecosystem for HPC Grand-Challenge Applications

Distributed European Infrastructure for Supercomputing Applications
DEISA dedicated high speed network on GEANT2 and the NRENs

1 Gb/s GRE tunnel
10 Gb/s wavelength
10 Gb/s routed
10 Gb/s switched
DEISA: Vision - Mission - Strategy

Vision:
Persistent European HPC ecosystem integrating Tier-1 (Tflop/s) centres and the new European Tier-0 (Pflop/s) centres.

Mission:
Enhance Europe’s capability in computing and science by integrating most powerful supercomputers into a European HPC e-infrastructure.

Built European Supercomputing Service on top of existing national services, based on the deployment and operation of a persistent, production quality, distributed supercomputing environment with continental scope.

Strategy:
Consolidate the existing DEISA1 HPC infrastructure and services.
Deliver a turnkey operational solution for the future persistent European HPC ecosystem.
new “petaflop” supercomputers

PRACE
petaflop supercomputers

EU

DEISA
virtual supercomputer

National

Local

Mario Campolargo
European Commission
OGF23, June 2008
Categories of DEISA services
DEISA Service Layers

- Multiple ways to access
- Single monitor system
- Data staging tools
- Unified AAA
- Workflow management
- Job rerouting
- Co-reservation and co-allocation
- Data transfer tools
- DEISA Sites
- Network connectivity
- Common production environment
- Presentation layer
- Job management layer and monitor
- Data management layer
- Network and AAA layers

InfoSys, April 20-25, 2009
Wolfgang Gentzsch, DEISA
DEISA Global File System

Global transparent file system based on the Multi-Cluster General Parallel File System (MC-GPFS of IBM)

InfoSys, April 20-25, 2009

Wolfgang Gentzsch, DEISA
Management of users in DEISA

- A dedicated LDAP-based distributed repository administers DEISA users
- Trusted LDAP servers are authorized to access each other (based on X.509 certificates) and encrypted communication is used to maintain confidentiality
Common User Administration

- Each partner is responsible for the registration of users affiliated to the partner (home organization)
- Other partners update local user administration (LDAP, NIS, /etc/passwd) with data from other sites on a daily basis. Based on trust between partners!
Next-Generation e-Infrastructures
A Peek at Intel’s Digital City Vision

Imagine a city...

That Transforms the way we Work, Live & Play...

Courtesy Robert Fogel, Intel
Facets of the Digital City

- Serving Citizens
- Digital Education
- Economic Vitality (Digital Office)
- Digital Govt GAPP Programs
- Digital Healthcare
- Digital Home
- Bridging the Digital Divide
- Safety & Security

The Digital City is the Fabric that Connects the Community
Today’s Digital Challenge

- Banking
- Working
- Shopping
- Taxes

- Multiple Identities
- Incompatible Networks
-Disconnected Agencies

- Business
- Citizen
- Govt
- Employee

- Wimax
- WiFi
- Wired
- Cellular
Building the Digital City Today

- **Value / Benefits**
  - Productivity
  - Economic Dev

- **Efficiency**
  - Interoperability

- **Access**
  - Convenience

**Modular & Scalable**

**E-Government Services**

- Security
- GIS
- SOA/SOI
- eForms
- Content Mgmt
- Wire-line
- Private Networks
- 3G WAN
- WiFi
- WiMAX

**Business**

**Government**

**Citizen**
Tomorrow's Integrated Digital City

A single, portable identity based on strong security

An intelligent infrastructure supporting seamless access

A gateway to integrated e-Services spanning multiple agencies
A Model for Sustainability
(a checklist)

Reduce or eliminate the barriers in all the different areas such technology, culture, legal, economics and politics!

Especially, incorporate existing sustainability already achieved with individual components!

Therefore, the DEISA sustainability model is based on ensuring sustainability of every individual component:

» Technology and infrastructure
» Operations and services
» Expertise
» Communities
» Collaborations
» Eco-political landscape
The DEISA Model for Sustainability

• **Technology and Infrastructure:**
  – DEISA infrastructure is built on existing, proven, sustainable technology components,
  – GEANT2, NRENs, Supercomputers, HPC services, global sw environment
  – deliver and operate a European supercomputing infrastructure and related services

• **Operations and Services**
  – benefit from the many-years operations of the individual European supercomputers centres
  – orchestrated by the partners after the end of the funded project
  – activities relevant for applications enabling, operation, and technologies have been developed

• **Expertise**
  – tight collaboration of the expert groups in the different HPC centres
  – provided in the future to the wider European HPC communities.
The DEISA Model for Sustainability

• **Communities**
  – annual DEISA Extreme Computing Initiative (DECI)
  – supporting single projects, Virtual European Communities, and international science communities across existing political boundaries

• **Collaboration**
  – Distributed Common Production Environment (DCPE)
  – Collaboration with new European and other international initiatives.
  – contacts to research infrastructure projects established by the ESFRI, and the European HPC and Grid projects such as PRACE and EGEE
  – European & international HPC centres; initiatives in Australia, China, Japan, Russia, US, and leading HPC projects worldwide
  – Participate in evaluation and implementation of interoperation standards

• **Eco-Political Landscape**
  – ESFRI European Strategy Forum on Research Infrastructures
  – PRACE: preparing installation of a limited number of leadership-class Tier-0 supercomputers in Europe.