Advances in Computational Sciences: From HPC to Grids to Clouds

Wolfgang Gentzsch
EU DEISA Project & Board of Directors of OGF
gentzsch at rzg.mpg.de
Content

- Performance requirements of scientific applications
- Components: HPC Centers, Grids, and Clouds
- Example: The DEISA Ecosystem for HPC Applications
- Cloud Computing
- HPC in the Cloud
- Applications in the Cloud
- Challenges in the Cloud
- Conclusions
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Climate and Earth System Research</td>
<td>20</td>
<td>50-100</td>
<td>&gt;500</td>
</tr>
<tr>
<td>Geophysics</td>
<td>1</td>
<td>10-100</td>
<td>&gt;1000</td>
</tr>
<tr>
<td>Nanostructure Physics</td>
<td>1</td>
<td>10-50</td>
<td>&gt;200</td>
</tr>
<tr>
<td>Solid-State Physics</td>
<td>1</td>
<td>50-100</td>
<td>&gt;1000</td>
</tr>
<tr>
<td>Computational Fluid Dynamics</td>
<td>2.5</td>
<td>25-100</td>
<td>&gt;1000</td>
</tr>
<tr>
<td>Astrophysics</td>
<td>10</td>
<td>50-100</td>
<td>&gt;500</td>
</tr>
<tr>
<td>Elementary Particle Physics and Physics of Hadrons and Nuclei</td>
<td>30</td>
<td>100</td>
<td>&gt;1000</td>
</tr>
<tr>
<td>Materials Science</td>
<td>10</td>
<td>50-100</td>
<td>&gt;500</td>
</tr>
<tr>
<td>Theoretical Chemistry</td>
<td>3</td>
<td>25-125</td>
<td>&gt;300</td>
</tr>
<tr>
<td>Soft Matter</td>
<td>3</td>
<td>30</td>
<td>&gt;200</td>
</tr>
<tr>
<td>Biophysics and Bioinformatics</td>
<td>3</td>
<td>15-80</td>
<td>&gt;1000</td>
</tr>
<tr>
<td>Plasma Physics</td>
<td>10</td>
<td>50</td>
<td>&gt;500</td>
</tr>
</tbody>
</table>

A. Bode, W. Hillebrandt, and Th. Lippert: Scientific Case for the German Government, 8/2005
**Terminology**

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

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

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

**Cloud Computing**
- IaaS, PaaS, SaaS
- Pay per use
- Public, Private, Hybrid
HPC Centers and Clusters - still our bread & butter -

- HPC Centers are **service providers**, for past 35 years
- IT Service: Computing, storage, applications, data, etc
- Serve (local) research, education, and industry
- Very professional: to end-users, they look (almost) like Cloud services, if compared with Amazon Cloud definition: easy, secure, flexible, on demand, pay per use, self serve)
- Challenges: peta/exa, software, scalability, multicore, GPUs, Green Computing, connected to Grids & Clouds,...
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
European HPC Eco-System

Tier-0 – PRACE1IP

Tier-1 DEISA/PRACE2IP

Tier-2 Grids / DCIs: EGI, and EMI, IGE and numerous other projects

Kostas Glinos
European Commission, 2010
Example:

The DEISA Ecosystem for HPC Grand-Challenge Applications

HPC Centers in the Grid

DEISA:
Distributed European Infrastructure for Supercomputing Applications
DEISA: Vision and Mission

Vision:
Persistent European **HPC ecosystem** integrating Tier-1 (Tflop/s) centres and 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.
DEISA Evolution

Grand Challenge projects performed on a regular basis

Most powerful European Supercomputers for most challenging projects

Top-level Europe-wide application enabling

Virtual Science Community Support

Six years of operation
Categories of DEISA services

- Operations
- Technologies
- Applications

- Distributed European Infrastructure for Supercomputing Applications

Requests development

Requests support

Offers product

Offers configuration

Offers technology
DEISA highly performant continental global file system

**Different Software Environments**

- **DEISA Common Production Environment**
- **Access via Internet**
  - single sign-on (based on X.509 ‘Grid’ certificates)
  - gsi-ssh -> D-ssh
  - Unicore, gridFTP

**Different SuperComputers - Compute elements and interconnect**

- Dedicated 10 Gb/s network – via GEANT2
DEISA Service Layers

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

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

Site A
- Deputy
- Network
- Data
- Computing
- AAA
  - User-environment
  - Application-support

Site B
- Deputy
- Network
- Data
- Computing
- AAA
  - User-environment
  - Application-support

Site C
- Deputy
- Network
- Data
- Computing
- AAA
  - User-environment
  - Application-support

Site D
- Deputy
- Network
- Data
- Computing
- AAA
  - User-environment
  - Application-support

Distributed European Infrastructure for Supercomputing Applications
Federated Operation of DEISA

Virtual European Supercomputing Centre
Federated Operation of DEISA

Head of Operations

Site deputies

Site A deputy

Site B deputy

Site C deputy

Site D deputy

Network-related Services (DEISA network, monitoring, support)

Data-related Services (GPFS, GridFTP, Data-staging, OGSA-DAI)

Compute-related Services (Batch systems, UNICORE, WS-GRAM)

AAA-related Services (AAI, User admin., Accounting, Security)

User-related Services (DCPE, User-documentation, User-support)

Task leaders

Service Categories

Virtual European Supercomputing Centre
Cloud... as a Service

Cloud: dynamically scalable and virtualized resources provided as a service over the Internet

- Accessible online, anytime, anywhere
- Pay for what you use
- Available on demand
- Service Level Agreements
- Automated:
  - Scalability
  - Failover
  - Concurrency management

Infrastructure (IaaS)
Platform (PaaS)
Software (SaaS)
## A Model for Delivering IT Capabilities

<table>
<thead>
<tr>
<th>What</th>
<th>Who</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Software as a Service</strong></td>
<td><strong>End-user</strong></td>
</tr>
<tr>
<td>On-demand access to any application</td>
<td>(does not care about hw or sw)</td>
</tr>
<tr>
<td><img src="skype.png" alt="skype" /> <img src="gmail.png" alt="gmail" /> <img src="facebook.png" alt="facebook" /></td>
<td></td>
</tr>
<tr>
<td><strong>Platform as a Service</strong></td>
<td><strong>Developer</strong></td>
</tr>
<tr>
<td>Platform for building and delivering web applications</td>
<td>(no managing of the underlying hw &amp; sw layers)</td>
</tr>
<tr>
<td><img src="windows_azure.png" alt="Windows Azure" /> <img src="force_com.png" alt="force.com" /></td>
<td></td>
</tr>
<tr>
<td><strong>Infrastructure as a Service</strong></td>
<td><strong>System Administrator</strong></td>
</tr>
<tr>
<td>Raw computer infrastructure</td>
<td>(complete management of the computer infrastructure)</td>
</tr>
<tr>
<td><img src="rackspace_hosting.png" alt="rackspace hosting" /> <img src="flexiscale.png" alt="flexiscale" /> <img src="amazon_web_services.png" alt="amazon web services" /></td>
<td></td>
</tr>
</tbody>
</table>

Courtesy: Ignacio Llorente
Gartner Hype Curve 2010

Figure 1 Hype Cycle for Emerging Technologies, 2010

*Source: Gartner (August 2010)*
How will your budget for cloud computing change in 2011 compared with 2010?

Source: John Barr, The 451 Group Cloud Adoption Survey 2010
Leading inhibitors to Cloud Adoption – hint, it’s not about the infrastructure!

Source: John Barr, The 451 Group Cloud Adoption Survey 2010
Grid versus Cloud

Why should my App run in the Grid?

- Closer collaboration with colleagues (VCs)
- Mapping workflows to resources (plumbing)
- Suitable resources => faster/more/accurate processing
- Different architectures serve different apps
- Failover: move jobs to another system

... and why in the Cloud?

- No upfront cost for additional resources
- CapEx => OpEx, pay-per-use
- Elasticity, scaling up and down
- Hybrid solution (private and public cloud)
FIGURE 1. NPB-OMP (CLASS B) RUNTIMES ON 8 CPUs ON EC2 AND NCSA CLUSTER COMPUTE NODES. OVERLAID IS THE PERCENTAGE PERFORMANCE DEGRADATION IN THE EC2 RUNS.

Ed Walker, Benchmarking Amazon EC2 for high-performance scientific computing, ;Login, October 2008.
FIGURE 2. NPB-MPI (CLASS B) RUNTIMES ON 32 CPUS ON THE NCSA AND EC2 CLUSTER. BT AND SP WERE RUN WITH 16 CPUS ONLY. OVERLAID IS THE PERCENTAGE DEGRADATION IN THE EC2 RUNS.

Ed Walker, Benchmarking Amazon EC2 for high-performance scientific computing, ;Login, October 2008.
Ed Walker, Benchmarking Amazon EC2 for high-performance scientific computing, ;Login, October 2008.
Ed Walker, Benchmarking Amazon EC2 for high-performance scientific computing, ;Login, October 2008.
Loosely coupled problems are suitable for the Cloud

- Ensemble runs to quantify *climate model uncertainty*
- Identify *potential drug targets* by screening database of ligand structures against target proteins
- Study *economic model sensitivity* to parameters
- Analyze *turbulence dataset* from many perspectives
- Numerical optimization to determine optimal resource assignment in energy problems
- Mine collection of data from *advanced light sources*
- Construct databases of *chemical compounds* properties
- Analyze data from the **Large Hadron Collider**
- Analyze *log data* from 100,000-node parallel computations
Clouds and supercomputers: Conventional wisdom?

- Clouds/Clusters:
  - Too slow
  - ✔
- Supercomputers:
  - Too expensive
  - ✔
- Loosely coupled applications
- Tightly coupled applications

Courtesy: Ian Foster
Introducing CCI Cluster Compute Instances

- New Amazon EC2 instance type
- Optimized for network intensive computing
  - Low latency
  - High bandwidth
  - New EC2 API: Placement groups
- Instance Configuration
  - 2 * Xeon 5570 (Intel “Nehalem” quad-core architecture)
    - 33.5 Elastic Compute Units
  - 23 GB 1333MHz DDR3 Registered ECC RAM
  - 1690 GB of local instance storage
  - 10 Gbps Ethernet interconnects CCI’s
Some EC2 CCI Results

- Some applications can expect 10x better performance
- LNBL NERSC saw 8.5x compared to similar clusters on standard EC2 instances
- Linpack benchmark
  - 880-instance CC1 cluster
  - Performance: 41.82 Tflops

EC ran #146 in the Top 500 rankings
MATLAB $Ax = b$ Benchmark, Sept 2010

Results
Finally: Pricing

- **On-demand**
  - $1.60/hr
  - Linpack: 880 CCIs (7040 cores)
  - $1.6\times24\times30\times880 = $1M per month, $12M per year

- **Reserved Instance**
  - 1 yr: $4290 one time + $0.56/hr
  - 3 yr: $6590 one time + $0.56/hr
  - Linpack: 880 CCIs (7040 cores)
  - $4.3M per year

- **In general, too expensive**
- **Best solution:** fire-drill problems, additional resources are needed immediately, for a restricted period of time.
- **It takes 6 months on average to procure, deploy and activate new (own) resources!**
- **All you need is some ‘Cloud Adapter‘ software (e.g. Oracle Grid Engine, Univa UniCloud, Eucalyptus, Nimbus, OpenNebula, etc)**
Case Study: Managed/Cloud Service Provider

Without Mellanox InfiniBand

With Mellanox

Savings

Customer profile: Web-based travel transactions
Volume exceeding Amazon.com transactions

256 VMs
4 racks, 4U servers
4X proc, 16GB
16 edge switches
192 I/O cards
$744K capital cost

256 VMs
1 rack, 1U servers
4X proc, 16GB
2 I/O Directors
32 I/O cards
$347K capital

Cap Ex: $397,000
Op Ex:
Floor space: $54,000
Power: $20,000
MAC*: $104,000

TOTAL SAVINGS $575,000

(32 servers over 5 years)
*2 moves/adds/changes per server per year

But:

Source: Xsigo
Courtesy: Gilad Shainer, Mellanox
Challenges in the Cloud

• Many HPC algorithms have to be optimized towards underlying computing architecture for best performance
• In the past, on vector, parallel, grid, and dataflow systems, the system architecture was known to the user
• Then, hand-optimization was possible by restructuring the core algorithms
• In the cloud, user has no information about individual system components, overall architecture, heterogeneity, etc
• In the cloud, in addition, virtualization hides the physical architecture from the user, and introduces additional performance uncertainties
Lessons Learned & Recommendations

- HPC Clusters for number crunching hand-optimized codes
- Grids for collaborative computing and for ‘plumbing’
- Fact is: Clouds are becoming the new Utility
- In the Cloud: no infos about system specifics, architecture, virtualization, network, etc...
- Successively, Clouds become inexpensive and will soon show better performance
- Engineering software ASPs will move from licenses-based models to service-oriented pay-per-use models.
Thank You for your attention

Gentzsch @ rzg.mpg.de