Grid Computing

Andreas Heiss
Steinbuch Centre for Computing
Computing - the past

Zuse 11
- One of the first commercial computers, available from 1956
- Price: 120000 DM
  - Taking inflation into account, this corresponds to 600000 € today.
- 654 relais, floating-point unit
- 10-20 operations per second

Drum storage (50s ~ 60s)
- <10 Mbit capacity
- 10 Mbit/s transfer rate (fast compared to the capacity!)

No networks

Quellen: http://de.wikipedia.org/wiki/Konrad_Zuse

Quelle: http://de.wikipedia.org/wiki/Trommelspeicher
Computing - today

Example: Intel Core i7 (”Nehalem”)
- 45nm structures
- 4 cores, 8 threads
- ~ 2.0 - 3.0 GHz
- ~12 GB/s memory bandwidth
- 1500 € for a well equipped system

Online storage: hard disks
- up to 2 TB per drive
- seek times < 5ms
- read/write performance <=100 MB/s
- ~ 100 € / TB (low performance, desktop)
- ~ 1000 € / TB (highest speed, 24x7, server)
Computing - today

- **Wide area networking**
  - **Arpanet (Advanced Research Projects Agency Network)**
    - started 1969, 4 institutes
    - ~ 50 kbit/s
  - **Internet today:**
    - ~ 10 Gbit/s backbone

- **Internal data links:**
  - PCI: < 5 Gbit/s
  - PCI-Xpress: < 64 Gbit/s
  - QPI: ~ 100 Gbit/s
Computing - development

- No. of transistors (and also compute power per chip) doubles every 18-24 months. (Moore's law, 1965/1975)

- Storage densities increase by a factor of 1.5 - 1.8 per year.

- Gilder's law: ”The total bandwidth of communication systems triples every 12 months.”
Available compute and storage resources are growing almost exponentially.

Network bandwidths come closer and closer to the speed of internal data links (e.g. PCI).

External resources are accessible almost as fast as local resources.

In such an environment, it is obvious to do certain computing tasks on external resources!

No need to buy and operate special devices for each special task.

Share resources and improve utilisation → minimise costs
Grid Computing

Ian Foster @ Supercomputing 2001, Denver, USA

Foster, Kesselman: The GRID: Blueprint for a new computing infrastructure (1999)
Grid Computing

”Grid computing is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations” (I. Foster)

”Grid” derives from ”power grid”

**The vision:**

- Get computing power from an (computing) power outlet
- Not only compute power, but also storage, access to measuring instruments, sensors, ...
- Just plug in and access resources worldwide
Grid Computing

- **Middleware** is the 'glue' software that pools together various resources and services and to create the Grid.
  - Interfaces to
    - access compute power
    - access data
    - authentication and authorization (PKI, X.509)
    - information service
    - accounting
    - ...

**Grid Computing**

![Diagram of Grid Computing layers]

- Application & serviceware layer
- Middleware layer
- Resource layer
- Network
  - Supercomputer
  - Storage
  - Servers
  - Grid Information Service
  - Collaboration and Remote Instrument
  - Security
  - Uniform Data Access
  - Grid and OGSA Hosting
  - Uniform Computing Access
  - Authentication Authorization
  - Communication Services
Grid Computing
Grid Computing

- operating system(s)
- available job slots
- installed software
- services provided
- supported VOs
- ...

Plot courtesy Martina Hardt / Desginal
Grid Computing

Plot courtesy Martina Hardt / Desginal
Grid Computing

Plot courtesy Martina Hardt / Desginal

large output
EGEE - world's largest Grid infrastructure

- July 2009:
  - ~150 VOs (virtual organisations)
  - ~ 17000 users
  - ~ 290 resource centres in 55 countries
  - ~ 140000 CPUs (cores)
  - > 25 Petabytes of online storage
  - ~330000 jobs per day

- gLite middleware
  - Compute Elemente (CE)
  - Storage Elements (SE)
  - Workload Management Systems (WMS)
  - File Catalogs (LFC)
  - File Transfer Service (FTS)
  - Information System (BDII)
  - ...

http://glite.web.cern.ch

~ 80 people in 12 academic and industrial research centres

http://www.eu-egee.org
**EGEE - world's largest Grid infrastructure**

- **EGEE-III**: 2 years
  - EU co-funding ~ 32M€
  - Total budget ~ 47 M€
  - Equipment ~ 50 M€
  - 9132 person months
  - ~ 382 FTE

---

### EGEE activities

#### Networking

- NA1: Project Management
- NA2: Dissemination, Communication and Outreach
- NA3: User Training and Induction
- NA4: User Community Support and Expansion
- NA5: Policy and International Cooperation

#### Services

- SA1: Grid Operations, Support and Management
- SA2: Networking Support
- SA3: Integration, Testing and Certification

#### Joint Research

- JRA1: Middleware Re-engineering
EGEE - world's largest Grid infrastructure

EGEE regional federations:
- Asia Pacific (Australia, Japan, Korea, Taiwan)
- Benelux (Belgium, the Netherlands)
- Central Europe (Austria, Croatia, Czech Republic, Hungary, Poland, Slovakia, Slovenia)
- France
- Germany/Switzerland
- Italy
- Nordic countries (Finland, Sweden, Norway)
- South West Europe (Portugal, Spain)
- South East Europe (Bulgaria, Cyprus, Greece, Israel, Romania, Serbia, Turkey)
- Russia
- United Kingdom/Ireland
- USA
EGEE - world's largest Grid infrastructure

no. of jobs in EGEE last 12 months

PRODUCTION Cumulative Total number of jobs by REGION and DATE
TOP10 VOs. August 2008 - August 2009

110,000,000
EGEE - world's largest Grid infrastructure

http://gridportal.hep.ph.ic.ac.uk/rtm/
Bandwidth vs. latency

- Network bandwidth scales almost indefinite (matter of money)
- Latency does not scale!

→ Not all computational problems are suited for the Grid (parallel computing: e.g. weather simulation / forecast)
→ Grids ideally suited for embarrassingly parallel workloads, small and large data volumes. (e.g. MC simulation and data analysis for (astro)particle physics.)
The Large Hadron Collider (LHC)
The Large Hadron Collider (LHC)
The Large Hadron Collider (LHC)
The Large Hadron Collider (LHC)

Tracks from a Higgs decay in the CMS tracker (76 mio. readout channels)

Alice:
40 MHz collisions of lead nuclei
The Worldwide LHC Computing Grid (WLCG)

Data-reduction 1/10 Mio.

Level 1 - special hardware
40 MHz x 25 MB = 1 PB/sec = 1000 TB/sec equivalent

Level 2 - Embedded Processors
75 KHz (75 GB/sec)

Level 3 – PC Farm (Linux)
5 KHz (5 GB/sec)

100 Hz (100 MB/sec)

~ 1-2 PB per year per experiment (+ MC data)

~ 1-2 PB per year per experiment (+ MC data)
The Worldwide LHC Computing Grid (WLCG)

10000 physicists worldwide ...

.... want to analyse the LHC data.
The Worldwide LHC Computing Grid (WLCG)

Tier-1 (11 sites)
- Manage permanent storage (RAW, simulated, processed)
- Capacity for re-processing and bulk analysis

Tier-2 (~120 sites)
- Monte Carlo event simulation
- User analysis

Tier-3
- Institute
- Institute
- Institute

10000 physicists worldwide

100 - 400 MB/s per Exp.
GridKa - the German WLCG Tier-1

- Started 2002, requested by the German High Energy Physics community
- Supports many international experiments
  - LHC: all 4 experiments, Alice, Atlas, CMS, LHCb
  - non-LHC HEP: Babar, (Super)Belle, CDF, Compass, D0
  - Astroparticle, other non-HEP:
    Auger, Magic, Medigrid, ...
- Provides Resources and services for EGEE and D-Grid
- 7820 CPU cores
- ~ 30000 jobs per day
GridKa - the German WLCG Tier-1

- 6300 TB disk
- 8500 TB tape
- > 50 Gbit/s WAN
GridKa - the German WLCG Tier-1

- Storage and computing resources

![Bar chart showing CPU, Disk, and Tape storage from 2008 to 2014]
GridKa WAN connections

Network failure of the LHCOPN link between CERN and FZK on April 26\textsuperscript{th} / 27\textsuperscript{th} 2007

automatic routing of T0-T1 traffic over the backup link via CNAF
GridKa services for WLCG and EGEE (regional core services)

- File transfer service
- File catalog
- Information system
- Workload management

- Used by Tier-2, university Grid centres and experiment-specific high level services (job submission robots, data management systems)
- Highest reliability necessary
Data management in WLCG

FTS service @ Tier-1 controls T1↔T2 traffic (both directions) and traffic from other Tier-1s.

- scheduling
- VO shares
- transfer parameters (#files, #streams)

User

'subscribe' dataset to a site

e.g. ATLAS/CMS Data Management Systems (DDM)

FTS, LFC

User

issue FTS job, check status, register files in LFC

CERN FTS, LFC

Tier-1

Tier-2

Tier-2

Tier-2
## CMS data management

### PhEDEx – CMS Data Transfers

**Info** | **Activity** | **Data** | **Requests** | **Components** | **Reports**
---|---|---|---|---|---
**Rate** | **Rate Plots** | **Queue Plots** | **Quality Plots** | **Routing** | **Transfer Details** | **Deletions** | **Recent Errors**

### Last hour

**Time span** | Last hour | Include links with nothing but errors | Update
---|---|---|---

<table>
<thead>
<tr>
<th>To</th>
<th>From</th>
<th>Files</th>
<th>Total Size</th>
<th>Rate</th>
<th>Errors</th>
<th>Expired</th>
<th>Avg. Est. Rate</th>
<th>Avg. Est. Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1_US_FINAL_Buffer</td>
<td>T0_CH_CERN_Export</td>
<td>63</td>
<td>125.1 GB</td>
<td>35.6 MB/s</td>
<td>-</td>
<td>-</td>
<td>37.0 MB/s</td>
<td>0h25</td>
</tr>
<tr>
<td>T2_CH_CAF</td>
<td>T0_CH_CERN_Export</td>
<td>93</td>
<td>108.7 GB</td>
<td>30.9 MB/s</td>
<td>-</td>
<td>-</td>
<td>36.4 MB/s</td>
<td>0h52</td>
</tr>
<tr>
<td>T1_IT_CNAF_Buffer</td>
<td>T0_CH_CERN_Export</td>
<td>47</td>
<td>103.9 GB</td>
<td>29.6 MB/s</td>
<td>-</td>
<td>-</td>
<td>35.7 MB/s</td>
<td>0h14</td>
</tr>
<tr>
<td>T1_US_FINAL_MSS</td>
<td>T1_US_FINAL_Buffer</td>
<td>48</td>
<td>101.1 GB</td>
<td>28.8 MB/s</td>
<td>-</td>
<td>-</td>
<td>86.0 MB/s</td>
<td>0h13</td>
</tr>
<tr>
<td>T1_IT_CNAF_MSS</td>
<td>T1_IT_CNAF_Buffer</td>
<td>43</td>
<td>91.7 GB</td>
<td>26.1 MB/s</td>
<td>-</td>
<td>-</td>
<td>36.0 MB/s</td>
<td>0h04</td>
</tr>
<tr>
<td>T2_IT_Rome</td>
<td>T1_FR_CCIN2P3_Buffer</td>
<td>62</td>
<td>57.3 GB</td>
<td>19.2 MB/s</td>
<td></td>
<td></td>
<td>18.5 MB/s</td>
<td>10h17</td>
</tr>
<tr>
<td>T2_IT_Rome</td>
<td>T1_US_FINAL_Buffer</td>
<td>111</td>
<td>58.4 GB</td>
<td>16.6 MB/s</td>
<td>-</td>
<td>-</td>
<td>15.9 MB/s</td>
<td>4d10h58</td>
</tr>
<tr>
<td>T2_FR_CCIN2P3</td>
<td>T1_IT_CNAF_Buffer</td>
<td>30</td>
<td>27.8 GB</td>
<td>7.9 MB/s</td>
<td>-</td>
<td>11</td>
<td>7.1 MB/s</td>
<td>15h47</td>
</tr>
<tr>
<td>T2_BR_UERJ</td>
<td>T1_DE_FZK_Buffer</td>
<td>10</td>
<td>21.8 GB</td>
<td>6.2 MB/s</td>
<td></td>
<td></td>
<td>6.0 MB/s</td>
<td>4d23h21</td>
</tr>
<tr>
<td>T2_FR_CCIN2P3</td>
<td>T1_DE_FZK_Buffer</td>
<td>18</td>
<td>17.4 GB</td>
<td>5.0 MB/s</td>
<td>-</td>
<td>-</td>
<td>7.4 MB/s</td>
<td>3d0h33</td>
</tr>
<tr>
<td>T1_UK_RAL_Buffer</td>
<td>T0_CH_CERN_Export</td>
<td>9</td>
<td>13.9 GB</td>
<td>4.0 MB/s</td>
<td>-</td>
<td>-</td>
<td>5.0 MB/s</td>
<td>0h16</td>
</tr>
<tr>
<td>T1_UK_RAL_MSS</td>
<td>T3_UK_RAL_Buffer</td>
<td>8</td>
<td>11.8 GB</td>
<td>3.4 MB/s</td>
<td>-</td>
<td>-</td>
<td>7.6 MB/s</td>
<td>0h06</td>
</tr>
<tr>
<td>T2_TW_Taiwan</td>
<td>T1_US_FINAL_Buffer</td>
<td>3</td>
<td>5.3 GB</td>
<td>1.5 MB/s</td>
<td>-</td>
<td>-</td>
<td>9.8 MB/s</td>
<td>0h19</td>
</tr>
<tr>
<td>T2_IT_Rome</td>
<td>T1_DE_FZK_Buffer</td>
<td>8</td>
<td>4.1 GB</td>
<td>1.2 MB/s</td>
<td>-</td>
<td>-</td>
<td>3.0 MB/s</td>
<td>6d6h25</td>
</tr>
<tr>
<td>T1_ES_PIC_Buffer</td>
<td>T1_US_FINAL_Buffer</td>
<td>1</td>
<td>2.5 MB</td>
<td>716.1 B/s</td>
<td>-</td>
<td>-</td>
<td>8.7 kB/s</td>
<td>0h04</td>
</tr>
<tr>
<td>T1_ES_PIC_MSS</td>
<td>T1_ES_PIC_Buffer</td>
<td>1</td>
<td>114.1 kB</td>
<td>32.5 B/s</td>
<td>-</td>
<td>-</td>
<td>3.7 kB/s</td>
<td>0h37</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>555</td>
<td>758.3 GB</td>
<td>215.7 MB/s</td>
<td>11</td>
<td>11</td>
<td>-/s</td>
<td>0h00</td>
</tr>
</tbody>
</table>
EGEE jobs last year

PRODUCTION Total elapsed time per VO
TOP 10 VOs (and Other VOS). August 2008 - August 2009

TIER1 Normalised Elapsed time per TIER1
ALL VOs. August 2008 - August 2009

ISAPP summer school
A Grid job example (1)

heiss@gridka25:~> voms-proxy-init --voms dteam
Enter GRID pass phrase:
Your identity: /O=GermanGrid/OU=FZK/CN=Andreas Heiss
Creating temporary proxy .............................................. Done
Contacting voms.cern.ch:15004 [/DC=ch/DC=cern/OU=computers/CN=voms.cern.ch]
"dteam" Done
Creating proxy .................................................. Done
Your proxy is valid until Fri Jul 24 02:11:55 2009

heiss@gridka25:~> cat job.jdl
Type = "Job";
JobType = "Normal";
Executable = "/bin/hostname";
StdOutput = "hello.out";
StdError = "hello.err";
OutputSandbox = {"hello.err","hello.out"};
RetryCount = 2;
VirtualOrganisation = "dteam";
A Grid job example (2)

heiss@gridka25:~> glite-wms-job-list-match -a job.jdl

Connecting to the service https://wms-3-fzk.gridka.de:7443/glite_wms_wmproxy_server

COMPUTING ELEMENT IDs LIST
The following CE(s) matching your job requirements have been found:

*CEId*
- agh2.atlas.unimelb.edu.au:2119/jobmanager-lcgpbs-dteam
- alice003.nipne.ro:2119/jobmanager-lcgpbs-dteam
- alice19.spbu.ru:2119/jobmanager-lcgpbs-dteam
- atlasce.physics.sinica.edu.tw:2119/jobmanager-lcgcondor-dteam
- atlasce01.na.infn.it:2119/jobmanager-lcgpbs-cert
- axon-g01.ieeta.pt:2119/jobmanager-lcgpbs-dteam
- bigmac-lcg-ce2.physics.utoronto.ca:2119/jobmanager-pbs-dteam
- bugaboo-hep.westgrid.ca:2119/jobmanager-lcgpbs-dteam
- ce-01.grid.sissa.it:2119/jobmanager-lcgpbs-cert
- ce-01.roma3.infn.it:2119/jobmanager-lcgpbs-cert
- ce-1-fzk.gridka.de:2119/jobmanager-pbspro-gLite3
- ce-2-fzk.gridka.de:2119/jobmanager-pbspro-gLite3
- ce-3-fzk.gridka.de:2119/jobmanager-pbspro-gLite3
- ce-4-fzk.gridka.de:2119/jobmanager-pbspro-gLite3
- ce-alice.sdfarm.kr:2119/jobmanager-lcgpbs-dteam
- ce-cyb.ca.infn.it:2119/jobmanager-lcglslf-ponetcert
- ....
A Grid job example (3)

```
heiss@gridka25:~> glite-wms-job-submit -a job.jdl

Connecting to the service https://wms-3-fzk.gridka.de:7443/glite_wms_wmproxy_server

====================== glite-wms-job-submit Success ========================

The job has been successfully submitted to the WMProxy
Your job identifier is:

https://lb-2-fzk.gridka.de:9000/P_NEqUnjyvz53uyIXDXTLw

============================================================================
```
A Grid job example (4)


******************************************************************************
BOOKKEEPING INFORMATION:

Status info for the Job : https://lb-2-fzk.gridka.de:9000/P_NEqUnjyvz53uyIXDXTLw
Current Status:     Scheduled
Status Reason:      Job successfully submitted to Globus
Destination:        grid-ce01.esrf.eu:2119/jobmanager-pbs-short
Submitted:          Thu Jul 23 14:50:49 2009 CEST
******************************************************************************


******************************************************************************
BOOKKEEPING INFORMATION:

Status info for the Job : https://lb-2-fzk.gridka.de:9000/P_NEqUnjyvz53uyIXDXTLw
Current Status:     Done (Success)
Logged Reason(s):
- Job terminated successfully
Exit code:          0
Status Reason:      Job terminated successfully
Destination:        grid-ce01.esrf.eu:2119/jobmanager-pbs-short
Submitted:          Thu Jul 23 14:50:49 2009 CEST
******************************************************************************

European synchrotron rad. facility, Grenoble
A Grid job example (5)

heiss@gridka25:~> glite-wms-job-output https://lb-2-fzk.gridka.de:9000/P_NEqUnjyvz53uy..
Connecting to the service https://wms-3-fzk.gridka.de:7443/glite_wms_wmproxy_server

================================================================================
JOB GET OUTPUT OUTCOME

Output sandbox files for the job: https://lb-2-fzk.gridka.de:9000/P_NEqUnjyvz53uyIXDXTLw have been successfully retrieved and stored in the directory: /tmp/jobOutput/heiss_P_NEqUnjyvz53uyIXDXTLw

================================================================================

heiss@gridka25:~> cat /tmp/jobOutput/heiss_P_NEqUnjyvz53uyIXDXTLw/hello.out
nuni07
Cloud Computing

- The ”Grid” hype is over.
  - Grid is an established technique.
  - Not much used in industry but established in science
  - Main reasons: data security, Grids built for special user communities, usability

- New Hype: ”Cloud Computing”

Quelle: Google Trends
Cloud Computing

Simply Explained - Part 17: Cloud Computing
Cloud Computing

- Wikipedia: "Cloud computing is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet."

- The concept generally incorporates combinations of the following:
  - infrastructure as a service (IaaS)
  - platform as a service (PaaS)
  - software as a service (SaaS)

- Amazon, Google, "OpenCirrus" Cloud testbed

- Has its roots more in industry than in science
  - Problems to set global (open) standards?
  - Probably: creation of several stable and flexible (but incompatible) solutions → survival of the fittest?
Cloud Computing

Cloud computing pursuits the same aims and visions as Grid computing

```
echo

"Grid computing is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations (I.Foster)"
```

```
| sed s/Grid/Cloud/g
```
Cloud Computing

- Ease of use: within minutes you can learn to use cloud resources
- Very flexible

- So far, automation is not on the same level than in Grids (global batch submission system)
- Large data volumes
- Service Levels? (Services distributed in the cloud / many cloud providers)
Questions?