Leibniz Supercomputing Centre
of the Bavarian Academy of Sciences and Humanities
Introduction to the usage of the Linux Cluster and the Compute Cloud@LRZ

Ferdinand.Jamitzky@lrz.de
The Leibniz Supercomputing Centre

Academy Institute of the Bavarian Academy of Science and Humanities

- IT Service Provider for the Munich Universities
- Regional Computing Centre for Research Institutions in Bavaria
- German National Supercomputing Centre
- European Supercomputing Centre
LRZ Overview by regions

**IT Service Provider for Munich Universities**
- Email, Web, Multimedia, IT Security, HelpDesk, Virtual Reality, Trainings, etc.

**Regional Computing Centre for Bavarian Universities and Research institutions**
- ~ 150 PByte Storage/Archive
- Digital Archive of the Bavarian State Library
- Munich Scientific Network

**German National Supercomputing Centre**
**SuperMUC-NG**
- 25 Pflop/s peak
- 300k compute cores
- 0.7 PB main memory
- 50 PB HDD

**European Supercomputing Centre**
- Participating in large European e-Infrastructures
- High Performance Computing
- High Speed Networks
- Grid/Cloud Computing

© 2019 | Ferdinand Jamitzky | Application Support
LRZ as IT Competence Centre: Providing Comprehensive IT Services for Science

- Service Desk
- Application Software Support
- Backup & Archive
- Authentication & Authorization
- IT Security
- Trainings & Workshops
- Tailored Solutions
- Personal Consulting

- High Speed Networking: Munich Scientific Network
- High Performance Computing: SuperMUC, LinuxCluster
- Virtual Reality & Visualisation: V2C (CAVE, Powerwall)
- Big Data: Bavarian State Library Digital Archive

© 2019 | Ferdinand Jamitzky | Application Support
LRZ by Hardware
Operating Cutting-edge IT Infrastructure

High Speed Networking: Munich Scientific Network

High Performance Computing: SuperMUC-NG, LinuxCluster

Big Data: Bavarian State Library Digital Archive

Virtual Reality & Visualisation: V2C (CAVE, Powerwall)
Linux Cluster
- Massively parallel Cluster (CoolMUC-2, CoolMUC-3, IvyMUC)
- Big shared memory system (Teramem)
- Serial Cluster
- Remote Visualisation

SuperMUC-NG
- Massively parallel Cluster (Worker Nodes)
- Big shared memory nodes (Big Nodes)

Cloud Systems (Compute and Storage)
- Compute Cloud (openNebula and OpenStack)
- Long running instances (vmware)
- Data Science Storage DSS
Access Interfaces from Internet

- ssh/putty
- https
- syncandshare
- NFS/cifs
- GridFTP

Login/Access/Mount

Uniform Authentication

User Provided Images/ Webservers/Interface to Internet

- Compute Cloud
- VMware
- Login Nodes
- Rstudio etc...

Special Purpose Hardware

- TeraMem
- GPU/Accelerator

DSS $HOME / $WORK

- Archive
- Scratch $TMP

Massively Parallel Hardware

- Linux Cluster/MPP
- SuperMUC-NG

 mount

mount

mount
LRZ "Linux Cluster" und "Cloud"

CoolMUC-2  CoolMUC-3  IvyMUC  teramem  DGX-1

Data Science Storage DSS  $WORK  $SCRATCH  $HOME  TSM-Archive

Compute Cloud  GPU Cloud

Firewall


User PC
## Linux Cluster Hardware

<table>
<thead>
<tr>
<th>System</th>
<th>CPU</th>
<th>#core</th>
<th>RAM GB</th>
<th>#nodes</th>
<th>#cores</th>
<th>#nodes</th>
<th>#cores</th>
<th>#time</th>
<th>RAM</th>
<th>queue</th>
<th>login node</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linux-Cluster</td>
<td>CoolMUC-2</td>
<td>28</td>
<td>64</td>
<td>384</td>
<td>10,752</td>
<td>60</td>
<td>1680</td>
<td>48h</td>
<td>3.8 TB</td>
<td>mpp2</td>
<td>lxlogin5-7 .lrz.de</td>
</tr>
<tr>
<td></td>
<td>Intel Xeon E5-2690 v3 (“Haswell”)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linux-Cluster</td>
<td>Serial</td>
<td>28</td>
<td>64</td>
<td>1</td>
<td>28</td>
<td>1</td>
<td>28</td>
<td>96h</td>
<td>64 GB</td>
<td>serial</td>
<td>lxlogin5-7 .lrz.de</td>
</tr>
<tr>
<td></td>
<td>Intel Xeon E5-2690 v3 (“Haswell”)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linux Cluster</td>
<td>CoolMUC-3</td>
<td>64</td>
<td>96</td>
<td>148</td>
<td>9,472</td>
<td>148</td>
<td>9472</td>
<td>48h</td>
<td>8.9 TB</td>
<td>mpp3</td>
<td>lxlogin8.lrz.de</td>
</tr>
<tr>
<td></td>
<td>Intel Xeon Phi (Knights Landing)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linux-Cluster</td>
<td>IvyMUC</td>
<td>16</td>
<td>64</td>
<td>31</td>
<td>496</td>
<td>12</td>
<td>192</td>
<td>72h</td>
<td>768 GB</td>
<td>ivymuc</td>
<td>lxlogin10.lrz.de</td>
</tr>
<tr>
<td></td>
<td>Intel Xeon E5-2660 v2 (“Sandy Bridge”)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linux-Cluster</td>
<td>Teramem</td>
<td>96</td>
<td>6,144</td>
<td>1</td>
<td>96</td>
<td>1</td>
<td>96</td>
<td>48h</td>
<td>6.1 TB</td>
<td>mpp4</td>
<td>any cluster login node</td>
</tr>
<tr>
<td></td>
<td>Intel Xeon E7-8890 v4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Cloud Resources

<table>
<thead>
<tr>
<th>System</th>
<th>Architecture</th>
<th>Total</th>
<th>Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine Learning System</td>
<td>Nvidia Pascal P100</td>
<td>8 GPU</td>
<td>128 GB</td>
</tr>
<tr>
<td>OpenNebula Compute Cloud</td>
<td>Intel Xeon E5540</td>
<td>1-20</td>
<td>1-512 GB</td>
</tr>
<tr>
<td></td>
<td>X5650 E5-2660v2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OpenStack Compute Cloud</td>
<td>Intel Xeon („Skylake“)</td>
<td>1-48</td>
<td>96 GB</td>
</tr>
<tr>
<td>LRZ Virtual Machines</td>
<td>Intel Xeon E5-2660 v2</td>
<td>1-8</td>
<td>1-32 GB</td>
</tr>
</tbody>
</table>
Moore’s Law (1965-2015)

Number of transistors doubles every 2 years
Why parallel programming?

End of the free lunch in 2000 (heat death)

Moore's law means not faster processors, only more of them.

But!

2 x 3 GHz < 6 GHz

(cache consistency, multi-threading, etc)
Supercomputer scaling
Evolution of Peak Performance and Memory (Sum over all LRZ systems)

1.000.000
100.000
10.000
1.000
100
10
1


GFlop/s, GByte

Cray Y-MP2
IBM SP2
KSR
Fujitsu
HLRB1: Hitachi SR8000 + Linux-Cluster
HLRB2: SGI Altix 4700 + Linux-Cluster
SuperMUC: IBM series X iDataplex + Linux-Cluster
SuperMUC Phase 1+2
SuperMUC-NG

1 Tera...
1 Peta...
10-fold every 3.5 years
Double every 13 Months

Cray Y-MP2
IBM SP2
SuperMUC Jobsizes 2012-2017
Supercomputer Layout

Supercomputer aka HPC Cluster

Pruned Tree

Switch

Fat Tree

Node

Accelerator: GPU, MIC
Core (4)
Socket (2)

Island
So... what is a Supercomputer?

- It has many off-the-shelf CPUs with vector instructions (AVX, AVX2, AVX512)
- The diskless nodes are connected by a high-speed internal network (Infiniband, OmniPath)
- The compute nodes (no ssh access) have to be programmed using Message Passing (MPI, GPI, ibverbs)
- All nodes are connected to a parallel file system (GPFS, Lustre) which needs special libraries for full speed (MPI-I/O)
- Programs cannot be run interactively, but have to be submitted to the batch scheduler (LoadLeveler, SLURM)
- The Operating System is a version of UNIX (e.g. Linux)
What is a Supercomputer (not)?

- It has overclocked high-speed processors? **NO**
- It has a big internal RAM? **NO (maybe)**
- It runs MS Windows? **NO**
- The CPU runs faster than a desktop PC? **NO**
- It will run my software without changes? **NO (maybe)**
- It will run my software with millions of threads? **NO**
- It will run my old trusted executable? **NO (maybe)**
- It will run my Excel spreadsheets? **NO**
- You can use it interactively? **NO (maybe)**
Levels of Parallelism

- Node Level (e.g. SuperMUC-NG has 6336 nodes)
- Accelerator Level (e.g. DGX-1 has 2 CPUs and 8 GPUs)
- Socket Level (e.g. teramem has 4 CPUs with 24 cores)
- Core Level (e.g. CoolMUC-3 has 64 cores with AVX512)
- Vector Level (e.g. AVX512 has 32 vector registers)

SuperMUC-NG Peak Performance: \(25.3 \text{ PFlop/s} = 6336 \text{ Nodes} \times 2 \text{ Sockets} \times 24 \text{ Cores} \times 32 \text{ Vectors} \times 2.6 \text{ GHz}\)
RStudio
Web GUI for R and Linux Shell

- read and edit code
- enter commands
- open data files
- plot data
- save plots to pdf/jpg

writing code to run and save as a script
loaded/open data files
entering and running commands directly
the current plot, which updates as you type in new commands
Hostname:
lxlogin5.lrz.de
or
lxlogin6.lrz.de
or
lxlogin7.lrz.de

Enter userid and password when asked for
WinSCP
File Manager

Explorer like GUI
Copy files from/to Linux Cluster
The LRZ Compute Cloud

Operational since March, 2015
LRZ Compute Cloud: OpenNebula/OpenStack

Dr. Matteo Lanati, LRZ, Distributed Resources Group
“Supercomputer in a box“: DGX-1 and Teramem1

**Teramem1 System**

Hardware Features:
- 4-way HP DL 580 Gen9
- 96 cores (“Broadwell”)
- 6.1 TB RAM
- Linux Cluster Integration

**Available Software**
- R
- Rstudio
- Redis
- LRZ Software Stack

**Nvidia DGX-1 and DGX-1v System**

Hardware Features:
- 8 x Nvidia Pascal P100/V100 GPUs
- 170/960 Teraflops (GPU FP 16)
- 128 GB GPU memory
- NVLink between GPUs
- 2 x Dual 20-core Intel Xeon E5-2698
- 512GB RAM main memory

**Available Software**
- Tensorflow
- Caffe
- Theano
- CNTK
- MXNet
- Torch
- DIGITS
Jupyter notebooks

Run your interactive program in the browser
Docker

$ docker run hello-world
Hello from Docker!

Advantages:
• reproducible
• portable
• lightweight
• Docker hub repository
• Jupyter out of the box
Terminal in the browser

VNC Connected (encrypted) to: QEMU (one-28986)

Ubuntu 14.04.4 LTS vm-10-155-208-215.cloud.mwn.de tty1

vm-10-155-208-215 login: root
Password:
Last login: Fri Oct 6 16:42:14 CEST 2017 from badulrz-cm43996.us.lrz.de on pts/0
Welcome to Ubuntu 14.04.4 LTS (GNU/Linux 3.13.0-85-generic x86_64)

* Documentation:  https://help.ubuntu.com/

System information as of Sat Oct 7 20:53:39 CEST 2017

System load: 0.32  Memory usage: 1%  Processes: 50
Usage of /: 71.1% of 19.37GB  Swap usage: 0%  Users logged in: 0

Graph this data and manage this system at:
https://landscape.canonical.com/

145 packages can be updated.
88 updates are security updates.

root@vm-10-155-208-215:~# _