



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

IT Service Provider for **Munich** Universities

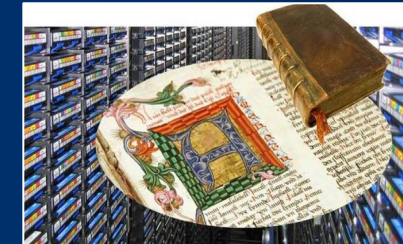
Email, Web,
Multimedia,
IT Security,
HelpDesk,
Virtual Reality,
Trainings, etc.



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



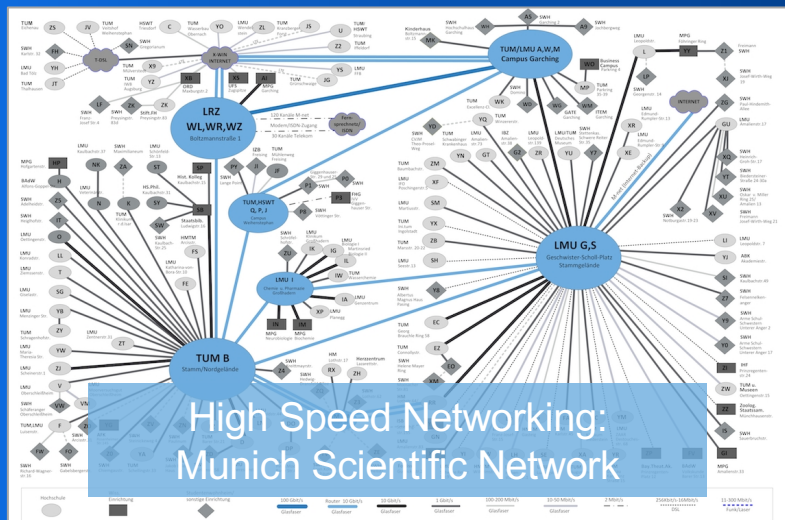
Grid/Cloud
Computing

LRZ as IT Competence Centre: Providing Comprehensive IT Services for Science



LRZ by Hardware

Operating Cutting-edge IT Infrastructure



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
- Machine Learning Systems (<https://datalab.srv.lrz.de>) Single GPU DGX-1, DGX-1v

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...

mount

Special Purpose Hardware

TeraMem

GPU/Accelerator

mount

DSS
\$HOME / \$WORK

Archive

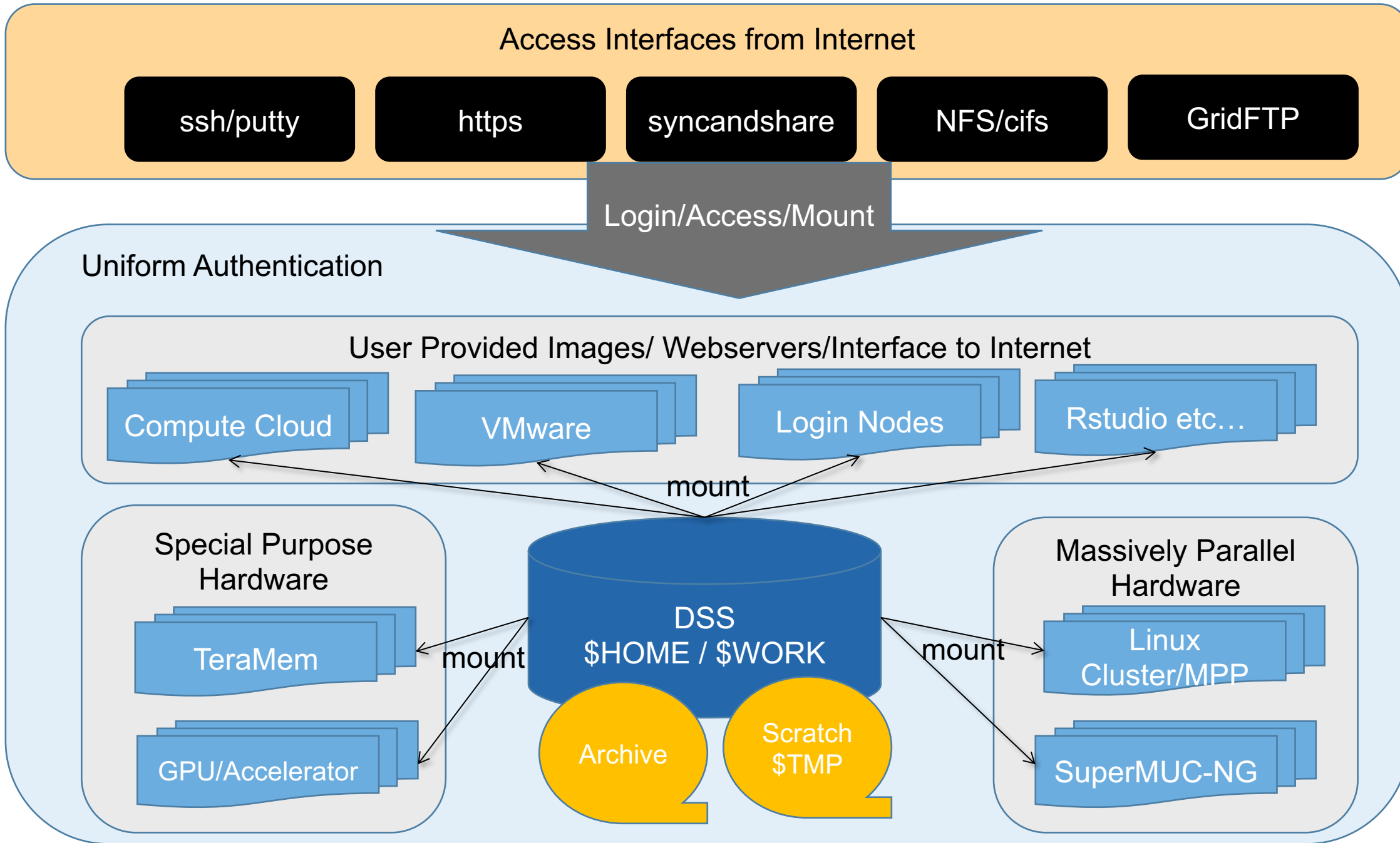
Scratch
\$TMP

mount

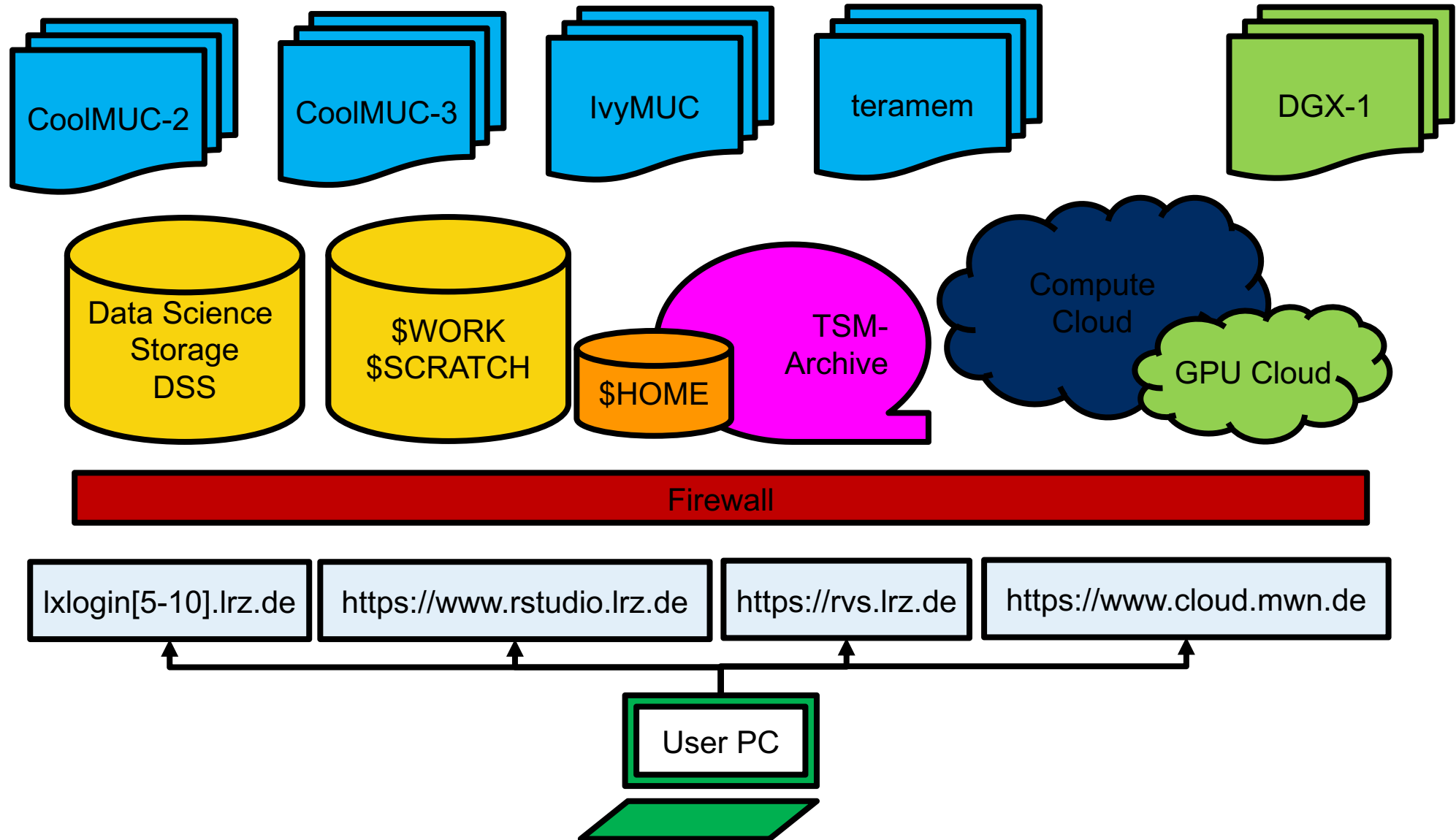
Massively Parallel Hardware

Linux Cluster/MPP

SuperMUC-NG



LRZ "Linux Cluster" und "Cloud"



Linux Cluster Hardware



Architecture				Total		Max Jobs Limits				Access	
System	CPU	#core	RAM GB	#nodes	#cores	#nodes	#cores	#time	RAM	queue	login node
Linux-Cluster CoolMUC-2	Intel Xeon E5-2690 v3 ("Haswell")	28	64	384	10,752	60	1680	48h	3.8 TB	mpp2	lxlogin5-7.lrz.de
Linux-Cluster Serial	Intel Xeon E5-2690 v3 ("Haswell")	28	64	1	28	1	28	96h	64 GB	serial	lxlogin5-7.lrz.de
Linux Cluster CoolMUC-3	Intel Xeon Phi (Knights Landing)	64	96	148	9,472	148	9472	48h	8,9 TB	mpp3	lxlogin8.lrz.de
Linux-Cluster IvyMUC	Intel Xeon E5-2660 v2 ("Sandy Bridge")	16	64	31	496	12	192	72h	768 GB	ivymuc	lxlogin10.lrz.de
Linux-Cluster Teramem	Intel Xeon E7-8890 v4	96	6,144	1	96	1	96	48h	6.1 TB	teramem_inter	any cluster login node

Cloud Ressources



Architecture				Total		Access
System	CPU	#core	RAM GB	#nodes	#cores	login node
Machine Learning System	Nvidia Pascal P100	8 GPU	128	1	n.a.	contact servicedesk
OpenNebula Compute Cloud	Intel Xeon E5540 X5650 E5-2660v2	1-20	1-512	95	852	www.cloud.mwn.de
OpenStack Compute Cloud	Intel Xeon („Skylake“)	1-48	96	64	3072	cc.lrz.de
LRZ Virtual Machines	Intel Xeon E5-2660 v2 ("Sandy Bridge")	1-8	1-32	90	1800	http://www.lrz.de/services/serverbetrieb/

Why parallel programming?

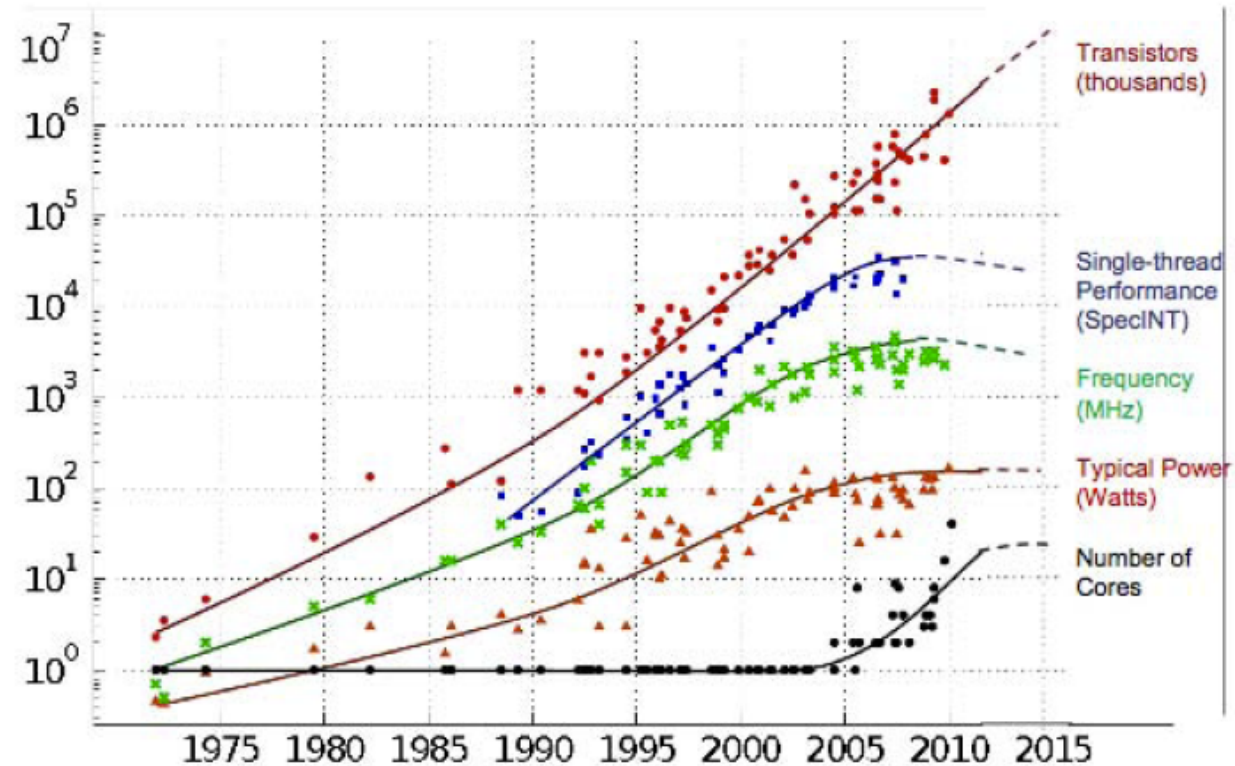
End of the free lunch
in 2000 (heat death)

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

But!
 $2 \times 3 \text{ GHz} < 6 \text{ GHz}$

(cache consistency,
multi-threading, etc)

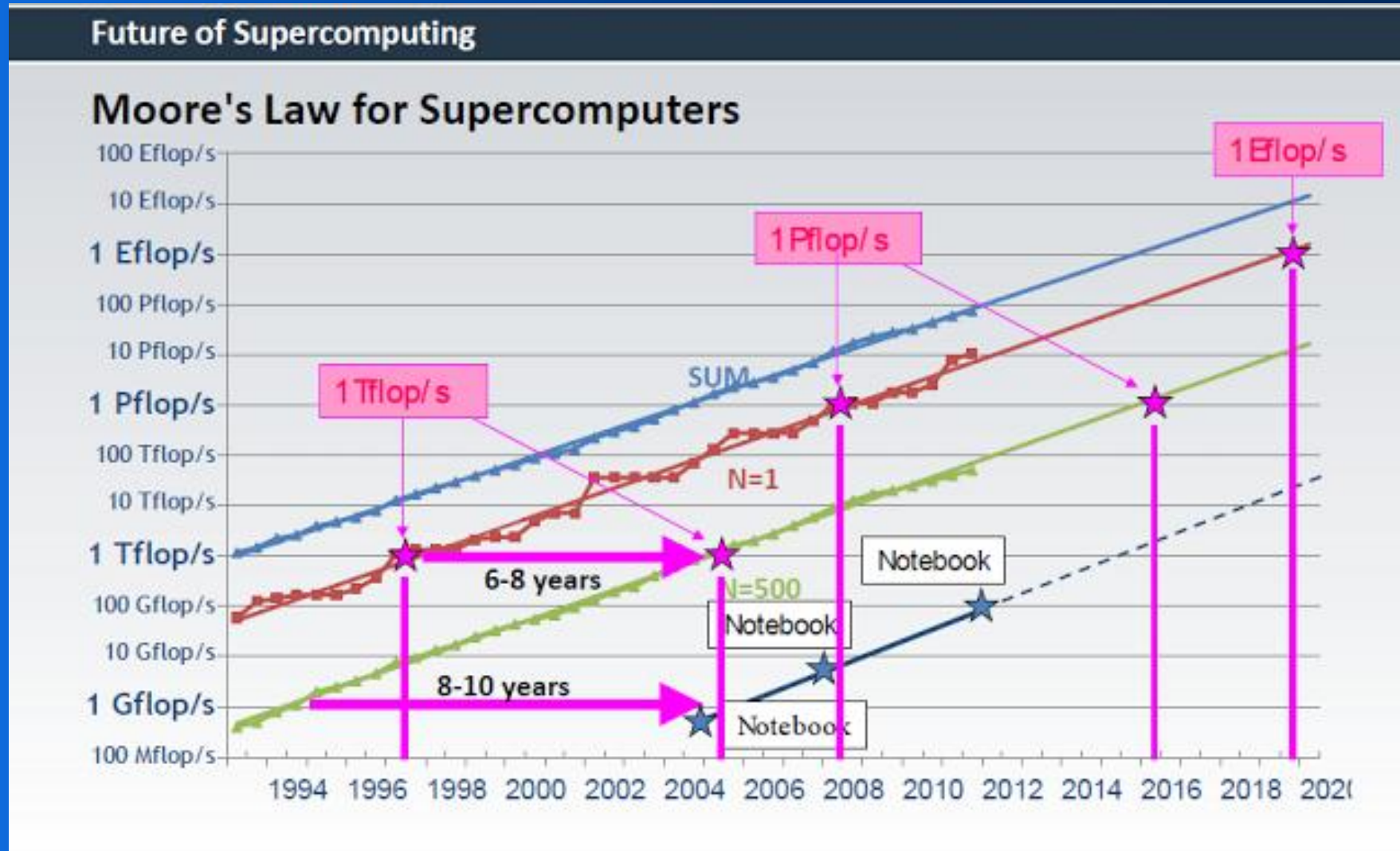
Result: The End of Historic Scaling



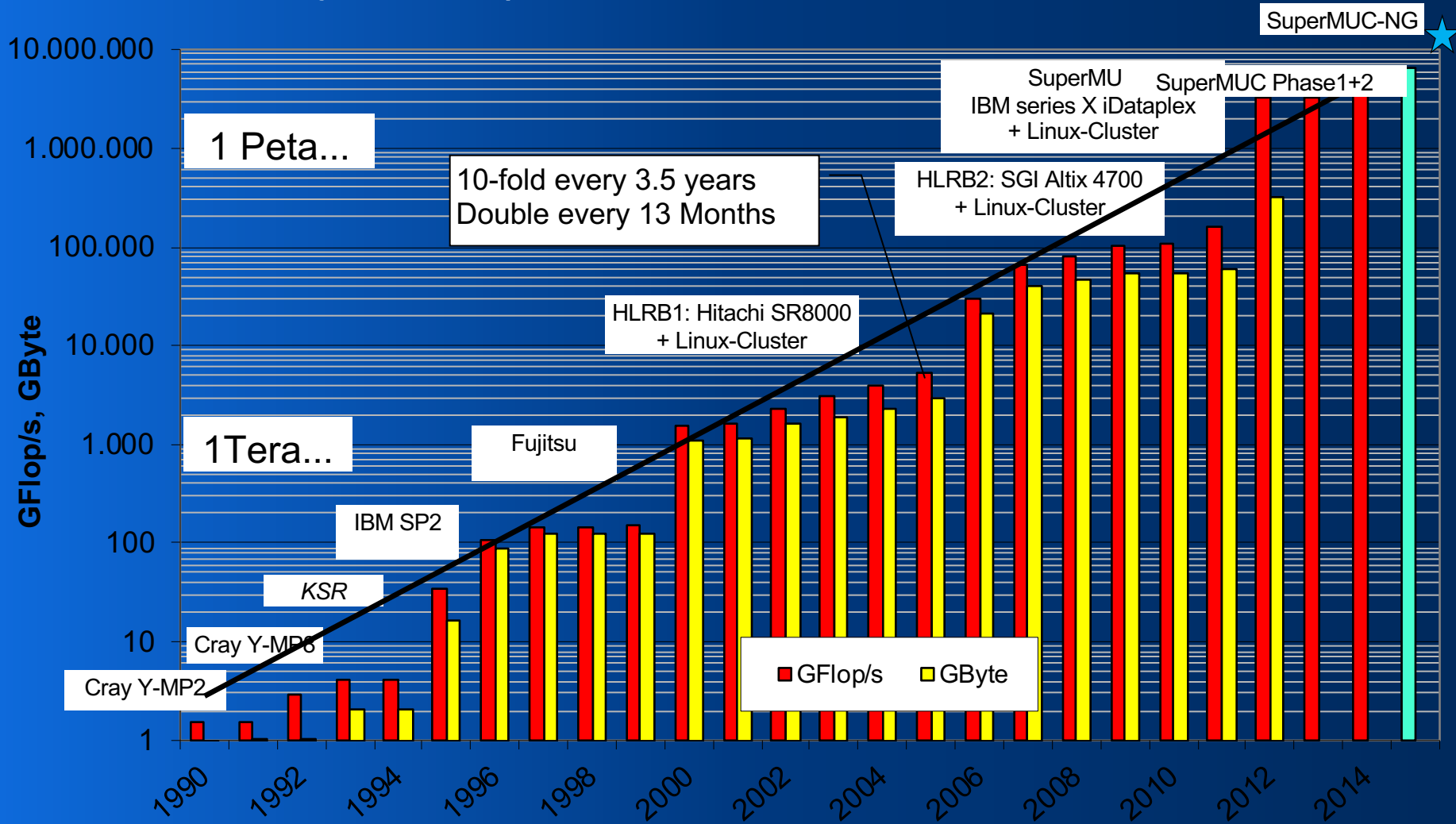
Original data collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond and C. Batten
Dotted line extrapolations by C. Moore

C Moore, *Data Processing in ExaScale-Class Computer Systems*, Salishan, April 2011

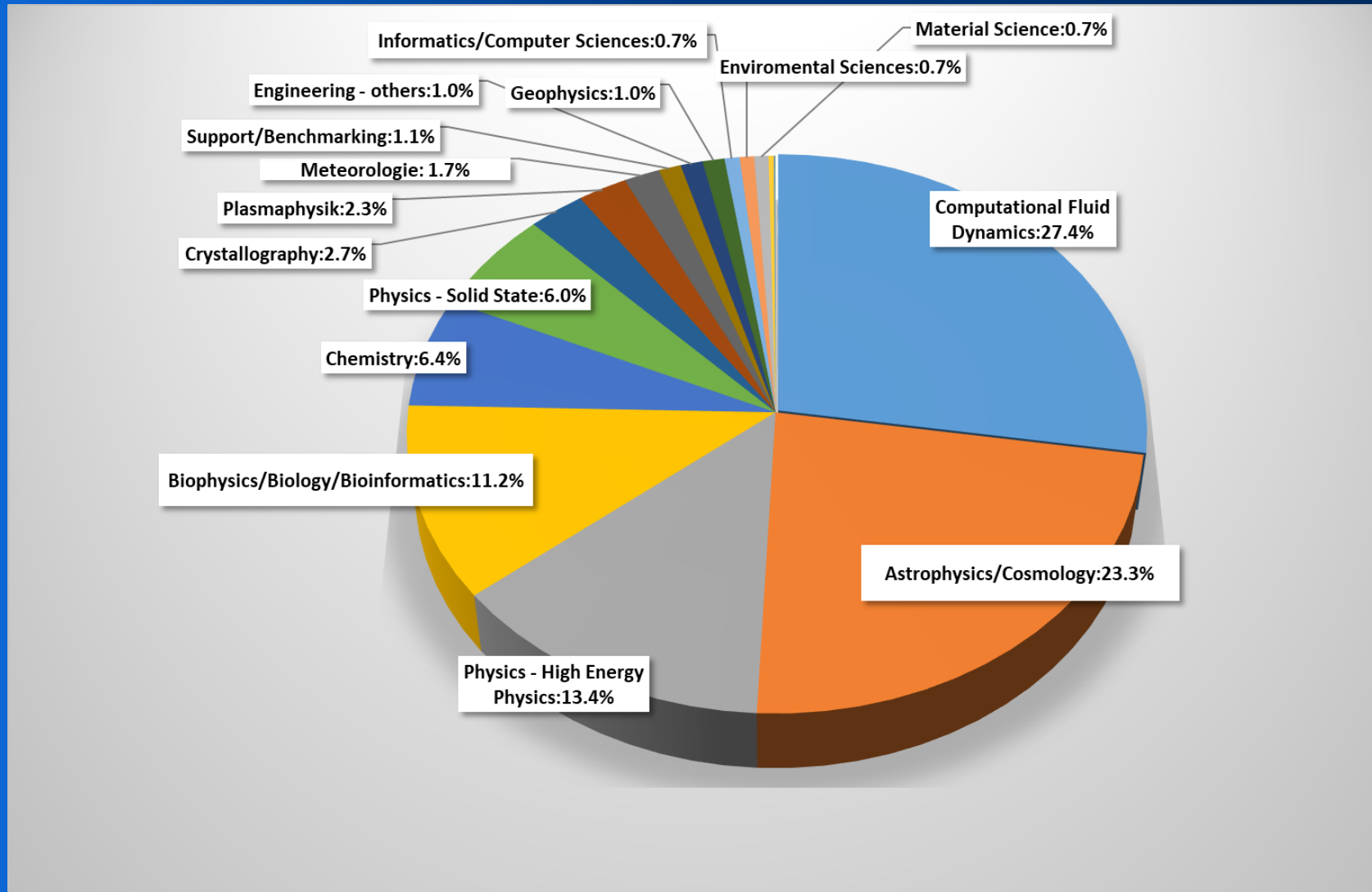
Supercomputer scaling



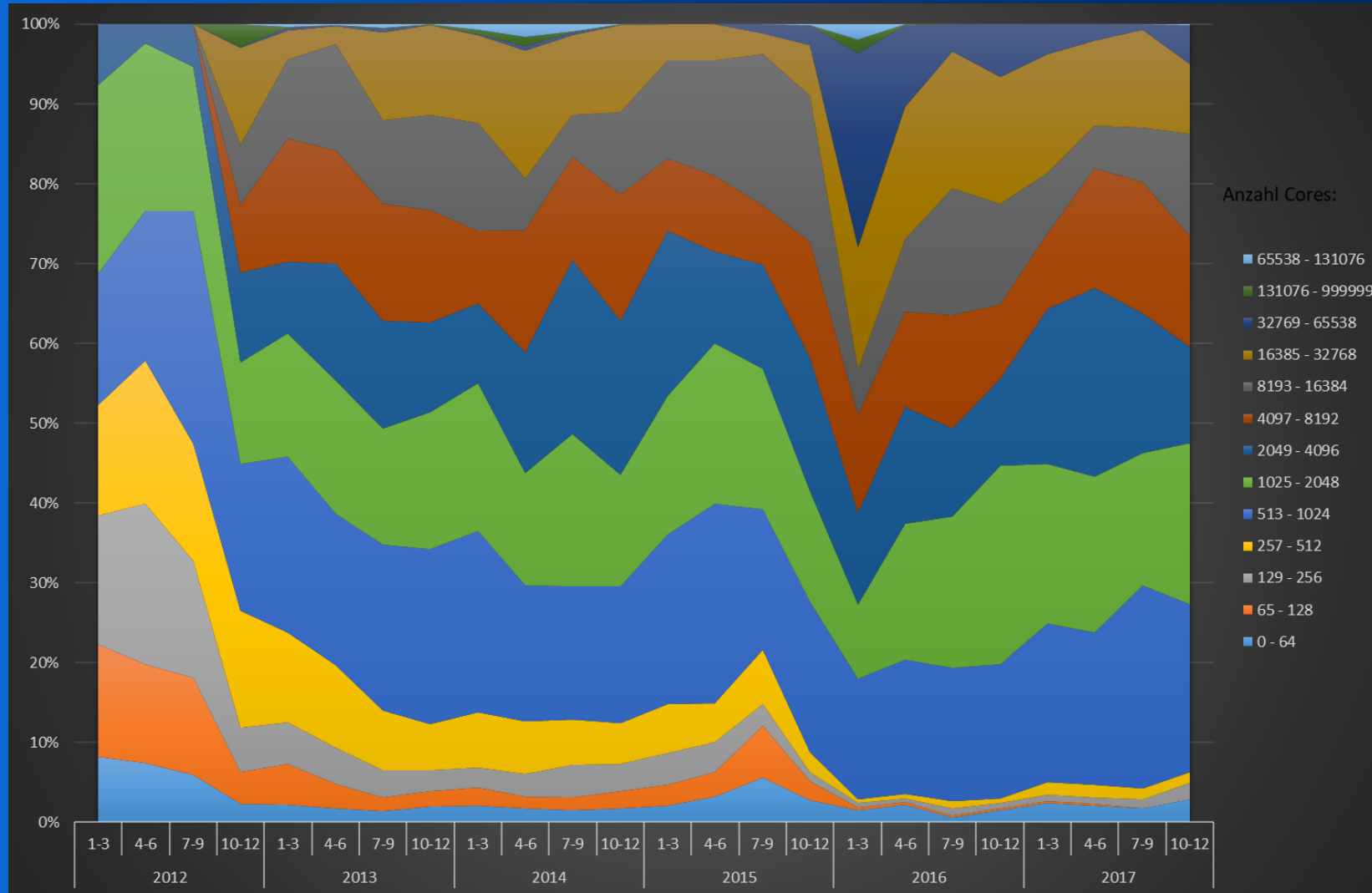
Evolution of Peak Performance and Memory (Sum over all LRZ systems)



SuperMUC Utilization 2017

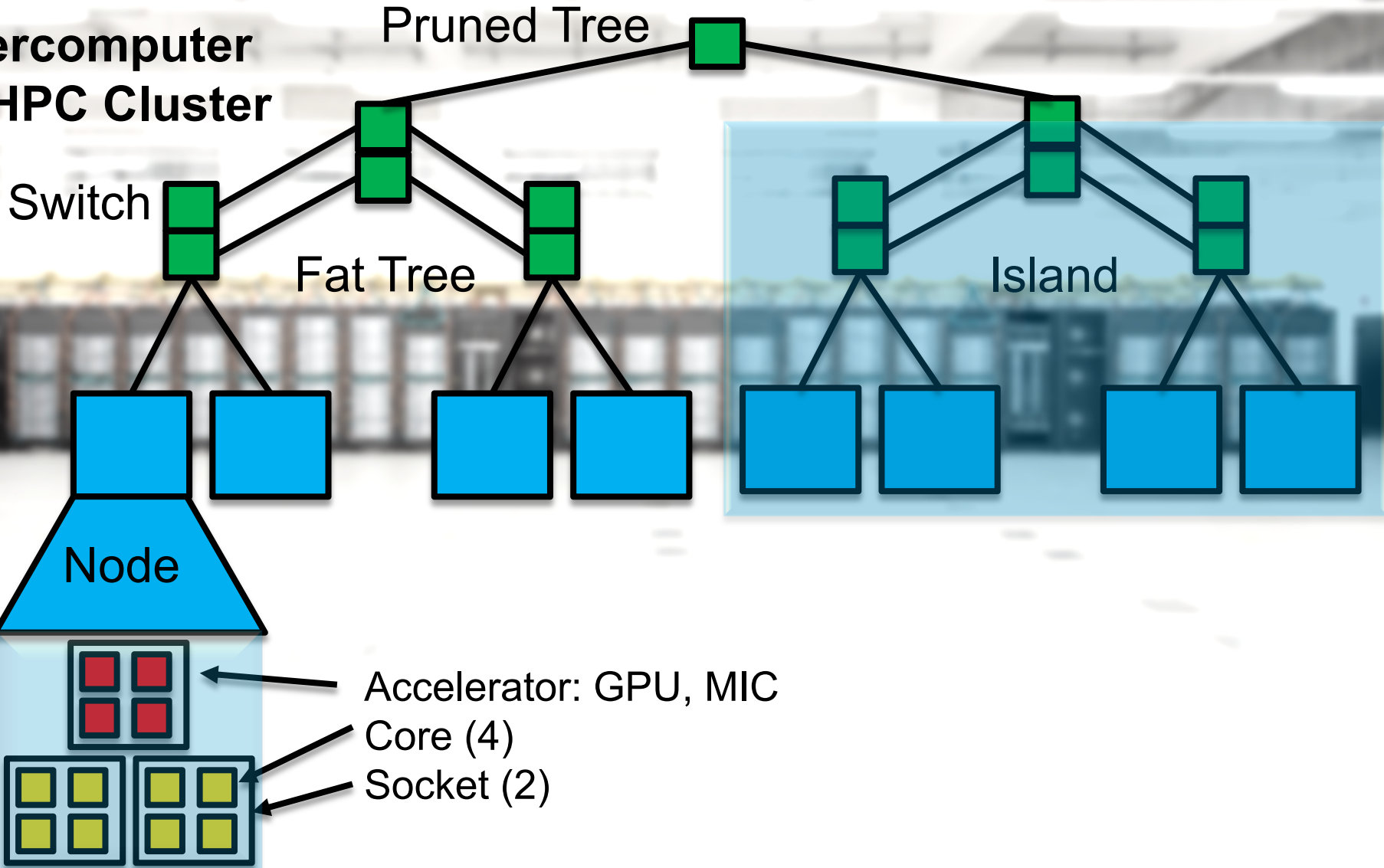


SuperMUC Jobsizes 2012-2017



Supercomputer Layout

Supercomputer
aka HPC Cluster



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

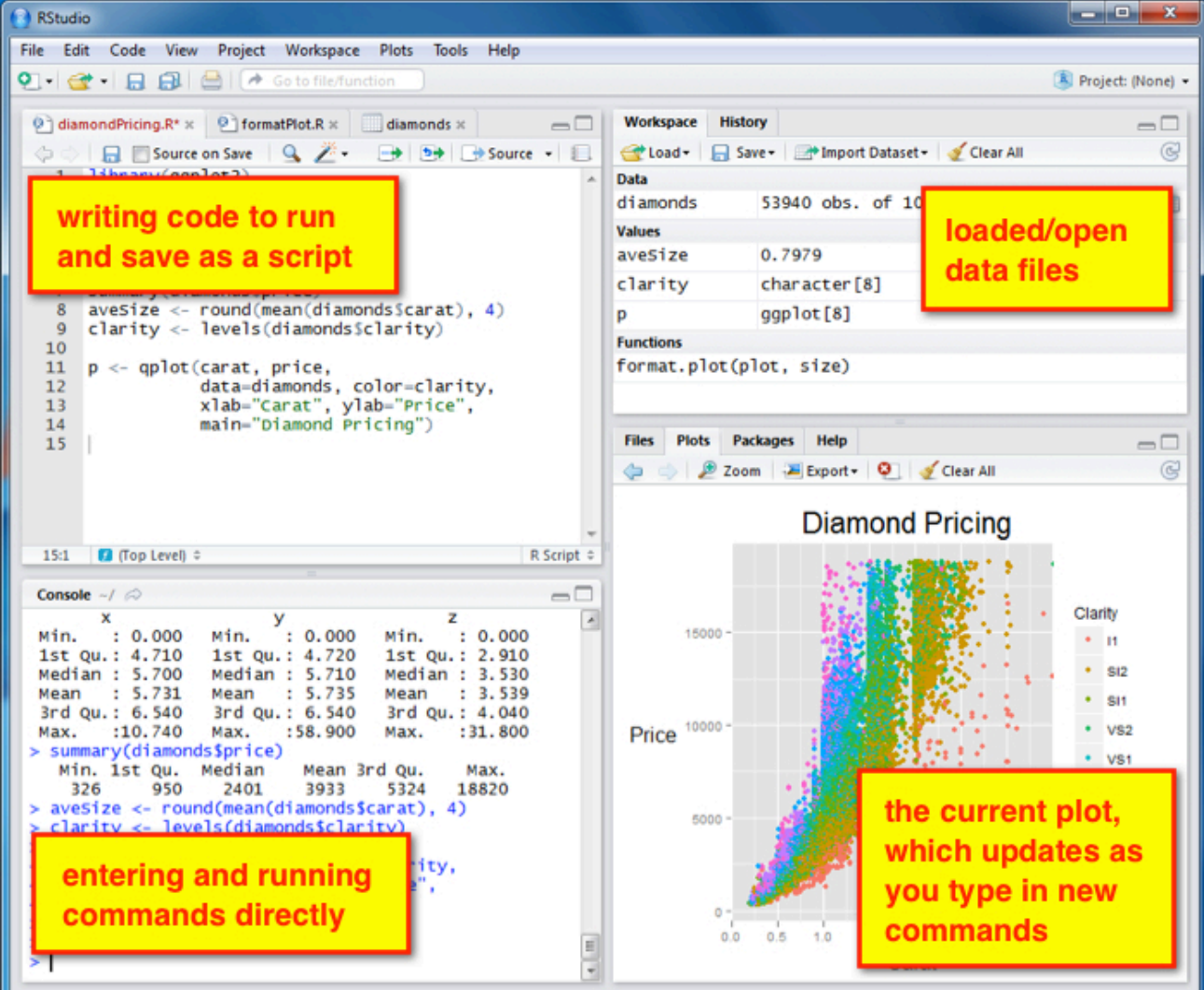


- Node Level (e.g. SuperMUC-NG has 6336 nodes)
- Accelerator Level (e.g. DGX-1 has 2 CPUs and 8 GPUs)
- Socket Level (e.g. teramem has 4 CPUs with 24 cores)
- Core Level (e.g. CoolMUC-3 has 64 cores with AVX512)
- Vector Level (e.g. AVX512 has 32 vector registers)

SuperMUC-NG Peak Performance: **25.3 PFlop/s** =
6336 Nodes x 2 Sockets x 24 Cores x 32 Vectors x 2.6 GHz



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



The screenshot displays the RStudio environment with several key components:

- Source Editor:** Contains R code for data manipulation and plotting. A yellow callout box highlights the text: "writing code to run and save as a script".
- Workspace:** Shows the loaded data object 'diamonds' with 53940 observations. A yellow callout box highlights the text: "loaded/open data files".
- Console:** Displays the output of the executed code, including a summary of the 'diamonds' data frame and the execution of the 'aveSize' and 'clarity' assignment commands. A yellow callout box highlights the text: "entering and running commands directly".
- Plots:** A scatter plot titled "Diamond Pricing" is shown, with 'Price' on the y-axis and 'carat' on the x-axis. Points are colored by 'clarity' (I1, SI2, SI1, VS2, VS1). A yellow callout box highlights the text: "the current plot, which updates as you type in new commands".

Hostname:

lxlogin5.lrz.de

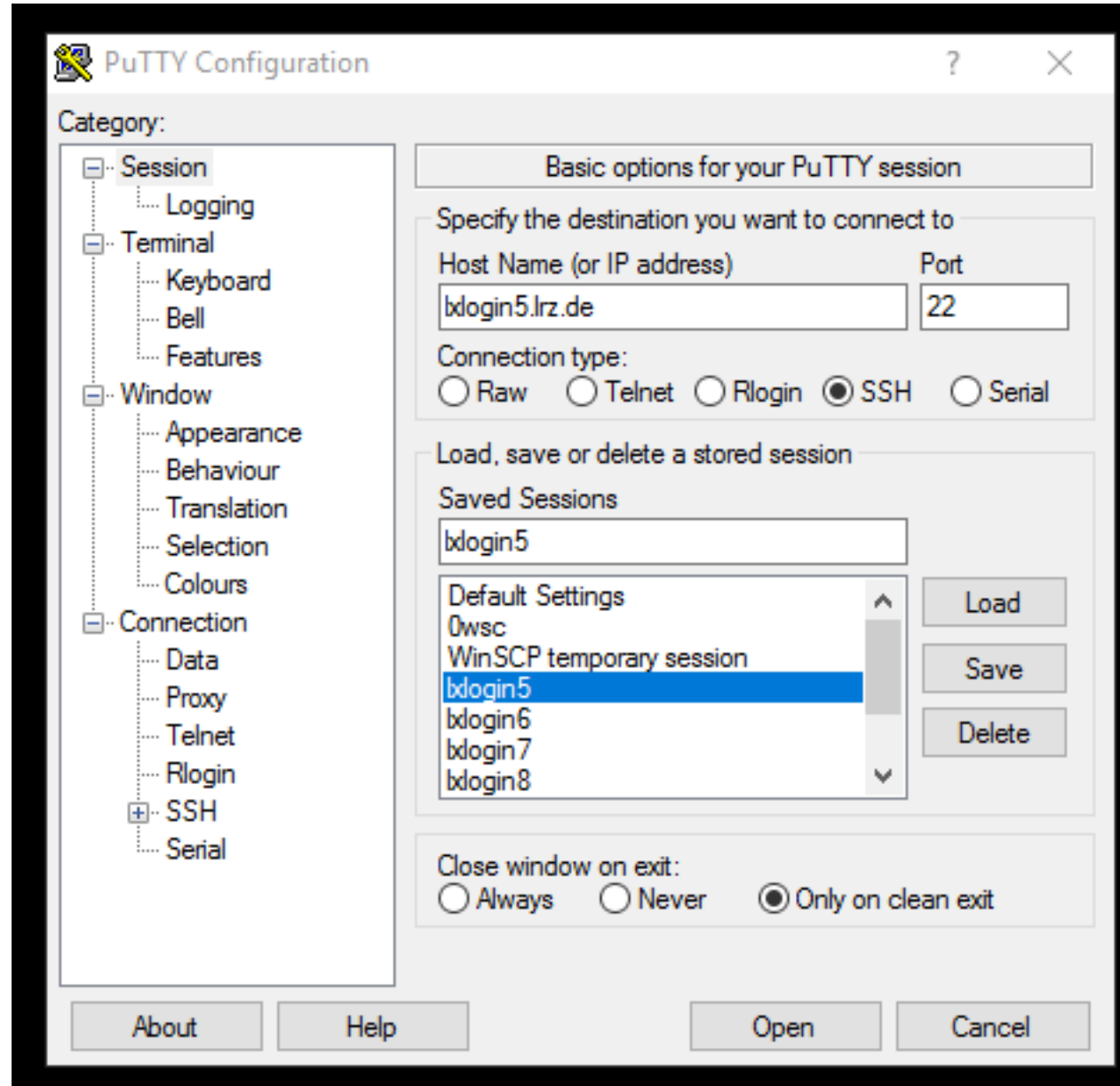
or

lxlogin6.lrz.de

or

lxlogin7.lrz.de

Enter userid and
password when
asked for

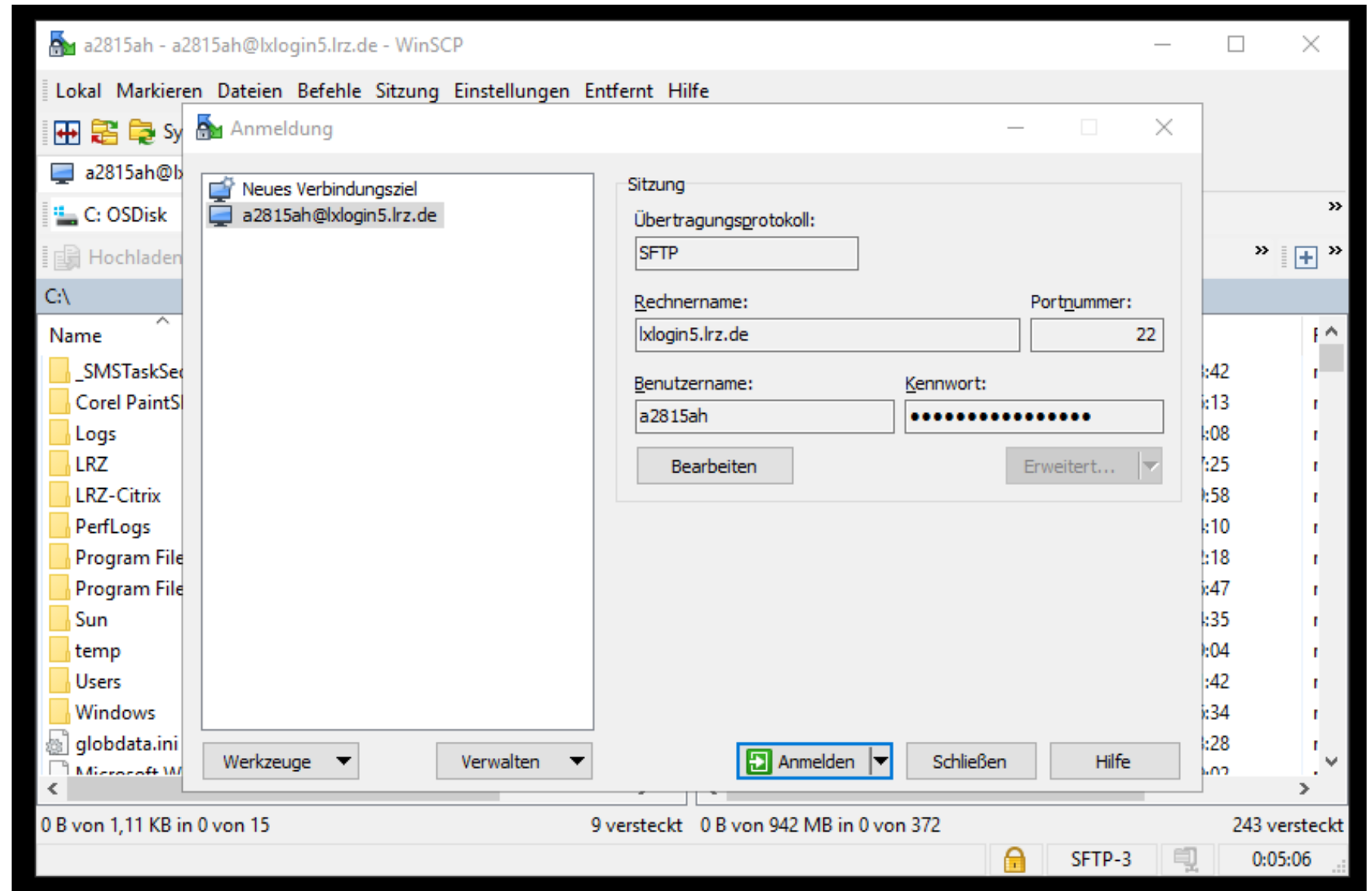


WinSCP File Manager



Explorer like GUI

Copy files from/to
Linux Cluster

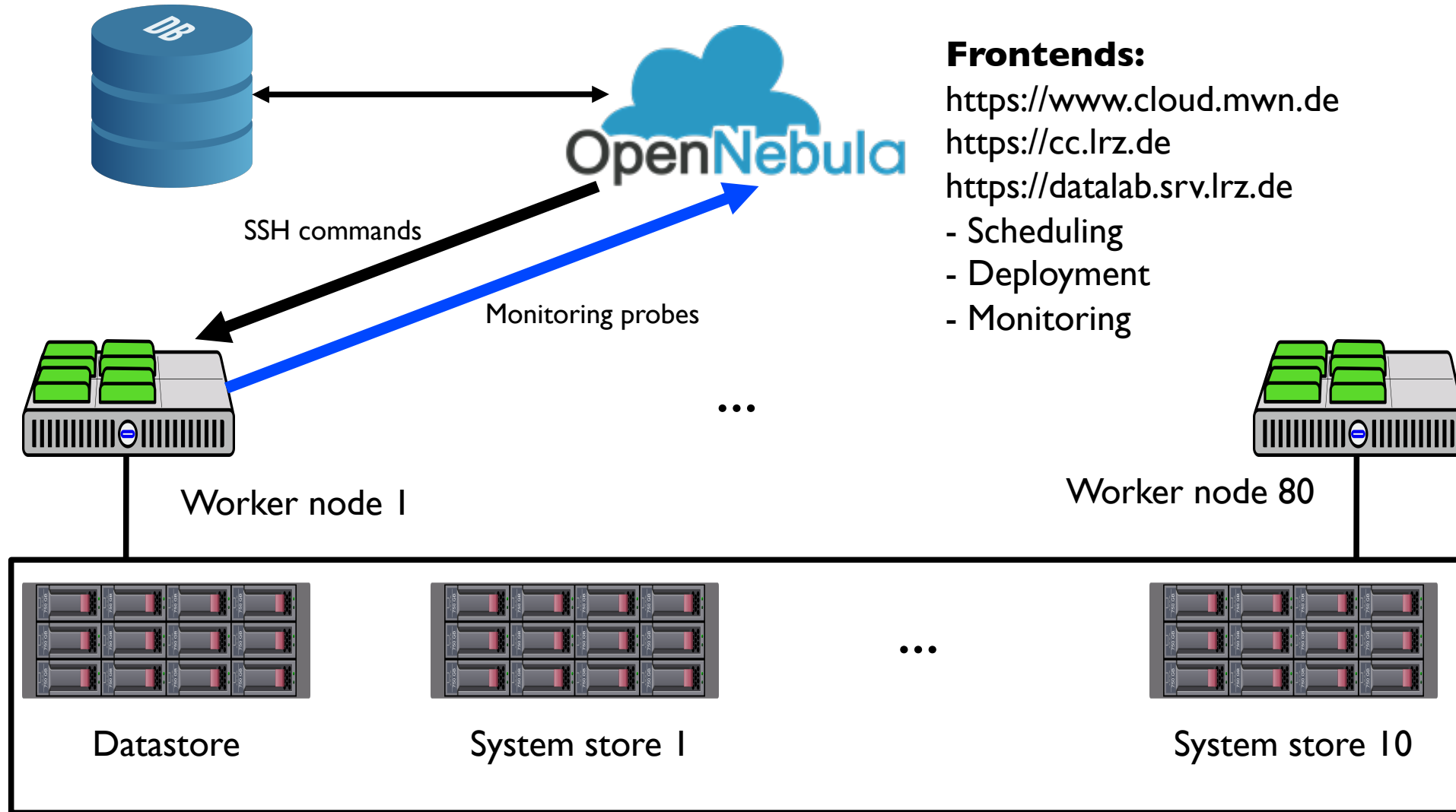


The LRZ Compute Cloud



Operational since March, 2015

LRZ Compute Cloud: OpenNebula/OpenStack



Frontends:

<https://www.cloud.mwn.de>

<https://cc.lrz.de>

<https://datalab.srv.lrz.de>

- Scheduling
- Deployment
- Monitoring

“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

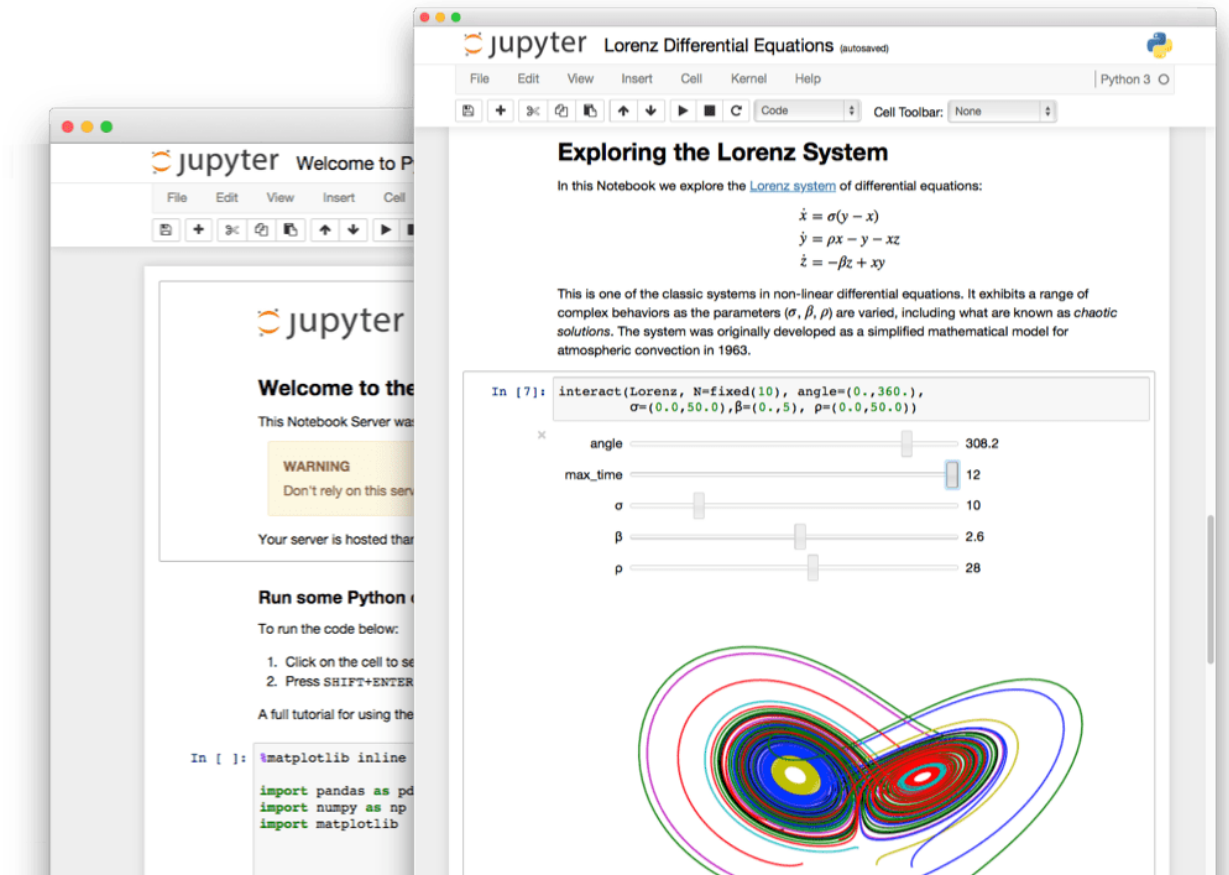
Available Software Teramem1
R/Rstudio
Redis
LRZ Software Stack

Nvidia DGX-1 and DGX-1v System

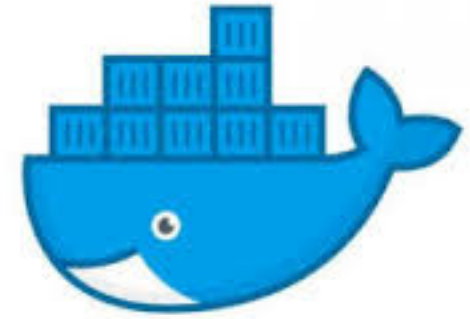
Hardware Features:
8 x Nvidia Pascal P100/V100 GPUs
170/960 Teraflops (GPU FP 16)
128 GB GPU memory
NVLink between GPUs
2 x Dual 20-core Intel Xeon E5-2698
512GB RAM main memory

Available Software DGX-1
Tensorflow
Caffe
Theano
CNTK
MXNet
Torch
DIGITS

Run your interactive program in the browser



Docker

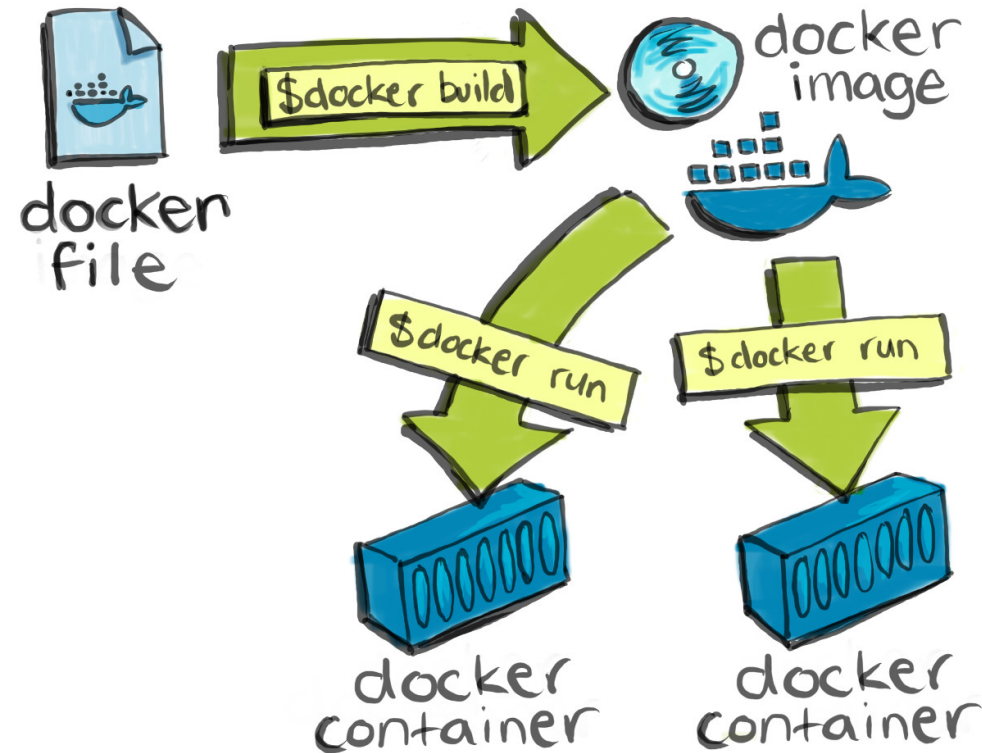


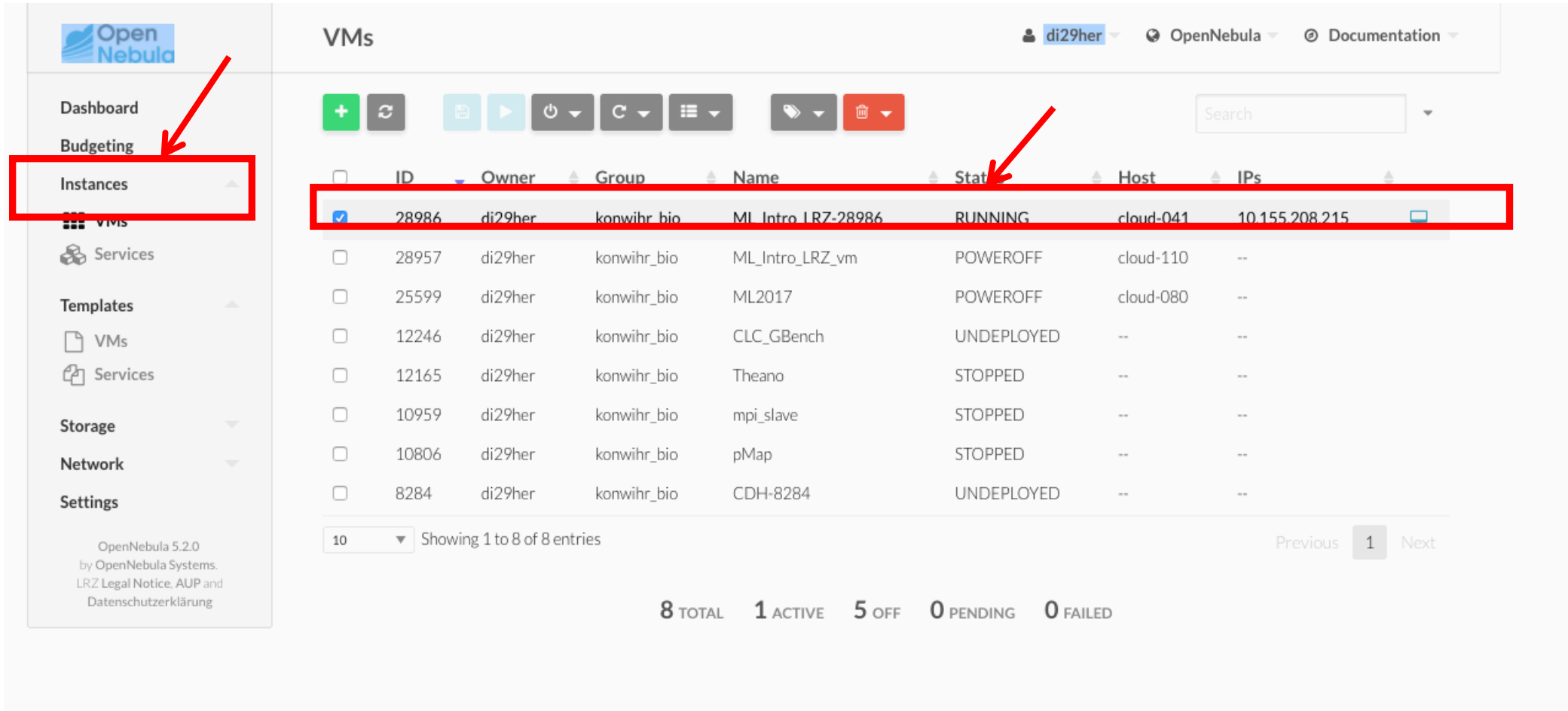
```
$ docker run hello-world  
Hello from Docker!
```

Advantages:

- reproducible
- portable
- lightweight
- Docker hub repository
- Jupyter out of the box

docker





The screenshot shows the OpenNebula web interface for managing VMs. On the left is a navigation sidebar with categories like Dashboard, Budgeting, Instances, VMs, Services, Templates, Storage, Network, and Settings. The 'Instances' menu item is highlighted with a red box and a red arrow. The main area displays a table of VMs with columns for ID, Owner, Group, Name, Status, Host, and IPs. The first row, representing a VM with ID 28986 and status 'RUNNING', is highlighted with a red box and a red arrow. Below the table, there is a pagination control showing 'Showing 1 to 8 of 8 entries' and a summary bar indicating '8 TOTAL', '1 ACTIVE', '5 OFF', '0 PENDING', and '0 FAILED'.

ID	Owner	Group	Name	Status	Host	IPs
28986	di29her	konwihr_bio	ML_Intro_LRZ-28986	RUNNING	cloud-041	10.155.208.215
28957	di29her	konwihr_bio	ML_Intro_LRZ_vm	POWEROFF	cloud-110	--
25599	di29her	konwihr_bio	ML2017	POWEROFF	cloud-080	--
12246	di29her	konwihr_bio	CLC_GBench	UNDEPLOYED	--	--
12165	di29her	konwihr_bio	Theano	STOPPED	--	--
10959	di29her	konwihr_bio	mpi_slave	STOPPED	--	--
10806	di29her	konwihr_bio	pMap	STOPPED	--	--
8284	di29her	konwihr_bio	CDH-8284	UNDEPLOYED	--	--

Terminal in the browser



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

Send CtrlAltDel



```
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 badwlrz-cm43996.ws.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:      58
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:~# _
```