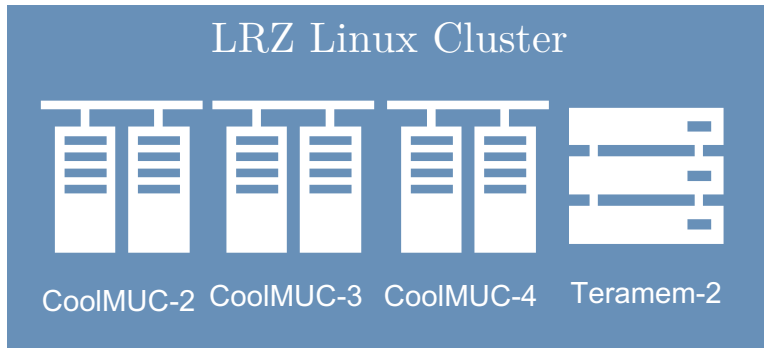
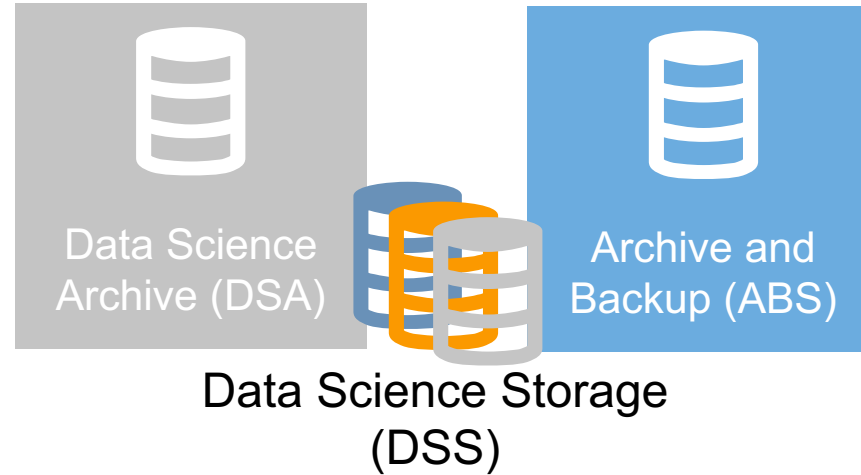


# Introduction to Multiuser Cluster Systems at LRZ

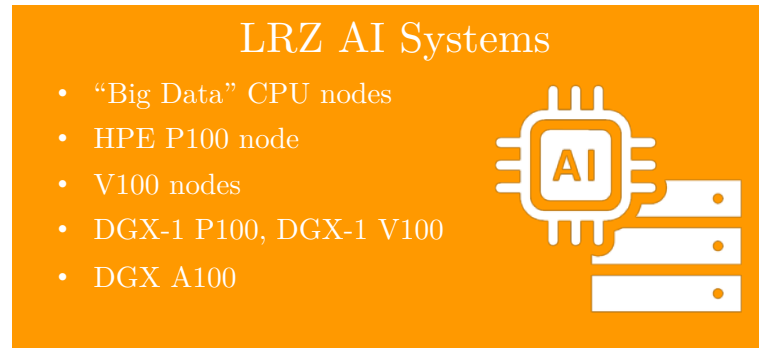
Panorama of Systems at LRZ & User perspective

April, 10<sup>th</sup> 2024

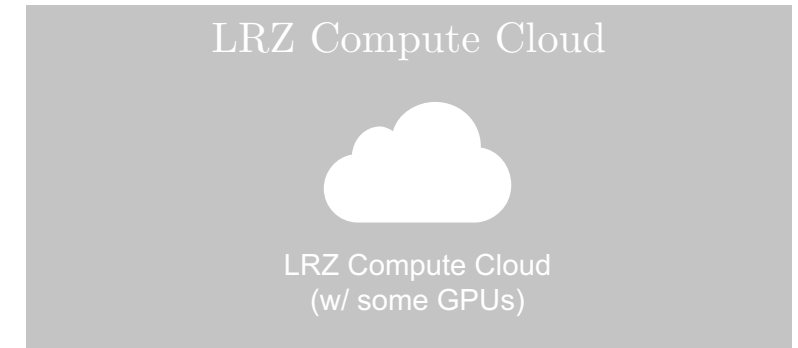
# HPC & BDAI Systems for Bavarian Universities



[lxlogin<X>.lrz.de](https://login.lrz.de)



<https://login.ai.lrz.de>  
`ssh login.ai.lrz.de`



<https://cc.lrz.de>

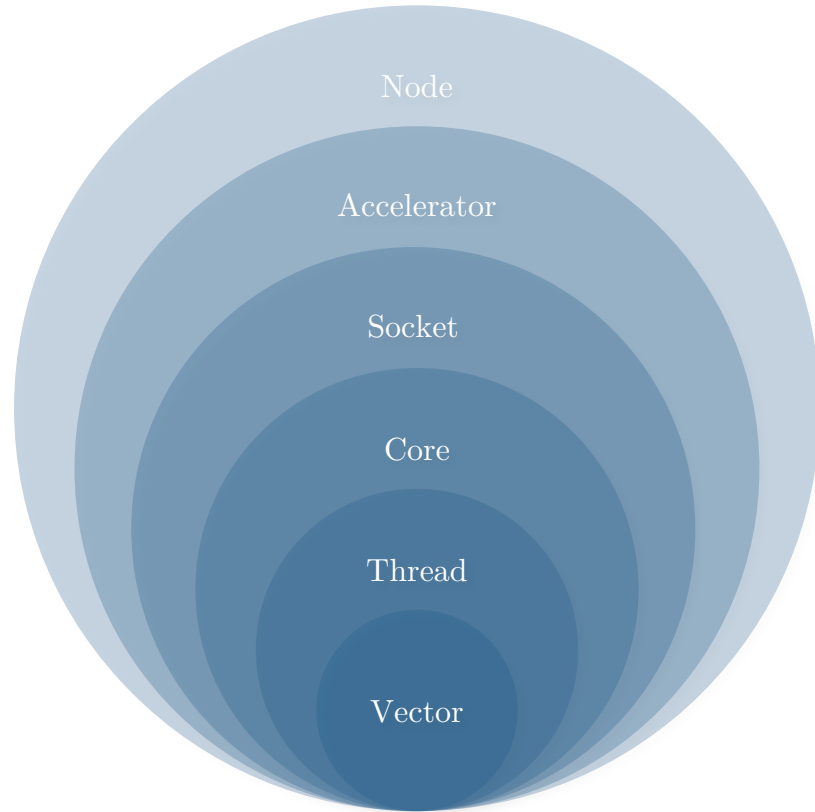
<https://doku.lrz.de/linux-cluster-10745672.html>   <https://doku.lrz.de/lrz-ai-systems-11484278.html>   <https://doku.lrz.de/display/PUBLIC/Compute+Cloud>





# SuperMUC-NG

SUPERMUC-  
NG



- **Node Level** (*e.g.*, SuperMUC-NG has 6480 nodes)
- **Accelerator Level** (*e.g.*, a Nvidia DGX A100 has 8 GPUs)
- **Socket Level** (*e.g.*, Linux Cluster Teramem has 4 sockets [with 24 cores each])
- **Core Level** (*e.g.*, Linux Cluster CoolMUC-3 nodes have 64 cores [on a single socket])
- **Thread Level** (*e.g.*, Linux Cluster CoolMUC-2 nodes allow 2 threads per core)
- **Vector Level** (*e.g.*, AVX-512 has 32 512-bit vector registers)

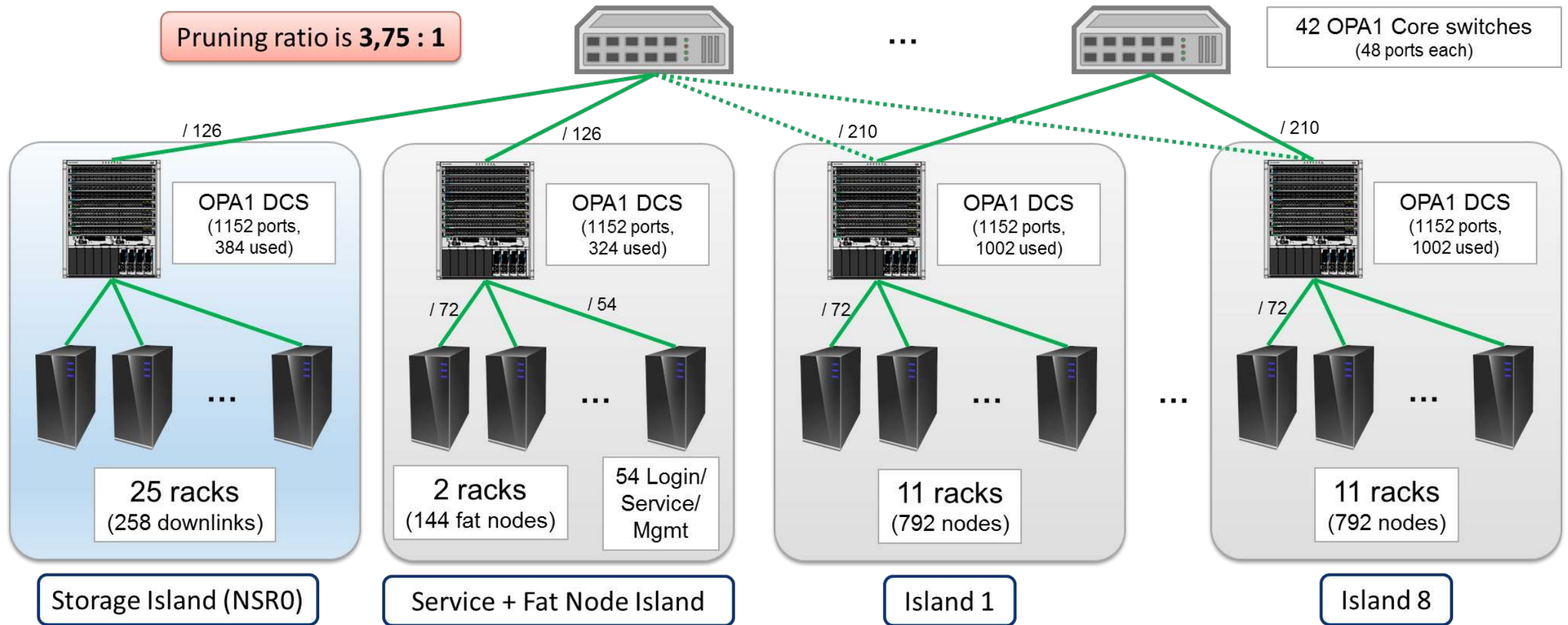
SuperMUC-NG theoretical peak performance:

$$6480 \text{ Nodes} \times 2 \text{ Sockets} \times 24 \text{ Cores} \times 32 \text{ Vectors} \times 2,7 \text{ GHz} \\ = 26\,873\,856\,000\,000\,000 \text{ Flop/s}$$





# High-Level System Architecture

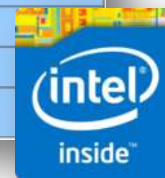


# SuperMUC-NG Hardware Overview



## Phase 1

Compute Nodes	Thin Nodes	Fat Nodes	Total (Thin + Fat)
Processor Type	Intel Skylake Xeon Platinum 8174	Intel Skylake Xeon Platinum 8174	Intel Skylake Xeon Platinum 8174
Cores per Node	48	48	48
Memory per Node [GByte]	96	768	N/A
Number of Nodes	6,336	144	6,480
Number of Cores	304,128	6,912	311,040
Peak Performance @ nominal [PFlop/s]	26.3	0.6	26.9
Linpack [PFlop/s]	-	-	19.476
Memory [TByte]	608	111	719
Number of Islands	8	1	9
Nodes per Island	792	144	N/A
<b>Filesystems</b>			
High Performance Parallel Filesystem	50 PiB @ 500 GB/s		
Data Science Storage	20 PiB @ 70 GB/s		
Home Filesystem	256 TiB		
<b>Infrastructure</b>			
Cooling	Direct warm water cooling		
Waste Heat Reuse	For producing cold water with adsorption coolers		
<b>Software</b>			
Operating System	Suse Linux Enterprise Server (SLES)		
Batch Scheduling System	SLURM		
High Performance Parallel Filesystem	IBM Spectrum Scale (GPFS)		
Programming Environment	Intel Parallel Studio XE, GNU compilers		
Message Passing	Intel MPI, (OpenMPI)		



## (Phase 2)

Nodes	
Processor	Intel Sapphire Rapids Intel Xeon Platinum 8480+
CPUs per Node	2
Cores per Node	112
Memory per Node	512 GByte DDR5
GPUs	Intel Ponte Vecchio Intel Data Center GPU Max 1550
GPUs per Node	4
Memory per GPU	128 GByte HBM2e
Number of Nodes	240 (incl. 4 login nodes)
Total CPU Cores	26,880
Total Memory	122.88 TByte DDR5
Total GPUs	960
Total GPU Memory	122.88 TByte HBM2e
PEAK (fp64; PFlop/s)	27.96 PFlop/s
Linpack (fp64; PFlop/s)	17.19 PFlop/s
<b>Compute network</b>	
Fabric	NVIDIA/Mellanox HDR Infiniband (200 Gbit/s)
Topology	fat tree
Interconnects per Node	2
Number of Islands	1
<b>Filesystems</b>	
HPPFS (same as Phase 1)	50 PB @ 500 GByte/s
DSS (same as Phase 1)	20 PB @ 70 GByte/s
Home Filesystem	256 TByte
DAOS	1 PB @ 750 GByte/s
<b>Infrastructure</b>	
Cooling	Direct warm water cooling
<b>Software</b>	
Operating System	Suse Linux (SLES)
Batch Scheduling System	SLURM
High Performance Parallel Filesystem (HPPFS)	IBM Spectrum Scale (GPFS)
Programming Environment	Intel OneAPI
Message Passing	Intel MPI, (OpenMPI)



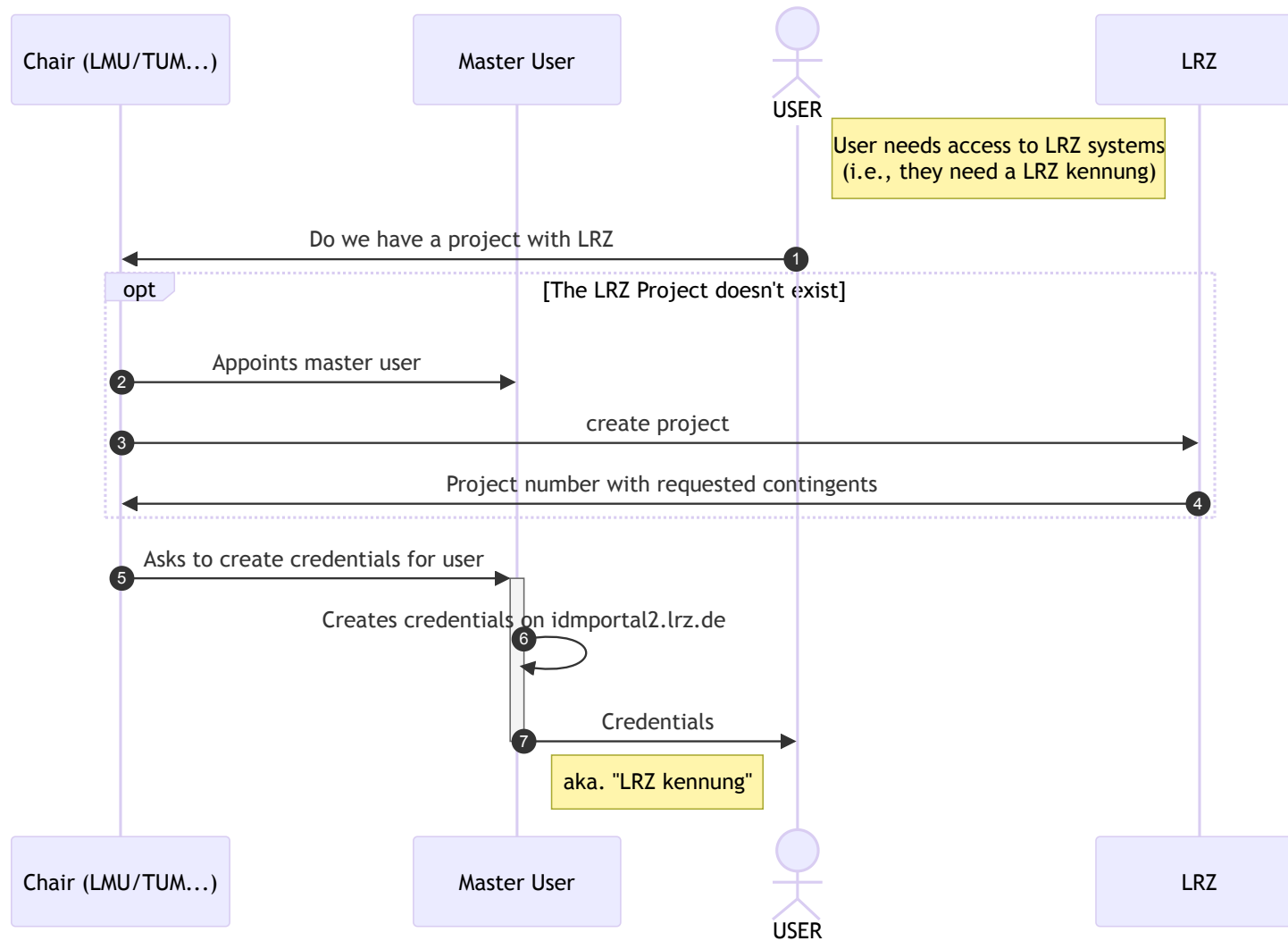
There are three (well, four) ways to apply for using SuperMUC-NG:

1. GCS test project: rolling call, fast review (short abstract), < 300.000 core-h
2. GCS regular project: rolling call, technical & scientific review, < 45m core-h
3. GCS large scale project: biannual, technical & scientific review, > 45m core-h
4. (biannual PRACE calls for academic users from any European country)

For further details, see <https://doku.lrz.de/x/XAAbAQ>

# LRZ User Management System

## The complete Perspective





# Linux Cluster

# Linux Cluster: Hardware Overview



Name	CPU	Cores/Node	RAM/Node (GB)	Nodes (total)	Cores (total)
CoolMUC-2	Intel Xeon E5-2690 v3 ("Haswell")	28	64	812	22736
CoolMUC-3	Intel Xeon Phi ("Knights Landing")	64	96	148	9472
Teramem	Intel Xeon E7-8890 v4 ("Broadwell")	96	6144	1	96
CoolMUC-4	-	-	-	-	-

<https://doku.lrz.de/linux-cluster-10745672.html>





# User Perspective: Environment & Workspace

- These are systems shared by many users, i.e. other people will be working on the same (login) node at the same time.
- Be aware of your surroundings and considerate of your fellow colleagues!

```

ssh lxlogin1.lrz.de ~
di67pif@cm2login1:~$ w
18:50:22 up 27 days, 8:55, 46 users, load average: 4.51, 5.31, 5.81
USER      TTY      FROM          LOGIN@      IDLE   JCPU   PCPU   WHAT
ra35fud pts/6    th-ws-7010m59.th Thu15 10:38m 2:00m 23:23 /dss/ds
ra35fud pts/13   th-ws-7010m59.th Thu15 10:40m 1:45m 21:15 /dss/ds
di57ril pts/14   dynamic-002-215- Sun10 51:02 5.90s 5.90s -bash
ra78wan2 pts/19   244-152-163-10.l Sat23 43:46m 0.27s 0.27s -bash
ra35fud pts/28   th-ws-7010m59.th 020ct23 4days 1:41m 0.36s /dss/ds
ga26kes2 pts/33   10.152.188.171 26Sep23 3days 60.96s 2.35s -tcsh
t388110 pts/36   f166.tum.vpn.lrz 18:22 59.00s 0.11s 0.11s -bash
ge72xes2 pts/38   10.162.92.110 Fri17 2:45m 1.38s 0.26s bash
t388110 pts/41   f166.tum.vpn.lrz 18:22 14:24 0.32s 0.32s -bash
ge49hid2 pts/42   onat.frm2.tum.de 18Sep23 20days 0.14s 0.14s -bash
ra68mop pts/44   10.163.213.247 Sun11 4:02m 1.04s 1.04s -bash
ga38qon3 pts/47   129.187.45.149 Fri12 2:48m 42:20 42:03 /dss/ds
ge52wid2 pts/58   ip139188.forst.w 26Sep23 25:24m 2.15s 2.15s -bash
ga92ziv3 pts/17   hirusako.aer.ed. Wed14 1:30m 0.31s 0.31s -bash
di68miy pts/8    pd9fe2ea2.dip0.t 28Sep23 4:49m 0.85s 0.85s -bash
ga84qec2 pts/10   10.162.204.183 22Sep23 6:23m 1.24s 1.24s -bash
ga26kes2 pts/54   10.152.188.171 020ct23 7days 0.66s 0.66s -tcsh
di98mug2 pts/66   p4fca8d79.dip0.t 18:24 19:34 0.16s 0.16s -bash
di67kah pts/68   10.153.163.46 09:07 9:20m 0.14s 0.14s -bash
di67pif pts/73   i59f7e60d.versan 18:35 3.00s 0.13s 0.02s w
ga38qon3 pts/74   129.187.45.149 Fri13 2:44m 44:27 44:26 /dss/ds
di93xej pts/75   10.156.37.219 13:48 5:01m 4.56s 4.43s /usr/bi
ga26kes2 pts/80   10.152.188.171 26Sep23 3days 8.30s 1.94s -tcsh
ga38qon3 pts/82   129.187.45.149 Fri13 2:47m 39:39 39:39 /dss/ds
ra57dut pts/84   lmb1dp1-wxrob08. 25Sep23 9:17m 0.18s 0.18s -bash
ra98fif pts/77   10.153.191.141 16:38 3:50 0.30s 0.30s -bash
di67kah pts/88   10.153.163.46 09:07 4:33m 0.79s 0.79s -bash
ka641ot pts/89   gw-acgd1.net.fh- 22Sep23 12days 0.51s 0.51s -bash
di39dux pts/90   10.153.163.218 09:09 6:01m 2.80s 2.80s -bash
di67kah pts/92   10.153.163.46 09:19 7:08m 0.58s 0.58s -bash

```

```

ssh lxlogin1.lrz.de ~
di67pif@cm2login1:~$ ls
a2832ba di39yol di82sos ga84coc2 gu92dot2 ra43cob ru47qah
atlas001 di46jof di82tun ga84coc3 gu92vih2 ra46jim ru48fak2
atlas051 di46puy di98tap ga84wug2 gu95lun2 ra52fef2 ru54vax2
atlas055 di46sap dteam007 ga86ket2 h039uaa ra52hen ru57maj
atlas066 di46taf ga26buq2 ga92wes2 h039uac ra52mer ru58guj2
atlas096 di49jat ga27rug2 ga92wof2 h039y36 ra52wos ru62guf
atlas103 di49mir ga34hed2 ga92yuh2 h039y45 ra56dat ru64nib
atlas104 di49qap ga34kat2 ga95nik2 ka85bup ra56yol ru64waf2
atlas107 di49suf ga34nox2 ga95xaf2 ka97kuk ra57biv ru67ban
atlas115 di49tom ga35hiw2 ga98dig2 lmu29425 ra57laj2 ru67yuf
atlas130 di52doh ga35piw2 ge23jiq2 lu26mur2 ra57lon ru68qum
atlas135 di52doz2 ga38coq2 ge24por2 lu28fam ra75kuw ru73jas
atlas137 di52mit ga38lix2 ge25don2 lu28tej ra75pan ru74jac
atlas139 di52qaw ga39dig2 ge29xac2 lu43fup2 ra78wuh ru74mon
atlas175 di67hal ga42jol2 ge29yig2 lu57gup9 ra96buh ru76qiq
atlas192 di67pif ga46luh2 ge34ket3 lu65cug ra98cit ru76tap
atlasprd di68tek ga48zoj2 ge35pom2 lu79hip3 ri32bet ru78zob
biokurs102 di68vad ga49cen2 ge37tiq2 lu79hun2 ri32bor ru83pey2
biokurs110 di69heg ga53vuj2 ge38qox2 lu96mah6 ri35xob ru84xox
biokurs125 di69pun ga54ger2 ge39duw2 nmmda009 ri42bof2 ru85kil2
biokurs157 di72mer ga58qes2 ge45cix2 nmmda012 ri47pih ru86wed
biokurs197 di72run ga58roj3 ge45set2 nmqc011 ri58huc ru87cir4
biokurs220 di72zuy2 ga58sur2 ge46tov2 ne53qez2 ri58mey ru94puk
biokurs257 di73gov ga58yec2 ge58yic2 ne65nib2 ri65cal ru95mof
biokurs283 di73wor3 ga58zer2 ge69sid2 ne85lif2 ri83xep t388110
di25mip2 di73wux ga59mer2 ge73woy2 ngscourse03 ri85voq t5112ae
di25seqq di73yux ga62kuy2 ge86gis2 ngscourse12 ri96kit t5431ad
di25muv di75gem ga62sed2 ge89sih2 ge89sih2 ngscourse14 ru23qir2 t7846ac
di29wad di75nef ga62tan2 ge98bej2 ngscourse15 ru27qad uh101ai
di29waj di76dan ga63yep2 ge98hun2 ngscourse26 ru23kel2 uh341ae
di34god di76dax ga67dij2 ge98sig2 ra35pim ru32yiv uh351bp
di34jag di76ral ga68jov6 genomics06 ra36jip ru36mij uj311ci

```

```

ssh lxlogin1.lrz.de ~
root 27591 0.0 0.0 0 0 ? S Sep28 0:00 [kworker/10:0]
di93qiz 27648 0.0 0.0 22688 2436 ? Ss Oct06 0:00 tmux
root 52566 0.0 0.0 0 0 ? S Oct07 0:03 [kworker/23:0]
di93sig 52918 0.0 0.0 23116 672 ? Ss Sep12 0:00 tmux
di93sig 52919 0.0 0.0 30908 8 pts/39 Ss+ Sep12 0:00 -bash
root 53091 0.0 0.0 127464 8332 ? Ss 18:35 0:00 sshd: di67pif [priv]
di67pif 53105 0.0 0.0 127464 5124 ? R 18:35 0:00 sshd: di67pif@pts/73
di67pif 53107 0.0 0.0 31844 10812 pts/73 Ss 18:35 0:00 -bash
root 53656 0.0 0.0 0 0 ? S 18:37 0:00 [kworker/29:0]
di93xej 53747 0.0 0.0 68700 3592 ? S 11:24 0:00 dbus-daemon --nofork --print-address 4 --session
root 53995 0.0 0.0 0 0 ? S Sep14 0:00 [kworker/19:2]
ga26kes2 53997 0.0 0.0 59296 0 pts/80 S Sep27 0:00 dbus-launch --autolaunch 473632f0f9e04159814ae522ae309b5
ga26kes2 53998 0.0 0.0 68712 80 ? Ss Sep27 0:00 /usr/bin/dbus-daemon --syslog-only --fork --print-pid 6
di93qiz 54001 0.0 0.0 8340 4 pts/70 T Sep27 0:00 less run_norm.slurm
root 54254 0.0 0.0 0 0 ? S 11:26 0:00 [mmkproc]
di39tel 54554 0.0 0.0 31952 8 ? Ss Sep20 0:00 SCREEN -S Tim
di39tel 54555 0.0 0.0 32100 8 pts/46 Ss+ Sep20 0:00 /bin/bash
ga26kes2 54650 0.0 0.0 298400 13048 pts/99 SL+ Oct06 0:00 emacs -nw PPP.PCF
root 54765 0.0 0.0 127472 0 ? Ss Sep26 0:00 sshd: ge52wid2 [priv]
root 54944 0.0 0.0 127472 3508 ? Ss Oct06 0:00 sshd: ga38qon3 [priv]
ga38qon3 54951 0.0 0.0 129296 3508 ? S Oct06 0:00 sshd: ga38qon3@notty
ga38qon3 54952 0.0 0.0 35220 2020 ? Ss Oct06 0:00 /usr/lib/ssh/sftp-server
ge52wid2 54955 0.0 0.0 128784 1912 ? S Sep26 0:20 sshd: ge52wid2@pts/58
ge52wid2 54956 0.0 0.0 34212 2760 pts/58 Ss+ Sep26 0:02 -bash
di93sig 55169 0.0 0.0 15912 0 ? Ss Sep12 0:00 ssh-agent
di93xej 55253 0.0 0.0 68700 0 ? S Oct04 0:00 dbus-daemon --nofork --print-address 4 --session
di93qiz 55305 0.0 0.0 8340 4 pts/70 T Sep27 0:00 less run_norm.slurm
di93qiz 55512 0.0 0.0 102064 0 pts/69 TL Sep27 0:00 salloc --clusters=serial --partition=serial_std --mem=55
root 55602 0.0 0.0 0 0 ? S Sep21 0:20 [kworker/54:2]
root 56155 0.0 0.0 0 0 ? S Oct02 0:43 [mmkproc]
root 56536 0.0 0.0 0 0 ? S Sep29 0:14 [kworker/20:1]
root 56585 0.0 0.0 0 0 ? S Sep28 3:07 [mmkproc]
di93xej 57065 0.0 0.0 68700 3536 ? S 15:05 0:00 dbus-daemon --nofork --print-address 4 --session

```

## User Perspective: Environment & Workspace



You don't have administrative rights on these systems, i.e. no root access.

You will not be able to use the `sudo` command

You're prohibited from making system-wide modifications

Disk access is restricted to your home directory (and possibly other storage areas accessible to your account, e.g., your DSS containers)

→ That said, your home (directory) is your castle – there, anything goes!



## User Perspective: Environment & Workspace

- If available on the system, modules allow for the dynamic modification of environment variables, e.g., they provide a flexible way to access various applications and libraries available on the system
- List the currently active modules (loaded by default):  
`$ module list`
- Search for available modules:  
`$ module available <module>` or  
`$ module av <module>`
- Get more information about a specific module:  
`$ module show <module>`
- Use `$ module load <module>` to apply the changes of a module to the environment

# User Perspective: Package Managers and Binaries

- **Conda** (<https://conda.io>) is “a package, dependency and environment management for any language – Python, R, Ruby, Lua, Scala, Java, JavaScript, C/ C++, FORTRAN, and more”.
- **pip** (<https://pip.pypa.io>) is “the package installer for Python. You can use it to install packages from the Python Package Index and other indexes”.

Make sure to install packages to the home directory instead of the system-wide default location:

```
~$ pip install --user <package>
```

- **wget** a binary from the internet (be careful!)

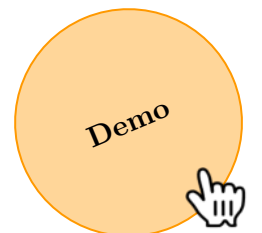
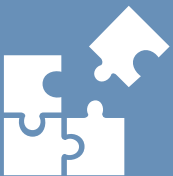
```
~$ wget http://free-software.ru/download/not-malware.bin
```

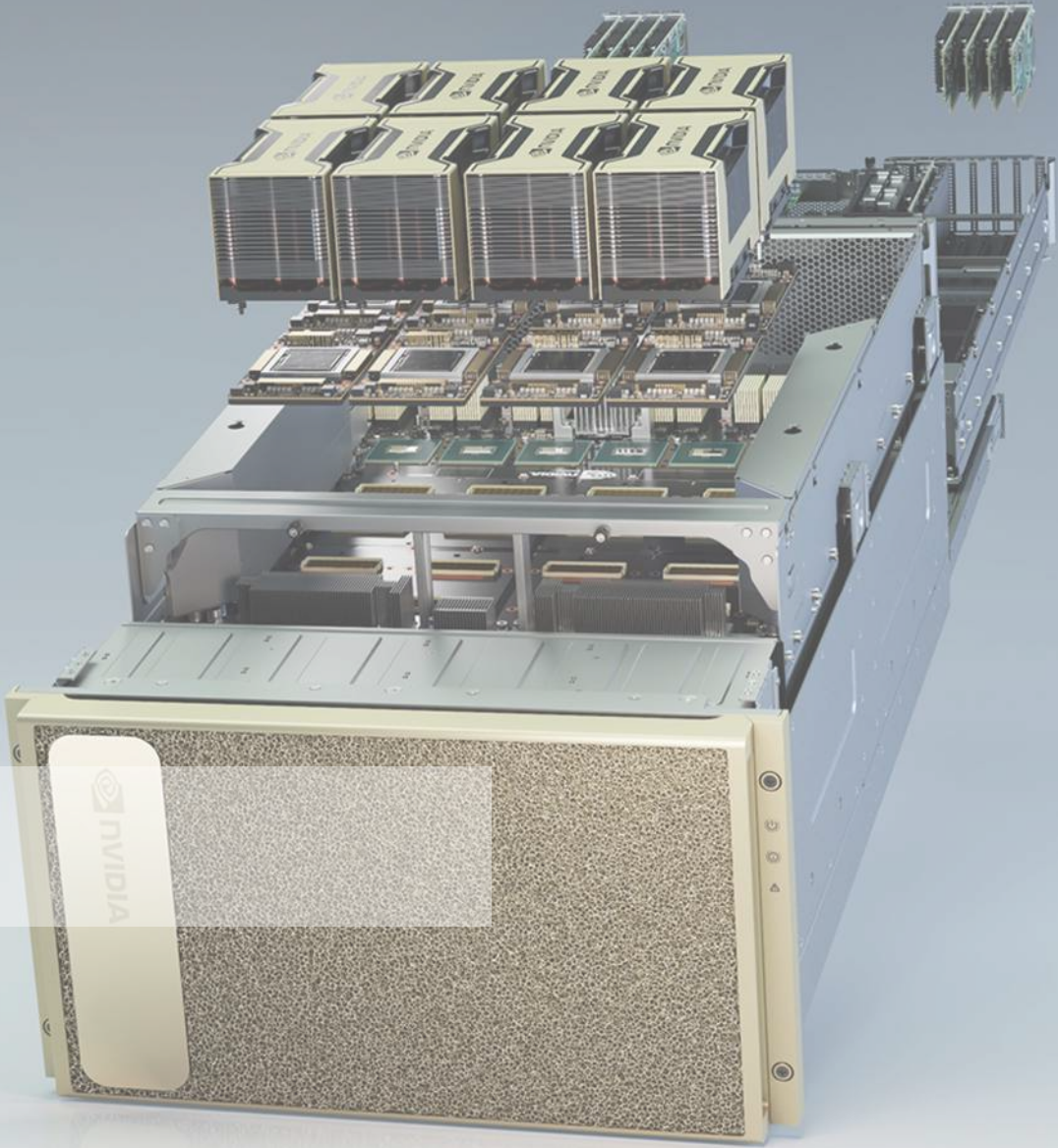
- **Compile** yourself

```
~$ git clone https://github.com/ggerganov/whisper.cpp
```

```
~$ cd whisper.cpp
```

```
~$ make
```



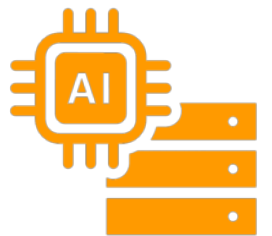


# LRZ AI Systems

# (BD)AI Systems: Hardware Overview



Type	Nodes	CPUs (Node)	Memory (Node)	GPUs (Node)	Memory (GPU)
CPU Nodes	9	up to 20	up to 850GB	-	-
HPE P100 Node	1	64	256 GB	4x P100	16 GB
V100 Nodes	4	40	368 GB	2x V100	16 GB
DGX-1 P100	1	80	512 GB	8x P100	16 GB
DGX-1 V100	1	80	512 GB	8x V100	16 GB
DGX A100/40	1	256	1 TB	8x A100	40 GB
DGX A100/80	4	256	2 TB	8x A100	80 GB

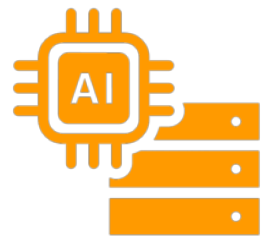




# (BD)AI Systems: Hardware Overview



- 3 NVIDIA A100 rack mounted at the Argonne National Lab
- 143kg / node
- 8 GPUs / node
- 400 W
- (not actually made of gold)



LRZ AI Systems Web UI (TEST INSTANCE) Files Jobs Clusters Interactive Apps My Interactive Sessions Help Logged in as di67pif

Very important notice: The previous home directories have been superseded by the default Linux Cluster home directories. Please see <https://doku.lrz.de/display/PUBLIC/LRZ+AI+Systems>

Home / My Interactive Sessions / Jupyter Notebook

Interactive Apps

- Servers
- Jupyter Notebook**
- RStudio Server

### Jupyter Notebook

Jupyter Notebook Access.

Choose the partition where the job will run

lrz-dgx-1-v100x8

Check available partitions <https://doku.lrz.de/x/sQCuAw>

Choose an Nvidia NGC container image or "Custom" to provide the container info in the next field

Tensorflow v2

Check <https://tinyurl.com/3uscc23c> to configure your Nvidia NGC access

Number of hours

6

Desired number of GPUs for your job

8

Comma separated list of mounts to perform from the host inside the container in the format <path-in-home><:path-in-container>

Make it Jupyter Lab!

If selected a Jupyter Lab will be started; otherwise a Jupyter Notebook will start

Launch

```
ssh datablab2 ~ (ssh) #2
Welcome to Ubuntu 20.04.5 LTS (GNU/Linux 5.4.0-122-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

System information as of Mon 10 Oct 2022 11:04:31 PM CEST

System load: 0.25          Processes:           1243
Usage of /:  90.2% of 39.99GB    Users logged in:    29
Memory usage: 69%          IPv4 address for eth0: 10.156.116.8
Swap usage:  0%

=> / is using 90.2% of 39.99GB

#####
#
#      *** VERY IMPORTANT NOTICE ***
#
# The previous home directories have been superseded by the default Linux
# Cluster home directories. Please see:
#
#      https://doku.lrz.de/display/PUBLIC/LRZ+AI+Systems
#
#####

# LRZ AI System
# For Help/Support please see:
# https://doku.lrz.de/display/PUBLIC/LRZ+AI+Systems
#
# Some notes:
# - Please stop calling the 'sudo' command, it will never work.
# - When submitting a job, you must specify the number of GPUs
#   you are planing to use, i.e. --gres=gpu:XX .
# Otherwise the job will stay in the state pending, look for
# ST = (PD) and REASON = (QOSMinGRES) when calling 'squeue'.
#
#####

Last login: Fri Oct  7 15:51:46 2022 from 129.187.49.87
di67pif@datablab2:~$ sinfo
PARTITION      AVAIL  TIMELIMIT  NODES  STATE NODELIST
lrz-v100x2*    up 14-00:00:0  2    mix gpu-[002-003]
lrz-v100x2*    up 14-00:00:0  2    alloc gpu-[001,005]
lrz-hpe-p100x4 up 14-00:00:0  1    idle p100-001
lrz-dgx-1-p100x8 up 14-00:00:0  1    alloc dgx-001
lrz-dgx-1-v100x8 up 14-00:00:0  1    alloc dgx-002
lrz-dgx-a100-80x8 up 14-00:00:0  4    mix lrz-dgx-a100-[001-002,004-005]
lrz-dgx-a100-40x8-mig up 14-00:00:0  1    idle lrz-dgx-a100-003
lrz-cpu        up 14-00:00:0  2    mix cpu-[005,007]
lrz-cpu        up 14-00:00:0  5    alloc cpu-[001-004,006]
mcm1-dgx-a100-40x8 up 14-00:00:0  5    mix mcm1-dgx-[001-003,006-007]
mcm1-dgx-a100-40x8 up 14-00:00:0  3    alloc mcm1-dgx-[004-005,008]
test-v100x2    up 14-00:00:0  1    idle gpu-004
test-amd-mi50x8 up 14-00:00:0  5    idle mankai[01-05]
di67pif@datablab2:~$
```

# User Perspective: OS-level Virtualization, Containers

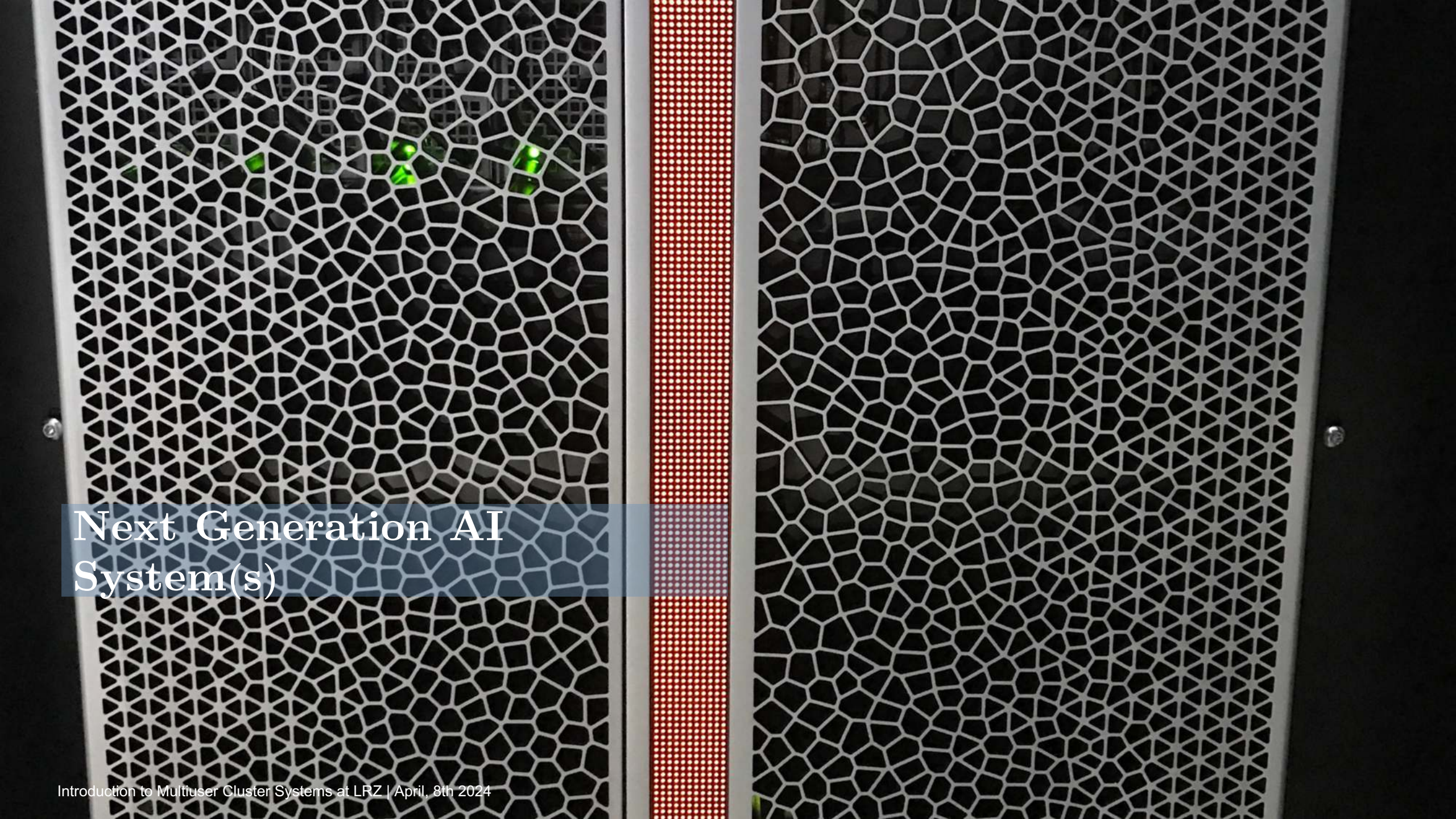
- Isolated **user space** instances, called containers, allow programs running inside to only see the container's contents and devices assigned to the container.
- Thus, the environment inside a container can essentially be modified freely, typically **providing (encapsulated) root privileges**
- The most prominent container runtime, Docker, is typically not available on multiuser systems, but you will encounter alternatives
  - Charliecloud (<https://hpc.github.io/charliecloud/>)
  - Enroot (<https://github.com/NVIDIA/enroot>)
- Containers imposes no noticeable overhead, i.e. there should be no performance impact and parallelization, GPU access, etc. should – if set up correctly – work as expected
- Containers are UDSS: User Defined Software Stacks: you're basically independent from the environment created by system administrators, but you will only receive limited support for the environment created instead (inside the container).



Demo





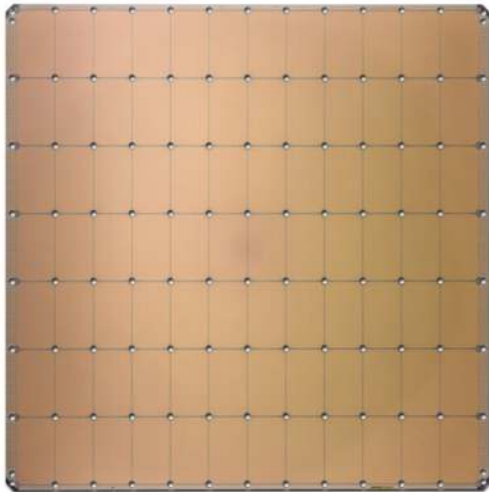


# Next Generation AI System(s)



# (BD)AI Systems: Hardware Overview

## Cerebras CS-2 Wafer Scale Engine (WSE2) in Numbers



**Cerebras WSE-2**  
2.6 Trillion Transistors  
46,225 mm<sup>2</sup> Silicon



**Largest GPU**  
54.2 Billion Transistors  
826 mm<sup>2</sup> Silicon

**A Systems Approach to Deep Learning**  
**“Cluster-scale acceleration on a single chip”**

	Cerebras WSE-2	NVIDIA A100	WSE 2 Advantage
<b>Chip Size</b>	46,225 mm <sup>2</sup>	826 mm <sup>2</sup>	56 X
<b>Cores</b>	850,000	6,912 + 432	123 X
<b>On Chip memory</b>	40 Gigabytes	40 Megabytes	1,000 X
<b>Memory B/W</b>	20 Petabytes/sec	1,555 Gigabytes/sec	12,862 X
<b>Fabric B/W</b>	220 Petabits/sec	4.8 Terabytes/sec	45,833 X

Data Source: <https://www.cerebras.net/whitepapers/>

# LRZ Compute Cloud

# LRZ Compute cloud: Hardware Overview



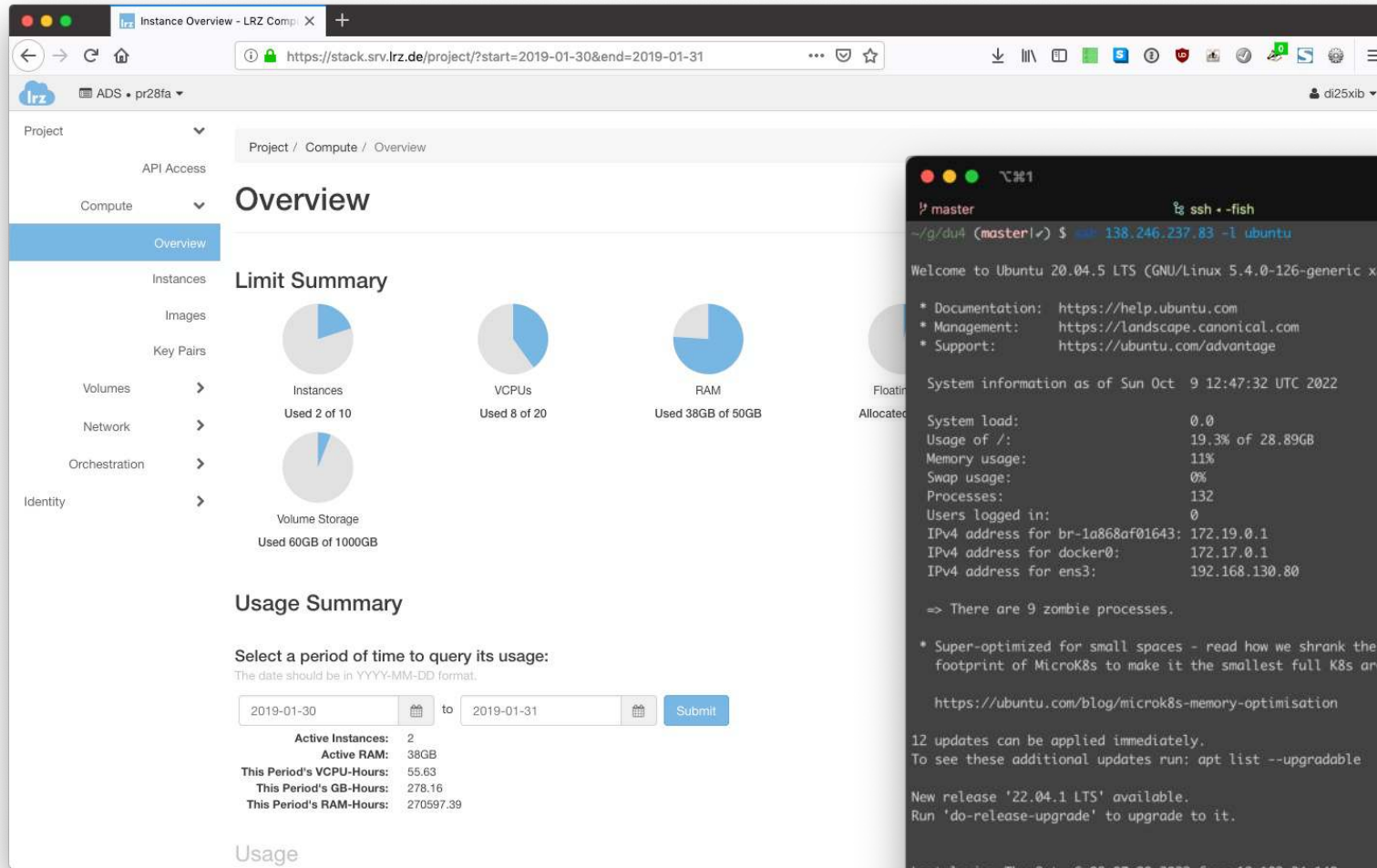
Compute	200 Nodes 192 GB to 1024 GB RAM Intel® Xeon® ~2.40 GHz
	32 x 2 GPUs Nodes 2x Nvidia Tesla V100 16 GB/node 768GB RAM/node
Storage	15 nodes 2 PB Raw Storage
Networking	100G Intel OmniPath
Software	OpenStack & CEPH

Access to more than 10 vCPUs and/or other restricted resources can be requested by contacting the cloud support team: <https://servicedesk.lrz.de/ql/create/105>

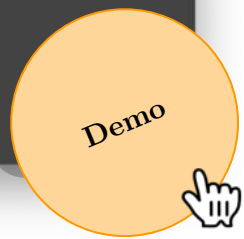
40000 vCPU capacity with overcommitment  
2000 users and 1500 active VMs



# Compute Cloud: Hardware Overview



```
root@course-node: /home/ubuntu
master ssh -fish 6% 9.6 GB 09.10., 2:48 PM
~/g/du4 (master|) $ ssh 138.246.237.83 -l ubuntu
Welcome to Ubuntu 20.04.5 LTS (GNU/Linux 5.4.0-126-generic x86_64)
* Documentation:  https://help.ubuntu.com
* Management:    https://landscape.canonical.com
* Support:       https://ubuntu.com/advantage
System information as of Sun Oct  9 12:47:32 UTC 2022
System load:            0.0
Usage of /:             19.3% of 28.89GB
Memory usage:          11%
Swap usage:            0%
Processes:             132
Users logged in:       0
IPv4 address for br-1a868af01643: 172.19.0.1
IPv4 address for docker0: 172.17.0.1
IPv4 address for ens3: 192.168.130.80
=> There are 9 zombie processes.
* Super-optimized for small spaces - read how we shrank the memory footprint of MicroK8s to make it the smallest full K8s around.
https://ubuntu.com/blog/microk8s-memory-optimisation
12 updates can be applied immediately.
To see these additional updates run: apt list --upgradable
New release '22.04.1 LTS' available.
Run 'do-release-upgrade' to upgrade to it.
Last login: Thu Oct  6 08:07:28 2022 from 10.183.34.140
ubuntu@course-node:~$ sudo su
root@course-node:/home/ubuntu# whoami; uname -a
root
Linux course-node 5.4.0-126-generic #142-Ubuntu SMP Fri Aug 26 12:12:57 UTC 2022 x86_64 x86_64 x86_64 GNU/Linux
root@course-node:/home/ubuntu#
```



Get access: <https://doku.lrz.de/display/PUBLIC/FAQ#FAQ-HowtogetaccesstotheComputeCloud?>



# LRZ Data Storage

## Data Storage: Overview

- The LRZ HPC/HPDA/HPAI Infrastructure is backed by the Data Science Storage (DSS)
  - Long-term storage solution for potentially vast amounts of data
  - Directly connected to the LRZ computing ecosystem
  - Flexible data sharing among LRZ users
  - Web interface for world-wide access and transfer
  - Data sharing with external users (invite per e-mail, access per web interface)
- Additionally, we also provide a new type of Data Archive, based on the DSS Solution stack, called Data Science Archive (DSA) (this basically relates to DSS like AWS Glacier relates to AWS S3).
- Disk space and access is managed (as DSS projects and containers) by data curators. This can be LRZ personnel (e.g., Linux Cluster \$HOME directories) or PIs/master users/dedicated data curators (e.g., project storage).



# Data Storage: Linux Cluster & AI Systems



- `$HOME` (DSS-backed home directory, managed by LRZ)
  - 100GB per user
  - Access: `/dss/dsshome1/lxc###/<user>`
  - Automatic tape backup and file system snapshots (see “`/dss/dsshome1/.snapshots/`” directory)
  - All your important files/anything you invested a lot of work into should be here
  - BUT Not suitable for heavy and/or high-frequency I/O operations, i.e. most machine learning applications. Use the AI Systems DSS instead.



## Data Storage: AI Systems

- AI DSS

- Up to 5 TB per project **upon request**, shared among project members
- Access: `$ dssusrinfo all`
- Configuration (e.g., exports, quota) to be managed by data curator
- Use this for e.g., high bandwidth, low latency I/O
- Can not (yet) be accessed from Linux Cluster





## Data Storage: Linux Cluster



- DSS [project storage](#)
  - Up to 10 TB per project **upon request**, shared among project members
  - Access: `$ dssusrinfo all`
  - Configuration (e.g., exports, backup, quota) to be managed by data curator
  - Use this for e.g., large raw data (and consider backup options)
  - Can be accessed from the AI systems





## Data Storage: Linux Cluster

- Legacy `$SCRATCH` (scratch file system, “temporary file system”)
  - 1.4 PB, shared among all users
  - Access: `/gpfs/scratch/<group>/<user>`
- New `$SCRATCH_DSS` (not yet available on CoolMUC-2 compute nodes)
  - 3.1 PB, shared among all users
  - Access: `/dss/lxclscratch/##/<user>`
- No backup (!) and sliding window file deletion, i.e. old files will eventually be deleted (!!)
  - a data retention time of approx. 30 days may be assumed, but is not guaranteed
- This is the place for e.g., very large, temporary files or intermediate results, directly feeding into additional analyses
- Data integrity is not guaranteed. Do not save any important data exclusively on these file systems! Seriously, don’t do it!



## Data Storage: Compute Cloud

- The storage backend of the Compute Cloud is used to host the virtual disks belonging to the VMs in the cloud. It is not meant to store large data sets. No backups are created.
- DSS containers can be made available for VMs running in the LRZ Compute Cloud without the need to copy data into the VM.
  - The data curator of the data project, to which the relevant container belongs, needs to export the container to the IP address used by your VM via NFS.
  - You should only export DSS containers to IPs that are statically assigned to and trusted by you. NFS exports follow a "host based trust" semantic, which means the DSS NFS server will trust any IP/system to which a DSS container is exported. There is no additional user authentication between NFS server and client enforced.

