REGIONALES HOCHSCHUL-
RECHENZENTRUM KAISERSLAUTERN

TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

RHRK-Seminar

High Performance Computing Cluster „Elwetritsch“ - Part II

Course instructor : Dr. Josef Schüle, RHRK

Overview

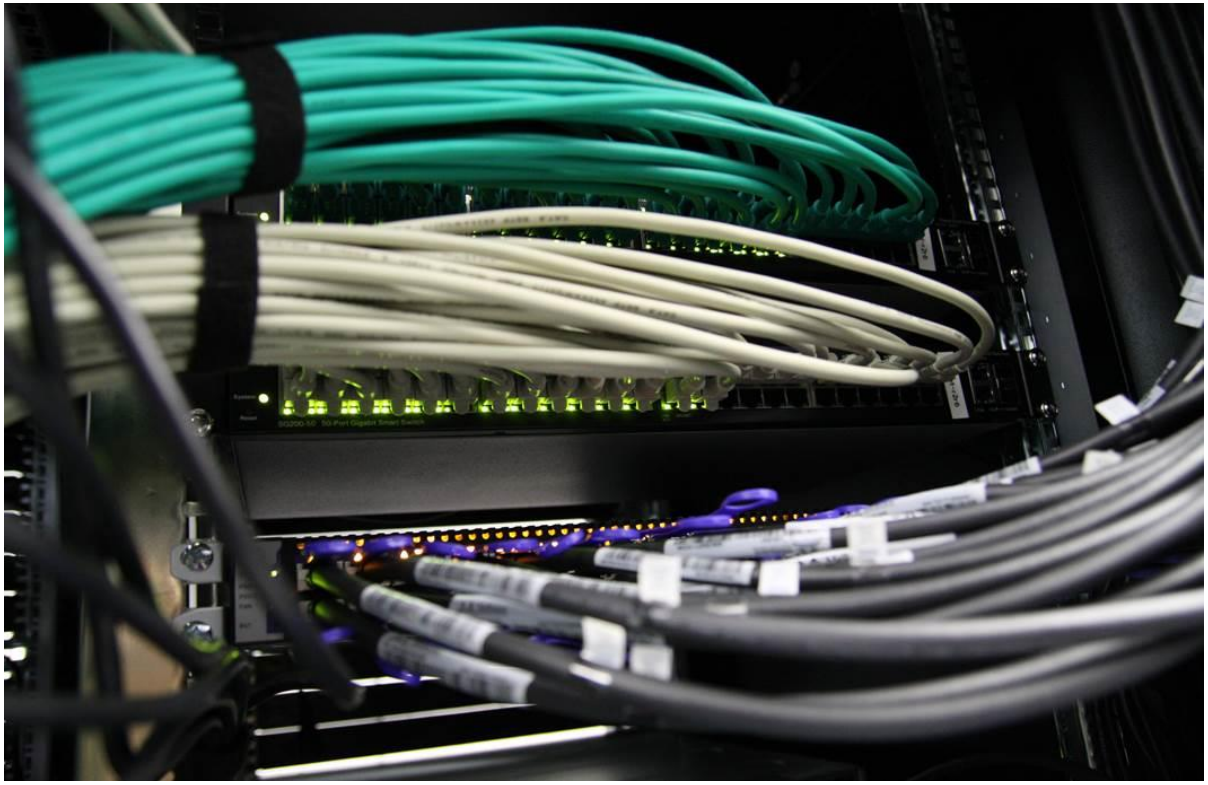
Course I

- **Login to cluster**
  - SSH
  - RDP / NX
- **Desktop Environments**
  - GNOME (default)
  - other desktops
- **Linux Basics**
  - Terminal / Shell
  - file systems
  - Tipps & Tricks

Course II

- **Cluster Basics**
  - Hardware
  - Software
  - Brainware
- **Batch system**
  - submitting jobs
  - Monitoring / Accounting
- **AHRP**
  - Projects
  - Working on MOGON /Mainz

Cluster Basics



Hardware

- **Nodes**
- **Network**
- **Filesystems**

Software

- **System Software**
- **Module**
- **Batch system**

Brainware

Hardware

Nodes

- login nodes
  - elwex, headx, x=1-4, rarp1,2
  - dcvx (nice-dv)
- compute nodes
  - interactive available
  - not interactive available
- special
  - different GPU versions/number
  - 1 TB RAM
  - AMD/Intel

Infiniband network QDR, Omnipath

Ethernet network 1/10Gb/s

File systems

- /scratch Fraunhofer BeeGFS (1.3 PB)
- /home NetApp-NFS for HOME (10 TB)
- /work many small files on request
- /tmp short-time temporary storage

Hardware Overview

Complex	Server Properties					Summary		
	Host Type	CPUs	Cores	Memory	Interconnect	# Servers	∑ Cores	∑ Memory
ims	Intel Xeon E5520	2	8	24 GB	1 GE	6	48	144 GB
smp	AMD Opteron 6140	4	32	256 GB	10 GE	7	224	1792 GB
mpi32	Intel E5-2670	2	16	32 GB	IB QDR	158	2528	5056 GB
mpi64	Intel E5-2670	2	16	64 GB	IB QDR	116	1856	7424 GB
mpi128	Intel E5-2670	2	16	128 GB	IB QDR	18	288	2304 GB
mpi64v3	Intel E5-2640v3	2	16	64 GB	IB QDR	96	1536	6144 GB
mpi256v3	Intel E5-2640v3	2	16	256 GB	IB QDR	24	384	6144 GB
single64v3	Intel E5-2637v3	2	8	64 GB	IB QDR	24	192	1536 GB
accelerator	Intel E5-2670 + 2x NVIDIA Tesla M2090	2	16	32 GB	IB QDR	3	48	96 GB
	Intel E5-2670 + 2x NVIDIA Tesla M2090	2	16	64 GB	IB QDR	3	48	192 GB
	Intel E5-2670 + 2x NVIDIA Tesla K20x	2	16	64 GB	IB QDR	2	32	128 GB
	Intel E5-2670 + 4x NVIDIA Tesla K20x	2	16	64 GB	IB QDR	1	16	64 GB
	Intel E5-2670 + 2x Intel XEON Phi 5110P	2	16	32 GB	IB QDR	1	16	32 GB
	Intel E5-2670 + 2x Intel XEON Phi 5110P	2	16	64 GB	IB QDR	2	32	128 GB
bigmem	Intel XEON E5-4650	4	32	1024 GB	IB QDR	1	32	1024 GB
vgl	Intel E3-1270 + NVIDIA Quadro 4000	1	4	16 GB	IB QDR	9	36	144 GB

Software

- System software
  - Scientific Linux
  - periodical updates
- Module
  - software installed from source ->/software
  - commercial software
  - available in several versions -> software modules
  - command: `module`
  - is something missing?:
    E-Mail to hotline@rhrk.uni-kl.de
- Batch system
  - SLURM
  - jSLURM – graphical user interface

Software

- **Compiler & Tools**
  - Intel Cluster Suite
  - Portland PGI
  - GNU Compiler Collection
  - CUDA Toolkit & GPU Programming SDK
  - Totalview Parallel Debugger
  - Various developer environments
- **Libraries**
  - Intel Math Kernel Library (MKL)
  - AMD Core Math Library (ACML)
- **Python**
  - Anaconda
- **MPI**
  - Intel MPI
  - Open MPI
  - IBM Platform MPI

Brainware

- Competence in using the cluster at it‘s best is important for HPC
- Don‘t hesitate to contact us and ask questions
  hotline@rhrk.uni-kl.de
- Persons behind
  - Dr. Josef Schüle
  - Dr. Markus Hillenbrand
  - Sven Daxinger
- .. if you need help
  - installing programs
  - select best programs
  - compile your programs
  - using the batch system (select parameters and resources)
  - developing your own software (OpenMP, MPI, CUDA)
  - writing scripts to support your work

Batch System

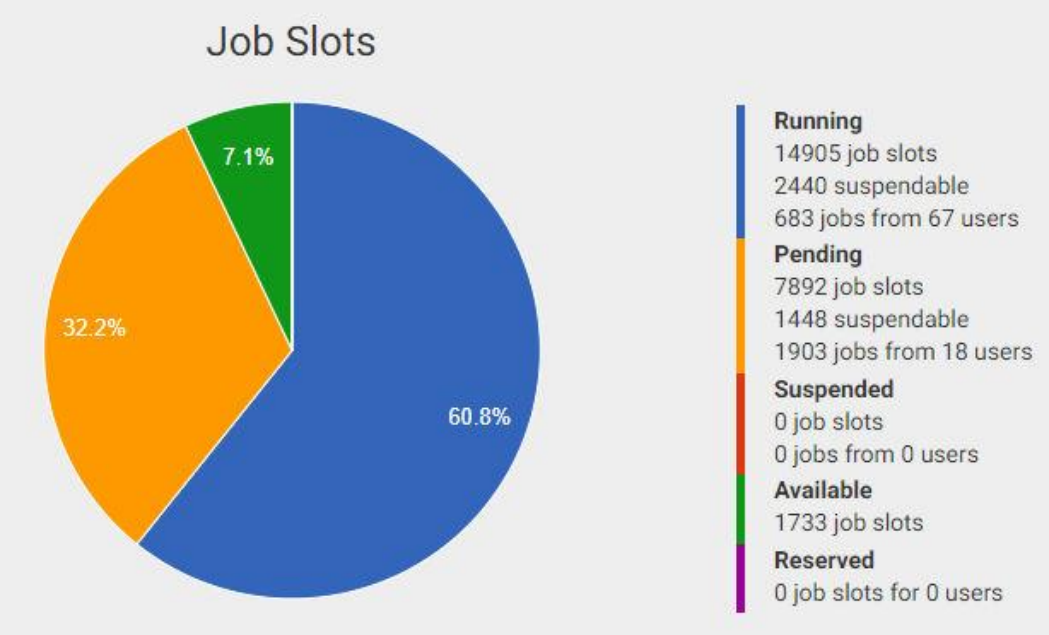
Job Management

- **submission**
- **query**
- **ending**

Monitoring

Accounting

jSLURM/jHPC



Extra seminar

Wording

Cluster: Number of servers to be seen as unity

- Bundling compute power
- resources and load adjusting

Job: work unit executed by the batch system

- typical a command or command sequence with options and arguments
- batch system schedules for execution, controls and protocols

Wording

Jobslot: Smallest execution unit.

- on Elwetritsch a jobslot corresponds to a single CPU core

Queue: sorting jobs according to resource requirements

- after submission, jobs are sorted into waiting queues
- remain waiting until all resource requirements are available

Scheduling: algorithm to select and start next jobs

- Fair share Scheduling
- each user has a fair amount
  - according to project priority
  - of cluster resources

target hardware

basic clock - 2.6 GHz

- Intel E5-2670 (codename Sandybridge)
  - > 200 nodes
  - 16 cores per node
  - 32, 64, 128, 1024 GB RAM
  - out of service -> IV/2021, 2022
- Intel E5-2640v3 (codename Haswell)
  - > 110 nodes
  - 16 cores per node
  - 64, 256 GB RAM
  - out of service IV/2022, 2023
- Intel E5-2637v3 (codename Haswell)
  - 24 nodes
  - 8 cores, 64 GB, 3.5 GHz

target hardware

basic clock 2.6

- Intel SP-6126 (codename Skylake)
  - > 300 nodes
  - 24 cores per node
  - 96, 384 GB RAM

basic clock 2.4

- AMD Epyc 7351
  - 2 nodes
  - 64 cores
  - 128 GB RAM

GPU - zoo

- 2xK20, 2xK20x, 4xK20x, 1xK80 - 9 nodes
- 1xV100, 2xV100, 4xV100 -10 nodes
- DGX-1, DGX-2

interactive nodes

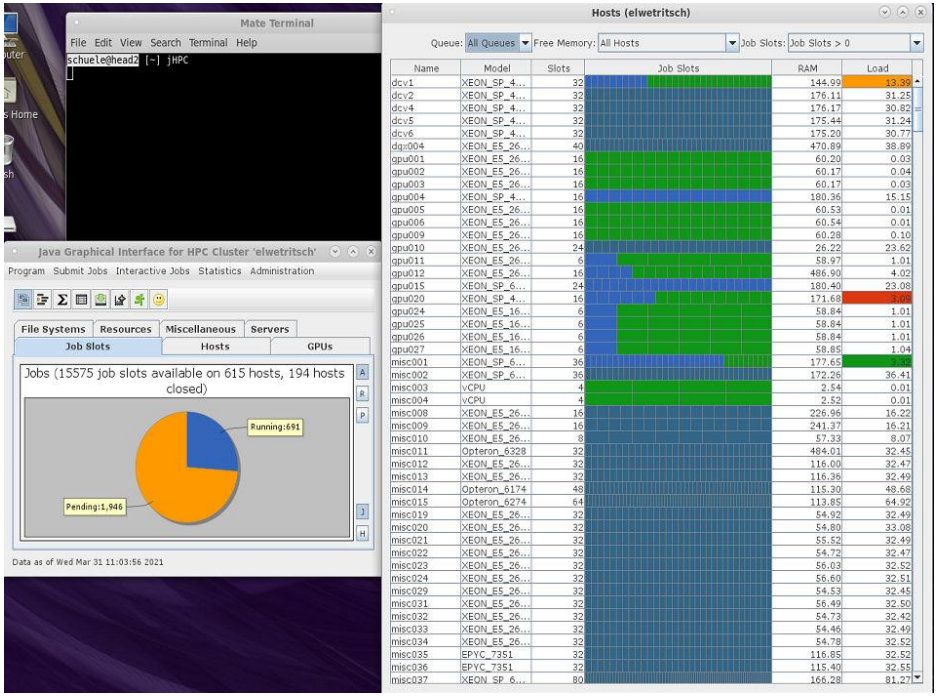
- graphics nodes
  - vgl[002-020]
  - Sandybridge, Q4000, 8 nodes
  - Broadwell (E5v4) P2000, 10 nodes
  - dcv[1-4], Skylake, RTX 8000, 4 nodes
- interactive (debugging, code developing)
  - Sandybridge, 7 nodes

REGIONALES HOCHSCHUL-
RECHENZENTRUM KAISERSLAUTERN

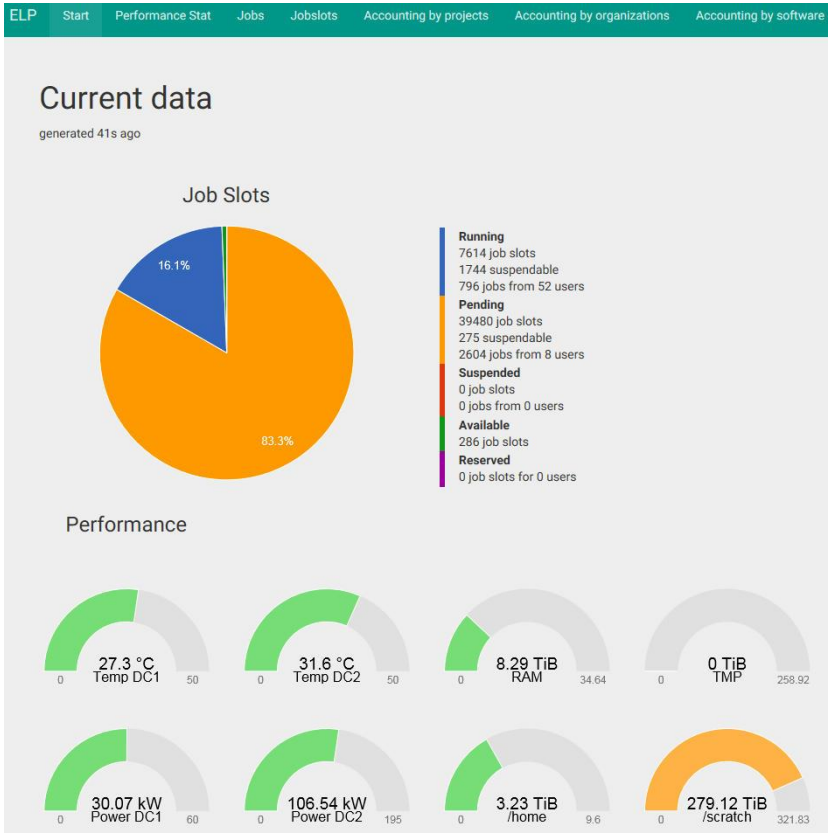
Monitoring

TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

jSLURM/jHPC



ELP



File Systems

/home

visible on all nodes
quota (command: `quota`)
archived (-> tape robot)
private data - not visible to others

intention:

like your wallet - absolute necessary information

File Systems

/tmp

- local to each node
- privacy according to umask
- faster than HOME
- no lifetime (but please clean up after you)

intention:

- small frequently accessed temporary files
  e.g. when you are compiling

File Systems

/scratch

visible on all nodes
no quota
limited lifetime - but clean up behind yourself
private data - not visible to others

/scratch is realized as so called parallel file system, there

- data and directory/file information is separated
- data is cut into pieces that are handed to different servers for reading/writing

File System - /scratch

1. reading/writing large files achieves a high bandwidth (up to 12 GB/s)
   - please inform us if you write single files > 1GB
   - bandwidth optimization requires some interaction
2. reading/writing small files (< 512K) uses only 1 server and duplicates the packets to be written.
3. reading/writing many small files (> 1000) requires many directory entries
   - slowing down the corresponding directory servers
   - and in addition point 2. above

Especially applications for AI require a different file system (2. and 3.) and are NOT suited for /scratch - please inform us.

File Systems

/work

NFS-based file system.

- limited capacity
- quota
- fast cached access to many small files
- poor for large files (cache is flooded)

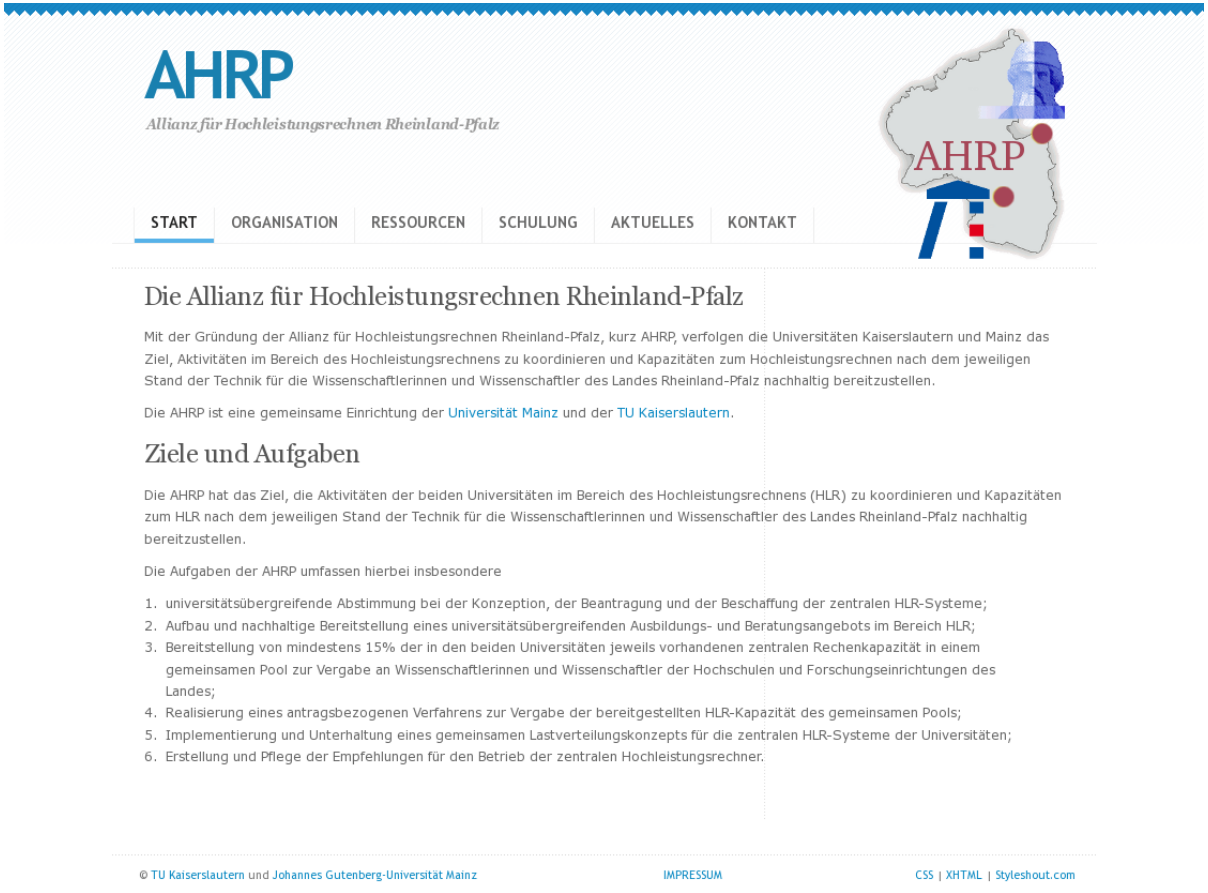
Access to /work requires interaction with HPC Team

Mail : hotline@rhrk.uni-kl.de

REGIONALES HOCHSCHUL-
RECHENZENTRUM KAISERSLAUTERN

TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

AHRP



General

- **history**
- **aims**
- **tasks**

Projects

Interaction with MOGON

Overview

founded April 2010 by

- TU Kaiserslautern
- Johannes Gutenberg Universität Mainz

aims

- coordinating HPC activities
- provisioning of state-of-the-art HPC resources for research in RLP

tasks

- install, maintain and purchase resources in common
- coordinated teaching and support for HPC
- provisioning of 15% of the resources for research to all in RLP, load balancing among the sites
- application based request for resources

REGIONALES HOCHSCHUL-
RECHENZENTRUM KAISERSLAUTERN

Projects

TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

- three steps
  - fill-in form
  - wait for review
  - follow the sent instruction
- project sizes
  - XS: < 5 NE / month
  - S:< 30 NE / month
  - M: < 100 NE / month
  - L:< 500 NE / month
  - XL: > 500 NE / month
- NE = CPU cores * h / 1000
- Link: http://www.ahrp.info

Antrag auf Nutzung eines rheinland-pfälzischen Hochleistungsrechner

Mit Ausfüllen und Absenden des nachfolgenden Formulars stellen Sie einen auf ein Forschungsprojekt bezogenen Hochleistungsrechners. Bei Fragen können Sie sich zu den üblichen Geschäftszeiten an die Geschäftsführung der

1. Angaben zum Antragsteller

Name *

Vorname Nachname

Hochschule / Institut / Einrichtung *

Abteilung / Arbeitsgruppe / Fachgebiet *

Telefon *

E-Mail *

Webseite

http://

2. Angaben zum Projekt

Titel *

- High Performance Computing on Elwetritsch
- Part II

Vielen Dank
Thank You