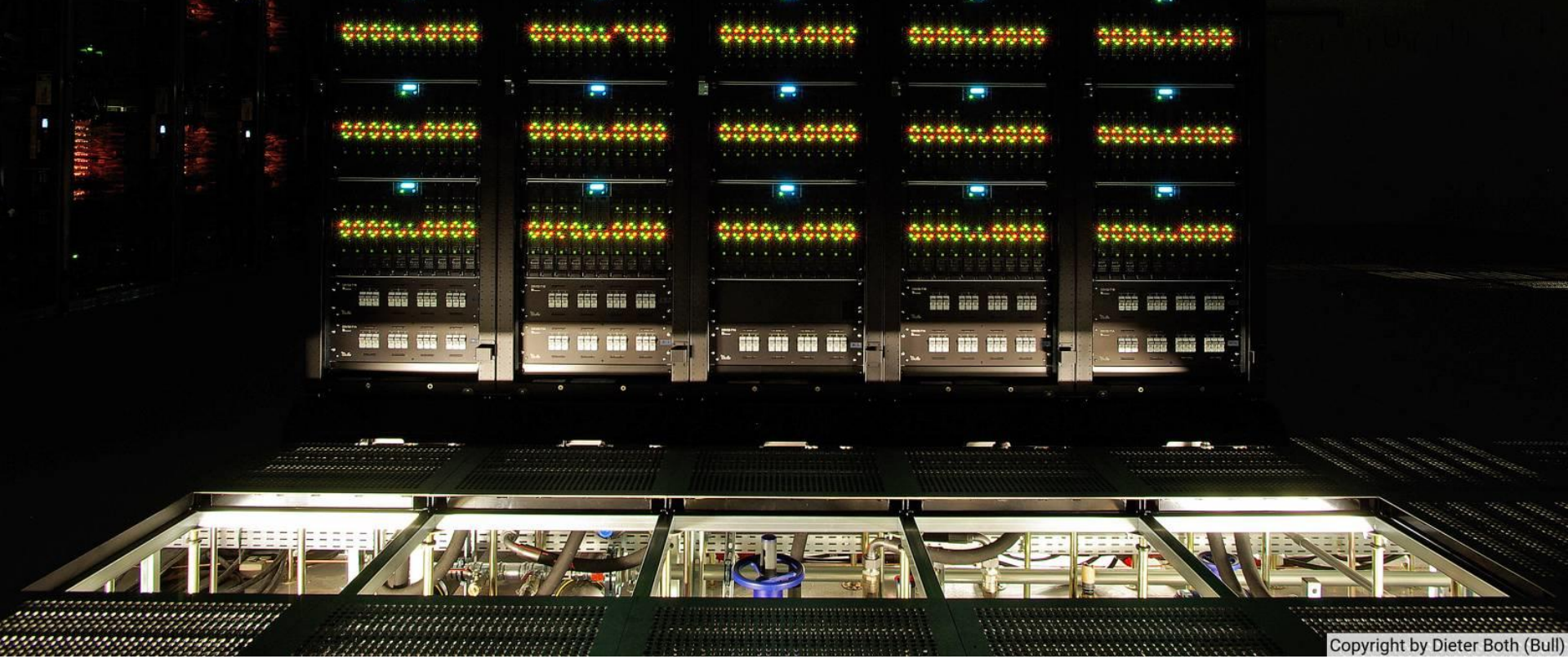


HPC Resources and Computing Projects

Dieter an Mey
IT Center, RWTH Aachen University

HPC Intro, March 12, 2021

Overview

- **Access to HPC Resources**
- **HPC Resources: Computers, Storage, Software**
- **Computing Projects**
- **Links + Upcoming Events**

Access to HPC Resources

Who can use the RWTH Compute Cluster?

Access to the RWTH Compute Cluster

- **Members of RWTH Aachen University**
  can use the RWTH Compute Cluster
  for education and research without any fees
  up to certain default limits (compute and storage resources)

  use the RWTH selfservice to get an account
  (Choose: Accounts und Kennwörter - Account anlegen - Hochleistungsrechnen)

- **Employees of RWTH and of Research Center Jülich (FZJ)**
  for education and publicly funded research projects without any fees.
  an application for a computing project has to be submitted for more resources.
  FZJ: one-time registration in the RWTH Identity Management
- **Students of RWTH Aachen University**
  lecturers can apply for resources for participating students.
  Theses advisers can apply for resources for the candidates.
- **Scientists of other German institutions in education and research** can submit proposals for large computing projects
- **Anyone** can take part in a computing project as a project member under the responsibility of a PI (principal investigator of an existing computing project)

Please find more information here

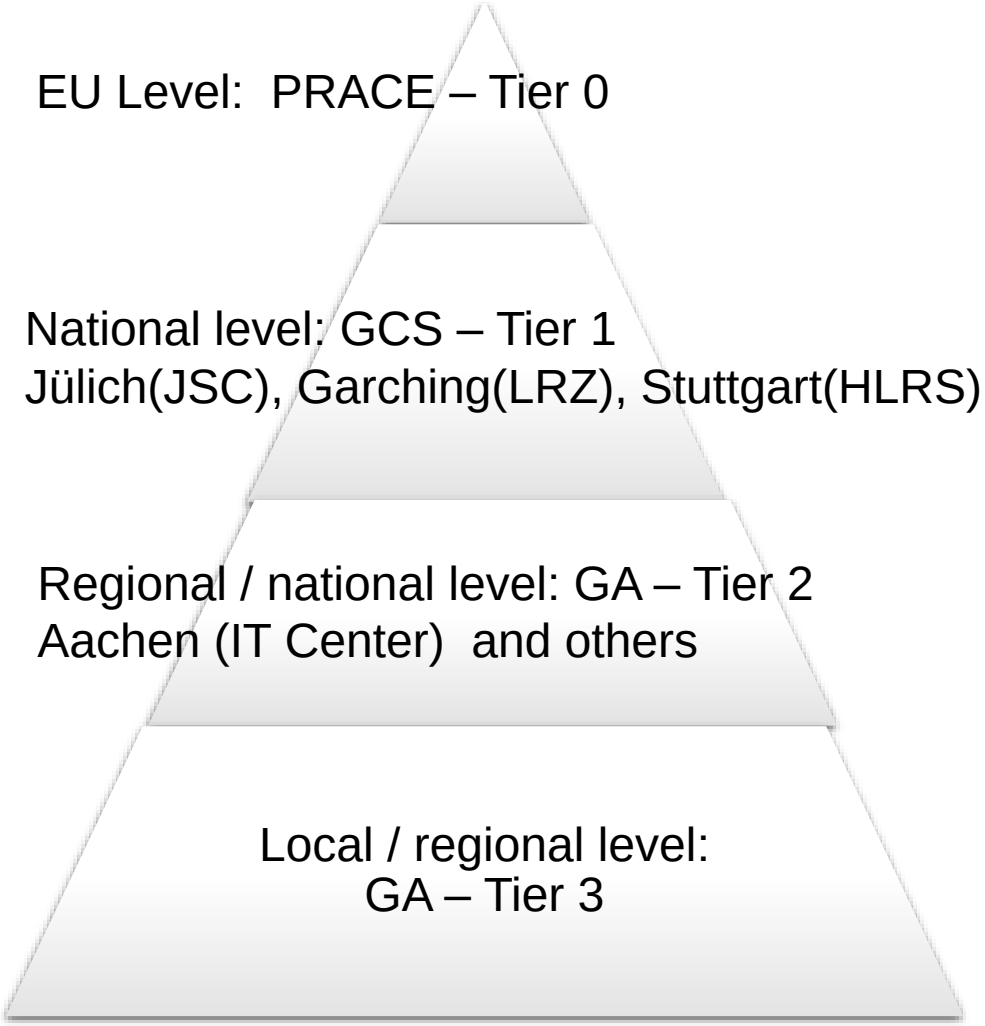
Access to other HPC Resources

Apart from using local compute resources scientists at RWTH can apply for resources on

- **machines of Research Center Jülich (mainly an Intel Xeon Phi based Booster system) in the context of the JARA-Partition.**
- **national HPC facilities organized in the Gauss Centre for Supercomputing**
- **European facilities in the context of the Partnership for Advanced Computing in Europe (PRACE).**
- **Such an application is carefully reviewed taking technical and scientific aspects into account.**

Citizens of countries that are subject to the export control policy of the German Federal Government may need additional authorization from the German Federal Office for Economic Affairs and Export Control (BAFA) before they are allowed to use the RWTH Compute Cluster.

HPC Pyramid – Structure of the Supply of HPC Resources



- Tier 0: European Level: Partnership for Advanced Computing in Europe (PRACE)
- Tier 1: National Level: Gauss Centre for Supercomputing
- Tier 2: Regional-National Level: Gauß Allianz (GA) => **new structure in preparation Nationales Hochleistungsrechnen (NHR)** ***https://www.nhr-gs.de/***
- Tier 3: Local (university) - Regional Level: Gauß Allianz (GA)

Access to HPC Resources

- File applications for PRACE high-end projects on resources currently located at France, Germany, Italy, Spain and Switzerland
  via ***www.prace-ri.eu***

- File applications for GCS large-scale projects (> 35 Mio Coreh) on resources at Jülich, Garching or Stuttgart twice a year directly at one of the 3 national centers, see ***www.gauss-centre.eu***

- File applications for GCS regular projects (< 35 Mio Coreh) at any time (Garching, Stuttgart) resp. twice a year (Jülich) directly at one of the 3 national centers, see ***www.gauss-centre.eu***

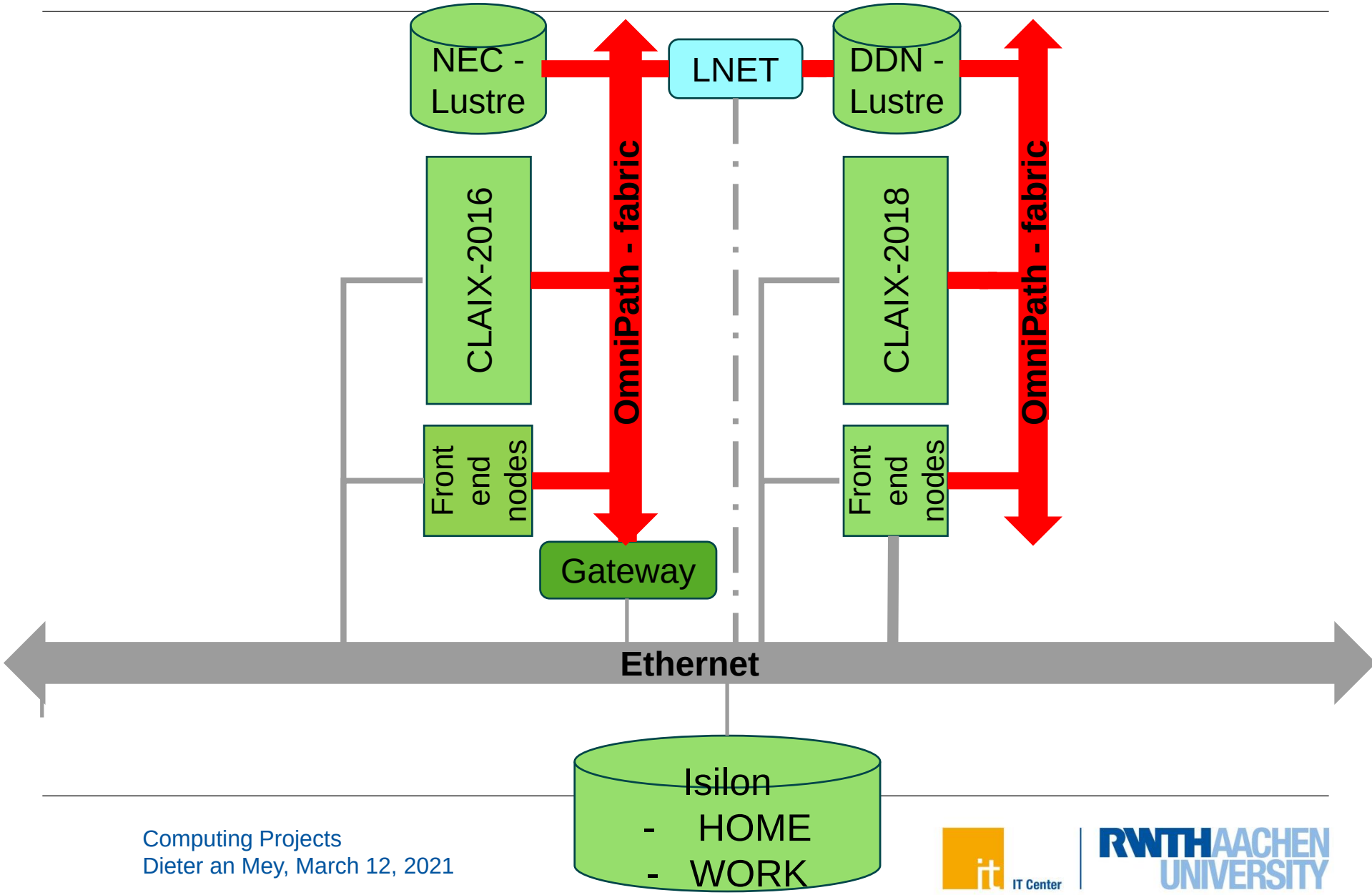
- File applications for JARA-HPC resources twice a year (researchers of FZJ and RWTH only) (>2.4 Mio Coreh) – ***the last call just closed in February***
  see ***https://www.jara.org/de/654***

- File applications for the RWTH Compute Cluster any time
  see ***www.itc.rwth-aachen.de/hpc-projects***

HPC Resources

Computing – Storage – Software (– Networks)

RWTH Compute Cluster Configuration Overview



CLAIX-2016 (Tier 2) versus CLAIX-2018 (Tier 2 + Tier 3)

- **609 MPI nodes**

2-socket Intel Xeon E5-2650v4 (Broadwell-EP)
12 cores each (24 cores total per node)
2.2 GHz, 2.9 GHz turbo mode
128 GB main memory
Local SSD: 120 GB

- **10 MPI-GPU nodes**

MPI nodes w/ 2 NVIDIA P100 SXM2 GPUs
NVLink GPU-GPU interconnect:

- **8 SMP nodes**

8-socket Intel Xeon E7-8860v4 (Broadwell-EX)
18 cores each (144 cores total per node)
2.2 GHz, 3.2 GHz turbo mode
1024 GB main memory

- **Frontend and service nodes**

- **1032 + 216 MPI nodes**

2-socket Intel Xeon Platinum 8160 (Skylake)
24 cores each (48 cores total per node)
2.1 GHz, 3.7 GHz turbo mode
192 GB main memory
Local SSD: 480 GB

- **48 + 6 MPI-GPU nodes**

MPI nodes w/ 2 NVIDIA V100 SXM2 GPUs
NVLink GPU-GPU interconnect

- **Frontend and service nodes**

- **Vendor:** NEC

(Mio) Core hours – the Metric for Computing Resources

- **Core hour = usage / reservation of one core for one hour**

- **Compute resources** are available up to a certain limit (measured in Core-Hours per year)
- Technically everyone with an account on the cluster has a certain budget

- 8760 hours per year / 720 hours per month
- 48 cores per compute node (CLAIX-2018-MPI)

=> 0.42 Mio Coreh per node per year / 35.000 Coreh per node per month

Personal default quota on the RWTH Compute Cluster

- **Monthly quota for a RWTH scientist: 2000 hours**
  (nearly 1 PC / laptop continuously)
- **Monthly quota for a RWTH student: 500 hours**
  (nearly 1 core continuously)

Inquiring project quota

```
r_wlm_usage -p jara1234 -q
Account:                                    jara1234
Type:                                           jara
Start of Accounting Period:               01.11.2020
End of Accounting Period:                 31.10.2021
State of project:                             active
-------------------------------------------------------
Quota monthly (core-h):                       234157
Remaining  core-h of prev. month:             -28868
Consumed   core-h current month:               68228
Consumed   core-h last 4 weeks:               248283
Consumable core-h (%):                            59
Consumable core-h:                            371217
-------------------------------------------------------
Total quota (core-h):                       2.800 Mio
Total consumed core-h so far:               1.383 Mio
-------------------------------------------------------
Default partition:                              c18m
Allowed partitions:                        c18m,c18g
Max. allowed wallclocktime:                24.0 hours
Max. allowed cores per job:                       384
```

```
 - 28868
+ 234157
+ 234157
 - 68228
= 371217
```

Inquiring project usage

```
r_wlm_usage -p jara1234 -m 3   # display 3 months

batchusage from Fri 01.01.2021 00:00:00 to Mon 08.03.2021 23:59:59
 of project jara1234 in corehours:
+===================+============+============+============+============+============+
#         Partition +       User +   Mar 2021 +   Feb 2021 +   Jan 2021 +      total #
+===================+============+============+============+============+============+
#     c18m - normal | user1234   |            |            |      4.29K |      4.29K #
#                   | user1234   |     54.50K |    167.94K |     50.42K |    272.85K #
#                   | user1234   |      2.29K |      1.87K |            |      4.16K #
#                   | user1234   |            |      5.81K |            |      5.81K #
#                   | user1234   |      5.44K |     79.77K |     23.51K |    108.72K #
#                   | user1234   |      6.00K |      7.47K |    164.26K |    177.73K #
#                   | user1234   |            |     28.03  |      0.44  |     28.47  #
+-------------------+------------+------------+------------+------------+------------+
# c18m_low - normal | user1234   |            |            |    666.53  |    666.53  #
#                   | user1234   |            |    139.57  |     49.98K |     50.12K #
#                   | user1234   |            |            |     14.10K |     14.10K #
#                   | user1234   |            |            |     13.69K |     13.69K #
#                   | user1234   |            |            |      1.83  |      1.83  #
+===================+============+============+============+============+============+
#              total |            |     68.23K |    263.02K |    320.92K |    652.17K #
+===================+============+============+============+============+============+
```

GigaBytes – The Metric for Storage Resources

- **Where to store your data?**
  How much data?
  Where is your data located?

1 Byte = 8 bits = space for 1 character
1 GigaByte (MB) = 10^6 Bytes
1 GigaByte (GB) = 10^9 Bytes
1 GigaByte (TB) = 10^12 Bytes
1 GigaByte (PB) = 10^15 Bytes

- Data to be computed has to reside in the computer‘s **main memory**:
  CLAIX-2016: 128 GigaBytes per 24 cores
  CLAIX-2018: 192 GigaBytes per 48 cores

- **Scratch files** can be put into the compute node‘s scratch (/tmp) file system for the life time of a batch job
  CLAIX-2016 / CLAIX-2018 : 120 / 480 GigaByte of fast SSD
  when using multiple compute nodes at a time, the SSDs can be combined to a temporary parallel filesystem *check for BeeOND here*
- **Permanent files** reside on the
  WORK-Fileserver (no back-up, limited life time) or on the
  HOME-Fileserver (back-up**)** or on
  Lustre parallel Filesystem(s) (HPCWORK) (no back-up, limited life time)
  please, find more information here

Storage Hierarchy

Storage	Life time of data	Total Size	Default quota	Peak Bandwidth	remark	usage
memory	program	128 GB / 192 GB		40 GB / 140 GB/s	volatile	CLAIX-2016 / CLAIX-2018
NVME	*job*	*2 TB*		*> 1 GB/s*	*Automatically cleaned*	*only few CLAIX-2016 nodes*
SSD (solid state disk)	job	120 GB / 480 GB		~ 1 GB/s	local disk for each node, automatically cleaned	CLAIX-2016 / CLAIX-2018 $TMP
BEEOND	job	#nodes * 480 GB		#nodes * 1 GB/s	temporary parallel scratch file system for a parallel job running on #nodes nodes	CLAIX-2018 (mid 2020)
Lustre parallel filesystem	Job / short term medium storage	3 PB / 10 PB	default 1TB max 30 TB	30 / 100 GB/s	shared between all cluster nodes, no backup suitable for large blocks of IO	CLAIX-2016 / CLAIX-2018 $HPCWORK
WORK filesystem	medium term storage	< 1 PB	default 250 GB max 1 TB	<1 GB/s	shared between all cluster nodes, no backup	$WORK
HOME filesystem	permanent storage	< 1 PB	default 150 GB max 500 GB	<1 GB/s	shared between all cluster nodes, backup	$HOME
ARCHIVE	long term storage	> 1 PB		< 1 GB/s	shared between all users of RWTH Research data has to be stored 10y (DFG requirement)	

(Application) Software

The bad news:

- **Users are responsible for providing licensies for software they need**
  - unless the software usage is free
  - or the software has been licensed centrally
- Unfortunately, regulations vary considerably between different software packages.
- Users from external institutions may need to provide licences

The good news:

- **The operating system – CentOS Linux**
- **the batch system – Slurm**
- **the development environment – compilers, tools, MPI library**
- **and many other software packages**

are free or are provided by the IT Center

Please, find more information here …

Computing Projects

Management of Computing (and Storage) Resources

Slurm

- It‘s all about batch jobs …
- Slurm is the Batch System, software to manage a large amount of parallel batch jobs
- Interactive usage is not accounted for - it is limited anyhow
- You are not supposed to comsume a large amount of compute resources interactively
- The sysadmins are taking measures to prevent you from doing this.

please, find more information here

What is a Computing Project ?

Organisationally

- A measure to manage large amounts of compute resources to fullfill the demands of the financing agencies (DFG, WR) for scientifically soundness of machine usage.

- the more compute cycles, the higher the quality requirements

Technically

- A linux group and associated file space

- A list of users (Linux user names) in that group

- A category for (Slurm) batch jobs for these users having higher limits(concerning file space and compute cycles) in addition to your personal default resources

- A bunch of (sysadmins‘) scripts and databases for book keeping

- Data stored in the online submission and reviewing system JARDS and related software packages

Filing a proposal for a computing project

Check out *www.itc.rwth-aachen.de/hpc-projects*

- **What is project-based resource management?**
- **Who is (not) affected?**
- **Support for research and education**
- **Roles and tasks related to any computing project**
- **Getting an HPC account for the RWTH Compute Cluster**
- **Filing a compute time application**
  1. Provide a description of the project
  2. fill out the JARDS online form
     a) A PI from a trusted domain (RWTH, FZJ, UKA) can finalize the JARDS online form himself (paperless submission)
     b) The PI has to sign the generated pdf-file, and scan it and send it by email to ***hpc-projects@itc.rwth-aachen.de***
- **Add Acknowledgements to related publications**
- **Compute Project Categories including Links to JARDS**
- **More information (technical details, reviewing)**

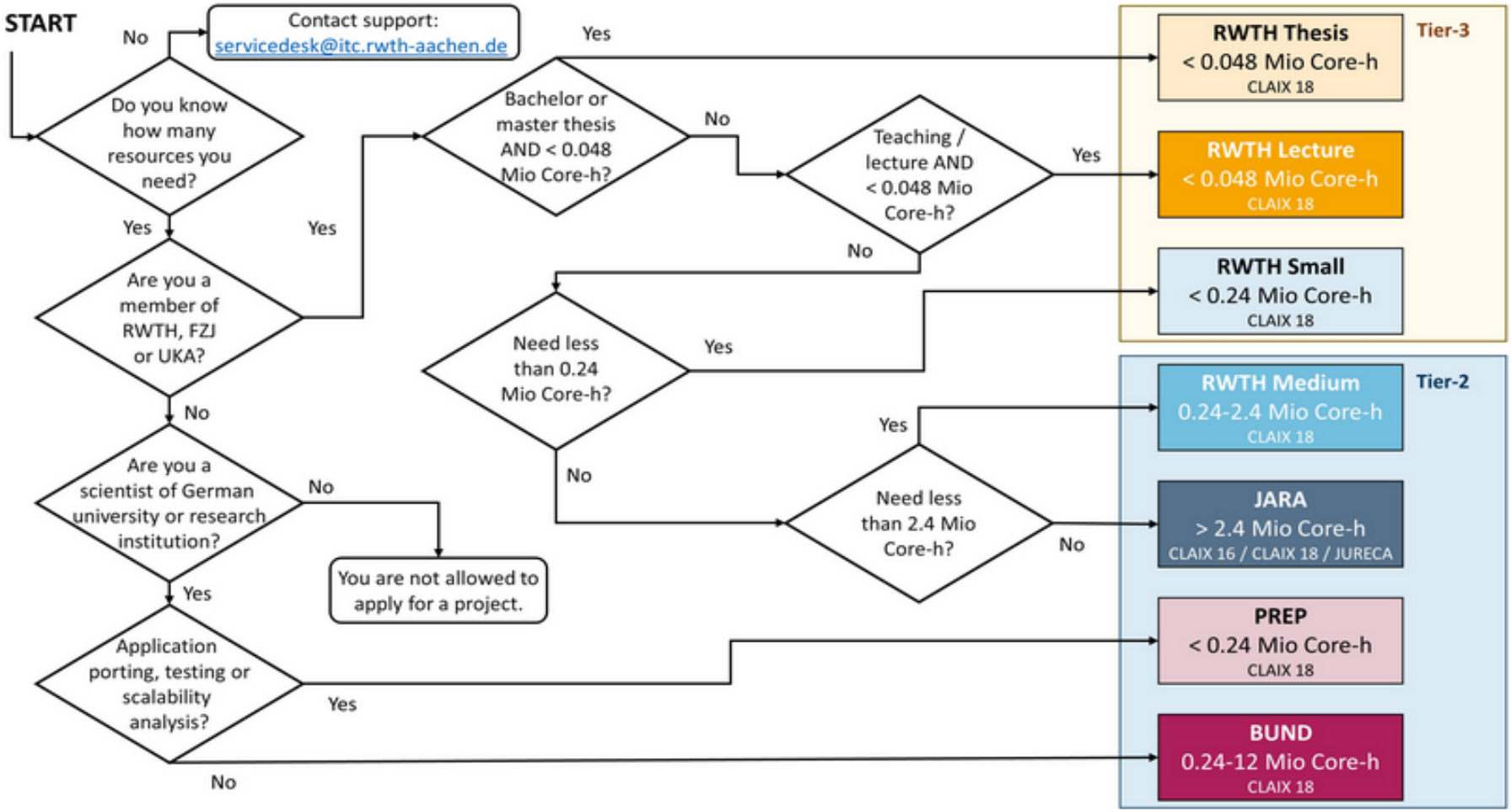
Roles and tasks related to a computing project

- **Principal investigator (PI)**
  Leading scientist, senior researcher
  (for projects related to education: Student's Advisor / Lecturer)
  Takes any responsibility for the computing project
  has to "sign" the project application
  is responsible for reporting about the project
  takes part in reviewing other scientists' applications
  (or nominates a replacement reviewer)

- **Person to contact (PC**) (previously called technical contact, TC)
  the IT Center's contact person for any questions
  technical management of the project
  adding project members
  PI = PC is OK.

- **Project members (PM)**
  Anyone whom the PI trusts
  performs the computations / submits the project's batch jobs
  PI and PC can act as project members as well.

Project Categories



please, find more information here including links to the submission forms

Links

Upcoming Events

Weblinks

- HPC at IT Center of RWTH
- Online Documentation at IT Center of RWTH
- HPC Access
- HPC Resources at IT Center of RWTH (Claix)
- JARA Partition
- Gauss Centre for Supercomputing (GCS)
- Partnership for Advanced Computing in Europa (PRACE)
- HPC Wiki
- Computing Projects at RWTH
- This Event: HPC Intro 2021
- Upcoming Event: PPCES 2021
- Software

Upcoming Event: PPCES March 22-26, 2021

The first two days will focus on performance programming of single compute nodes using Shared Memory Programming w/ OpenMP.

Monday, March 22, 9:00 - 17:00

Tuesday, March 23, 9:00 - 17:00

Days 3 and 4 will focus on performance programming of multiple compute nodes using Message Passing with MPI

Wednesday, March 24, 9:00 - 17:30

Thursday, March 25, 9:00 - 17:30

The last day will focus on machine learning, particularly using NVIDIA GPGPUs.

Friday, March 26, 9:00 – 15:00

www.itc.rwth-aachen.de/ppces

Please, register until Tuesday, March 16

Thank you for your attention