11^{th} Uni.lu HPC School 2021

Overview and Challenges of the ULHPC facility at the EuroHPC Horizon

---

Dr. S. Varrette & UL HPC Team

University of Luxembourg (UL)

https://hpc.uni.lu

Nov. 12-19, 2021, Luxembourg

https://hpc.uni.lu/education/hpcschool

- **11th edition** of this training initiated in 2014
  - ↪ **Hybrid** event (MSA and WebEx), **6** days, **3** keynotes, **12** practical sessions
  - ↪ *dedicated* Slack workspace for live support
    - ✓ ulhpc-school-2021.slack.com
- **Requirement** (outside coffee or tea)
  - ↪ your favorite laptop with your favorite OS
  - ↪ basic knowledge in Linux command line and ability to take notes (Markdown etc.)
    - ✓ (**new**) UNIX/Linux shell-novice sessions was held on Friday

Online UL HPC Tutorials
ulhpc-tutorials.rtfd.io

hpc.uni.lu

ULHPC Technical Docs
hpc-docs.uni.lu

Uni.lu HPC School 2021 Agenda

Friday, November 12, 2021			WebEx + MSA 2.130*
Sessions Objectives : ***Introduction to UNIX/Linux Shell and Command lines***			*Speaker(s)*
08:30 - 09:00	**Remote setup and registration**		
09:00 - 10:45	PS1 [beginners]	Preliminary: Shell setup on your local Laptop (Windows, Mac OS, Linux) Introduction; Navigating Files and Directories	T. Valette, A. Olloh
10:45 - 11:00	**BREAK**		
11h00 - 12:30	PS1 [beginners]	Working With Files and Directories Pipes and Filters Loops	T. Valette, A. Olloh
12:30 - 14:00	**LUNCH**		
14:00 - 15:30	PS1 [beginners]	Shell Scripts Finding Things Concluding remarks	H. Cartiaux

Uni.lu HPC School 2021 Agenda

Monday, November 15, 2021			WebEx + MSA 2.240*
Sessions Objectives :	***ULHPC Ecosystem Overview; HPC Access Setup***		*Speaker(s)*
08:30 - 09:00	**Remote setup and registration**		
09:00 - 10:30	Keynote	Welcome Overview and Challenges of the UL HPC Facility at the EuroHPC Horizon	S. Varrette
10:30 - 11:00	**BREAK**		
11h00 - 12:30	PS2 [beginners]	Preliminaries (SSH - OpenOnDemand)	T. Valette, A. Olloh

Uni.lu HPC School 2021 Agenda

Tuesday, November 16, 2021			WebEx + MSA 2.240*
Sessions Objectives :	***Getting Started on ULHPC***		*Speaker(s)*
08:00 - 08:30	**Remote setup and registration**		
08:30 - 10:15	PS3 [beginners]	Getting Started 2.0 Introduction to the SLURM Job Scheduler, basic launchers	H. Cartiaux
10:15 - 10:30	**BREAK**		
10h30 - 11:30	PS4 [intermediate]	HPC Management of Sequential and Embarrassingly parallel jobs	S. Varrette
11:40 - 12:30	PS5 [advanced]	HPC Software Building: optimizing and complementing the ULHPC software set	S. Varrette

Uni.lu HPC School 2021 Agenda

Wednesday, November 17, 2021			WebEx + MSA 3.330*
Sessions Objectives :	***Python, AI and Machine Learning***		*Speaker(s)*
08:00 - 08:30	**Remote setup and registration**		
08:30 - 10:00	PS6 [beginners]	Prototyping with Python	S. Peter
10:00 - 10:15	**BREAK**		
10h15 - 11:00	PS7 [intermediate]	HPC Containers with Singularity	E. Kieffer
11:10 - 12:45	PS8 [advanced]	Advanced distributed computing with Python	E. Kieffer

Uni.lu HPC School 2021 Agenda

Thursday, November 18, 2021			WebEx + MSA 3.200*
Sessions Objectives :	***Scalable Science and Parallel Programming***		*Speaker(s)*
08:00 - 08:30	**Remote setup and registration**		
08:30 - 09:50	PS9 [advanced]	Scalable Science with OpenMP/MPI	S. Varrette
10:00 - 10:45	Keynote	Advanced Performance Engineering for Efficient Parallel Debugging	X. Besseron
10:45 - 11:00	**BREAK**		
11h00 - 12:45	PS10 [advanced]	Introduction to GPU programming with OpenACC and OpenCL	E. Ezhilmathi, L. Koutsantonis, T. Pessoa

Uni.lu HPC School 2021 Agenda

Friday, November 19, 2021			WebEx + MSA 4.150*
Sessions Objectives :	***Statistical Computing and Big Data Analytics***		*Speaker(s)*
08:00 - 08:30	**Remote setup and registration**		
08:30 - 10:00	PS11 [intermediate]	R - statistical computing	A. Ginolhac
10:00 - 10:15	**BREAK**		
10:15 - 11:15	Keynote	Data management (backup, security, ...)	S. Peter
11:25 - 12:30	PS12 [intermediate]	Big Data Analytics: Batch, Stream and Hydrid processing engines	S. Varrette
12:30 - 12:45	**Closing Remarks / Take Away messages**		

Uni.lu HPC School 2021 Contributors

... in alphabetical order

Dr. Xavier Besseron
Research Scientist

Hyacinthe Cartiaux
Infra. & HPC Arch. Engineer

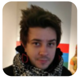
Dr. Aurelien Ginohac
Research Scientist

Dr. Emmanuel Kieffer
Research Scientist

Dr. Loizos Koutsantonis
Postdoctoral Researcher

Dr. Ezhilmathi Krishnasamy
Postdoctoral Researcher

Abatcha Olloh
Infra. & HPC Arch. Engineer



Dr. Tiago C. Pessoa
Postdoctoral Researcher

Sarah Peter
Infra. & Arch. Engineer

Teddy Valette
Infra. & HPC Arch. Engineer

Dr. Sebastien Varrette
Research Scientist

... and additional help (Survey, session tests)

Arlyne Vandeventer
Project Manager

Uni.lu HPC School 2021 Contributors

... in alphabetical order

Dr. Xavier Besseron
Research Scientist

Hyacinthe Cartiaux
Infra. & HPC Arch. Engineer

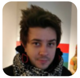
Dr. Aurelien Ginohac
Research Scientist

Dr. Emmanuel Kieffer
Research Scientist

Dr. Loizos Koutsantonis
Postdoctoral Researcher

Dr. Ezhilmathi Krishnasamy
Postdoctoral Researcher

Abatcha Olloh
Infra. & HPC Arch. Engineer



Dr. Tiago C. Pessoa
Postdoctoral Researcher

Sarah Peter
Infra. & Arch. Engineer

Teddy Valette
Infra. & HPC Arch. Engineer

Dr. Sebastien Varrette
Research Scientist

... and additional help (Survey, session tests)

Arlyne Vandeventer
Project Manager

Summary

Summary

Prerequisites: Metrics

- **HPC**: **H**igh **P**erformance **C**omputing


Main HPC Performance Metrics

- **Computing Capacity**: often measured in **flops** (or **flop/s**)
  - ↪ **Fl**oating **p**oint **o**perations per **s**econds (often in DP)
  - ↪ **GFlops** $= 10^9$ **TFlops** $= 10^{12}$ **PFlops** $= 10^{15}$ **EFlops** $= 10^{18}$

Prerequisites: Metrics

- **HPC**: **H**igh **P**erformance **C**omputing

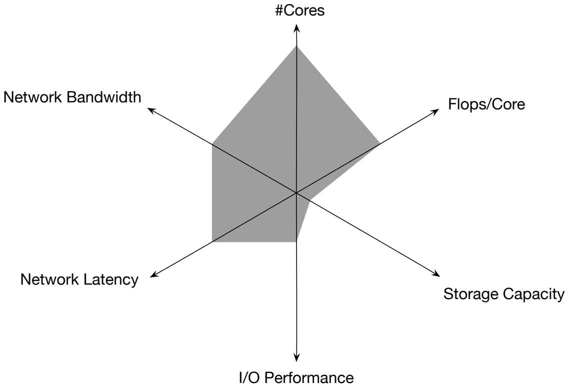
BD: **B**ig **D**ata

Main HPC Performance Metrics

- **Computing Capacity**: often measured in **flops** (or **flop/s**)
  - ↪ **Fl**oating **p**oint **o**perations per **s**econds (often in DP)
  - ↪ **GFlops** $= 10^9$ **TFlops** $= 10^{12}$ **PFlops** $= 10^{15}$ **EFlops** $= 10^{18}$
- **Storage Capacity**: measured in multiples of **bytes** = 8 **bits**
  - ↪ **GB** $= 10^9$ bytes **TB** $= 10^{12}$ **PB** $= 10^{15}$ **EB** $= 10^{18}$
  - ↪ **GiB** $= 1024^3$ bytes **TiB** $= 1024^4$ **PiB** $= 1024^5$ **EiB** $= 1024^6$
- Transfer rate on a medium measured in **Mb/s** or **MB/s**
- Other metrics: Sequential vs Random **R/W speed**, **IOPS** ...

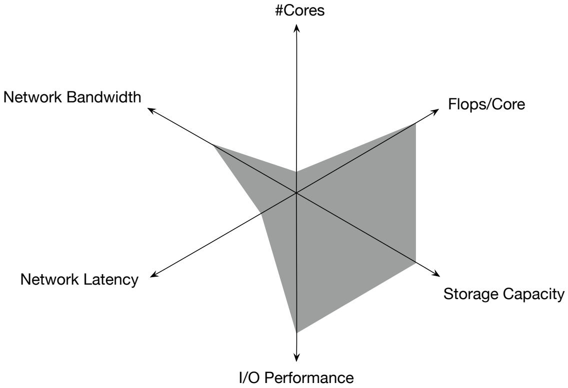
Different Needs for Different Domains

Material Science & Engineering



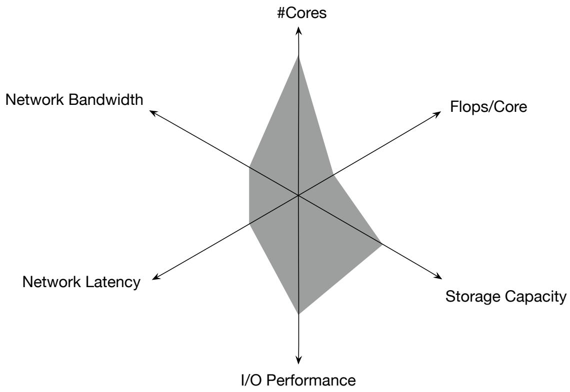
Different Needs for Different Domains

Biomedical Industry / Life Sciences



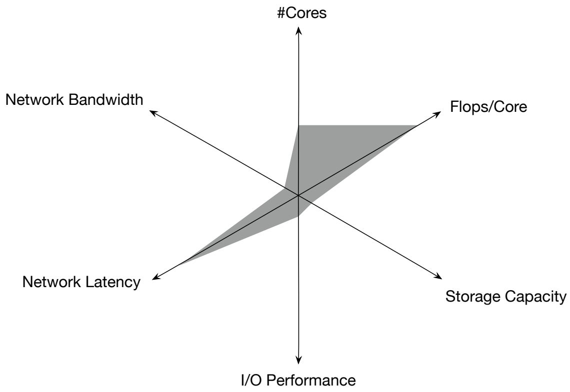
Different Needs for Different Domains

Deep Learning / Cognitive Computing



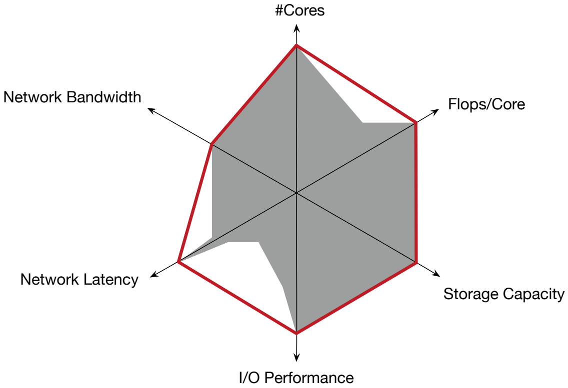
Different Needs for Different Domains

IoT, FinTech



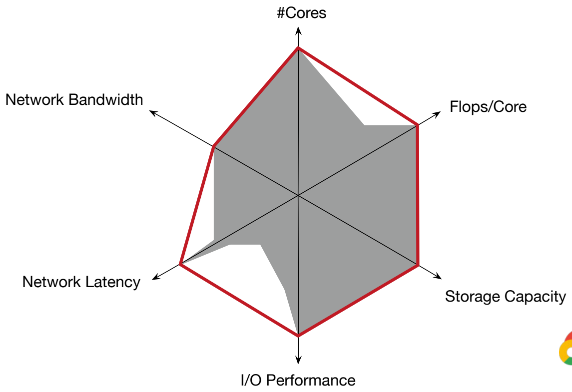
Different Needs for Different Domains

ALL Research Computing Domains



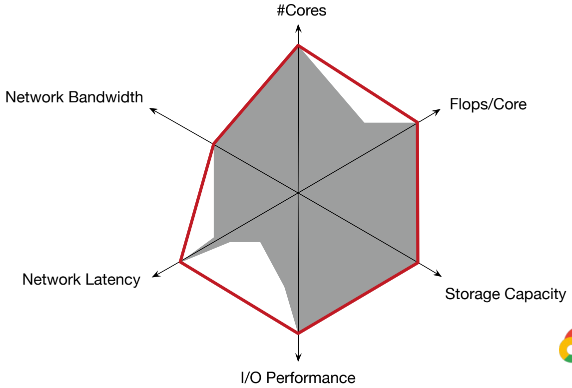
Different Needs for Different Domains

ALL Research Computing Domains



Different Needs for Different Domains

ALL Research Computing Domains



High Performance Computing & Big Data Services

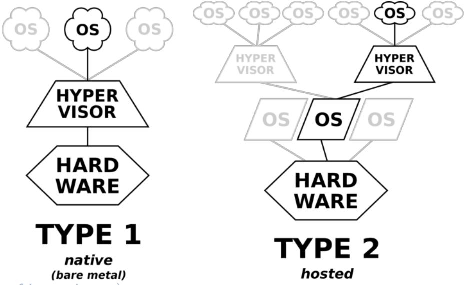
hpc.uni.lu
hpc@uni.lu
@ULHPC

Computing for Researchers: Laptop

- **Regular PC / Local Laptop / Workstation**
  - ↪ **Native OS** (Windows, Linux, Mac etc.)

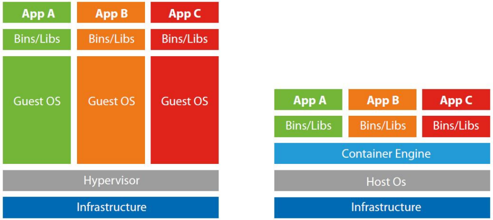
Computing for Researchers: Laptop

- **Regular PC / Local Laptop / Workstation**
  - ↪ **Native OS** (Windows, Linux, Mac etc.)
- Virtualized OS (**VM**) through an **hypervisor**
  - ↪ *Hypervisor*: core virtualization engine / environment
    - ✓ Ex: Xen, VMWare ESXi, KVM, VirtualBox – **Non-negligible Performance loss**: $\geq 20\%$



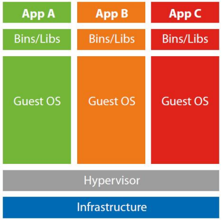
Computing for Researchers: Laptop

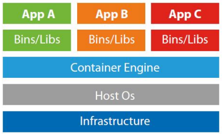
- **Regular PC / Local Laptop / Workstation**
  - ↪ **Native OS** (Windows, Linux, Mac etc.)
- Virtualized OS (**VM**) through an **hypervisor**
  - ↪ *Hypervisor*: core virtualization engine / environment
    - ✓ Ex: Xen, VMWare ESXi, KVM, VirtualBox – **Non-negligible Performance loss**: $\geq 20\%$
- **Container**-based Virtualization
  - ↪ similar to VMs, **yet** containers **share** the system kernel of the host with others
    - ✓ Ex: Docker, Singularity, Shifter
  - ↪ Orchestration: Kubernetes



Computing for Researchers: Laptop

- **Regular PC / Local Laptop / Workstation**
  - ↪ **Native OS** (Windows, Linux, Mac etc.)
- Virtualized OS (**VM**) through an **hypervisor**
  - ↪ *Hypervisor*: core virtualization engine / environment
    - ✓ Ex: Xen, VMWare ESXi, KVM, VirtualBox – **Non-negligible Performance loss**: $\geq$ 20%
- **Container**-based Virtualization
  - ↪ similar to VMs, **yet** containers **share** the system kernel of the host with others
    - ✓ Ex: Docker, Singularity, Shifter
  - ↪ Orchestration: Kubernetes



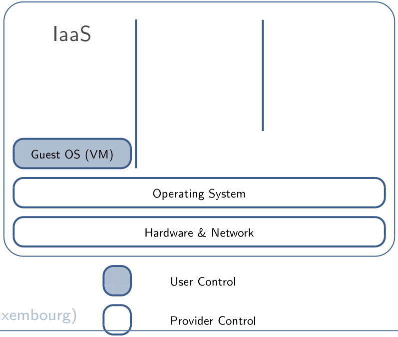


HPC-compliant Containers

- No docker for security reasons
- **Singularity**, Sarus

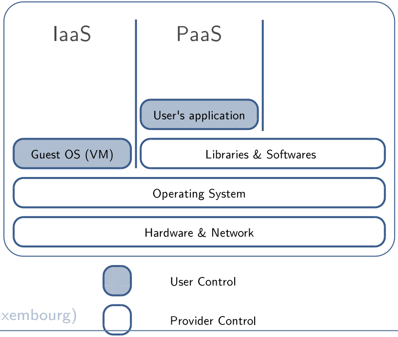
Computing for Researchers: Cloud

- **Cloud Computing**
  - ↪ access to shared *(generally virtualized)* resources
  - ↪ pay-per-use approach
  - ↪ **Infrastructure** as a Service (**I**aaS)



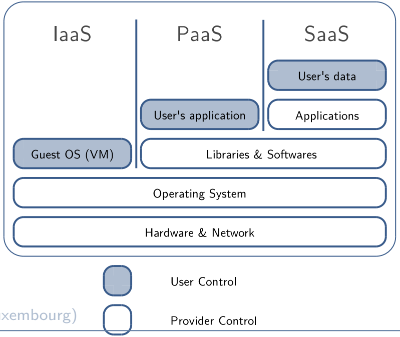
Computing for Researchers: Cloud

- **Cloud Computing**
  - ↪ access to shared *(generally virtualized)* resources
  - ↪ pay-per-use approach
  - ↪ **Platform** as a Service (**P**aaS)



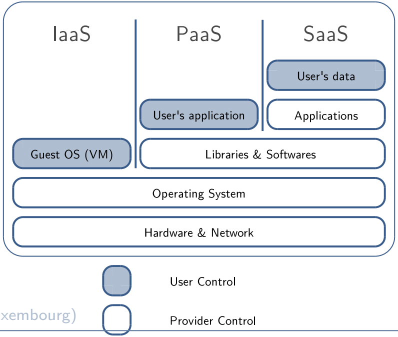
Computing for Researchers: Cloud

- **Cloud Computing**
  - ↪ access to shared *(generally virtualized)* resources
  - ↪ pay-per-use approach
  - ↪ **Software** as a Service (**S**aaS)



Computing for Researchers: Cloud

- **Cloud Computing**
  - ↪ access to shared *(generally virtualized)* resources
  - ↪ pay-per-use approach
  - ↪ **XXX** as a Service (<**X**>aaS)



Computing for Researchers: HPC

- High Performance Computing (HPC) platforms
  ↪ For **Speedup**, **Scalability** and **Faster Time to Solution**

Computing for Researchers: HPC

- High Performance Computing (HPC) platforms
  - ↪ For **Speedup**, **Scalability** and **Faster Time to Solution**
  - ↪ Available peta-scale facilities in Luxembourg
    - ✓ Tier 0/1: MeluXina (part of EuroHPC network)

Computing for Researchers: HPC

- High Performance Computing (HPC) platforms
  - ↪ For **Speedup**, **Scalability** and **Faster Time to Solution**
  - ↪ Available peta-scale facilities in Luxembourg
    - ✓ Tier 0/1: MeluXina (part of EuroHPC network)
    - ✓ Tier 2: Uni.lu HPC

Computing for Researchers: HPC

- High Performance Computing (HPC) platforms
  - ↪ For **Speedup**, **Scalability** and **Faster Time to Solution**
  - ↪ Available peta-scale facilities in Luxembourg
    - ✓ Tier 0/1: MeluXina (part of EuroHPC network)
    - ✓ Tier 2: Uni.lu HPC

YET...

PC $\neq$ Cloud $\neq$ HPC

Computing for Researchers: HPC

- High Performance Computing (HPC) platforms
  - ↪ For **Speedup**, **Scalability** and **Faster Time to Solution**
  - ↪ Available peta-scale facilities in Luxembourg
    - ✓ Tier 0/1: MeluXina (part of EuroHPC network)
    - ✓ Tier 2: Uni.lu HPC

YET...

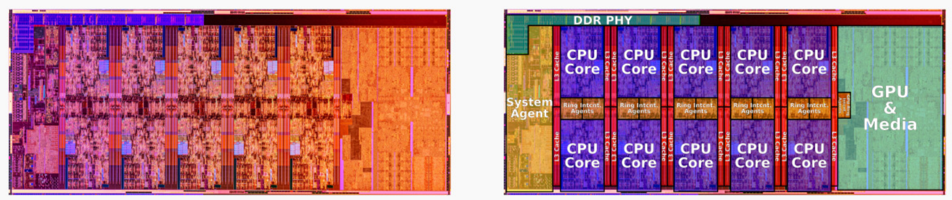
PC ≠ Cloud ≠ HPC

- **HPC ≃ Formula 1** *(permits to quickly detect* ***real*** *HPC experts)*
  - ↪ relies on ultra efficient hardware / interconnect (IB HDR. . .)
  - ↪ . . . when Cloud has to stay standard (10 GbE etc. . .)
- **Does not mean the 3 approaches cannot work together**

HPC Computing Hardware

Base

- **CPU** (Central Processing Unit) *Highest software flexibility*
  - ↪ High performance across all computational domains
  - ↪ Ex: Intel Core i9-10900K *(Q2'20)* $R_{peak} \simeq 1{,}18$ TFlops (DP)
    - ✓ 10 cores @3.7GHz (14nm, 125W, $\simeq$ 7 billion transistors) + integ. graphics



Intel Comet Lake die (2020)

HPC Computing Hardware

Base

- **CPU** (Central Processing Unit) — *Highest software flexibility*
  - ↪ High performance across all computational domains
  - ↪ Ex: Intel Core i9-10900K *(Q2'20)* — $R_{peak} \simeq$ 1,18 TFlops (DP)
    - ✓ 10 cores @3.7GHz (14nm, 125W, $\simeq$ 7 billion transistors) + integ. graphics

Accelerators

- **GPU** (Graphics Processing Unit): — *Ideal for ML/DL workloads*
  - ↪ Ex: Nvidia Tesla A100 *(Q1'20)* — $R_{peak} \simeq$ 9.7 TFlops (DP)
    - ✓ 6912 cores @ 1.41GHz — (7nm, 400W, 54,2 billion transistors)

HPC Computing Hardware

Base

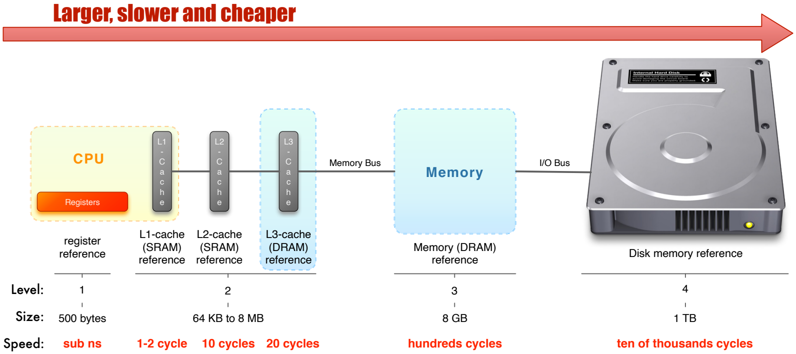
- **CPU** (Central Processing Unit) *Highest software flexibility*
  - ↪ High performance across all computational domains
  - ↪ Ex: Intel Core i9-10900K *(Q2'20)* $R_{peak} \simeq$ 1,18 TFlops (DP)
    - ✓ 10 cores @3.7GHz (14nm, 125W, $\simeq$ 7 billion transistors) + integ. graphics

Accelerators

- **GPU** (Graphics Processing Unit): *Ideal for ML/DL workloads*
  - ↪ Ex: Nvidia Tesla A100 *(Q1'20)* $R_{peak} \simeq$ 9.7 TFlops (DP)
    - ✓ 6912 cores @ 1.41GHz (7nm, 400W, 54,2 billion transistors)
- ~~Intel MIC (Many Integrated Core) Accelerator~~
- **ASIC** (Application-Specific Integrated Circuits), **FPGA** (Field Programmable Gate Array)
  - ↪ least software flexibility, highest performance for specialized problems
    - ✓ Ex: AI, Mining, Sequencing...

$\Longrightarrow$ toward hybrid platforms w. DL enabled accelerators

HPC Components: Local Memory

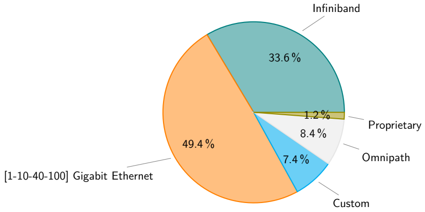


- SSD (SATA3) R/W: 550 MB/s; 100000 IOPS **450 €/TB**
- HDD (SATA3 @ 7,2 krpm) R/W: 227 MB/s; 85 IOPS **54 €/TB**

HPC Components: Interconnect

- **latency**: time to send a minimal (0 byte) message from A to B
- **bandwidth**: max amount of data communicated per unit of time

Technology	Effective Bandwidth		Latency
Gigabit Ethernet	1 Gb/s	125 MB/s	40μs to 300μs
10 Gigabit Ethernet	10 Gb/s	1.25 GB/s	4μs to 5μs
Infiniband QDR	40 Gb/s	5 GB/s	1.29μs to 2.6μs
Infiniband EDR	100 Gb/s	12.5 GB/s	0.61μs to 1.3μs
Infiniband HDR	200 Gb/s	25 GB/s	0.5μs to 1.1μs
100 Gigabit Ethernet	100 Gb/s	1.25 GB/s	30μs
Intel Omnipath	100 Gb/s	12.5 GB/s	0.9μs

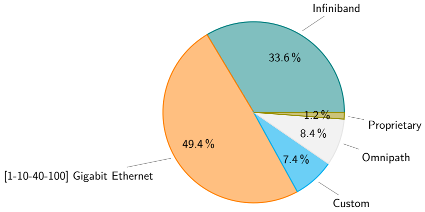


[Source : www.top500.org, Jun 2021]

HPC Components: Interconnect

- **latency**: time to send a minimal (0 byte) message from A to B
- **bandwidth**: max amount of data communicated per unit of time

Technology	Effective Bandwidth		Latency
Gigabit Ethernet	1 Gb/s	125 MB/s	40μs to 300μs
10 Gigabit Ethernet	10 Gb/s	1.25 GB/s	4μs to 5μs
Infiniband QDR	40 Gb/s	5 GB/s	1.29μs to 2.6μs
Infiniband EDR	100 Gb/s	12.5 GB/s	0.61μs to 1.3μs
Infiniband HDR	200 Gb/s	25 GB/s	0.5μs to 1.1μs
100 Gigabit Ethernet	100 Gb/s	1.25 GB/s	30μs
Intel Omnipath	100 Gb/s	12.5 GB/s	0.9μs



[Source : www.top500.org, Jun 2021]

Network Topologies

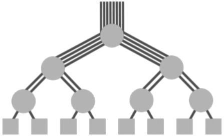
- **Direct** vs. **Indirect** interconnect
  - ↪ *direct*: each network node attaches to at least one compute node
  - ↪ *indirect*: compute nodes attached at the edge of the network only
    - ✓ many routers only connect to other routers.

Network Topologies

- **Direct** vs. **Indirect** interconnect
  - ↪ *direct*: each network node attaches to at least one compute node
  - ↪ *indirect*: compute nodes attached at the edge of the network only
    - ✓ many routers only connect to other routers.

Main HPC Topologies

- **CLOS Network / Fat-Trees** [Indirect]
  - ↪ can be fully non-blocking (1:1) or blocking (x:1)
  - ↪ typically enables **best performance**
    - ✓ Non blocking bandwidth, lowest network latency

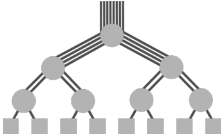


Network Topologies

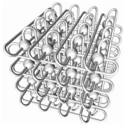
- **Direct** vs. **Indirect** interconnect
  - ↪ *direct*: each network node attaches to at least one compute node
  - ↪ *indirect*: compute nodes attached at the edge of the network only
    - ✓ many routers only connect to other routers.

Main HPC Topologies

- **CLOS Network / Fat-Trees** [Indirect]
  - ↪ can be fully non-blocking (1:1) or blocking (x:1)
  - ↪ typically enables **best performance**
    - ✓ Non blocking bandwidth, lowest network latency

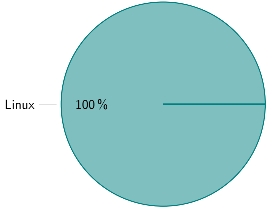


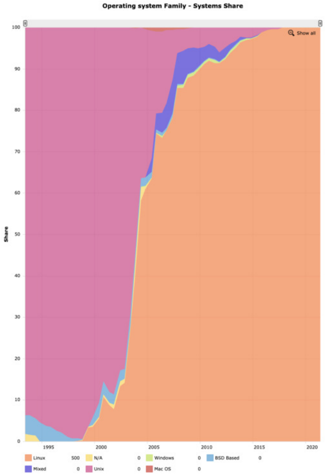
- **Mesh or 3D-torus** [Direct]
  - ↪ Blocking network, cost-effective for systems at scale
  - ↪ Great performance solutions for applications with locality
  - ↪ Simple expansion for future growth



HPC Components: Operating System

- Exclusively Linux-based (**really** 100%)
  - ↪ Note: Used to be Unix before
  - ↪ **better to become familiar with Linux environments**
    - ✓ interaction can be done from **ANY** OS
- Reasons:
  - ↪ stability
  - ↪ development flexibility





[Source : `www.top500.org`, Jun 2021]

HPC Components: Software Stack

- **Remote connection to the platform** SSH
- **Identity Management / SSO**: LDAP, Kerberos, IPA...
- **Resource management**: job/batch scheduler
  - ↪ SLURM, OAR, PBS, MOAB/Torque...
- (Automatic) **Node Deployment**:
  - ↪ FAI, Kickstart, Puppet, Chef, Ansible, Kadeploy...
- (Automatic) **User Software Management**:
  - ↪ Easybuild, Environment Modules, LMod
- **Platform Monitoring**:
  - ↪ Nagios, Icinga, Ganglia, Foreman, Cacti, Alerta...

[Big]Data Management: Disk Encl.

- $\simeq$ **150 K€**/encl. - 60-84 disks (4U)
  - ↪ incl. redundant (i.e. 2) RAID controllers (master/slave)

[Big]Data Management: FS Summary

- **File System (FS)**: Logical manner to *store, organize & access* data
  - ↪ (local) **Disk FS** : `FAT32`, `NTFS`, `HFS+`, `ext4`, `{x,z,btr}fs`...
  - ↪ **Networked FS**: `NFS`, `CIFS/SMB`, `AFP`
  - ↪ **Parallel/Distributed FS**: `SpectrumScale/GPFS`, `Lustre`
    - ✓ typical FS for HPC / HTC (High Throughput Computing)

[Big]Data Management: FS Summary

- **File System (FS)**: Logical manner to *store, organize & access* data
  - ↪ (local) **Disk FS** : `FAT32, NTFS, HFS+, ext4, {x,z,btr}fs...`
  - ↪ **Networked FS**: `NFS, CIFS/SMB, AFP`
  - ↪ **Parallel/Distributed FS**: `SpectrumScale/GPFS, Lustre`
    - ✓ typical FS for HPC / HTC (High Throughput Computing)

Main Characteristic of Parallel/Distributed File Systems

Capacity and Performance increase with #servers

[Big]Data Management: FS Summary

- **File System (FS)**: Logical manner to *store, organize & access* data
  - ↪ (local) **Disk FS** : `FAT32, NTFS, HFS+, ext4, {x,z,btr}fs...`
  - ↪ **Networked FS**: `NFS, CIFS/SMB, AFP`
  - ↪ **Parallel/Distributed FS**: `SpectrumScale/GPFS, Lustre`
    - ✓ typical FS for HPC / HTC (High Throughput Computing)

Main Characteristic of Parallel/Distributed File Systems

Capacity and Performance increase with #servers

Name	Type	Read* [GB/s]	Write* [GB/s]
`ext4`	Disk FS	0.426	0.212
`nfs`	Networked FS	0.381	0.090
`gpfs` (iris/aion)	Parallel/Distributed FS	**22.58**	**19.02**
`lustre` (iris/aion)	Parallel/Distributed FS	**12.97**	**16,15**

* maximum **random** read/write, per IOZone or IOR measures, using concurrent nodes for networked FS.

HPC Components: Data Center

Definition (Data Center)

- Facility to house computer systems & components
  - ↪ Basic storage component: **rack** (height: 42 RU)

$$PUE = \frac{\textit{Total facility power}}{\textit{IT equipment power}}$$

HPC Components: Data Center

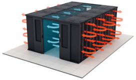
Definition (Data Center)

- Facility to house computer systems & components
  - ↪ Basic storage component: **rack** (height: 42 RU)

$$PUE = \frac{Total\ facility\ power}{IT\ equipment\ power}$$

Challenges: Power (UPS, battery), Cooling, Fire protection, Security

- Power/Heat dissipation per rack:
  - ↪ HPC **computing** racks: **30-120 kW**
  - ↪ **Storage** racks: **15 kW**
  - ↪ **Interconnect** racks: **5 kW**
- Various **Cooling** Technology
  - ↪ Airflow with In-Row cooling



  - ↪ Direct-Liquid Cooling, Immersion...

Summary

University of Luxembourg & HPC

- *With regards to HPC*, University of Luxembourg offers:
  - ↪ **People**
    - ✓ **Domain experts**, Computational and Data scientists
    - ✓ Specialists in parallel algorithmics

University of Luxembourg & HPC

- *With regards to HPC*, University of Luxembourg offers:
  - ↪ **People**
    - ✓ **Domain experts**, Computational and Data scientists
    - ✓ Specialists in parallel algorithmics
  - ↪ **Services**
    - ✓ **HPC Compute & Data services**
    - ✓ IT services (SIU)

University of Luxembourg & HPC

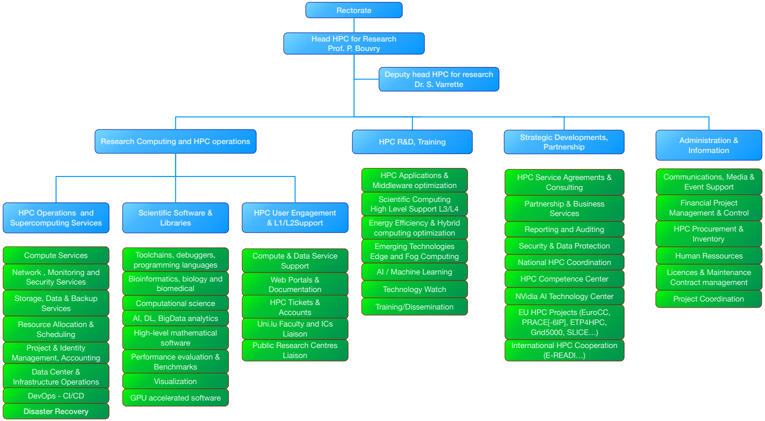
- *With regards to HPC*, University of Luxembourg offers:
  - ↪ **People**
    - ✓ **Domain experts**, Computational and Data scientists
    - ✓ Specialists in parallel algorithmics
  - ↪ **Services**
    - ✓ **HPC Compute & Data services**
    - ✓ IT services (SIU)
  - ↪ **Infrastructure**
    - ✓ **State-of-the-art HPC systems**, **2.7 PFlops** compute capacity
    - ✓ Highly capable Data Center (*Centre De Calcul* **CDC**)
    - ✓ Cutting-edge energy-efficient Direct Liquid Cooling capability

University of Luxembourg & HPC

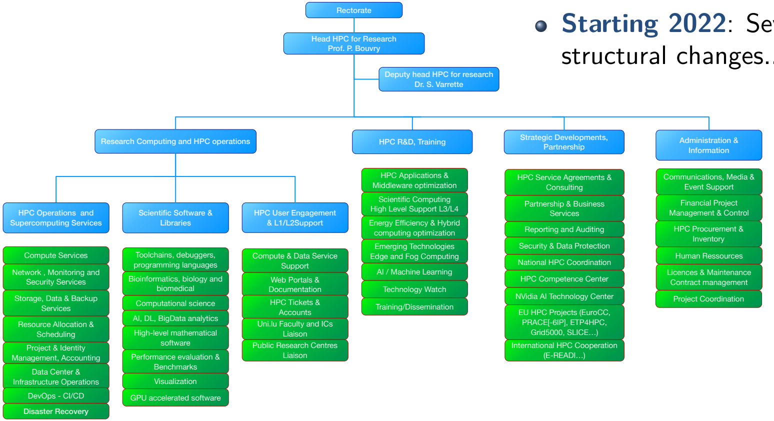
- *With regards to HPC*, University of Luxembourg offers:
  - ↪ **People**
    - ✓ **Domain experts**, Computational and Data scientists
    - ✓ Specialists in parallel algorithmics
  - ↪ **Services**
    - ✓ **HPC Compute & Data services**
    - ✓ IT services (SIU)
  - ↪ **Infrastructure**
    - ✓ **State-of-the-art HPC systems**, **2.7 PFlops** compute capacity
    - ✓ Highly capable Data Center (*Centre De Calcul* **CDC**)
    - ✓ Cutting-edge energy-efficient Direct Liquid Cooling capability
  - ↪ **Education & Training**
    - ✓ MICS Parallel Computing lecture, ULHPC School, *(new)* EuroHPC Master
    - ✓ **Technology Transfer HPC workshops & seminars**

      ... in collaboration with UL / National HPC Competence Center

UL HPC Governance & Pillars



UL HPC Governance & Pillars



- **Starting 2022**: Several structural changes...

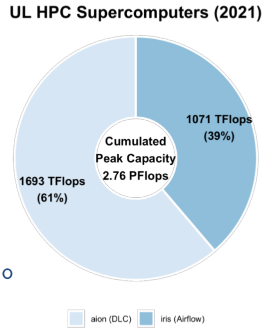
UL HPC Facility

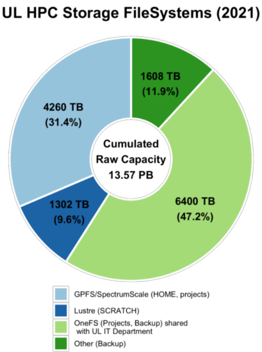
- **Managed and operated since 2007** (Dr. S. Varrette & Co.)
  - ↪ 2nd Largest HPC facility in Luxembourg after EuroHPC MeluXina

hpc.uni.lu

Technical Docs:
hpc-docs.uni.lu

ULHPC Tutorials:
ulhpc-tutorials.rtfd.io





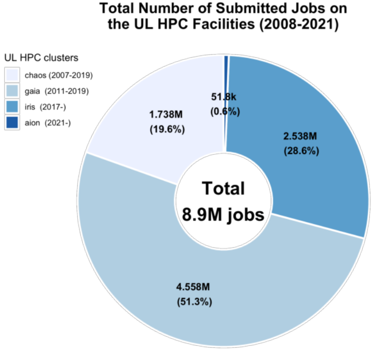
HPC in Luxembourg and Around in EU

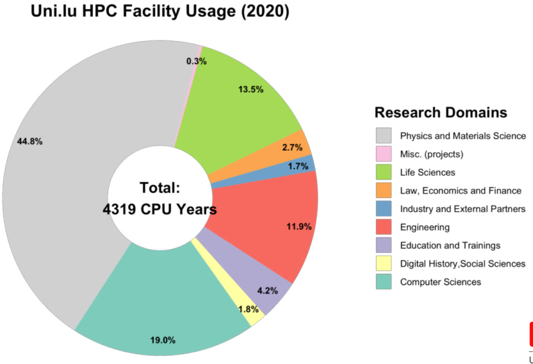
Tier 0: EU
Tier 1: National
Tier 2: Regional | Univ.

Country	System(s)	Type	Institute	(CPU) #Nodes	#Cores	#[GPU]Accelerators	R_{peak}	Shared Storage
	MeluXina (2021)	Euro-HPC **Petascale** *Tier 0/1 (EU,Nat)*	LuxProvide	824	≃ 88 000	764 NVidia A100	**17.57 PF**	≃ 20 PB
Luxembourg	aion,iris	*Tier 2 (Univ)*	**Uni.lu HPC**	**514**	46528	**96** NVidia V100	**2.76 PF**	**13.60 PB**
		Tier 2 (local)	LIST	40	1280	8 Nvidia V100	0.126 PF	0.58 PB
	TGCC (Joliot-Curie)	*Tier 0 (EU)*	GENCI/CEA	4808	430 448	828 Xeon Phi, 128 NVidia V100	22.26 PF	35PB
France	JeanZay	*Tier 1 (Nat.)*	GENCI/Idris	1 528	61 120	1292 NVidia V100	14.97 PF	31.2 PB
	ROMEO	*Tier 2 (Reg.)*	Univ. Reims	115	3 220	280 NVidia P100	1.75 PF	0.634
	Vlaams	*Tier 1 (Nat.)*	VSC	988	27 664	n/a	1.63 PF	1.3PB
Belgium	zenobe	*Tier 1 (Nat.)*	Cenaero	584	14 016	4 NVidia K40	0.41 PF	0.356PB
	Hortense	*Tier 2 (Reg.)*	Gent Univ.	n/a	≃ 40 000	88 NVidia V100	3.3PF	3PB
	JUWELS	*Tier 0 (EU)*	JSC	2571	122 768	224 Nvidia V100	12.3 PF	130.3PB
	JURECA	*Tier 0 (EU)*	JSC	3524	156 736	1640 Xeon Phi	7.24 PF	(as above)
Germany	Hawk	*Tier 0 (EU)*	HLRS, Univ. Stuttgart	5632	720 896	n/a	26 PF	≃25PB
	SuperMUC-NG	*Tier 0 (EU)*	LRZ, Munich	6480	311 040	n/a	26.9 PF	70.16PB
	CLAIX-2018	*Tier 2 (Univ)*	Univ. Aachen	1307	61 200	108 Nvidia V100	4.11 PF	3PB
Bulgaria	PetaSC (2021)	Euro-HPC **Petascale** *Tier 0/1 (EU,Nat)*	SofiaTech	n/a	n/a	n/a	**4.5 PF**	n/a
Czech Republic	Barbora	*Tier 1 (Nat.)*	IT4Innovation	201	7232	32 NVidia V100	0.85 PF	≃ 1PB
	Karolina (2021)	Euro-HPC **Petascale** *Tier 0/1 (EU,Nat)*	IT4Innovation	826	≃ 100K	560 NVidia A100	**9.4 PF**	1PB
Finland	LUMI (2021)	Euro-HPC **Pre-exascale** *Tier 0 (EU)*	CSC	n/a	≃ 200K (LUMI-C)	n/a	**375 PF**	127PB
	Marconi-A3	*Tier 0 (EU)*	Cineca	3216	154 368	n/a	10.37 PF	10PB
Italy	Galileo	*Tier 1 (Nat.)*	Cineca	1022	36792	n/a	1.35 PF	1.92PB
	Leonardo (2021)	Euro-HPC **Pre-exascale** *Tier 0 (EU)*	Cineca	4992	n/a	13824 Nvidia A100	**249.5 PF**	100PB
Portugal	Deucalion (2021)	Euro-HPC **Petascale** *Tier 0/1 (EU,Nat)*	MACC	n/a	n/a	n/a	**7.2 PF**	n/a
Slovenia	VEGA (2021)	Euro-HPC **Petascale** *Tier 0/1 (EU,Nat)*	Maribor SC	960	122,8K	240 NVidia A100	**10.1 PF**	24 PB
Spain	MareNostrum 4	*Tier 0 (EU)*	BSC	3456	165 888	n/a	11.15 PF	14PB
	MareNostrum 5 (2021)	Euro-HPC **Pre-exascale** *Tier 0 (EU)*	BSC	n/a	n/a	n/a	**≃ 200 PF**	n/a
Switzerland	Piz-Daint	*Tier 0 (EU)*	CSCS, ETH Zürich	7517	387 872	5704 NVidia P100	29.34 PF	8.8PB

Uni.lu HPC Users Activity and Profile

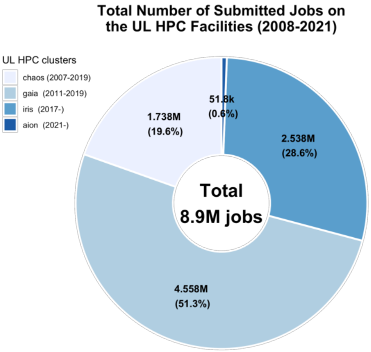
- **23 computational domains accelerated on UL HPC**
  - ↪ ULHPC Annual Report 2020: 630 active *(submitting)* users

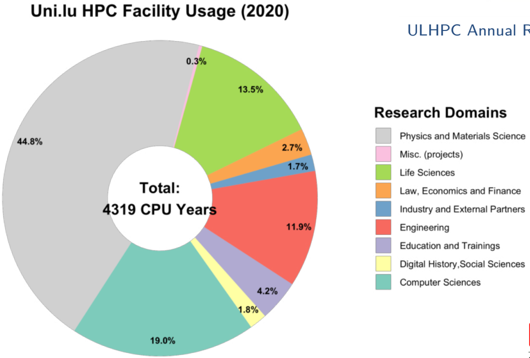


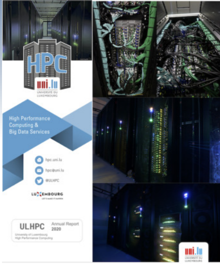


Uni.lu HPC Users Activity and Profile

- **23 computational domains accelerated on UL HPC**
  - ↪ ULHPC Annual Report 2020: 630 active *(submitting)* users



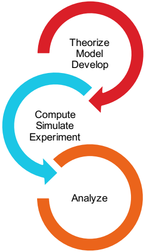


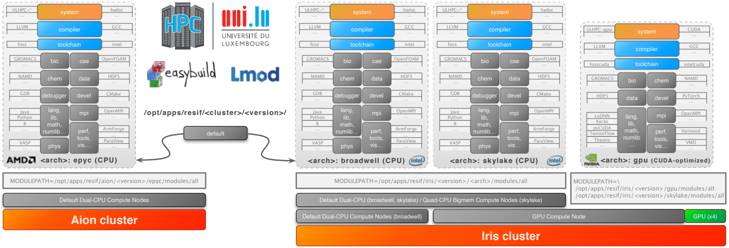


ULHPC Annual Report 2020

Accelerating Research - User Software Sets

- **Over 270 software packages** available for researchers
  - ↪ software environment generated using RESIF 3.0 framework over Easybuild
    - ✓ **optimized builds** organized by architecture, exposed through Environment Modules/Lmod
    - ✓ Categorized Naming Scheme <category>/<name>/<version>-<toolchain><versionsuffix>
  - ↪ containerized applications delivered with Singularity system





[ACM PEARC'21] S. Varrette, E. Kieffer, F. Pinel, E. Krishnasamy, S. Peter, H. Cartiaux, and X. Besseron. *RESIF 3.0: Toward a Flexible & Automated Management of User Software Environment on HPC facility.* In Practice & Experience in Advanced Research Computing (PEARC'21), July 2021

UL HPC Bundles and Software Set Versioning

- **[bi-]Yearly** release of the ULHPC Software Set
  - ↪ following on Easybuid release of toolchains
    - ✓ see Component versions in foss and intel toolchains.
    - ✓ ≃ 6 months of validation/import delay after EB release
- Public export of **RESIF 3 code** on Github: ULHPC/sw

Toolchain Component	Software set release <version> 2019a (*deprecated*)	2019b old	2020a prod	2021a* devel
GCCCore	8.2.0	8.3.0	9.3.0	10.3.0
foss	2019a	2019b	2020a	2021a
intel	2019a	2019b	2020a	2021a
binutils	2.31.1	2.32	2.34	2.36
Python	3.7.2 (2.7.15)	3.7.4 (2.7.16)	3.8.2 (2.7.18)	3.9.2
LLVM	8.0.0	9.0.1	10.0.1	11.1.0
OpenMPI	3.1.4	3.1.4	4.0.3	4.1.1
RESIF version	2.0 (*old*)	3.0	3.0	3.1
#Modules:	229	<arch>: 269 gpu: 135	<arch>: 274 gpu: 151	<arch>: n/a gpu: n/a

Bundle Name	Description	Featured applications
ULHPC-<version>	Default global bundle for 'regular' nodes	ULHPC-*-<version> (root bundle)
ULHPC-toolchains-<version>	Toolchains, compilers, debuggers, programming languages, MPI suits, Development tools and libraries	GCCcore, foss, intel, LLVM, OpenMPI, CMake, Go, Java, Julia, Python, Spack...
ULHPC-bd-<version>	Big Data	Apache Spark, Flink, Hadoop...
ULHPC-bio-<version>	Bioinformatics, biology and biomedical	GROMACS, Bowtie2, TopHat, Trinity...
ULHPC-cs-<version>	Computational science, incl. CAE, CFD, Chemistry, Earth Sciences, Physics and Materials Science	ANSYS, OpenFOAM, ABAQUS, NAMD, GDAL, QuantumExpresso, VASP...
ULHPC-dl-<version>	AI / Deep Learning / Machine Learning	TensorFlow, PyTorch, Horovod...
ULHPC-math-<version>	High-level mathematical software and Optimizers	R, MATLAB, CPLEX, GEOS, GMP, Gurobi...
ULHPC-perf-<version>	Performance evaluation / Benchmarks	ArmForge, PAPI, HPL, IOR, Graph500...
ULHPC-tools-<version>	General purpose tools	DMTC, Singularity, gocryptfs...
ULHPC-visu-<version>	Visualization, plotting, documentation & typesetting	OpenCV, ParaView...
ULHPC-gpu-<version>	Specific GPU/CUDA-accelerated software	{foss,intel}cuda, cuDNN, TensorFlow, PyTorch, GROMACS...

Uni.lu Data Center

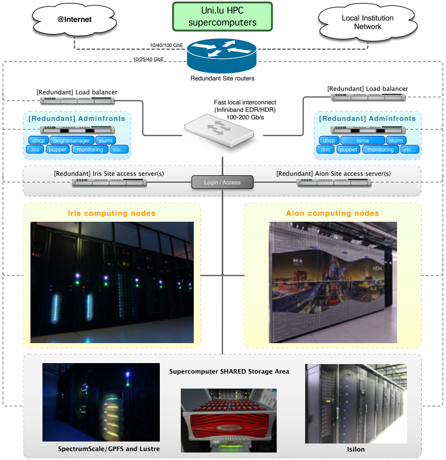
Belval Campus

Centre De Calcul (CDC)

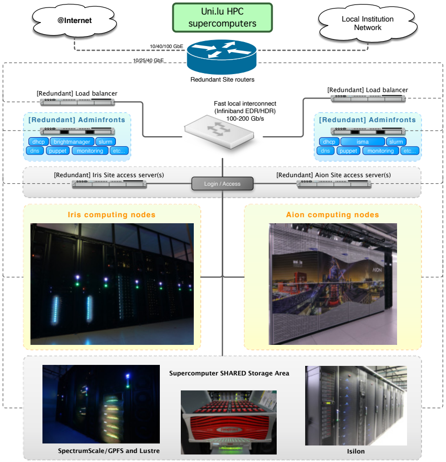
- Power generation station for HPC floor:
  - ↪ up to **3 MW of electrical power**
  - ↪ **2.4 MW of cold water** at a 12-18°C regime
    - ✓ used for traditional Airflow with In-Row cooling.
  - ↪ Separate hot water circuit (between 30 and 40°C)
    - ✓ used for Direct Liquid Cooling (DLC): `aion`

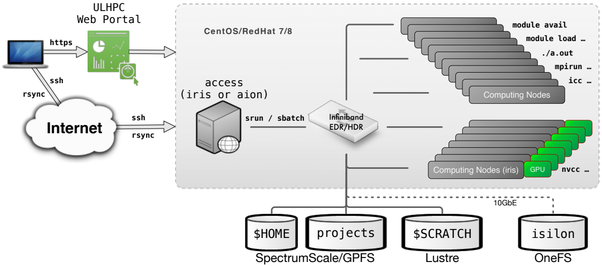
Location	Cooling	Usage
CDC S-02-001	Airflow	**Future extension**
CDC S-02-002	Airflow	**Future extension**
CDC S-02-003	DLC	**Future extension** - High Density/Energy efficient HPC
CDC S-02-004	**DLC**	**High Density/Energy efficient HPC**: `aion`
CDC S-02-005	**Airflow**	**Storage / Traditional HPC**: `iris` and common equipment

UL HPC Supercomputers: General Architecture



UL HPC Supercomputers: General Architecture



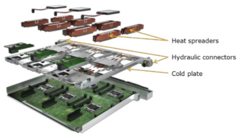


UL HPC Supercomputers: iris cluster

hpc-docs.uni.lu/systems/iris/

- **Dell/Intel** supercomputer *Air-flow cooling*
  - ↪ 196 compute nodes, **5824 cores**, 52.2 TB RAM
  - ↪ R_{peak}: **1,07 PetaFlop/s**
- Fast InfiniBand (IB) EDR network
  - ↪ **Fat-Tree** Topology blocking factor 1:1.5

Rack ID	Purpose	Description
D02	Network	Interconnect equipment
D04	Management	Management servers, Interconnect
D05	Compute	iris-[001-056], interconnect
D07	Compute	iris-[057-112], interconnect
D09	Compute	iris-[113-168], interconnect
D11	Compute	iris-[169-177,191-193](gpu), iris-[187-188](bigmem)
D12	Compute	iris-[178-186,194-196](gpu), iris-[189-190](bigmem)



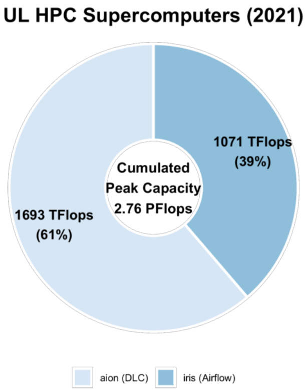
UL HPC Supercomputers: aion cluster

hpc-docs.uni.lu/systems/aion/

- **Atos/AMD** supercomputer, DLC cooling
  - ↪ 4 BullSequana XH2000 adjacent racks
  - ↪ 318 compute nodes, **40704 cores**, 81.4 TB RAM
  - ↪ R_{peak}: **1,693 PetaFLOP/s**
- Fast InfiniBand (IB) HDR network
  - ↪ **Fat-Tree** Topology — blocking factor 1:2

	Rack 1	Rack 2	Rack 3	Rack 4	TOTAL
Weight [kg]	1872,4	1830,2	1830,2	1824,2	**7357 kg**
#X2410 Rome Blade	28	26	26	26	**106**
#Compute Nodes	84	78	78	78	**318**
#Compute Cores	10752	9984	9984	9984	**40704**
R_{peak} **[TFlops]**	447,28 TF	415,33 TF	415,33 TF	415,33 TF	**1693.29 TF**

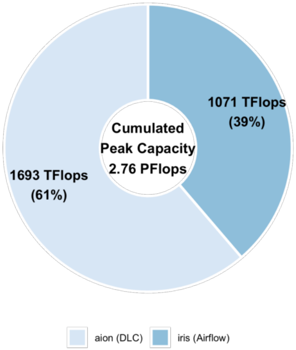
UL HPC Computing Capacity



hpc-docs.uni.lu/systems/

UL HPC Computing Capacity

UL HPC Supercomputers (2021)



`hpc-docs.uni.lu/systems/`

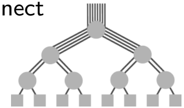
	#N	#C	R_{peak}
Uni.lu HPC TOTAL (production supercomputers):	**514**	**46528**	**2764.3 TFlops**
			(incl. 748.8 GPU TFlops)

Cluster	Date	Vendor	Processors Type and Model		#N	#C	R_{peak}
aion	2021	Atos	AMD EPYC 7H12 @2.6 GHz	2 × 64c, 256GB	318	40704	1693,29 TFlops
				`aion` **TOTAL:**	**318**	**40704**	**1693.3 TFlops**
iris	2017	Dell	Intel Xeon E5-2680 v4@2.4GHz	2 × 14C,128GB	108	3024	116,12 TFlops
	2018	Dell	Intel Xeon Gold 6132 @ 2.6 GHz	2 × 14C,128GB	60	1680	123,65 TFlops
	2018	Dell	Intel Xeon Gold 6132 @ 2.6 GHz	2 × 14C,768GB	24	672	49,45 TFlops
	2019		*Per node*: 4x NVIDIA Tesla V100 SXM2 16/32GB		96 GPUs	491520	748,8 GPU TFlops
	2018	Dell	Intel Xeon Platinum 8180M @ 2.5 GHz	4 × 28C,3072GB	4	448	32.97 TFlops
				`iris` **TOTAL:**	**196**	**5824**	**322.20 TFlops**
					96 GPUs	491520	**+748.8 GPU Tflops**

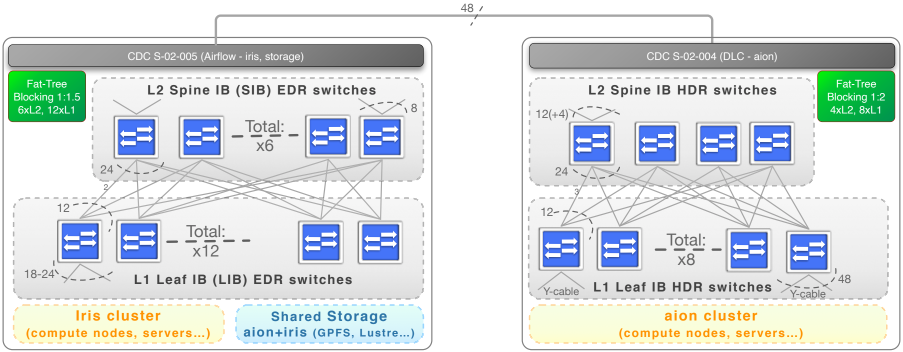
Interconnect Networks (Infiniband and Ethernet)

- HPC interconnect technologies nowadays divided into three categories
  1. **Ethernet**: dominant interconnect standard yet underlying protocol has inherent limitations
     - ✓ preventing low-latency deployments expected in real HPC environment
  2. **InfiniBand**: predominant interconnect technology in the HPC market
  3. Vendor specific interconnects: Cray/HPC Slingshot, Intel Omni-Path, Bull BXI...

- **On ULHPC Supercomputers**:
  - ↪ **InfiniBand (IB)** in a **Fat-Tree** Topology as *Ultra-Fast* local interconnect
    - ✓ `iris`: IB **EDR** Fabric
    - ✓ `aion`: IB **HDR100** Fabric
  - ↪ Complementary **Ethernet network**
    - ✓ Consolidated as a 2-layers topology (Gateway / Switching Layers)

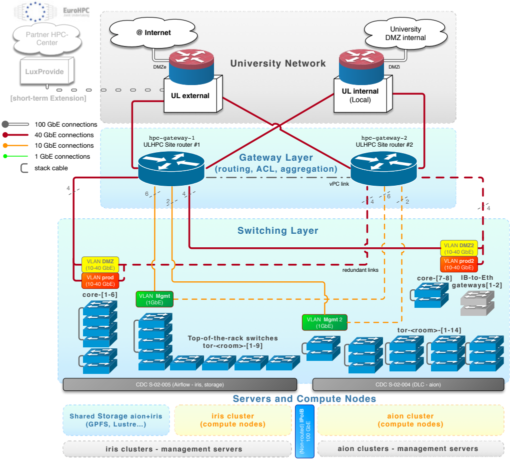


Fast Local Infiniband Interconnect Network



hpc-docs.uni.lu/interconnect/ib/

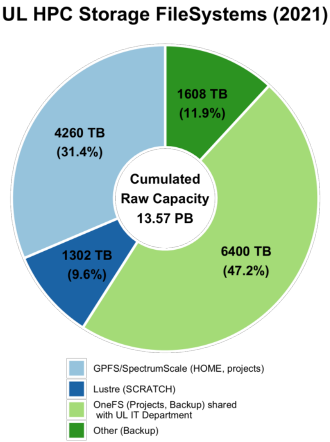
Ethernet Network



hpc-docs.uni.lu/interconnect/ethernet/

- Flexibility of Ethernet-based networks still required
- **2-layers** topology
  - ↪ Upper level: **Gateway Layer**
    - ✓ routing, switching features, network isolation and filtering (ACL) rules
    - ✓ meant to interconnect only switches.
    - ✓ allows to interface the University network (LAN/WAN)
  - ↪ bottom level: **Switching Layer**
    - ✓ composed by [stacked] core switches as well as the TOR (Top-the-rack) switches,
    - ✓ meant to interface HPC servers and compute nodes

UL HPC Storage Systems



UL HPC Software Stack

Operating System: Linux CentOS/Redhat

- **User Single Sign-on**: Redhat IdM/IPA
- **Remote connection & data transfer**: SSH/SFTP
  - ↪ **User Portal**: Open OnDemand
- **Scheduler/Resource management**: Slurm
- **(Automatic) Server / Compute Node Deployment**:
  - ↪ BlueBanquise, Bright Cluster Manager, Ansible, Puppet and Kadeploy
- **Virtualization and Container Framework**: KVM, Singularity
- **Platform Monitoring** (User level): Ganglia, SlurmWeb, OpenOndemand...
- **ISV software**: ABAQUS, ANSYS, MATLAB, Mathematica, Gurobi Optimizer, Intel Cluster Studio XE, ARM Forge & Perf. Report, Stata, ...

Summary

Notable Events/Activities (2021) - RFP 190027

- **Main activity**: release of a new HPC supercomputer Aion (European Tender RFP190027)

RFP 190027 Release

2019

- **Sept, 2019**: Official Public release of Aion cluster tenders on TED European tender and PMP Portal (Portail des Marchés Publics)
  - TED European tender: TED72/2019-608787
  - PMP Reference: 1901442

- **July-August 2019**: RFP 190027 Preparation
  - Tender description (***116 pages***)
  - Criteria Weighting: ~***550 evaluated criterias***
  - Budget: **3.5 M€**


RFP 190027 - TENDER FOR [THE ACQUISITION OF A COMPLEMENTARY HIGH PERFORMANCE COMPUTING (HPC) CLUSTER 'AION']

Notable Events/Activities (2021) - RFP 190027

- **Main activity**: release of a new HPC supercomputer Aion (European Tender RFP190027)

RFP 190027 Release

2019

- **Sept, 2019**: Official Public release of Aion cluster tenders on TED European tender and PMP Portal (Portail des Marchés Publics)
  - TED European tender: TED72/2019-608787
  - PMP Reference: 1901442

Attribution & Kickoff

2020

- **Dec, 2019**: RFP awarded to Atos (MOM)
  - ***Lot 1***: DLC supercomputer aion
  - ***Lot 2***: iris storage extension and consolidation
  - ***Lot 3***: Infiniband/Ethernet interconnect consolidation
- **Jan, 2020**: Project Kickoff
  - Planned completion: Q3 2020

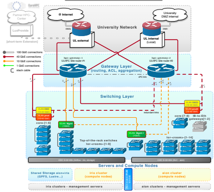
- **July-August 2019**: RFP 190027 Preparation
  - Tender description (***116 pages***)
  - Criteria Weighting: ~***550 evaluated criterias***
  - Budget: **3.5 M€**


RFP 190027 - TENDER FOR THE ACQUISITION OF A COMPLEMENTARY HIGH PERFORMANCE COMPUTING (HPC) CLUSTER 'AION'

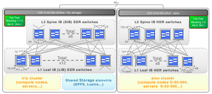
Lot 1: Aion Supercomputer

Lot 2 / Storage extension

Lot 3 / Ethernet



Lot 3 / Infiniband



Notable Events/Activities (2021) - RFP 190027

- **Main activity**: release of a new HPC supercomputer Aion (European Tender RFP190027)

RFP 190027 Release

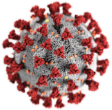
2019

- **Sept, 2019**: Official Public release of Aion cluster tenders on TED European tender and PMP Portal (Portail des Marchés Publics)
  - TED European tender: TED72/2019-608787
  - PMP Reference: 1901442

Attribution & Kickoff

2020

- **Dec, 2019**: RFP awarded to Atos (MOM)
  - ***Lot 1***: DLC supercomputer aion
  - ***Lot 2***: iris storage extension and consolidation
  - ***Lot 3***: Infiniband/Ethernet interconnect consolidation
- **Jan, 2020**: Project Kickoff
  - Planned completion: ~~Q3 2020~~ (COVID crisis)



COVID-19 Pandemic and worldwide crisis

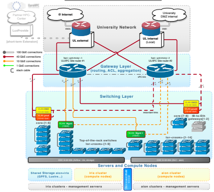
- **July-August 2019**: RFP 190027 Preparation
  - Tender description (***116 pages***)
  - Criteria Weighting: ~***550 evaluated criterias***
  - Budget: **3.5 M€**

RFP 190027 - TENDER FOR [THE ACQUISITION OF A COMPLEMENTARY HIGH PERFORMANCE COMPUTING (HPC) CLUSTER 'AION']

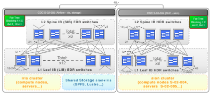
Lot 1: Aion Supercomputer

Lot 2 / Storage extension

Lot 3 / Ethernet



Lot 3 / Infiniband



Notable Events/Activities (2021) - RFP 190027

- **Main activity**: release of a new HPC supercomputer Aion (European Tender RFP190027)

Delivery & Install

2021

- **Jan-Feb, 2021**: Start of installation Lot 1
- **Mar, 2021**: Installation Lot 2 GPFS/SpectrumScale: GS7990 expansion installation, setup, and integration ; Lustre upgrade
  Slurm upgrade

- **Oct, 2020**: CDC S02-004 (Power & Hydraulic Work) completed
- **Dec, 2020**: Partial Delivery of intermediate equipment (Servers, DDN part)





Notable Events/Activities (2021) - RFP 190027

- **Main activity**: release of a new HPC supercomputer Aion (European Tender RFP190027)

2021

Delivery & Install

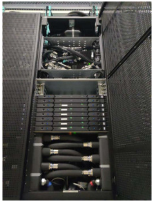
- **Jan-Feb, 2021**: Start of installation Lot 1
- **Mar, 2021**: Installation Lot 2 GPFS/SpectrumScale: GS7990 expansion installation, setup, and integration ; Lustre upgrade Slurm upgrade

Setup & Evaluation

- **Apr, 2021**: Start of installation Lot 3
- **May, 2021**: Merge Iris/Aion Infiniband Island
- **Jun-Jul, 2021**: Stabilization and Performance Evaluation

- **Oct, 2020**: CDC S02-004 (Power & Hydraulic Work) completed
- **Dec, 2020**: Partial Delivery of intermediate equipment (Servers, DDN part)







Notable Events/Activities (2021) - RFP 190027

- **Main activity**: release of a new HPC supercomputer Aion (European Tender RFP190027)

2021

Delivery & Install

- **Jan-Feb, 2021**: Start of installation Lot 1
- **Mar, 2021**: Installation Lot 2 GPFS/SpectrumScale: GS7990 expansion installation, setup, and integration ; Lustre upgrade
  Slurm upgrade

Setup & Evaluation

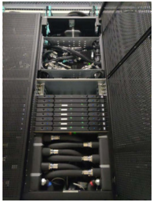
- **Apr, 2021**: Start of installation Lot 3
- **May, 2021**: Merge Iris/Aion Infiniband Island
- **Jun-Jul, 2021**: Stabilization and Performance Evaluation

Beta Opening

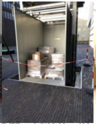
- **July 30, 2021**: Aion supercomputer opened for beta testers
  Aion Pre-acceptance
  Benchmark results submissions (HPL, HPCG, Green500, Graph500, IOR, IO500...)

- **Oct, 2020**: CDC S02-004 (Power & Hydraulic Work) completed
- **Dec, 2020**: Partial Delivery of intermediate equipment (Servers, DDN part)









Notable Events/Activities (2021) - RFP 190027

- **Main activity**: release of a new HPC supercomputer Aion (European Tender RFP190027)

2021

Delivery & Install

- **Jan-Feb, 2021**: Start of installation Lot 1
- **Mar, 2021**: Installation Lot 2 GPFS/SpectrumScale: GS7990 expansion installation, setup, and integration ; Lustre upgrade
  Slurm upgrade

Setup & Evaluation

- **Apr, 2021**: Start of installation Lot 3
- **May, 2021**: Merge Iris/Aion Infiniband Island
- **Jun-Jul, 2021**: Stabilization and Performance Evaluation

Beta Opening

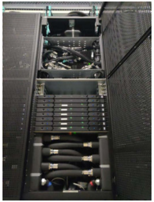
- **July 30, 2021**: Aion supercomputer opened for beta testers
  Aion Pre-acceptance
  Benchmark results submissions (HPL, HPCG, Green500, Graph500, IOR, IO500...)

Production Release

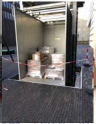
- **Oct 3, 2021**: Aion supercomputer opened
- **Nov 10, 2021**: Official inauguration
- **Nov 12-19, 2021**: 11th UL HPC School 2021
  Practical sessions on Aion
  3 keynotes, 12 practical sessions

- **Oct, 2020**: CDC S02-004 (Power & Hydraulic Work) completed
- **Dec, 2020**: Partial Delivery of intermediate equipment (Servers, DDN part)









Aion would NOT be here Today without. . .

Rectorate & Financial Support


Prof. S. Pallage


Prof. J. Kreisel

Uni.lu HPC for research (86.66%)


Prof. P. Bouvry


Prof. J.M. Schlenker

FSTM (3.33%)


Prof. R. Balling

LCSB (3.33%)

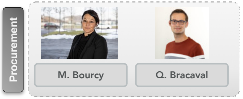

Prof. B. Ottersten

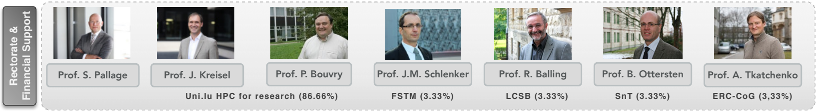
SnT (3.33%)


Prof. A. Tkatchenko

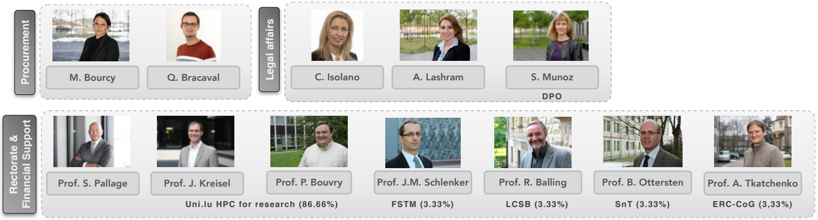
ERC-CoG (3,33%)

Aion would NOT be here Today without. . .



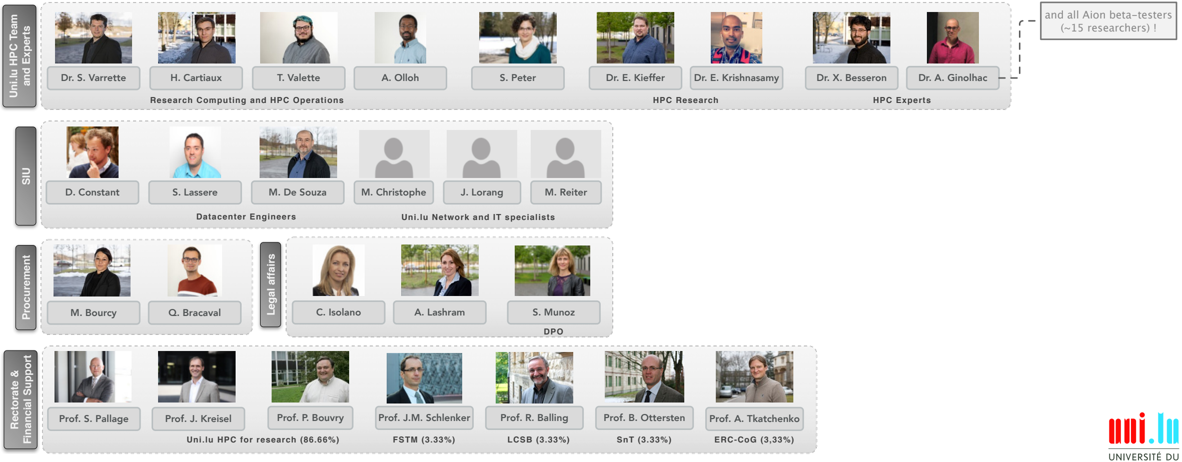


Aion would NOT be here Today without. . .

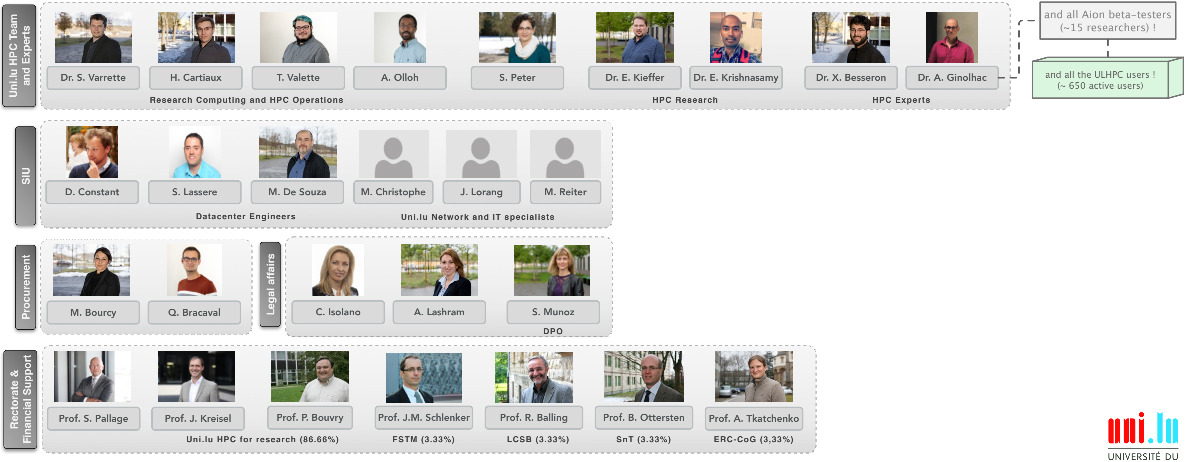


Aion would NOT be here Today without. . .

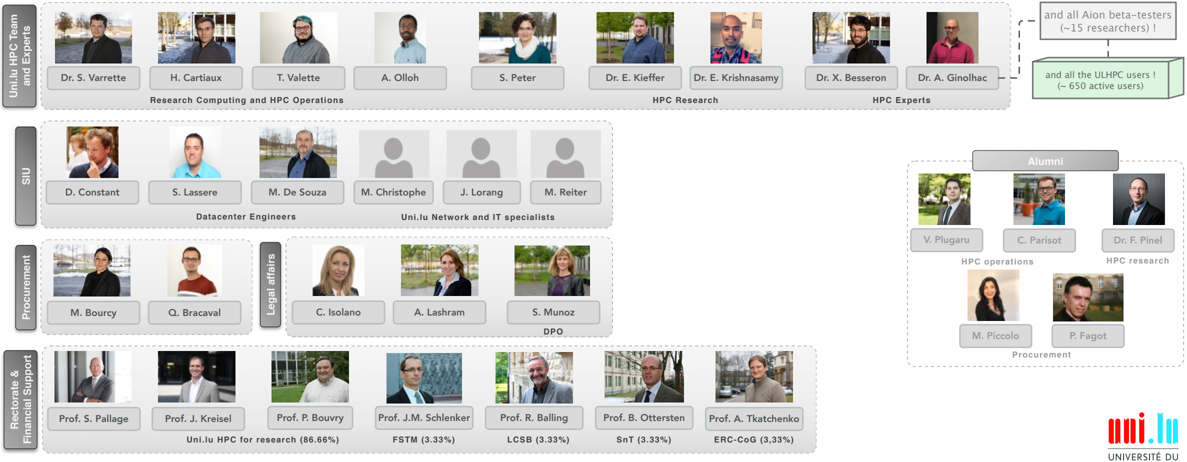
Aion would NOT be here Today without. . .



Aion would NOT be here Today without. . .



Aion would NOT be here Today without. . .



Notable Events/Activities (2021) - Websites

- **Jan 2021**: Official Release for the ULHPC Websites 2.0
  - ↪ Main website used to rely on deprecated technology, with mixed content
    - ✓ Octopress 2 / Jekyll 2.x – my first commit dates back from Dec 20, 2012
    - ✓ content to be fully reviewed to align with new standards and HPC technologies
    - ✓ Still many thanks to the wide range of contributors between Dec 20, 2012 and Dec 31, 2020

Statistics based on git-stats on the OLD hpc website source repository commits

Sebastien Varrette	(36.28%)
Hyacinthe Cartiaux	(25.57%)
Valentin Plugaru	(23.15%)
Fotis Georgatos	(7.81%)
Clement Parisot	(3.52%)
Sarah Diehl / Peter	(1.79%)
Clément Parisot	(1.04%)
Teddy Valette	(0.35%)
Emmanuel Kieffer	(0.21%)
Ezhilmathi Krishnasamy	(0.14%)
Maxime Schmitt	(0.14%)

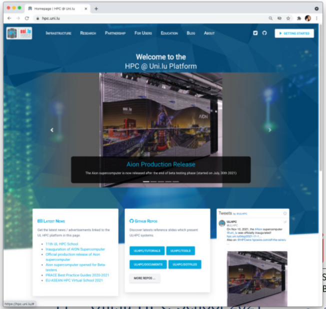
hpc.uni.lu

Notable Events/Activities (2021) - Websites

- **Jan 2021**: Official Release for the ULHPC Websites 2.0
  - ↪ Main website used to rely on deprecated technology, with mixed content
    - ✓ Octopress 2 / Jekyll 2.x – my first commit dates back from Dec 20, 2012
    - ✓ content to be fully reviewed to align with new standards and HPC technologies
    - ✓ Still many thanks to the wide range of contributors between Dec 20, 2012 and Dec 31, 2020

- New website (VueJS-based) bootstrapped by Teddy Valette
  - ↪ now used by the HPC operation team

hpc.uni.lu

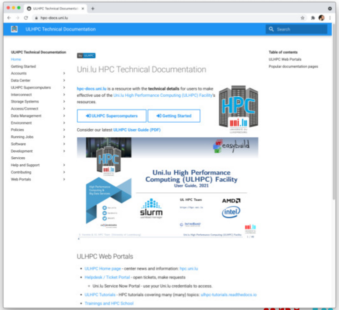


Notable Events/Activities (2021) - Websites

- **Jan 2021**: Official release of the ULHPC Technical Documentation
  - ↪ bootstrapped from mkdocs-material theme and the PyMdown Extensions
    - ✓ inspired from the **excellent** NERSC Technical documentation
  - ↪ now inspiring other HPC sites *(credits/links to ULHPC may come one day)*

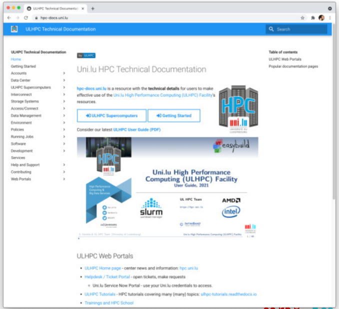
ULHPC Technical Docs

`hpc-docs.uni.lu`



Notable Events/Activities (2021) - Websites

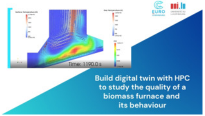
- **Jan 2021**: Official release of the ULHPC Technical Documentation
  - ↪ bootstrapped from mkdocs-material theme and the PyMdown Extensions
    - ✓ inspired from the **excellent** NERSC Technical documentation
  - ↪ now inspiring other HPC sites *(credits/links to ULHPC may come one day)*



ULHPC Technical Docs

`hpc-docs.uni.lu`

- **Jul 2021**: Uni.lu HPC Use Case Demos (EuroCC Luxembourg)

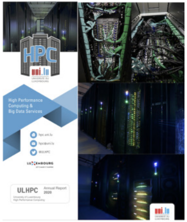


Uni.lu HPC Demonstrators

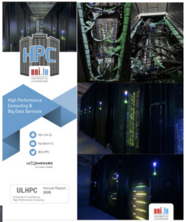
`hpc-demos.uni.lu`

Notable Events/Activities (2021) - Communication

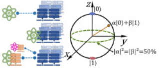
- **Mar-May, 2021**: Annual Reports
  - ↪ ULHPC Annual Report 2020 (Apr. 2021) [1]
  - ↪ SnT Annual report 2020:
    - ✓ "Supercomputer's Crucial Role In Fighting COVID-19" (May, 2021)
  - ↪ FSTM Highlights 2019-2020



Notable Events/Activities (2021) - Communication



- **Mar-May, 2021**: Annual Reports
  - ↪ ULHPC Annual Report 2020 (Apr. 2021) [1]
  - ↪ SnT Annual report 2020:
    - ✓ "Supercomputer's Crucial Role In Fighting COVID-19" (May, 2021)
  - ↪ FSTM Highlights 2019-2020
- **PRACE** Best Practice Guides and White Papers
  - ↪ [2] Best Practice Guide on Modern Accelerators
  - ↪ [3] Quantum Computing: A European Perspective
  - ↪ [4] Security in an evolving European HPC Ecosystem



[1] S. Varrette, "*ULHPC Annual Report 2020*", Apr. 2021 - (pdf)
[2] J. Bispo, J. G. Barbosa, P. F. Silva, C. Morales, M. Myllykoski, P. Ojeda-May, M. Bialczak, M. Uchronski, A. Wlodarczyk, P. Wauligmann, **E. Krishnasamy, S. Varrette**, and S. Lürs, "*PRACE Best Practice Guide 2021: Modern Accelerators*" PRACE aisbl, Jun. 2021 - (pdf)
[3] M.P. Johansson, **E. Krishnasamy**, N. Meyer and C. Piechurski, "*Quantum Computing: A European Perspective*", PRACE aisbl, Oct. 2021 (pdf)
[4] D. Pleiter, **S. Varrette, E. Krishnasamy**, E. Ozdemir, M. Pilc, "*Security in an evolving European HPC Ecosystem*", PRACE aisbl, Dec 2021

Notable Events/Activities (2021)

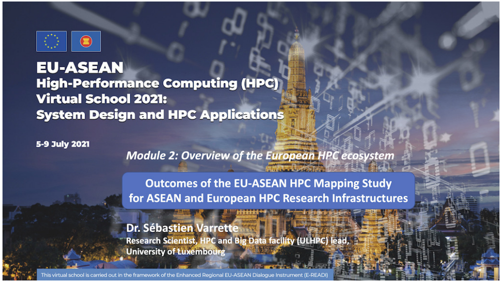
- **Jun 2021**: Inauguration of Luxembourg's ~~first~~ supercomputer MeluXina
  - ↪ early access released (Nov)
  - ↪ EuroHPC/PRACE Access Calls for researchers also released (Nov 2021) - submission portal
    - ✓ you need to demonstrate code scalability - use ULHPC Supercomputers for that.

Notable Events/Activities (2021) - Training

- **Mar 2021**: NVAITC Data Science Webinar

Notable Events/Activities (2021) - Training

- **Mar 2021**: NVAITC Data Science Webinar
- **Jul 2021**:
  - ↪ Elixir Workshop: Reproducible Analysis workflow with Snakemake/Conda
  - ↪ EU-ASEAN HPC Virtual School 2021: System Design & HPC Applications



Notable Events/Activities (2021) - Training

- **Mar 2021**: NVAITC Data Science Webinar
- **Jul 2021**:
  - ↪ Elixir Workshop: Reproducible Analysis workflow with Snakemake/Conda
  - ↪ EU-ASEAN HPC Virtual School 2021: System Design & HPC Applications
- **Aug-Sep, 2021**: PRACE Summer of HPC 2021

Notable Events/Activities (2021) - Training

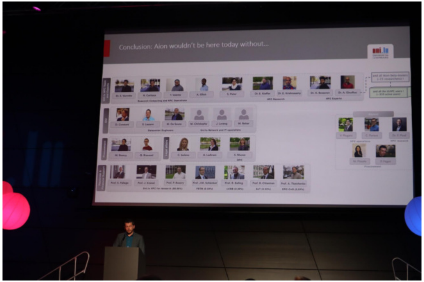
- **Mar 2021**: NVAITC Data Science Webinar
- **Jul 2021**:
  - ↪ Elixir Workshop: Reproducible Analysis workflow with Snakemake/Conda
  - ↪ EU-ASEAN HPC Virtual School 2021: System Design & HPC Applications
- **Aug-Sep, 2021**: PRACE Summer of HPC 2021
- **Oct 2021**: 1st European Master's programme in HPC, led by University of Luxembourg
  - ↪ EUMaster4HPC Proposals – call EuroHPC-2020-03 "Training and Education on HPC"
  - ↪ 7 M€ budget (cumul. for the consortium)
  - ↪ Planned start Autumn 2022 with existing Masters
  - ↪ Autumn 2023: New master based on EU curriculum

Notable Events/Activities (2021) - Training

- **Mar 2021**: NVAITC Data Science Webinar
- **Jul 2021**:
  - ↪ Elixir Workshop: Reproducible Analysis workflow with Snakemake/Conda
  - ↪ EU-ASEAN HPC Virtual School 2021: System Design & HPC Applications
- **Aug-Sep, 2021**: PRACE Summer of HPC 2021
- **Oct 2021**: 1st European Master's programme in HPC, led by University of Luxembourg
  - ↪ EUMaster4HPC Proposals – call EuroHPC-2020-03 "Training and Education on HPC"
  - ↪ 7 M€ budget (cumul. for the consortium)
  - ↪ Planned start Autumn 2022 with existing Masters
  - ↪ Autumn 2023: New master based on EU curriculum
- **Nov 2021**: 11^{th} UL HPC School 2021
  - ↪ 6 days, 3 keynotes, 12 practical sessions

Notable Events/Activities (2021) - Event

- **Nov 2021**: Aion Inauguration *(following production release Oct 3, 2021)*
  - ↪ "*The largest supercomputer ever deployed at the university*"
  - ↪ Event organized by VRR Prof. Jens Kreisel



Summary

General Guidelines

Acceptable Use Policy (AUP) 2.0

`Uni.lu-HPC-Facilities_Acceptable-Use-Policy_v2.0.pdf`

- **UL HPC is a shared** *(and expansive)* **facility: you must practice good citizenship**
  - ↪ **Users are accountable for their actions**
    - ✓ Users are allowed **one account per person** - **user credentials sharing is strictly prohibited**
    - ✓ Use of UL HPC computing resources for personal activities is prohibited
    - ✓ limit activities that may impact the system for other users.
  - ↪ **Do not abuse the shared filesystems**
    - ✓ Avoid too many simultaneous file transfers
    - ✓ regularly clean your directories from useless files
  - ↪ **Do not run programs or I/O bound processes on the login nodes**
  - ↪ Plan large scale experiments during night-time or week-ends
- Resource allocation is done on a **fair-share** principle, with **no guarantee** of being satisfied

General Guidelines

Acceptable Use Policy (AUP) 2.0

- **Data Use / GDPR**
  - ↪ **You** are responsible to ensure the appropriate level of protection, backup & integrity checks
    - ✓ Data Authors/generators/owners are responsible for its correct categorization as sensitive/non-sensitive
    - ✓ Owners of sensitive information are responsible for its secure handling, transmission, processing, storage, and disposal on the UL HPC systems
    - ✓ Data Protection inquiries can be directed to the Uni.lu Data Protection Officer
  - ↪ We make **no guarantee** against loss of data
- We provide [project] **usage report** to user/PI **on-demand** and *(by default)* on a **yearly basis**

- For **ALL** publications having results produced using the UL HPC Facility
  - ↪ **Acknowledge** the UL HPC facility and **cite** reference ULHPC article
    - ✓ using official banner
  - ↪ Tag your publication upon registration on ORBiLu.

ULHPC Websites 2.0 and Documentation

Main Website

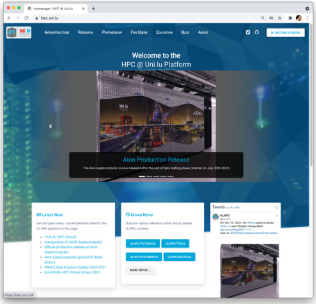
`hpc.uni.lu`

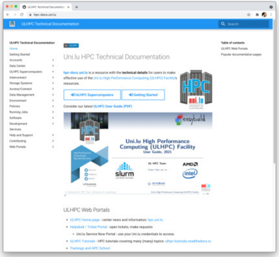
ULHPC Technical Docs

`hpc-docs.uni.lu`

ULHPC HelpDesk

`hpc.uni.lu/support`



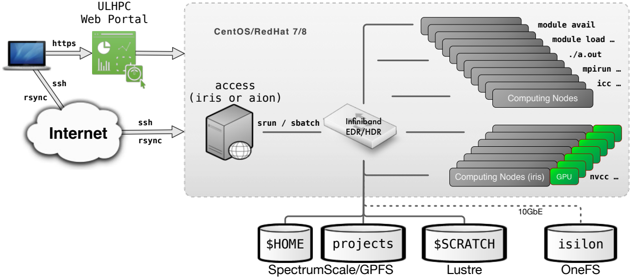


ULHPC Tutorials

`ulhpc-tutorials.rtfd.io`

- **Fallback Support**:
  - ↪ `hpc-team@uni.lu`
  - ↪ ULHPC Community: `hpc-users@uni.lu`
    - ✓ moderated

Compute Nodes / Storage Environment



- **Storage usage**: `df-ulhpc [-i]`
  - ↪ `$HOME`: regular backup policy
  - ↪ `$SCRATCH` **NO** backup & purged
    - ✓ 60 days retention policy
  - ↪ Project quotas attached to group
    - ✓ **not** *(default)* `clusterusers` group
    - ✓ Commands writing in project dir: `sg <group> -c "<command>"`
- **LMod/Environment modules**
  - ↪ **Not** on access, **only** on compute nodes

Directory	FileSystem	Max size	Max #files	Backup
`$HOME` (iris)	GPFS	500 GB	1.000.000	YES
`$SCRATCH`	Lustre	10 TB	1.000.000	**NO**
Project	GPFS	*per request*		PARTIALLY (/backup subdir)
Project	OneFS	*per request*		PARTIALLY

Reporting Problems

- **First checks**
  1. My issue is probably documented — https://hpc-docs.uni.lu
  2. An event is on-going: **check ULHPC Live status page** — https://hpc.uni.lu/live-status/motd/
     - ✓ Planned maintenance are announced *at least* 2 weeks in advance
     - ✓ The proper SSH banner is displayed during **planned** downtime
  3. check the state of your nodes
     - ✓ `{ scontrol show job <jobid> | sjoin <jobid>}; htop` — *on active jobs*
     - ✓ `{ slist <jobid> | sacct [-X] -j <jobid> -l }` — *post-mortem*

Reporting Problems

- **First checks**
  1. My issue is probably documented — https://hpc-docs.uni.lu
  2. An event is on-going: **check ULHPC Live status page** — https://hpc.uni.lu/live-status/motd/
     - ✓ Planned maintenance are announced *at least* 2 weeks in advance
     - ✓ The proper SSH banner is displayed during **planned** downtime
  3. check the state of your nodes
     - ✓ `{ scontrol show job <jobid> | sjoin <jobid>}; htop` — *on active jobs*
     - ✓ `{ slist <jobid> | sacct [-X] -j <jobid> -l }` — *post-mortem*

- **ONLY NOW**, consider the following depending on the severity:
  - ↪ Open an new issue on https://hpc.uni.lu/support — (**preferred**)
    - ✓ Uni.lu Service Now Helpdesk Portal: relies on **Uni.lu (≠ ULHPC) credentials**
  - ↪ Mail (only now) us — `hpc-team@uni.lu`
  - ↪ **Ask the help of other users** — `hpc-users@uni.lu`

Reporting Problems

- **First checks**
  1. My issue is probably documented — https://hpc-docs.uni.lu
  2. An event is on-going: **check ULHPC Live status page** — https://hpc.uni.lu/live-status/motd/
     - ✓ Planned maintenance are announced *at least* 2 weeks in advance
     - ✓ The proper SSH banner is displayed during **planned** downtime
  3. check the state of your nodes
     - ✓ `{ scontrol show job <jobid> | sjoin <jobid>}; htop` — *on active jobs*
     - ✓ `{ slist <jobid> | sacct [-X] -j <jobid> -l }` — *post-mortem*

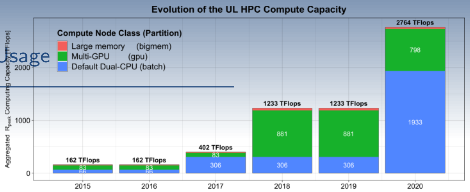
- **ONLY NOW**, consider the following depending on the severity:
  - ↪ Open an new issue on https://hpc.uni.lu/support — **(preferred)**
    - ✓ Uni.lu Service Now Helpdesk Portal: relies on **Uni.lu (≠ ULHPC) credentials**
  - ↪ Mail (only now) us — `hpc-team@uni.lu`
  - ↪ **Ask the help of other users** — `hpc-users@uni.lu`

In all cases: **Carefully describe the problem and the context** — Guidelines

Slurm on ULHPC clusters

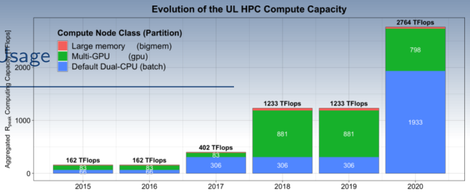
- ULHPC uses Slurm for cluster/resource management and job scheduling
  - ↪ Simple Linux Utility for Resource Management https://slurm.schedmd.com/
  - ↪ Handles submission, scheduling, execution, and monitoring of **jobs**
  - ↪ official documentation, tutorial, FAQ

- **User jobs** have the following key characteristics:
  - ↪ set of requested resources:
    - ✓ number of computing resources: **nodes** (including all their CPUs and cores) or **CPUs** (including all their cores) or **cores**
    - ✓ amount of **memory**: either per node or per CPU
    - ✓ **(wall)time** needed for the users tasks to complete their work
  - ↪ a requested node **partition** (job queue)
  - ↪ a requested **quality of service** (QoS) level which grants users specific accesses
  - ↪ a requested **account** for accounting purposes



Slurm on ULHPC clusters

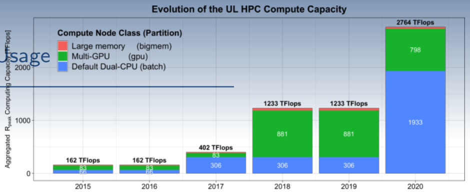
- Predefined **Queues/Partitions** depending on node type
  - ↪ `batch` (Default Dual-CPU nodes) — **Max**: 64 nodes, 2 days walltime
  - ↪ `gpu` (GPU nodes nodes) — **Max**: 4 nodes, 2 days walltime
  - ↪ `bigmem` (Large-Memory nodes) — **Max**: 1 node, 2 days walltime
  - ↪ In addition: `interactive` (for quicks tests) — **Max**: 2 nodes, 2h walltime
    - ✓ for code development, testing, and debugging



Slurm on ULHPC clusters

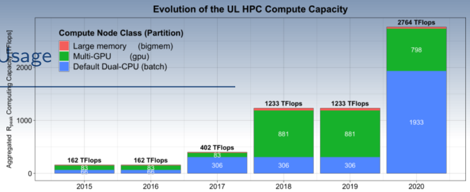
- Predefined **Queues/Partitions** depending on node type
  - ↪ `batch` (Default Dual-CPU nodes) — **Max**: 64 nodes, 2 days walltime
  - ↪ `gpu` (GPU nodes nodes) — **Max**: 4 nodes, 2 days walltime
  - ↪ `bigmem` (Large-Memory nodes) — **Max**: 1 node, 2 days walltime
  - ↪ In addition: `interactive` (for quicks tests) — **Max**: 2 nodes, 2h walltime
    - ✓ for code development, testing, and debugging
- Queue Policy: **cross-partition QOS**, mainly tied to **priority level** (`low` → `urgent`)
  - ↪ `long` QOS with extended Max walltime (`MaxWall`) set to **14 days**
  - ↪ special **preemptible QOS** for best-effort jobs: `besteffort`.

Slurm on ULHPC clusters



- Predefined **Queues/Partitions** depending on node type
  - ↪ `batch` (Default Dual-CPU nodes) — **Max**: 64 nodes, 2 days walltime
  - ↪ `gpu` (GPU nodes nodes) — **Max**: 4 nodes, 2 days walltime
  - ↪ `bigmem` (Large-Memory nodes) — **Max**: 1 node, 2 days walltime
  - ↪ In addition: `interactive` (for quicks tests) — **Max**: 2 nodes, 2h walltime
    - ✓ for code development, testing, and debugging
- Queue Policy: **cross-partition QOS**, mainly tied to **priority level** (`low` → `urgent`)
  - ↪ `long` QOS with extended Max walltime (`MaxWall`) set to **14 days**
  - ↪ special **preemptible QOS** for best-effort jobs: `besteffort`.
- Accounts associated to supervisor (multiple associations possible)
  - ↪ Proper group/user accounting

Slurm on ULHPC clusters



- Predefined **Queues/Partitions** depending on node type
  - ↪ `batch` (Default Dual-CPU nodes) — **Max**: 64 nodes, 2 days walltime
  - ↪ `gpu` (GPU nodes nodes) — **Max**: 4 nodes, 2 days walltime
  - ↪ `bigmem` (Large-Memory nodes) — **Max**: 1 node, 2 days walltime
  - ↪ In addition: `interactive` (for quicks tests) — **Max**: 2 nodes, 2h walltime
    - ✓ for code development, testing, and debugging
- Queue Policy: **cross-partition QOS**, mainly tied to **priority level** (`low` → `urgent`)
  - ↪ `long` QOS with extended Max walltime (`MaxWall`) set to **14 days**
  - ↪ special **preemptible QOS** for best-effort jobs: `besteffort`.
- Accounts associated to supervisor (multiple associations possible)
  - ↪ Proper group/user accounting
- Slurm Multi-cluster configuration between `iris` and `aion`
  - ↪ easily **query** state on remote cluster — `{squeue...} -M, --cluster aion|iris`

Main Slurm Commands: Submit Jobs

```
$> sbatch -p <partition> [--qos <qos>] [-A <account>] [...]  <path/to/launcher.sh>
```

Submitting Jobs

- `sbatch`: Submit batch **launcher script** for later execution **batch/passive mode**
  - ↪ allocate resources (nodes, tasks, partition, etc.)
  - ↪ runs a single **copy** of the batch script on the **first** allocated node

Main Slurm Commands: Submit Jobs

```
$> srun -p <partition> [--qos <qos>] [-A <account>] [...]  --pty bash
```

Submitting Jobs

- `sbatch`: Submit batch **launcher script** for later execution **batch/passive mode**
  - ↪ allocate resources (nodes, tasks, partition, etc.)
  - ↪ runs a single **copy** of the batch script on the **first** allocated node
- `srun`: initiate parallel **job steps within a job** **OR** **start an interactive job**
  - ↪ allocate resources (number of nodes, tasks, partition, constraints, etc.)
  - ↪ launch a job that will execute on them.

Main Slurm Commands: Submit Jobs

```
$> salloc -p <partition> [--qos <qos>] [-A <account>] [...]  <command>
```

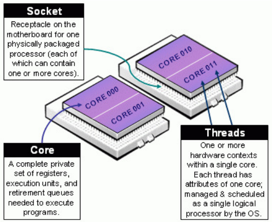
Submitting Jobs

- `sbatch`: Submit batch **launcher script** for later execution **batch/passive mode**
  - ↪ allocate resources (nodes, tasks, partition, etc.)
  - ↪ runs a single **copy** of the batch script on the **first** allocated node
- `srun`: initiate parallel **job steps within a job** **OR** **start an interactive job**
  - ↪ allocate resources (number of nodes, tasks, partition, constraints, etc.)
  - ↪ launch a job that will execute on them.

- `salloc`: : request interactive jobs/allocations
  - ↪ allocate resources (nodes, tasks, partition, etc.), either run a command or start a shell.

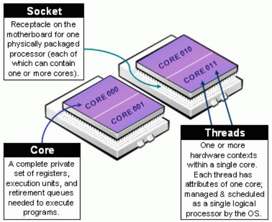
Specific Resource Allocation

- **Beware of Slurm terminology** in Multicore Architecture!
  - ↪ Slurm Node = Physical node `-N <#nodes>`
    - ✓ **Advice**: explicit number of expected tasks **per node** `--ntasks-per-node <n>`



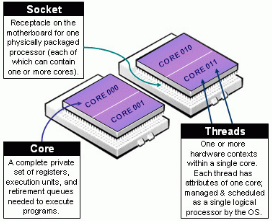
Specific Resource Allocation

- **Beware of Slurm terminology** in Multicore Architecture!
  - ↪ Slurm Node = Physical node — `-N <#nodes>`
    - ✓ **Advice**: explicit number of expected tasks **per node** — `--ntasks-per-node <n>`
  - ↪ Slurm Socket = Physical Socket/**CPU**/Processor — `--ntasks-per-socket <n>`



Specific Resource Allocation

- **Beware of Slurm terminology** in Multicore Architecture!
  - ↪ Slurm Node = Physical node `-N <#nodes>`
    - ✓ **Advice**: explicit number of expected tasks **per node** `--ntasks-per-node <n>`
  - ↪ Slurm Socket = Physical Socket/**CPU**/Processor `--ntasks-per-socket <n>`
  - ↪ **Slurm CPU** = Physical **Core** `-c <#threads>`
    - ✓ Hyper-Threading (HT) Technology is **disabled** on all the compute nodes
    - ✓ #cores = #threads `-c <N>` ⟶ `OMP_NUM_THREADS=${SLURM_CPUS_PER_TASK}`
    - ✓ Total number of tasks: `${SLURM_NTASKS}` ⟶ `srun -n ${SLURM_NTASKS} [...]`



Specific Resource Allocation

- **Beware of Slurm terminology** in Multicore Architecture!
  - ↪ Slurm Node = Physical node `-N <#nodes>`
    - ✓ **Advice**: explicit number of expected tasks **per node** `--ntasks-per-node <n>`
  - ↪ Slurm Socket = Physical Socket/**CPU**/Processor `--ntasks-per-socket <n>`
  - ↪ **Slurm CPU** = Physical **Core** `-c <#threads>`
    - ✓ Hyper-Threading (HT) Technology is **disabled** on all the compute nodes
    - ✓ #cores = #threads `-c <N>` ⟶ `OMP_NUM_THREADS=${SLURM_CPUS_PER_TASK}`
    - ✓ Total number of tasks: `${SLURM_NTASKS}` ⟶ `srun -n ${SLURM_NTASKS} [...]`

- **Important: Always align resource specs with physical NUMA characteristics**
  - ↪ **Ex (AION): 16 cores per socket, 8 sockets ("physical" CPUs) per node** (128c/node)
  - ↪ `[-N <N>] --ntasks-per-node <8n> --ntasks-per-socket <n> -c <thread>`
    - ✓ **Total**: `<N>`×8×`<n>` tasks, each on `<thread>` threads
    - ✓ **Ensure** `<n>`×`<thread>`= 16 on aion
    - ✓ Ex: `-N 2 --ntasks-per-node 32 --ntasks-per-socket 4 -c 4` (**Total**: 64 tasks)

Specific Resource Allocation

- **Beware of Slurm terminology** in Multicore Architecture!
  - ↪ Slurm Node = Physical node — -N <#nodes>
    - ✓ **Advice**: explicit number of expected tasks **per node** — --ntasks-per-node <n>
  - ↪ Slurm Socket = Physical Socket/**CPU**/Processor — --ntasks-per-socket <n>
  - ↪ **Slurm CPU** = Physical **Core** — -c <#threads>
    - ✓ Hyper-Threading (HT) Technology is **disabled** on all the compute nodes
    - ✓ #cores = #threads — -c <N> ⟶ OMP_NUM_THREADS=${SLURM_CPUS_PER_TASK}
    - ✓ Total number of tasks: ${SLURM_NTASKS} ⟶ srun -n ${SLURM_NTASKS} [...]

- **Important: Always align resource specs with physical NUMA characteristics**
  - ↪ **Ex (IRIS): 14 cores per socket, 2 sockets (“physical” CPUs) per node** (28c/node)
  - ↪ [-N <N>] --ntasks-per-node <2n> --ntasks-per-socket <n> -c <thread>
    - ✓ **Total**: <N>×2×<n> tasks, each on <thread> threads
    - ✓ **Ensure** <n>×<thread>= 14 on iris
    - ✓ Ex: -N 2 --ntasks-per-node 4 --ntasks-per-socket 2 -c 7 (**Total**: 8 tasks)

Specific Resource Allocation

- **Beware of Slurm terminology** in Multicore Architecture!
  - ↪ Slurm Node = Physical node `-N <#nodes>`
    - ✓ **Advice**: explicit number of expected tasks **per node** `--ntasks-per-node <n>`
  - ↪ Slurm Socket = Physical Socket/**CPU**/Processor `--ntasks-per-socket <n>`
  - ↪ **Slurm CPU** = Physical **Core** `-c <#threads>`
    - ✓ Hyper-Threading (HT) Technology is **disabled** on all the compute nodes
    - ✓ #cores = #threads `-c <N>` ⟶ `OMP_NUM_THREADS=${SLURM_CPUS_PER_TASK}`
    - ✓ Total number of tasks: `${SLURM_NTASKS}` ⟶ `srun -n ${SLURM_NTASKS} [...]`

Hostname	Node type	#Nodes	#Socket	#Cores	RAM	Features
`aion-[0001-0318]`	Regular	318	8	128	256 GB	`batch,epyc`
`iris-[001-108]`	Regular	108	2	28	128 GB	`batch,broadwell`
`iris-[109-168]`	Regular	60	2	28	128 GB	`batch,skylake`
`iris-[169-186]`	Multi-GPU	18	2	28	768 GB	`gpu,skylake,volta`
`iris-[191-196]`	Multi-GPU	6	2	28	768 GB	`gpu,skylake,volta32`
`iris-[187-190]`	Large Memory	4	4	112	3072 GB	`bigmem,skylake`

- List available features: `sfeatures`

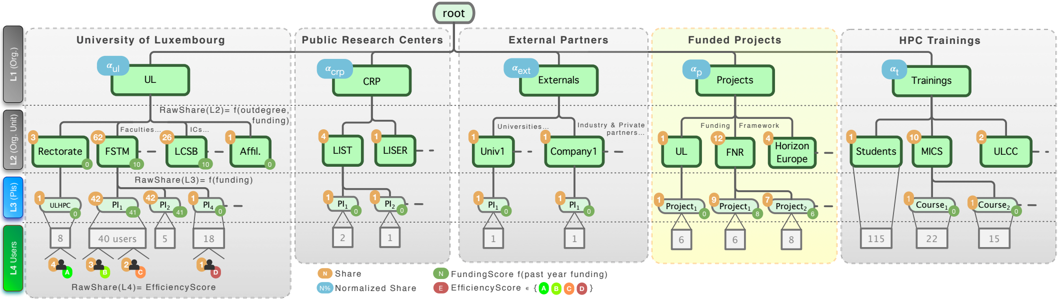
```
# OR: sinfo -o '%20N %.6D %.6c %15F %12P %f'
```

Account Hierarchy 2.0

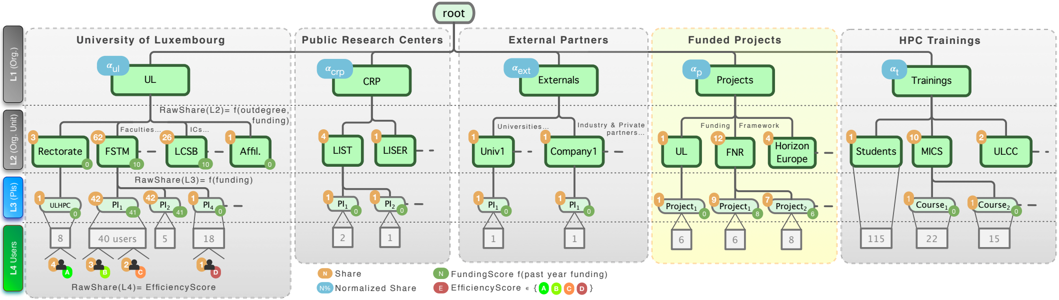
- Every user job runs under a group account (default: PI/line manager)
  - ↪ granting access to specific QOS levels, default raw share for accounts: 1
  - ↪ you **MUST** request a dedicated **Slurm project account** for accounting monitoring
    - ✓ HPC Support (Service Now) / HPC / Storage & projects / New Slurm project account

- **L1**: Organization Level: UL, CRPs, Externals, Projects, Trainings
  - ↪ guarantee 85% of the shares for core UL activities
- **L2**: Organizational Unit (Faculty, ICs, External partner, Funding program...)
  - ↪ Raw share depends on **outdegree** and **funding score**
- **L3**: Principal Investigator (PIs), Projects, Course
  - ↪ Raw share depends on **funding score** (different weight)
  - ↪ Eventually restricted **only** to projects and courses
- **L4**: End User (ULHPC login)
  - ↪ Raw share based on **efficiency score**

Account Hierarchy 2.0

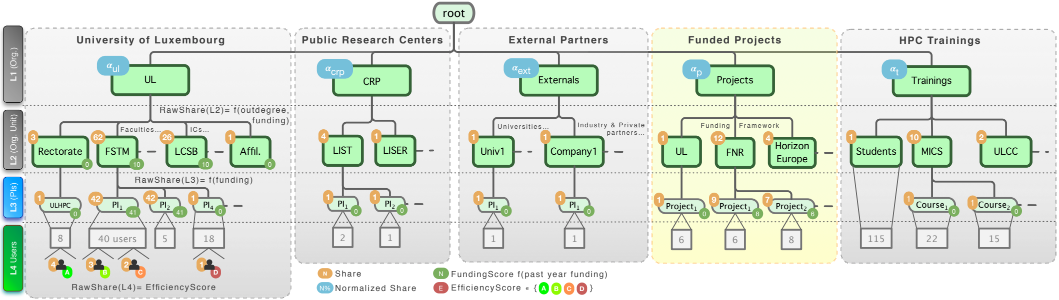


Account Hierarchy 2.0




```
# /!\ ADAPT <name> accordingly
sassoc <name>
```

Account Hierarchy 2.0




```
# Regular submission as End user L4
{srun | sbatch | salloc } [...]
# Accounting associated to **project account** <project> -- required for auditing
{srun | sbatch | salloc } -A <project> [...]
```

Funding Score (L2/L3)

- Associated with an account A belonging to a level L in the hierarchy
  - ↪ yearly updated at the beginning of the year
  - ↪ depreciation based on contribution type

$$\texttt{FundingScore}_L(A) = \left\lfloor \beta_L \frac{Investment_A(Year-1)}{\#months} \right\rfloor$$

- **Ex1**: Exceptional contribution of 120K€ performed in 2021 by a faculty (L2 account A)
  - ↪ depreciation: 12 months (*default*)
  - ↪ **funding score in 2022**: $\left\lfloor \beta_{L_2} \frac{120000}{12} \right\rfloor = \lfloor \beta_{L_2} \times 10000 \rfloor$.
- **Ex2**: let P be a project granted in 2021 to start in 2022 for a duration of 36 months
  - ↪ **budget**: 27K€ allocated for HPC costs
  - ↪ **funding score for the years 2022, 2023 and 2024**: $\left\lfloor \beta_{L_3} \frac{27000}{36} \right\rfloor = \lfloor \beta_{L_3} \times 750 \rfloor$

Efficiency Score (L4)

- **Updated every year based on past jobs efficiency.**
  - ↪ Similar notion of "nutri-score": A(very good - 3), B (good: 2), C (bad, 1), D(very bad - 0)
- Proposed Metric for **user** U: **Average Wall-time Accuracy (WRA)** (higher the better)
  - ↪ Defined for a given time period (past year)

```
sacct -u <U> -X -S <start> -E <end> [...] # --format User,JobID,state,time,elapsed
```

  - ↪ Reduction for N `COMPLETED` jobs:

$$S_{\text{efficiency}}(U, Year) = \frac{1}{N} \sum_{JobID \in (U, Year)} \frac{T_{\text{elapsed}}(JobID)}{T_{\text{asked}}(JobID)}$$

- Default thresholds

Score	Avg. WRA
A	$S_{\text{efficiency}} \geq 75\%$
B	$50\% \leq S_{\text{efficiency}} < 75\%$
C	$25\% \leq S_{\text{efficiency}} < 50\%$
D	$S_{\text{efficiency}} < 25\%$

- **WIP**: integrate other efficiency metrics (CPU, mem, GPU efficiency)

Job Priority, Fairsharing and Fair Tree

- **Fairsharing**: way of ensuring that users get their appropriate portion of a system
  - ↪ **Share**: portion of the system users have been granted.
  - ↪ **Usage**: amount of the system users have actually **used**.
  - ↪ **Fairshare score**: value the system calculates based off of user's usage.
    - ✓ difference between the portion of the computing resource that has been promised and the amount of resources that has been consumed
  - ↪ **Priority score**: priority assigned based off of the user's fairshare score.
- ULHPC Slurm configuration with Multifactor Priority Plugin and **Fair tree** algorithm
  - ↪ rooted plane tree (rooted ordered tree) being created then sorted by Level Fairshare
  - ↪ All users from a higher priority account receive a higher fair share factor than all users from a lower priority account

```
$> sshare -l                                                    # See Level FS
```

ULHPC Job Prioritization Factors

- **Age**: length of time a job has been waiting (PD state) in the queue
- **Fairshare**: difference between the portion of the computing resource that has been promised and the amount of resources that has been consumed
- **Partition**: factor associated with each node partition
  - ↪ Ex: privilege `interactive` over `batch`
- **QOS** A factor associated with each Quality Of Service (`low` ⟶ `urgent`)

```
Job_priority =
    PriorityWeightAge       * age_factor +
    PriorityWeightFairshare * fair-share_factor+
    PriorityWeightPartition * partition_factor +
    PriorityWeightQOS       * QOS_factor +
    - nice_factor
```

ULHPC Job Prioritization Factors

- **Age**: length of time a job has been waiting (PD state) in the queue
- **Fairshare**: difference between the portion of the computing resource that has been promised and the amount of resources that has been consumed
- **Partition**: factor associated with each node partition
  - ↪ Ex: privilege `interactive` over `batch`
- **QOS** A factor associated with each Quality Of Service (`low` ⟶ `urgent`)

```
Job_priority =
    PriorityWeightAge         * age_factor +
    PriorityWeightFairshare * fair-share_factor+
    PriorityWeightPartition * partition_factor +
    PriorityWeightQOS         * QOS_factor +
    - nice_factor
```

```
# Show current weights
sprio -w
# List pending jobs, sorted by jobid
sprio [-n]
# List pending jobs, sorted by  priority
sprio [-n] -S+Y
sprio [-n] | sort -k 3 -n
sprio [-n] -l | sort -k 4 -n
```

hpc-docs.uni.lu/policies/usage-charging/

ULHPC Cost Model

- ULHPC **free of charge for UL staff** for their **internal work and training activities**
  - ↪ Yet data storage extension **above** default capacity is charged
- HPC Resource Allocations for **funded research project** **MUST** comply with ULHPC Cost Model policy approved July 7, 2020 (Rectorate, FNR).
  - ↪ You must prepare your budget plan to support HPC costs within your FNR project proposals
    - ✓ UL HPC Cost Budget Plans for Project Proposals [xlsx] provided for help
- **When charged** ULHPC computing resources are billed as follows:
  - ↪ project PI will receive a usage report
  - ↪ **new** SAP Workflow to facilitate auditing and charging from HPC usage report

Cluster	**Node Type**	**#Cores/node [#GPUs]**	Billing Rate	**Hourly Price [€, HT]**
Aion	Regular	128	200.96	6.03€
Iris	Regular	28	56	1.68€
Iris	GPU	28 [+4 NVidia V100]	256	7.68€
Iris	Large-Mem	112	224	6.72€

Fairshare Factor and Job Billing

- Utilization of the University computational resources is charged in **Service Unit (SU)**
  - ↪ 1 SU $\simeq$ 1 hour on 1 physical processor core on regular computing node
  - ↪ Usage charged **0,03€ per SU (VAT excluded)** (external partners, funded projects etc.)
- A Job is characterized (and thus billed) according to the following elements:
  - ↪ T_{exec}: Execution time (in hours)
  - ↪ N_{Nodes}: number of computing nodes, and **per node**:
    - ✓ N_{cores}: number of CPU cores allocated per node
    - ✓ Mem: memory size allocated per node, in GB
    - ✓ N_{gpus}: number of GPU allocated per node
  - ↪ associated weighted factors $\alpha_{cpu}, \alpha_{mem}, \alpha_{GPU}$ defined as `TRESBillingWeight` in Slurm
    - ✓ account for consumed resources other than just CPUs
    - ✓ taken into account in fairshare factor
    - ✓ α_{cpu}: normalized relative perf. of CPU processor core (reference: skylake 73,6 GFlops/core)
    - ✓ α_{mem}: inverse of the average available memory size per core
    - ✓ α_{GPU}: weight per GPU accelerator

Job Billing and Budget Planning

Number of SU associated to a job

$$N_{\text{Nodes}} \times [\alpha_{cpu} \times N_{\text{cores}} + \alpha_{mem} \times Mem + \alpha_{gpu} \times N_{\text{gpus}}] \times T_{\text{exec}}$$

Cluster	Node Type	Partition	#Cores/node	CPU	α_{cpu}	α_{mem}	α_{GPU}
Iris, Aion	Regular	`interactive`	28/128	n/a	0	0	0
Iris	Regular	`batch`	28	`broadwell`	1.0*	$\frac{1}{4} = 0,25$	0
Iris	Regular	`batch`	28	`skylake`	1.0	$\frac{1}{4} = 0,25$	0
Iris	GPU	`gpu`	28	`skylake`	1.0	$\frac{1}{27}$	50
Iris	Large-Mem	`bigmem`	112	`skylake`	1.0	$\frac{1}{27}$	0
Aion	Regular	`batch`	128	`epyc`	0,57	$\frac{1}{1.75}$	0

```
scontrol show job <jobID> | grep -i billing   # running job
#  Billing rate for completed job <jobID>
sbill <jobID>
```

Job Billing and Budget Planning

Number of SU associated to a job

$$N_{\text{Nodes}} \times [\alpha_{cpu} \times N_{\text{cores}} + \alpha_{mem} \times Mem + \alpha_{gpu} \times N_{\text{gpus}}] \times T_{\text{exec}}$$

Cluster	Node Type	Partition	#Cores/node	CPU	α_{cpu}	α_{mem}	α_{GPU}
Iris, Aion	Regular	`interactive`	28/128	`n/a`	0	0	0
Iris	Regular	`batch`	28	`broadwell`	1.0*	$\frac{1}{4} = 0,25$	0
Iris	Regular	`batch`	28	`skylake`	1.0	$\frac{1}{4} = 0,25$	0
Iris	GPU	`gpu`	28	`skylake`	1.0	$\frac{1}{27}$	50
Iris	Large-Mem	`bigmem`	112	`skylake`	1.0	$\frac{1}{27}$	0
Aion	Regular	`batch`	128	`epyc`	0,57	$\frac{1}{1.75}$	0

- Continuous use of **2 regular skylake nodes** (56 cores, 224GB Memory) on `iris` cluster
  - ↪ 28 cores per node, 4 GigaByte RAM per core i.e., 112GB per node
  - ↪ **For 30 days**: 2 nodes $\times[\alpha_{cpu} \times 28 + \alpha_{mem} \times 4 \times 28 + \alpha_{gpu} \times 0] \times$ 30 days $\times$ 24 hours
    - ✓ Total: $2 \times [(1.0 + \frac{1}{4} \times 4) \times 28] \times 720 = 80640$ SU = **2419,2€ VAT excluded**

Job Billing and Budget Planning

Number of SU associated to a job

$$N_{\text{Nodes}} \times [\alpha_{cpu} \times N_{\text{cores}} + \alpha_{mem} \times Mem + \alpha_{gpu} \times N_{\text{gpus}}] \times T_{\text{exec}}$$

Cluster	Node Type	Partition	#Cores/node	CPU	α_{cpu}	α_{mem}	α_{GPU}
Iris, Aion	Regular	`interactive`	28/128	n/a	0	0	0
Iris	Regular	`batch`	28	`broadwell`	1.0*	$\frac{1}{4} = 0,25$	0
Iris	Regular	`batch`	28	`skylake`	1.0	$\frac{1}{4} = 0,25$	0
Iris	GPU	`gpu`	28	`skylake`	1.0	$\frac{1}{27}$	50
Iris	Large-Mem	`bigmem`	112	`skylake`	1.0	$\frac{1}{27}$	0
Aion	Regular	`batch`	128	`epyc`	0,57	$\frac{1}{1.75}$	0

- Continuous use of **2 regular epyc nodes** (256 cores, 448GB Memory) on `aion` cluster
  - ↪ 128 cores per node, 1,75 GigaByte RAM per core i.e., 224 GB per node
  - ↪ **For 30 days**: 2 nodes$\times[\alpha_{cpu} \times 128 + \alpha_{mem} \times 1.75 \times 128 + \alpha_{gpu} \times 0]\times$ 30 days$\times$ 24 hours
    - ✓ Total: $2 \times [(0.57 + \frac{1}{1.75} \times 1.75) \times 128] \times 720 = 289382{,}4$ SU = **8681,47€ VAT excluded**

Job Billing and Budget Planning

Number of SU associated to a job

$$N_{\text{Nodes}} \times [\alpha_{cpu} \times N_{\text{cores}} + \alpha_{mem} \times Mem + \alpha_{gpu} \times N_{\text{gpus}}] \times T_{\text{exec}}$$

Cluster	Node Type	Partition	#Cores/node	CPU	α_{cpu}	α_{mem}	α_{GPU}
Iris, Aion	Regular	`interactive`	28/128	n/a	0	0	0
Iris	Regular	`batch`	28	`broadwell`	1.0*	$\frac{1}{4} = 0,25$	0
Iris	Regular	`batch`	28	`skylake`	1.0	$\frac{1}{4} = 0,25$	0
Iris	GPU	`gpu`	28	`skylake`	1.0	$\frac{1}{27}$	50
Iris	Large-Mem	`bigmem`	112	`skylake`	1.0	$\frac{1}{27}$	0
Aion	Regular	`batch`	128	`epyc`	0,57	$\frac{1}{1.75}$	0

- Continuous use of **1 GPU nodes** (28 cores, 4 GPUs, 756GB Memory) on `iris` cluster
  - ↪ 28 cores per node, 4 GPUs per nodes, 27 GigaByte RAM per core, 756 GB per node
  - ↪ **For 30 days**: 1 node$\times[\alpha_{cpu} \times 28 + \alpha_{mem} \times 27 \times 28 + \alpha_{gpu} \times 4$ GPUS$]\times$ 30 days$\times$ 24 hours
    - ✓ Total: $1 \times [(1.0 + \frac{1}{27} \times 27) \times 28 + 50.0 \times 4] \times 720 = 184320$ SU $=$ **5529,6€ VAT excluded**

Job Billing and Budget Planning

Number of SU associated to a job

$$N_{\text{Nodes}} \times [\alpha_{cpu} \times N_{\text{cores}} + \alpha_{mem} \times Mem + \alpha_{gpu} \times N_{\text{gpus}}] \times T_{\text{exec}}$$

Cluster	Node Type	Partition	#Cores/node	CPU	α_{cpu}	α_{mem}	α_{GPU}
Iris, Aion	Regular	`interactive`	28/128	n/a	0	0	0
Iris	Regular	`batch`	28	`broadwell`	1.0*	$\frac{1}{4} = 0,25$	0
Iris	Regular	`batch`	28	`skylake`	1.0	$\frac{1}{4} = 0,25$	0
Iris	GPU	`gpu`	28	`skylake`	1.0	$\frac{1}{27}$	50
Iris	Large-Mem	`bigmem`	112	`skylake`	1.0	$\frac{1}{27}$	0
Aion	Regular	`batch`	128	`epyc`	0,57	$\frac{1}{1.75}$	0

- Continuous use of **1 Large-Memory nodes** (112 cores, 3024GB Memory) on `iris` cluster
  - ↪ 112 cores per node, 27 GigaByte RAM per core i.e. 3024 GB per node
  - ↪ **For 30 days**: 1 node$\times[\alpha_{cpu} \times 112 + \alpha_{mem} \times 27 \times 112 + \alpha_{gpu} \times 0]\times$ 30 days$\times$ 24 hours
    - ✓ Total: $1 \times [(1.0 + \frac{1}{27} \times 27) \times 112] \times 720 = 161280$ SU $=$ **4838,4€ VAT excluded**

Job Billing and Budget Planning

Number of SU associated to a job

$$N_{\text{Nodes}} \times [\alpha_{cpu} \times N_{\text{cores}} + \alpha_{mem} \times Mem + \alpha_{gpu} \times N_{\text{gpus}}] \times T_{\text{exec}}$$

Cluster	Node Type	Partition	#Cores/node	CPU	α_{cpu}	α_{mem}	α_{GPU}
Iris, Aion	Regular	interactive	28/128	n/a	0	0	0
Iris	Regular	batch	28	broadwell	1.0*	$\frac{1}{4} = 0, 25$	0
Iris	Regular	batch	28	skylake	1.0	$\frac{1}{4} = 0, 25$	0
Iris	GPU	gpu	28	skylake	1.0	$\frac{1}{27}$	50
Iris	Large-Mem	bigmem	112	skylake	1.0	$\frac{1}{27}$	0
Aion	Regular	batch	128	epyc	0,57	$\frac{1}{1.75}$	0

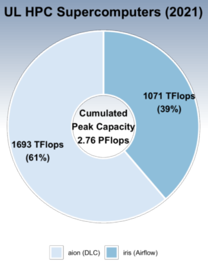
- **Not able to anticipate the type and amount of resources needed?**
  - ↪ we suggest a simple rule based on the total number of funded persons
    - ✓ account 5529.60€ for every 12 PM of funded personnel

guidelines

- **In all cases**: contact UL research support / facilitators for help and guidance

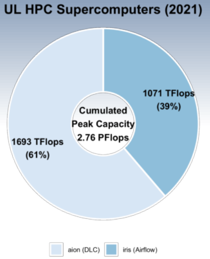
Summary

Conclusion

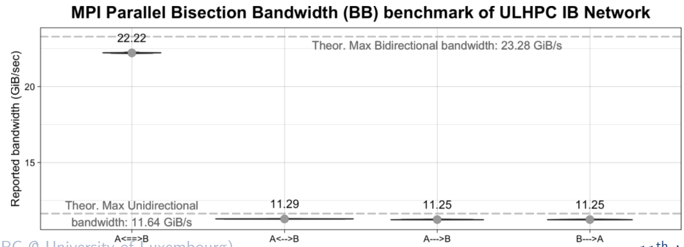


- Uni.lu HPC at the heart of the National Digital Strategy
  ↪ Several drastic changes ongoing/incoming
- **Major Achievement (2021)**: Aion supercomputer released for production

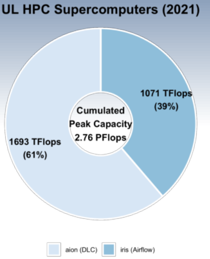
Conclusion



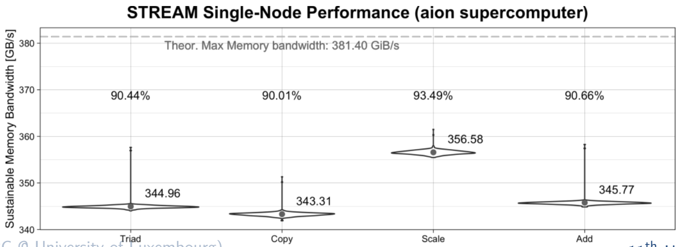
- Uni.lu HPC at the heart of the National Digital Strategy
  - ↪ Several drastic changes ongoing/incoming
- **Major Achievement (2021)**: Aion supercomputer released for production
  - ↪ Aion Supercomputer *[selected]* Performance Evaluations
    - ✓ Bisection Bandwidth (**BB**) benchmarks: **96,99% efficiency**



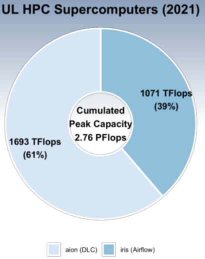
Conclusion



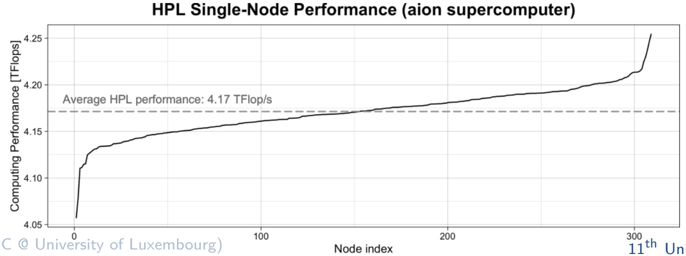
- Uni.lu HPC at the heart of the National Digital Strategy
  - ↪ Several drastic changes ongoing/incoming
- **Major Achievement (2021)**: Aion supercomputer released for production
  - ↪ Aion Supercomputer *[selected]* Performance Evaluations
    - ✓ Bisection Bandwidth (BB) benchmarks: **96,99% efficiency**
    - ✓ **STREAM** sustainable **Memory Bandwidth**: **above 90,01% efficiency**



Conclusion

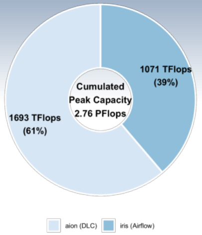


- Uni.lu HPC at the heart of the National Digital Strategy
  - ↪ Several drastic changes ongoing/incoming
- **Major Achievement (2021)**: Aion supercomputer released for production
  - ↪ Aion Supercomputer *[selected]* Performance Evaluations
    - ✓ Bisection Bandwidth (**BB**) benchmarks: **96,99% efficiency**
    - ✓ **STREAM** sustainable **Memory Bandwidth**: **above 90,01% efficiency**
    - ✓ High Performance Linpack (HPL) Single node (before large-scale run)



Conclusion

UL HPC Supercomputers (2021)



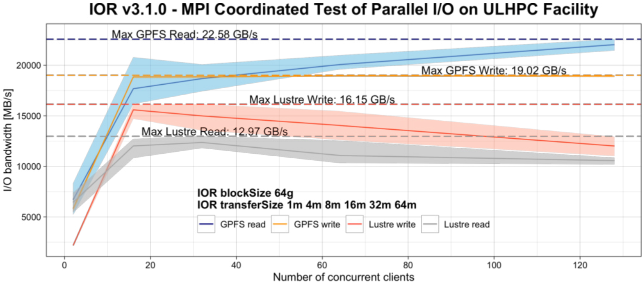
- Uni.lu HPC at the heart of the National Digital Strategy
  - ↪ Several drastic changes ongoing/incoming
- **Major Achievement (2021)**: Aion supercomputer released for production
  - ↪ Aion Supercomputer *[selected]* Performance Evaluations
    - ✓ Bisection Bandwidth (**BB**) benchmarks: **96,99% efficiency**
    - ✓ **STREAM** sustainable **Memory Bandwidth**: **above 90,01% efficiency**
    - ✓ High Performance Linpack (HPL) Single node (before large-scale run)
    - ✓ **Large-scale optimized [full] runs**: HPL, HPCG, Graph500, Green500, GreenGraph500...

Benchmark	#Node	Best Performance	Improvement*	Equivalent Worldwide Rank
HPL (Top500)	318	R_{max} = **1255.36 TFlops** (74,20% efficiency)	+ 1.9%	> 500 (Nov 2021), #490 (Jun 2020)
Green500	318	5.19 GFlops/W	+ 12,83%	#56 (Jun 2021)
HPCG	318	16.842 TFlops	+ 15,35%	#135 (Jun 2021)
Graph500 BFS	256	**975 GTEPS**	+ 64%	#23 (Jun 2021)
GreenGraph500	256	6.14 MTEPS/W	+ 180%	#36 (Jun 2021)

*: *performance improvement with the minimal acceptance threshold set in the tender*

Conclusion

- UL HPC Shared Storage Capacity and Performance Extensions
  ↪ **IOR** I/O benchmarks: **2x performance increase with Lot 2 entension**

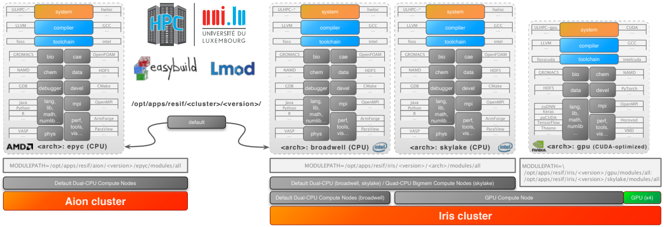


Conclusion

- UL HPC Shared Storage Capacity and Performance Extensions
  - ↪ **IOR** I/O benchmarks: **2x performance increase with Lot 2 entension**
  - ↪ **IO500** best score **11.345219**
    - ✓ Would rank ULHPC #42 in the latest IO500 list (Nov 2020)

Conclusion

- UL HPC Shared Storage Capacity and Performance Extensions
  - ↪ **IOR** I/O benchmarks: **2x performance increase with Lot 2 entension**
  - ↪ **IO500** best score **11.345219**
    - ✓ Would rank ULHPC #42 in the latest IO500 list (Nov 2020)
- UL HPC Network Consolidated
- New software set (2020b) and ULHPC Software Stack management published [ACM PEARC'21]



Conclusion

- UL HPC Shared Storage Capacity and Performance Extensions
  - ↪ **IOR** I/O benchmarks: **2x performance increase with Lot 2 entension**
  - ↪ **IO500** best score **11.345219**
    - ✓ Would rank ULHPC #42 in the latest IO500 list (Nov 2020)
- UL HPC Network Consolidated
- New software set (2020b) and ULHPC Software Stack management published [ACM PEARC'21]

[ACM PEARC'21] S. Varrette, E. Kieffer, F. Pinel, E. Krishnasamy, S. Peter, H. Cartiaux, and X. Besseron. *RESIF 3.0: Toward a Flexible & Automated Management of User Software Environment on HPC facility*. In Practice & Experience in Advanced Research Computing (PEARC'21), July 2021

- Continious opportunities for **enhanced HPC partnership** (PRACE, EuroHPC, LuxProvide, EuroCC...)

Perspectives (2022)

- **Aion Computing Capacity Expansion** (Q1 2022)
  - ↪ 12 blades, 36 computing nodes
  - ↪ Add **191,52 TFlops to ULHPC Capacity**
- New "experimental" platforms with more or less interest for UL researchers
- **HPC and Data Science** central structure
- ...

Thank you for your attention...

Questions?

High Performance Computing @ Uni.lu

University of Luxembourg, Belval Campus
Maison du Nombre, 4th floor
2, avenue de l'Université
L-4365 Esch-sur-Alzette
mail: hpc@uni.lu

hpc.uni.lu