

5 a 6  
NOVEMBRO

# encontro de computação avançada 2024



UBI, Universidade  
da Beira Interior



# Deucalion

5 Novembro 2024



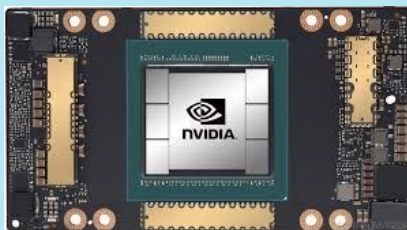
# Deucalion Compute Nodes



**Compute nodes** – 1632  
**Cores Number** – 78,336  
**Memory Capacity** – 52 TB  
**Rpeak** – 5.013 PFlops

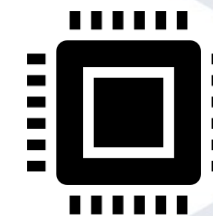


**Compute nodes** – 500  
**Cores Number** – 64,000  
**Memory Capacity** – 128 TB  
**Rpeak** – 2.304 PFlops



**Compute nodes** – 33  
**CPU Cores Number** – 4.224  
**Memory Capacity** – 16 TB  
**GPU Memory** – 8 TB  
**Rpeak CPU** – 152,064 GFlops  
**Rpeak GPU** – 2.572 PFlops

10 PFlops

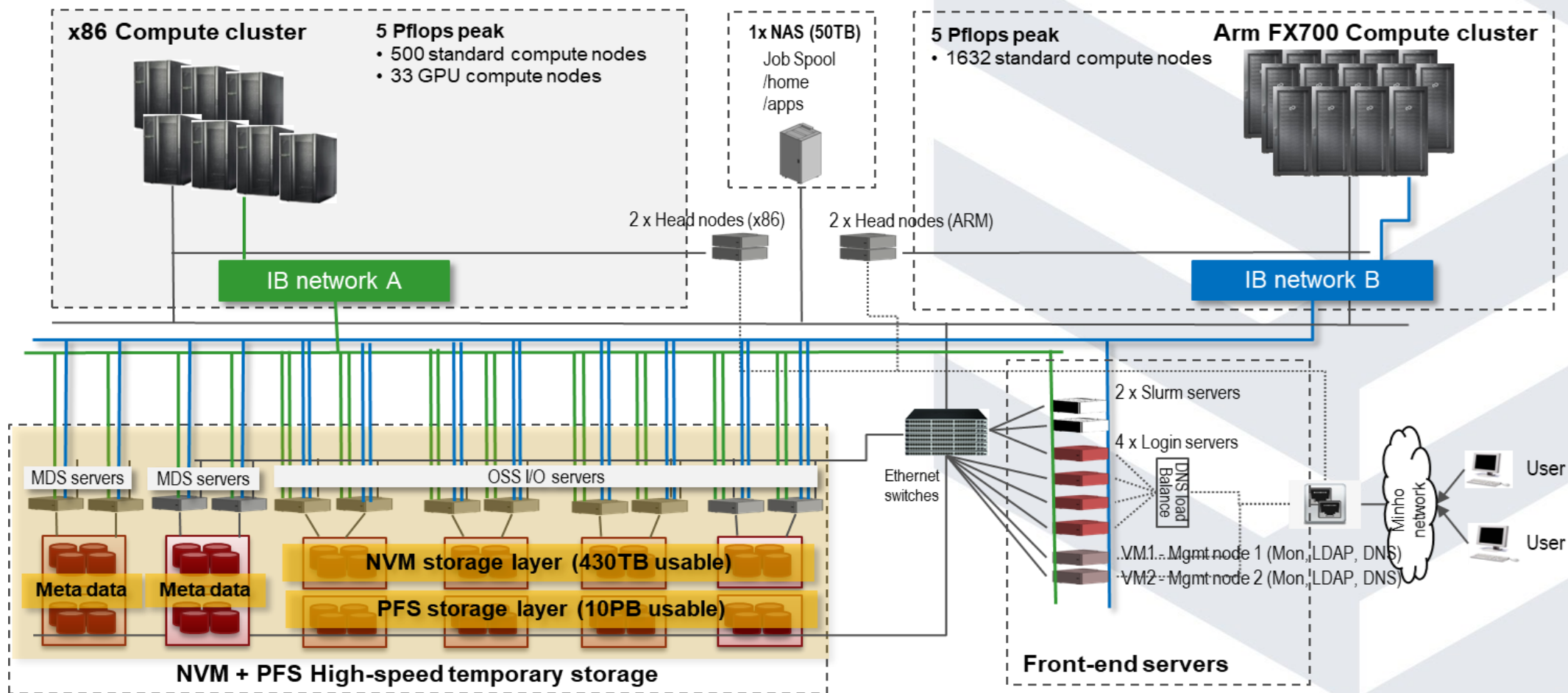


2 clusters / 3 partitions

- The ARM cluster is based on the Fujitsu A64FX processor with high levels of performance with **low energy consumption**
- The x86 cluster with AMD EPYC highly efficient processor with very good HPL efficiency and excellent energy
- The accelerator nodes have Ampere GPU from NVIDIA



# Overall Architecture

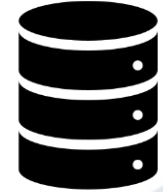


# Deucalion Storage

## High Speed Storage

**Metadata and Hot Pools NVMe** – 430 TB  
**HDD Datapools** – 10 PB usable  
**MDS Servers** – 8 Nodes  
**OSS I/O Servers** – 32 Nodes  
**Filesystem** – Lustre PFS

**Aggregated Performance**  
- 340GB/s in reads, 260GB/s in writes



- Building block architecture High speed storage with both an NVME tier and a traditional PFS disk-based tier



## NAS Storage

**Total SSD net Capacity** – 50 TB  
**Data Modules** – 2 for redundancy  
**Connection type** – 8x10 GbE

- NAS subsystem to store user homes (/home), application sources and binaries (/apps), as well as Job Spool and node images. Provides visibility of data across all clusters and servers.

# Software Stack



Functional Area	x86 cluster	ARM cluster
Operating System	Rocky Linux 8.4	
HPC software suite	OpenHPC 2.x version	
Provisioning	Warewulf	
User management	LDAP	
Job resource manager	SLURM (with MUNGE)	
Node health check	NHC	
End-user portal	Open OnDemand	
Cluster monitoring and usage	Cluster Efficiency Monitoring and Real-time Analysis (Grafana, Icinga, Kibana)	
Job usage reporting	Cluster Efficiency Monitoring and Real-time Analysis (Grafana, Icinga, Kibana)	
Other cluster tools	Lmod, pdsh, EasyBuild	
Containerization	Singularity	
PFS storage	DDN Lustre client	
Numerical/Scientific Libraries	OpenHPC libs + Intel MKL	OpenHPC libs + Fujitsu optimized BLAS, LAPACK, SCALAPACK, FFTW
I/O Libraries	HDF5 (pHDF5), NetCDF (including C++ and Fortran interfaces), Adios	
Compiler Families	GNU (gcc, g++, gfortran), Intel Parallel Studio Cluster Edition	GNU (gcc, g++, gfortran), Fujitsu Compiler suite
MPI Families	MVAPICH2, OpenMPI, Intel MPI	OpenMPI, Fujitsu MPI
Development Tools	GNU GDB, Intel Inspector, VTUNE ... etc	GNU GDB, Fujitsu debugger/profiler
Power/Energy monitoring management	MERIC (in deployment)	MERIC (in deployment)

# Batch Job Queues

## ARM

Partition name	Node range	Resource value
dev-arm	cna[0001-1632]	Min node = 1 node Max node = 16 nodes (768 cores) Default walltime = 5 min Max walltime = 4 hours DEFAULT = Yes
normal-arm	cna[0001-1632]	Min node = 1 node Max node = 128 nodes (6144 cores) Default walltime = 5 min Max walltime = 48 hours (2 days)
large-arm	cna[0001-1632]	Min node = 1 node Max node = 256 nodes (24576 cores) Default walltime = 5 min Max walltime = 72 hours (3 days)

## X86

Partition name	Node range	Resource value
dev-x86	cnx[001-500]	Min node = 1 node Max node = 8 nodes (1024 cores) Default walltime = 5 min Max walltime = 4 hours
normal-x86	cnx[001-500]	Min node = 1 node Max node = 64 nodes (8192 cores) Default walltime = 5 min Max walltime = 48 hours (2 days)
large-x86	cnx[001-500]	Min node = 1 node Max node = 64 nodes (16384 cores) Default walltime = 5 min Max walltime = 72 hours (3 days)
dev-a100-40	gnx[501-516]	1 node Default walltime = 5min Max walltime = 4 hours
normal-a100-40	gnx[501-516]	Min node = 1 node Max node = 2 node (8 GPU cards) Default walltime = 5min Max walltime = 48 hours (2 days)
normal-a100-80	gnx[517-533]	Min node = 1 node Max node = 1 node (8 GPU cards) Default walltime = 5min Max walltime = 48 hours (2 days)
dev-a100-80	gnx[517-533]	1 node Default walltime = 5min Max walltime = 4 hours

# Command Line Access

SSH Access / 2FA

Compile and Optimization Tools

Batch submission

Development Interactive Jobs

File handling and data movement





# Web Interface Access




## Message of the Day


### Changelog

- **26/09/2024:** We found a bug that does not let jobs run while using `srun`. While we correct it, please use `mpirun` instead. You will need to load an OpenMPI module beforehand (e.g. `ml OpenMPI`). EDIT: this has now been solved.
- **26/08/2024:** Corrected a bug that allowed accounts made for a certain architecture to use other architectures. To check slurm accounts associated with a user run `sacctmgr show Association where User=<username> format=Cluster,Account%30,User`. The last letter of the account shows the architecture your account can access (a=arm; x=x86; g=gpu [a100 partitions]).


### Pinned Apps A featured subset of all available apps




**Active Jobs**  
System Installed App




**Home Directory**  
System Installed App




**Job Composer**  
System Installed App




**Deucalion Cluster Shell Access**  
System Installed App




**Jupyter Notebook**  
System Installed App



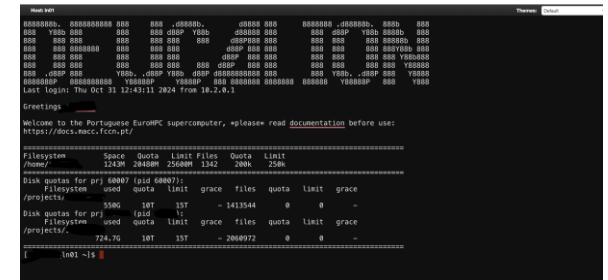
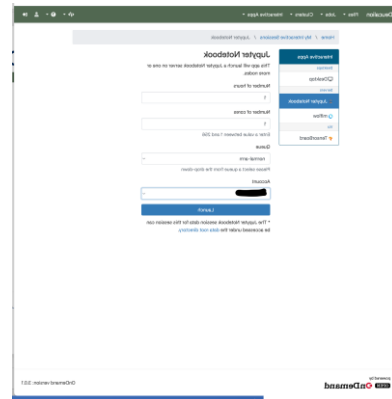
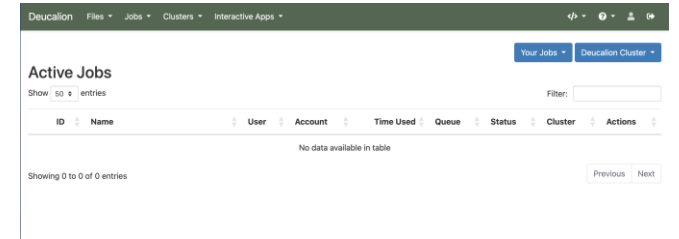
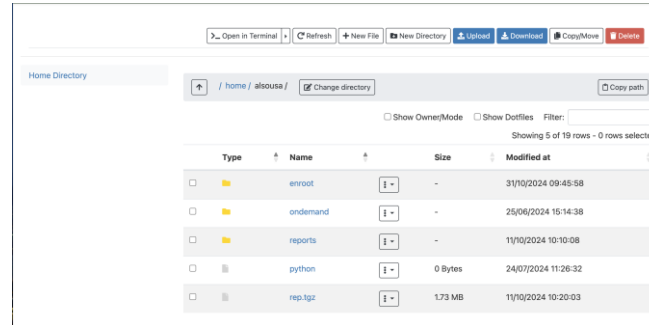
**mlflow**  
System Installed App



**Desktop**  
System Installed App



**TensorBoard**  
System Installed App



# User Software Environment



# Containerized Workloads

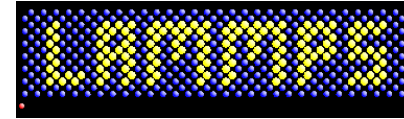


# Applications



- Molecular Dynamics

**GROMACS**  
FAST. FLEXIBLE. FREE.



- Quantum Chemistry



- Bioinformatics

SAMTools, BCFTTools, HTSLib

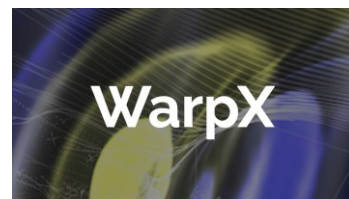
- Materials Science

Open  FOAM®

**Ans**ys



- Physics



**Smilei)**

**Osiris**

# Thank you.

For more information about FCCN services,  
see [fccn.pt/en](https://fccn.pt/en)