5 a 6 NOVEMBRO

encontro de computação avançada 2024



UBI, Universidade da Beira Interior



E para a Ciência e a Tecnologia













Financiado pela União Europeia NextGenerationEU

Deucalion

5 Novembro 2024



fct Fundação para a Cióncia UNIVERSIDADE e a Tecnologia UNIVERSIDADE







0







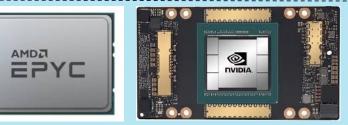
Deucalion Compute Nodes

ευβτςυ **Α64FX**™

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



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



ECT Fundação para a Cádecia UN UNIVERSIDADE e a Tecnologia UN BEIRA INTERIOR









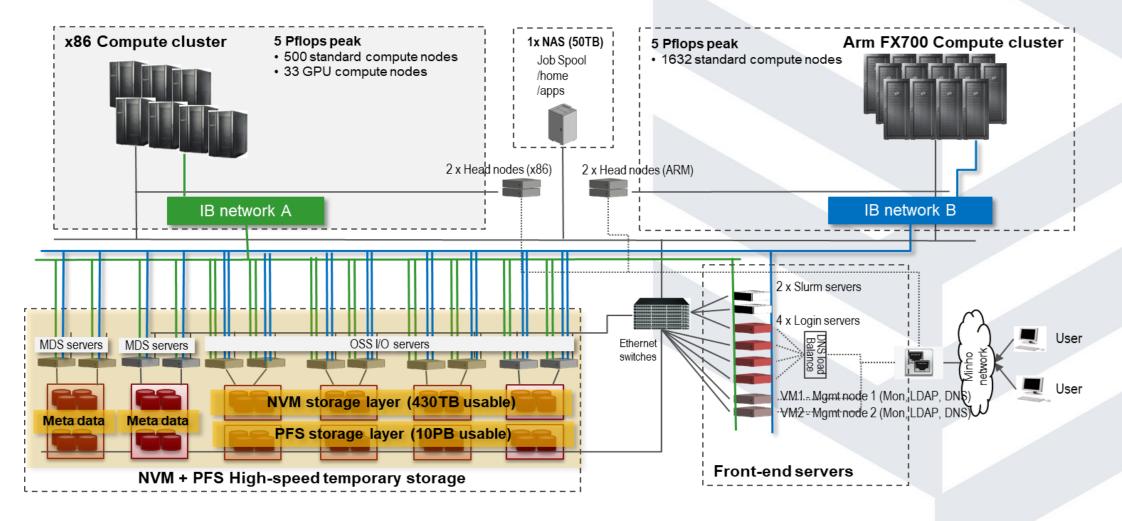


0

PFlops



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

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



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



 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.











EuroHPC







Software Stack



Functional Area	x86 cluster	ARM cluster					
Operating System	Rocky L	inux 8.4					
HPC software suite	OpenHPC	2.x version					
Provisioning	Ware	ewulf					
User management	LD	AP					
Job resource manager	SLURM (wi	th MUNGE)					
Node health check	Nł	HC					
End-user portal	Open Or	Demand					
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	n, EasyBuild					
Containerization	Singu	ularity					
PFS storage	DDN Lus	stre 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	normal-x86	C
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	large-x86	C
normal-arm	cna[0001-1632]	Min node = 1 node Max node = 128 nodes (6144 cores) Default walltime = 5 min Max walltime = 48 hours (2	dev-a100-40 normal-a100-40	ç
large-arm	cna[0001-1632]	days) Min node = 1 node Max node = 256 nodes (24576		
		cores) Default walltime = 5 min Max walltime = 72 hours (3 days)	normal-a100-80	Ç
			1 100 00	

	X8	6	
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





>_ Open in Terminal → C Refresh + New File In New Directory & Upload & Download I Copy/Move Delete

Show Owner/Mode Show Dotfiles Filter:

Size

1.73 MB

↑ / home / alsousa / 🖉 Change directory

≜ Name

reports

ovthor

rep toz

Type

.

1 -

1 - 0 Bytes



Copy pat

Showing 5 of 19 rows - 0 rows selected

Modified at

31/10/2024 09:45:58

25/06/2024 15:14:38

11/10/2024 10:10:08

24/07/2024 11:26:32

11/10/2024 10:20:03

Deucalion Files * Jobs * Clus	ters * Interactive Apps *	«/>• ⊕ - ≛ 0+
		Your Jobs * Deucation Cluster *
Active Jobs		Filter:
ID 🕆 Name	🗄 User 🌵 Account 👙 Time Used 🗄 (ueue 🗄 Status 🛱 Cluster 🗍 Actions
	No data available in table	
Showing 0 to 0 of 0 entries		Previous Nex

											Therees
888 Y885 888 888 888 888 888 888 888 888 888	88 88 98 78	8 888 8 888 8 888 8 888 8 888 8 888 8d88P Y88888P	d88P Y 888 888 888 888 888 Y885 d Y885 d	885 888 888 di 889 di 889 di	888888888 8 8888888	888 888 888 888 888 888 888	888888 888 888 888 888 888 888 888	888 888 888 888 888 888 888 888 988 888 9886	888 Y		
Greetings											
	-										
Welcome to the I				nputer,		 read d 	ocumenta	tion before	use:		
		t/ Quota 20480M	Limit 25600M	Files 1342	Quota 200k	Linit 250k					
Welcome to the M https://docs.mar Filesystem /home/	space 1243M	t/ Quota 28488M	Limit 25600M	Files 1342	Quota 200k	Linit 250k					
Welcome to the H https://docs.mar Filesystem /home/ Disk quotas for Filesystem	space 1243M	t/ Quota 28488M	Limit 25608M	Files 1342	Quota 200k	Linit 250k					
Welcome to the I https://docs.man Filesystem /home/ Disk quotas for Filesystem /projects/	Space 1243M prj 6000 used 550G	Quota 20480M 7 (pid 60 quota 10T	Linit 25600M 007): linit 15T	Files 1342 grace	Quota 200k	Linit 250k					
Welcome to the H https://docs.mar Filesystem /home/ Disk quotas for Filesystem	Space 1243M prj 6000 used 550G	Quota 20480M 7 (pid 60 quota	Linit 25608M 807): linit	Files 1342 grace	Quota 200k files 1413544	Linit 250k quota	limit				

OnDemand version: 3.0.1

User Software Environment





Containerized Workloads



Applications

ris

Molecular Dynamics

- Quantum Chemistry
- GROMACS FAST. FLEXIBLE. FREE.

BioInformatics

Materials Science

Physics



Thank you.

For more information about FCCN services, see <u>fccn.pt/en</u>















