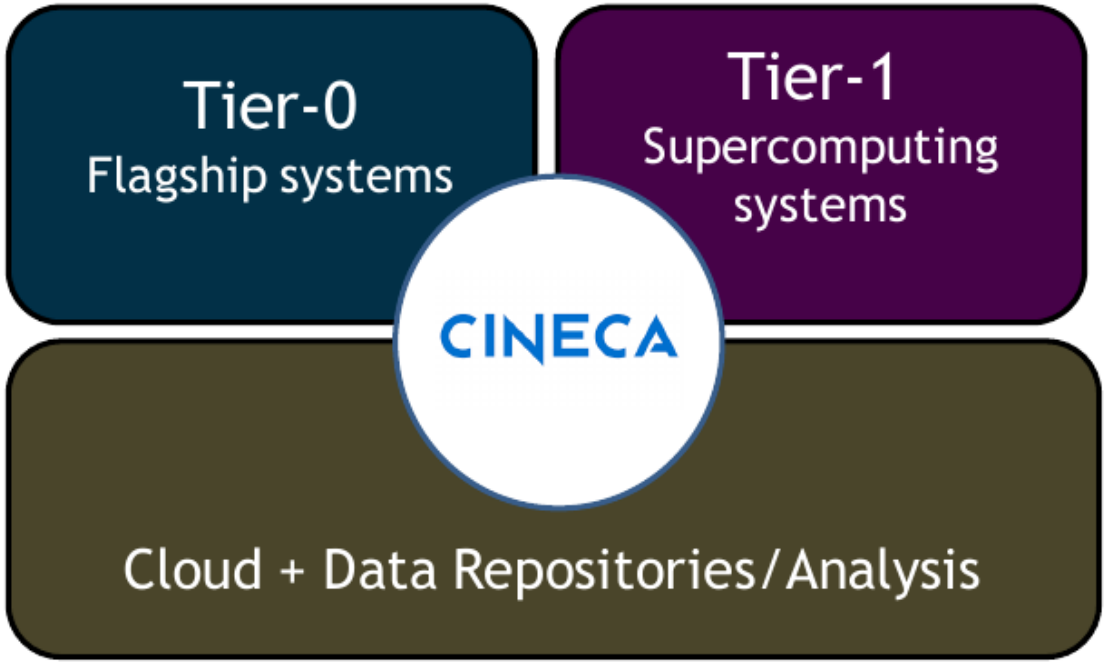
NEW HPC RESOURCES AND HOW TO ACCESS IN EUROPE AND ITALY

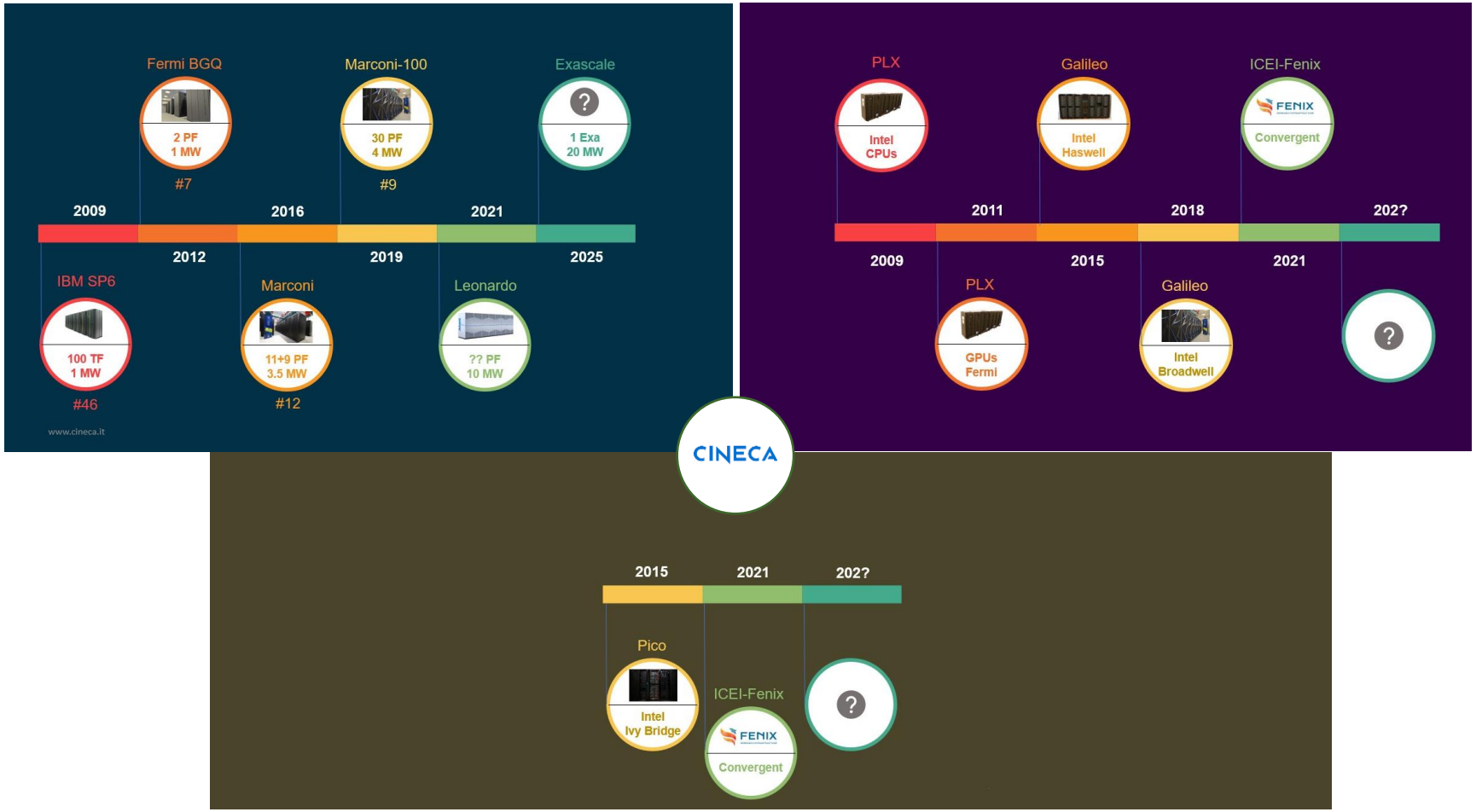
Dr. Massimiliano Guarrasi – CINECA
m.guarrasi@cineca.it

CINECA
consorzio interuniversitario

CINECA HPC Infrastructure



CINECA HPC Infrastructure



Tier0: Marconi100

- **IBM Power System AC922**
- 980 nodes
- Each equipped with:
  - 2 * [IBM POWER9 16C 3GHz]
  - 256 GB RAM
  - 4* [Nvidia V 100 16 GB HBM2]
- For a total:
  - 347'776 cores
  - 253 TB RAM
  - 3920 GPUs
- Dual-rail Mellanox EDR Infiniband
- Linpack Performance (Rmax): 21'640 TFlops
- Theoretical Peak (Rpeak): 29'354 TFlops

23/06/2021

Tier1: Galileo100

- Instrumented by Dell
- Ready for production by July 2021
- Two partitions (Scalable compute, Cloud compute)
- Scalable Compute (564 server):
  - 2x CPU 8260 Intel CascadeLake, 24 core, 2.4 GHz
  - 384 GB RAM DDR4 2933MT/s
  - 348 standard server:
    - 480 GB SSD
  - 80 Data processing server:
    - 2 TB SSD
    - 1,5 TB Intel Optane
  - 36 GPU server:
    - 2 TB SSD
    - 2x NVIDIA GPU V100
  - 20,5 PB active storage
  - 720 TB fast storage (IME DDN)
  - Infiniband 100 Gbs
- Cloud Compute:
  - 77 computing server OpenStack
    - 2x CPU 8260 Intel CascadeLake, 24 cores, 2.4 GHz
    - 768 GB RAM DDR4 2933MT/s
    - 2 TB SSD
  - 1 PB CEPH storage (full NVMe/SSD)
  - Ethernet 100 Gbs

LEONARDO

EuroHPC
Joint Undertaking

CINECA
consorzio interuniversitario

EuroHPC
Joint Undertaking

CINECA

LEONARDO

HELLENIC REPUBLIC
Ministry of Digital Governance

REPUBLIKA SLOVENIJA
REPUBLIC OF SLOVENIA
Ministrstvo za izobraževanje, znanost in šport
Ministry of education, science and sport

MINISTRY OF
INNOVATION AND TECHNOLOGY

Istituto Nazionale di Fisica Nucleare

Federal Ministry
Republic of Austria
Education, Science
and Research

23/06/2021

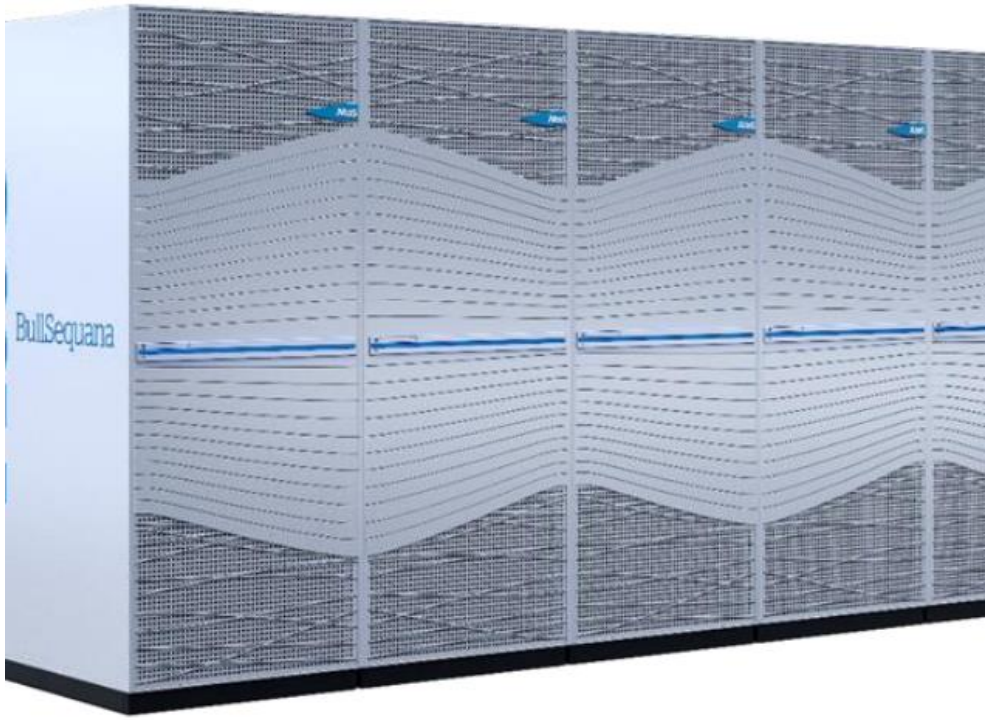
MINISTRY
OF EDUCATION, SCIENCE,
RESEARCH AND SPORT
OF THE SLOVAK REPUBLIC

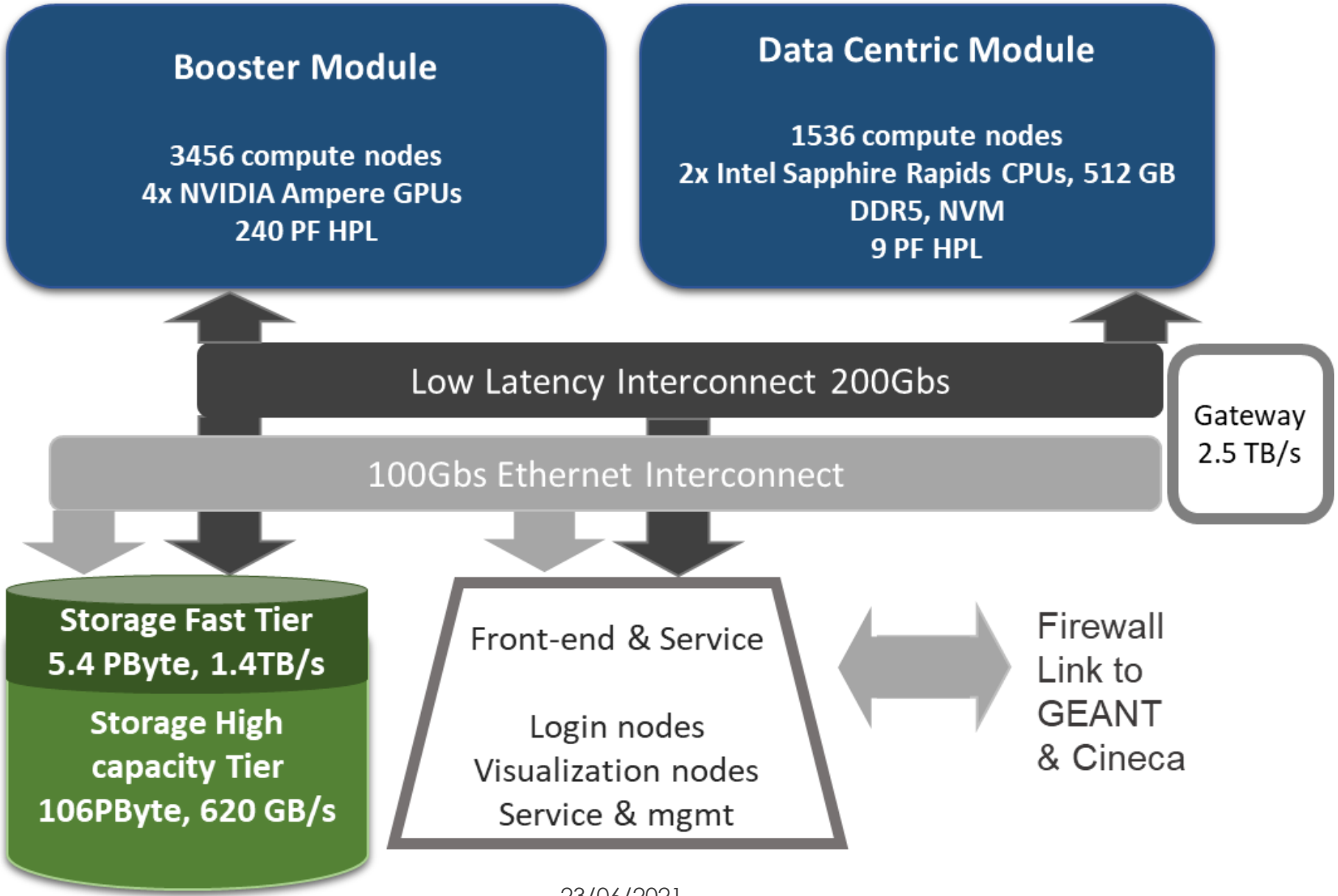
CINECA
consorzio interuniversitario

Leonardo Specification

- Available **Q1 2022**
- Based on **Atos XH2000** platform technology
- Computing racks **95% Direct Liquid Cooled**
- **Warm water**: Inlet temperature of 37 degrees
- NVIDIA Mellanox **HDR 200** interconnect
  - Dragonfly+ topology
  - 1.11:1 (intra-cell)
  - 0.8:1 globally

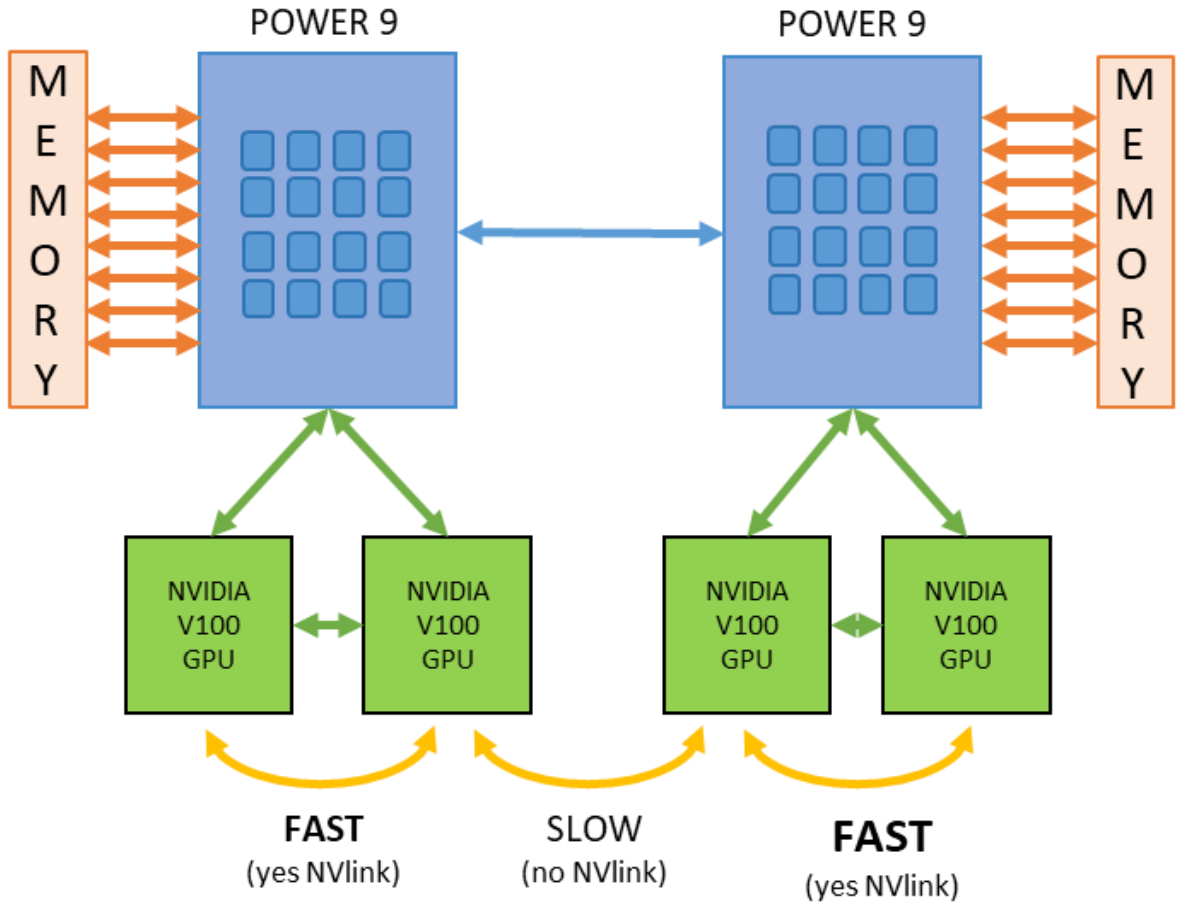
- **Bull Smart Energy management** suite
  - Bull Energy Optimizer
  - Bull Dynamic Power Optimizer




Booster Module
3456 compute nodes
4x NVIDIA Ampere GPUs
240 PF HPL
Data Centric Module
1536 compute nodes
2x Intel Sapphire Rapids CPUs, 512 GB
DDR5, NVM
9 PF HPL
Low Latency Interconnect 200Gbs
100Gbs Ethernet Interconnect
Gateway
2.5 TB/s
Storage Fast Tier
5.4 PByte, 1.4TB/s
Storage High
capacity Tier
106PByte, 620 GB/s
Front-end & Service
Login nodes
Visualization nodes
Service & mgmt
Firewall
Link to
GEANT
& Cineca

	Marconi100	Leonardo-Booster
CPU	2 POWER9	1 IceLake
Cores	32 (16 per P9)	32 (2.4 GHz - 250 W)
Memory	256 GB	256 GB
CPUs : Accelerators	2:4	1:4
Accelerators	4 Volta V100	4 Ampere based GPU
GPU-GPU bandwidth	150 GB/s	400 GB/s
Accelerator DP Flops	28 TF	NDA
Accelerator Memory	64 GB HBM2	256 GB HBM2e (4 x 64 GB)
Accelerator Memory Bandwidth	3.6 TB/s (900 GB/s x 4 GPUs)	6.5 TB/s (1.6 TB/s x 4 GPUs)
SSD Capacity	1.6 TB	-

2 sockets
x
16 cores/socket
x
4 threads/core
=
128 virtual cores



Each physical cores = 4 virtual cores (SMT / hyperthreading)

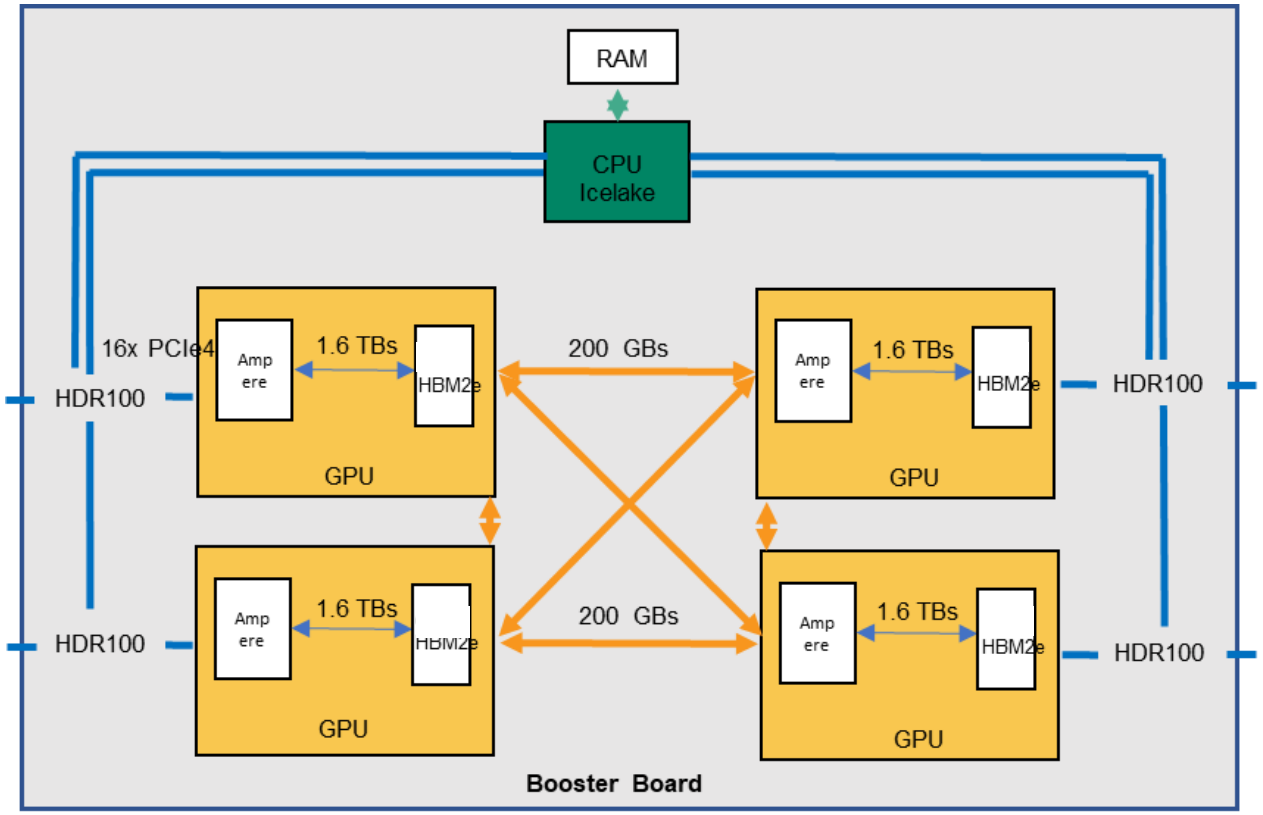
Xbus
(64 GBps)

NVlink 2.0
(3 x 50 GBps per channel)

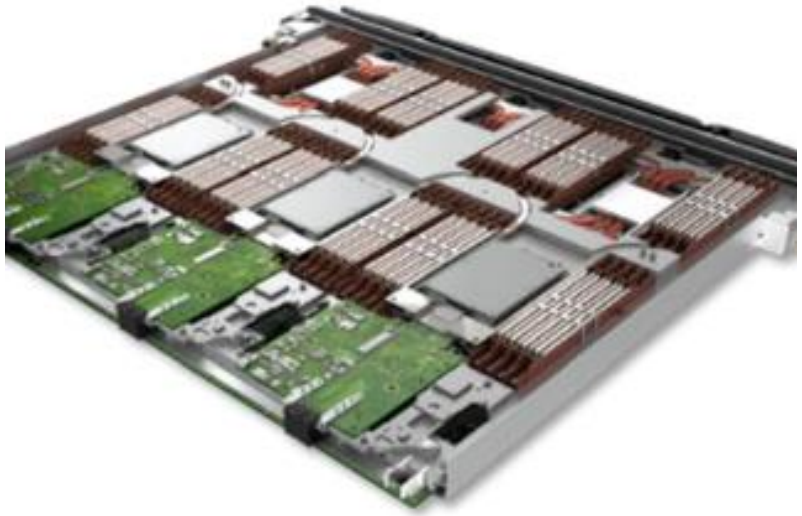
CPU memory channel
(17.5 GBps per channel)

Marconi 100 node and system architecture: *https://wiki.u-gov.it/confluence/pages/viewpage.action?pageId=336727645#UG3.2:MARCONI100UserGuide-SystemArchitecture*

Leonardo Booster: Da Vinci Blade



- ad hoc board from ATOS
- 4 NVidia Ampere GPUs (SXM)
- CPU-GPU connection via PCIe4 16x connection through HDR Connect6 HCA
  - PCI passthrough
  - 16 PCI links towards CPU, 16 links towards GPU
  - Bandwidth: 64 GBs duplex
- Full NV Link GPU-GPU connection
  - 200 GB/s bi-directional
- No PCI switch between host and external network
  - Low latency
- Out-of-band telemetry information
- GPUdirect



Leonardo Data Centric

- Based on BullSequana X2610 compute blade
- 2x Intel Sapphire Rapids
  - > 40 cores
- 512 GB DC
  - DIMM 16GB DDR5 4800 MTS
  - Bandwidth 250 GB/s x socket
- 3.8 TB NVM

Obtaining HPC resources in Italy and Europe

- Institution Level:
  - Agreement (e.g. INFN, INAF, SISSA, UniMI,…)
- Regional Level:
  - Calls promoted by regional administrations (e.g. LISA)
- National Level:
  - ISCRA (B,C,D, …)
- European level:
  - DECI
  - FENIX
  - PRACE
  - EUROHPC

PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE

EuroHPC Joint Undertaking

Italian SuperComputing Resource Allocation - ISCRA

- Open to all scientific researchers affiliated to an Italian research organization needing large allocations of computer time, supporting resources and data storage to pursue transformational advances in science.
- Projects' Principal Investigators are expected to be affiliated to an Italian institution, while no restriction is applied for the Co-PI and collaborators.
- Further information ad for applying: https://www.hpc.cineca.it/services/iscra

- ISCRA C:
  - Small Project (2000 Nh on M100, 3000 Nh on G100, 20K Ch on DGX)
  - Duration: 9 months
  - Easy to submit (1 page, few data needed)
  - Only technical evaluation
  - Continuous submission, 1 cut off per month
  - Several types of the project (HPC, Cloud, HPC + Cloud, Development & Benchmark, Quantum Computing) and the project focus (General purpose, Special Focus: AI & ML, Big Data/Bioinformatics, COVID-19, ...)
- ISCRA B:
  - Mid size projects (up to 50K Nh on M100, up to 75K Nh on G100)
  - Duration: 1 year
  - More detailed proposal (some pages, scalability plot, detailed budget estimation, technical and scientific details needed)
  - Tech and Scientific evaluation
  - 2 call per year
- ISCRA D:
  - Only for long term storage resources
  - Up to 50 TB on FS and/or 200 TB on Tape Library
  - Maximum duration 36 months (+ 6 for move the data)
- ISCRA A/Key projects (TBD)

Distributed European Computing Initiative - DECI

- The PRACE Distributed European Computing Initiative (DECI) programme provides access to Tier-1 level resources across Europe via a series of competitive calls.
- Projects' Principal Investigators are expected to be affiliated to an Eurepean Based institution, while no restriction is applied for the Co-PI and collabortors.
- Further information and for applying: https://prace-ri.eu/hpc-access/deci-access/deci-access-how-to-apply/

- Tier-1 resources i.e., Galileo100 in italy, Cartesius in Netherland, ARCHER2 in UK, …
- Typical size of projects: 1-2 M Core Hours (<ISCRA B, >ISCRA C)
- Aperiodic (Typically 1 call per year)
- Duration 1 year
- Detailed proposal (some pages, scalability plot, detailed budget estimation, technical and scientific details needed)
- Subject to Scientifical and Technical Evaluation
- Just return mechanism (85% of resources dedicated to Pis from hosting member's nations, 15% to others European Pis)
- Typically, 1-2 project per year and per nation are awarded.

Partnership for Advanced Computing in Europe - PRACE

- Project Access is intended for individual researchers and research groups and can be used for 1-year production runs, as well as for 2-year or 3-year (Multi-Year Access) production runs.
- Projects' Principal Investigators are expected to be affiliated to an institution (public or private) based in a PRACE member's country, while no restriction is applied for the Co-PI and collabortors.
- Also Non-PRACE PI can apply, but the approval of the proposal depends on the hosting centres.
- Further information and for applying: https://prace-ri.eu/hpc-access/calls-for-proposals/

- Tier-0 resources, i.e. Marconi100 in Italy, Joliot Curie in France, Piz Daint in Switzerland, Mare Nostrum 4 in Spain, Juwels, SuperMUC NG and HAWK in Germany
- Large scale project (typical scale 30-150 M core hours)
- But also dedicated calls for benchmark and code development are present
- 2 calls per years (every 6 months)
- Duration: from 1 to 3 years
- Very difficult application form (3 pages, more technical and scientific details are needed)
- Rate of success: >50% (depending on the call)
- Subject to Scientifical and Technical Evaluation
- Just return mechanism (75% of resources dedicated to Pis from PRACE hosting member's nations, 25% to PRACE General Partner Pis)

Fenix

- Projects' Principal Investigators are expected to be affiliated to an Eurepean Based institution, while no restriction is applied for the Co-PI and collabortors.
- Special calls for HBP reserchers and neuroscientis
- Further information and for applying: https://fenix-ri.eu/access

- Resources for storage and data processing:
  - Scalable computing
  - Interactive computing
  - Virtual machines/Container
  - Active storage
  - Object storage
- Up to some hundreds TB of storage, dedicated nodes for services and post processing, …
- 4 call per year(every 3 months)
- Subject to scientific and technical evaluation
- 62.5% of available resources for HBP, 37,5% for dedicated calls (by PRACE peer reiew team)
- Dedicated A&A procedure for using the same credentials in the entire infrastructure
- Hosted on Galileo100 in Italy

EuroHPC

- Open to all scientific researchers affiliated to an EU research organization and/or Industry needing large allocations of computer time, supporting resources and data storage to pursue transformational advances in science.
- Projects' Principal Investigators are expected to be affiliated to an EU institution and/or company, while no restriction is applied for the Co-PI and collaborators.

- Phase 1 (2020-21):
  - 3 pre-exascale systems:
    - Italy (Leonardo@CINECA)
    - Finland (LUMI@CSC)
    - SPAIN (T.B.D.@BSC)
  - 5 Petascale systems:
    - Bulgary (DISCOVERER@Sofia Tech Park)
    - Czech Repubblic (KAROLINA@IT4I)
    - Luxemborg (MeluXina@LuxProvide)
    - Slovenia (VEGA@IZUM)
    - Portugal (DEUCALION@Minho Advanced Computing Centre)
- Phase 2 (2023,TBD):
  - 2 Full Exascale Systems
- Several kinds of calls managed by PRACE:
  - Preparatory
  - Fast track
  - Industrial access
  - Standard call (PRACE Tier-0 size allocations, continuous submission, 3 cut-off per year, mid difficulty proposal, scientific and technical evaluation)
  - New Tier-0 calls (allocations > PRACE Tier-0 size, 2 cut-off per year, high difficulty proposal(similar to PRACE Tier-0, scientific and technical evaluation)
- Only preparatory on VEGA are available by now

Thank you

- Dr. Massimiliano Guarrasi
- Technical contact for PRACE projects @CINECA and HLST staff member

- CINECA, www.cineca.it
- Mail: m.guarrasi@cineca.it
- Phone: +39 091 6171 560