CINECA HPC Infrastructure

Dr. Massimiliano Guarrasi – CINECA

m.guarrasi@cineca.it

SoHPC Training Week - 1/7/2019

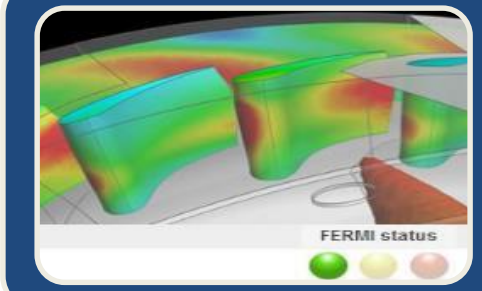
WELCOME

Thank you for joining us today!

CINECA at glance

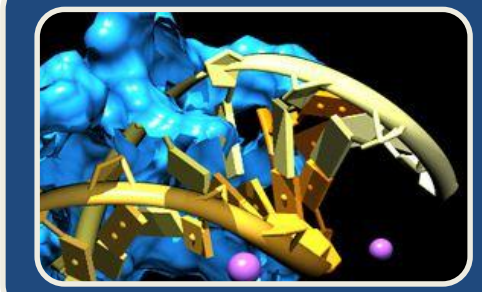
Services both for Universities and Ministry

- Solutions & Services for the University Administration
- Information system for the Ministry for the management, assessment, evaluation, funding of the research



Scientific Research

- Promote the use of the most advanced High Performance Computing systems to support public and private scientific and technological research



Innovation and technology transfer

- Numerical experiment, virtual prototyping
- Big data
- Scientific visualizations, computer graphics

About us

Not for profit

Founded in 1969

HPC

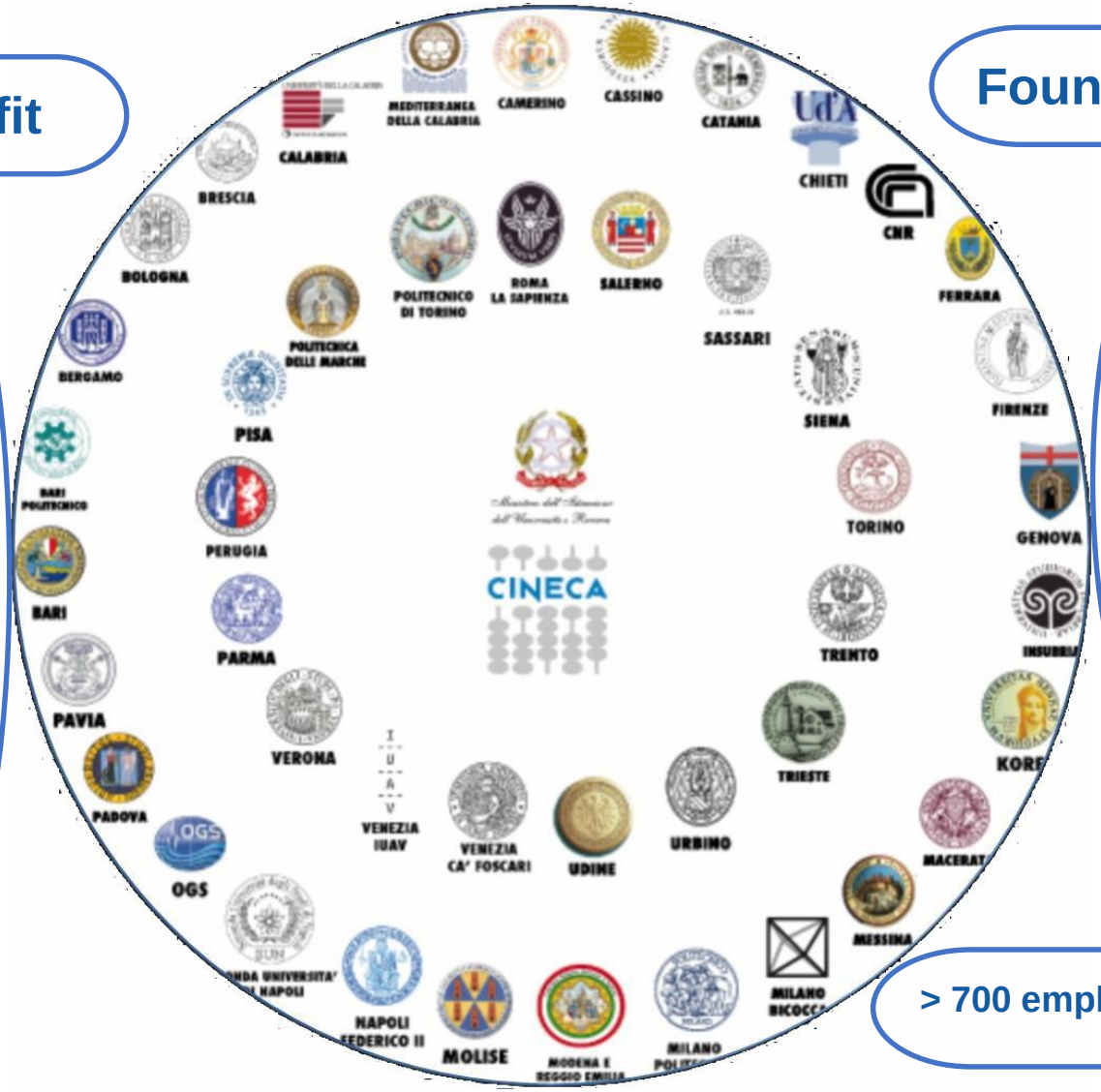
MARCONI: TBD

FERMI: Top 500

GALILEO : Top 500

PICO: Data Movement

EURORA: Green 500



76 Members

70 Universities

5 Research Institutes

Ministry of Education

> 700 employees (~80 in HPC)

Access to HPC resources: CINECA aims and basic principles

Our objectives:

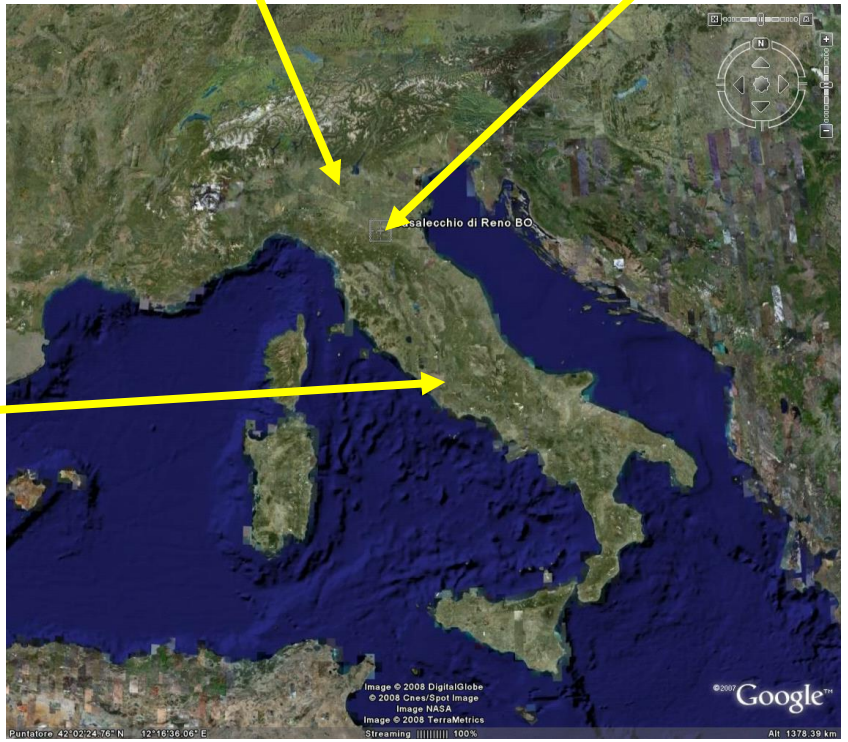
- ✓ Providing Italian and European researchers with an advanced computational environment
- ✓ Supporting Italian researcher for increasing their competitiveness
- ✓ Following Italian researchers in their path towards Tier 0
- ✓ Soliciting large-scale and computationally intensive projects

Basic principles:

- ✓ Transparency
- ✓ Fairness
- ✓ Conflict of Interest management
- ✓ Confidentiality

CINECA

CASPUR



Partnership for Advanced Computing in Europe - PRACE

- **The mission of PRACE is to enable high impact scientific discovery … to enhance European competitiveness.**

- **PRACE offers world class computing and data resources through a peer review process.**

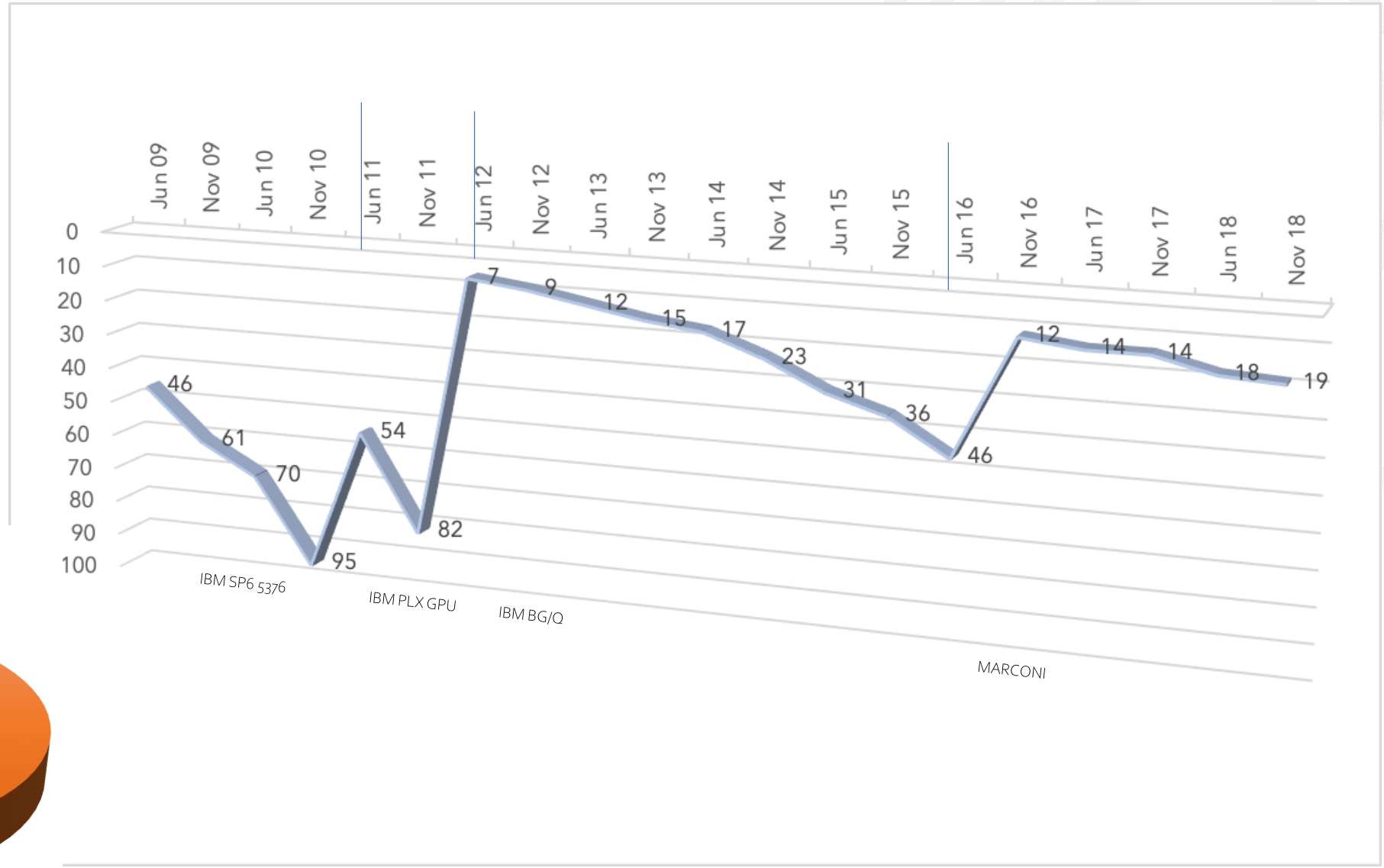
- **PRACE also seeks to strengthen the European users of HPC in industry through various initiatives.**

- **PRACE has a strong interest in improving energy efficiency of computing systems and reducing their environmental impact.**

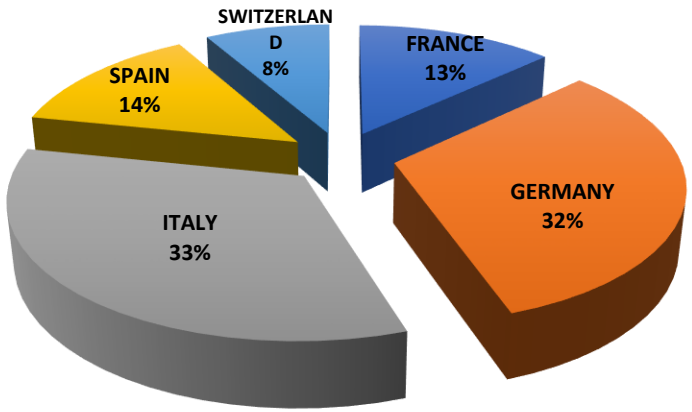
http://www.prace-ri.eu/call-announcements/

http://www.prace-ri.eu/prace-resources/

HPC Systems @ Cineca



Resources allocated by PRACE (core hours)



D.A.V.I.D.E.(Prototype)

Logical Name	D.A.V.I.D.E. (August 2017)
Model	E4 Cluster Open Rack
Architecture	OpenPower NVIDIA NVLink
Processor	OpenPower 8 NVIDIA Tesla P100 SXM2
# of core	~ 1000
# of node	45 x (2 Power8 + 4 Tesla P100)
# of rack	3
RAM per node	256GByte
Interconnection	Mellanox EDR
Operating System	GNU/Linux
Total Power	~ 90Kw
Peak Performance	~ 1 Pflops

PRACE pre-commercial procurement, “Whole System Design for Energy Efficient HPC”

Marconi - convergent HPC solution

It will be upgraded in the Q1 of 2020

Scale Out

MARCONI

- 3200 Lenovo Stark servers > 9 PFlops
  Intel SkyLake
  2x24 cores @ 2.1GHz. 196GByte x node
- 3600 Intel/ lenovo servers > 11PFlops
  Intel PHI code name Knight Landing
  68 cores @ 1.4GHz.
  single socket node: 96GByte DDR4
  + 16GByte MCDRAM
- 720 Lenovo NeXtScale servers
  Intel E5-2697 v4 Broadwell
  18 cores @ 2.3GHz.
  128GByte x node

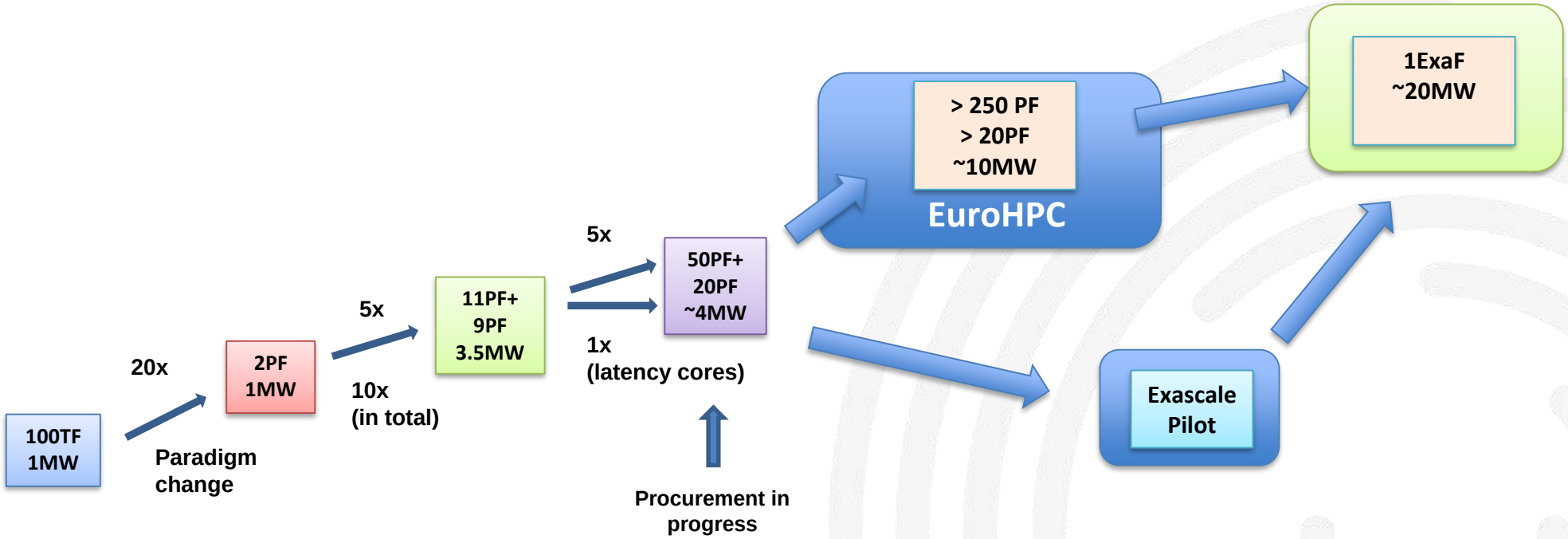
Cloud/Data Proc.

Lenovo NeXtScale servers
Intel E5-2697 v4 Broadwell
- 216 nodes eth x cloud HT INFN (CNAF)
- 216 nodes eth x cloud HPC/DP (MEUCCI)
- 360 nodes QDR x Tier 1 – HPC (GALILEO)
- 720 nodes OPa x Tier 1 – HPC (GALILEO 2)

40 Lenovo NeXtscale servers (GALILEO)
- Intel E5-2630 v3 Haswell
- QDR + Nvidia K80

Lenovo GSS + SFA12K + IBM Flash
>30PByte

RoadMap



2009	2012	2016	2019/2020	2021	2023-2025	2025-2027
IBM SP6 Power6	Fermi IBM BGQ PowerA2	Marconi Lenovo Xeon+KNL	Marconi PPI4HPC ICEI - PPIHBP	Pre-exasxle with EuroHPC contribution	Post-Marconi Exascale pilot technology (National Research Plan)	Exascale with EuroHPC

CINECA

Ministero dell'Istruzione Università e Ricerca

MIUR

SISSA - ISAS

INFN
Istituto Nazionale di Fisica Nucleare

REPUBLIKA SLOVENIJA
REPUBLIC OF SLOVENIA
Ministrstvo za izobraževanje, znanost in šport
Ministry of education, science and sport
Masarykova cesta 16, SI - 1000 Ljubljana

EuroHPC JU approved 840 million euro funding for eight supercomputers

Hosting Sites

- 3 Pre-Exascale:
  - Bologna (Italy): Leonardo
  - Kajaani (Finland): LUMI
  - Barcellona (Spain): Mare Nostrum 5
- 5 Petascale:
  - Sofia (Bulgaria)
  - Ostrava (Czech Republic)
  - Bissen (Luxemburg)
  - Minho (Portugal)
  - Maribor (Slovenia)

Partnership

INFN
Istituto Nazionale di Fisica Nucleare

Ministero dell'Istruzione
Università e Ricerca
MIUR

SISSA - ISAS

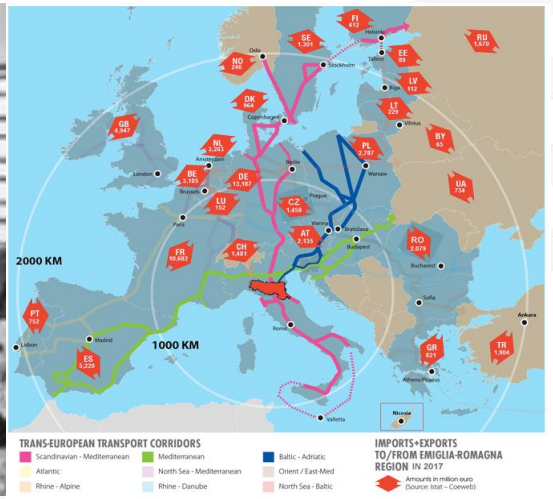
CINECA

+

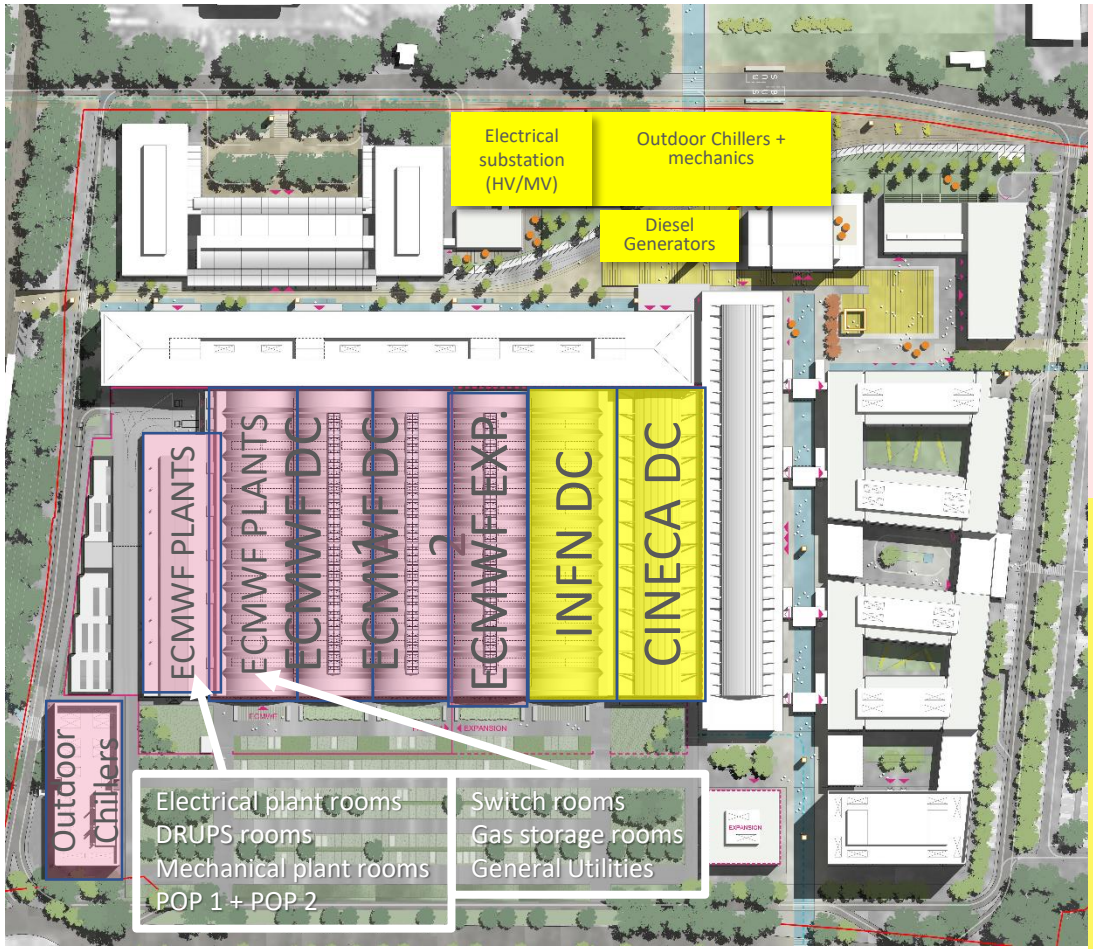
Consortium

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

Bologna Science Park



The data centers at the Science Park



ECMWF DC main characteristics

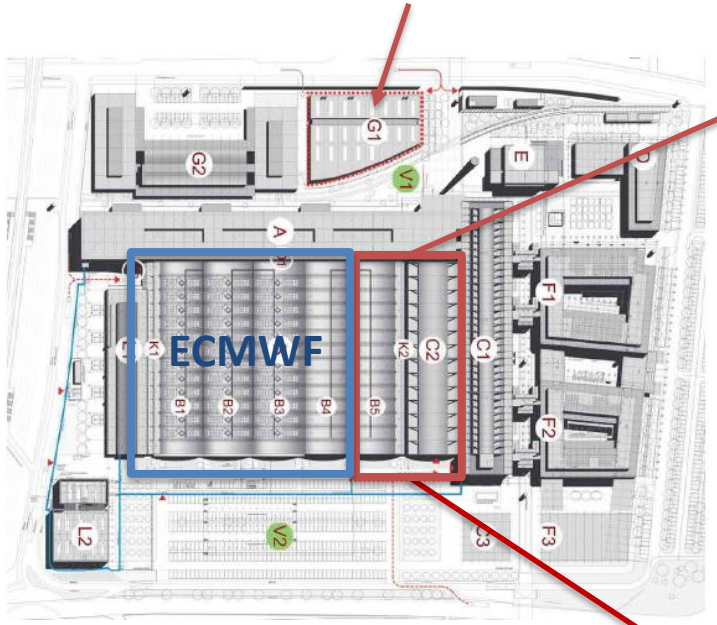
- 2 power line up to 10 MW (one bck up of the other)
- Expansion to 20 MW
- Photovoltaic cells on the roofs (500 MWh/year)
- Redundancy N+1 (mechanics and electrical)
- 5 x 2 MW DRUPS
- Cooling
  - 4 dry coolers (1850 kW each)
  - 4 groundwater welles
  - 5 refrigerator units (1400 kW each)

INFN – CINECA DC main characteristics

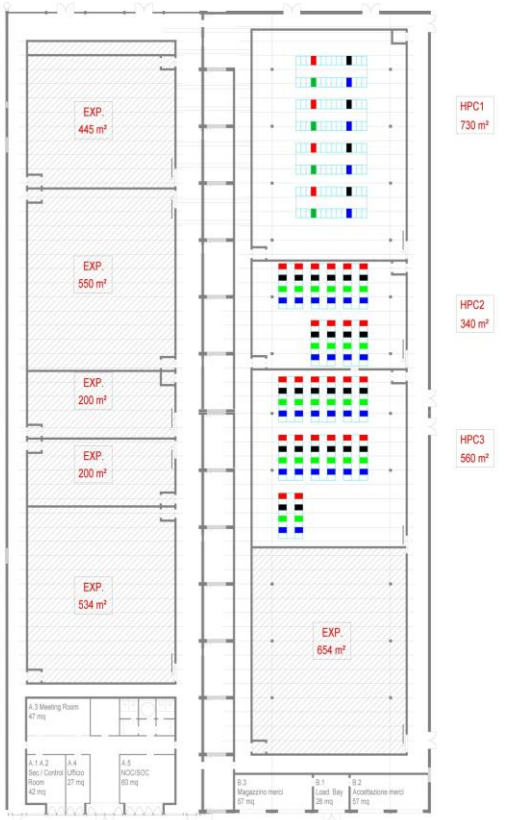
- up to 20 MW (one bck up of the other)
- Possible use of Combined Heat and Power Fuel Cells Technology
- Redundancy strategy under study
- Cooling, still under study
  - dry coolers
  - groundwater welles
  - refrigerator units
- PUE < 1.2 – 1.3

EuroHPC hosting @ Bologna Science Park

Cooling equipment
3MW (2020) -> 5MW(2023)



Computer Rooms
10MW (2020) -> 20MW (2023)

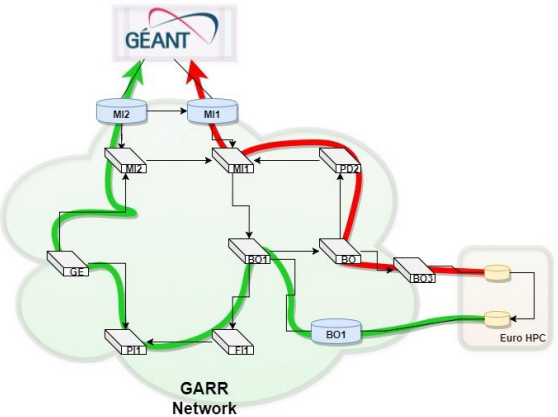


DATA ROOM STAGE 1: 1600 sqm
DATA ROOM STAGE 2: 2600 sqm
ANCILLARY SPACES: 900 sqm

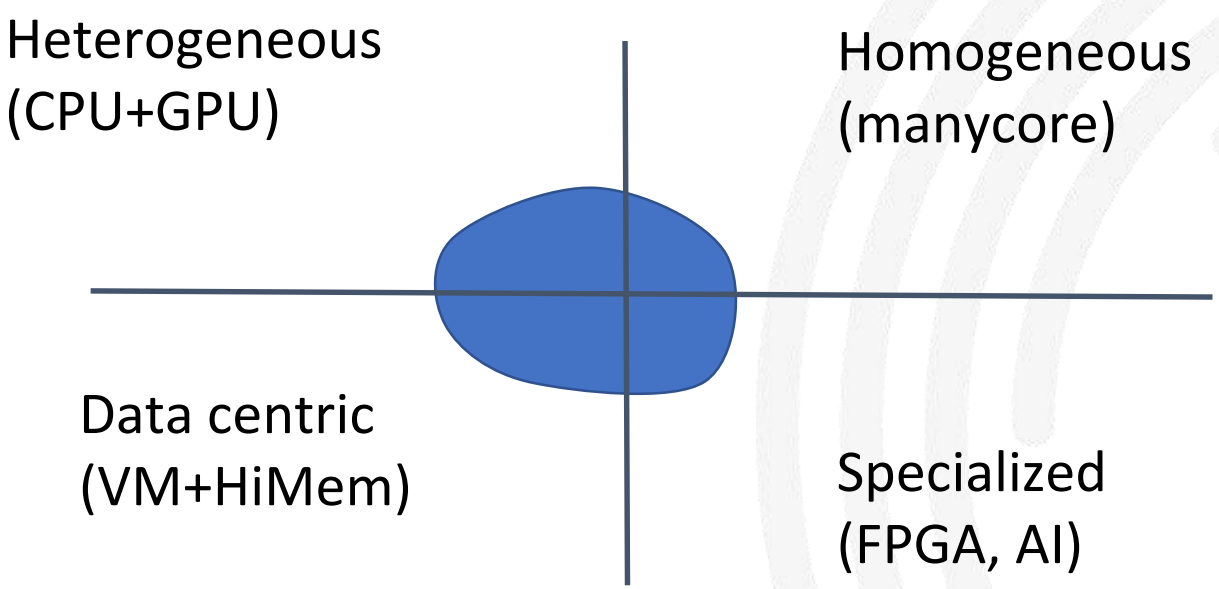
8MW hot water DLC
Compute nodes

2MW AIR Cooled
Storage + Ancillary

PUE < 1.1

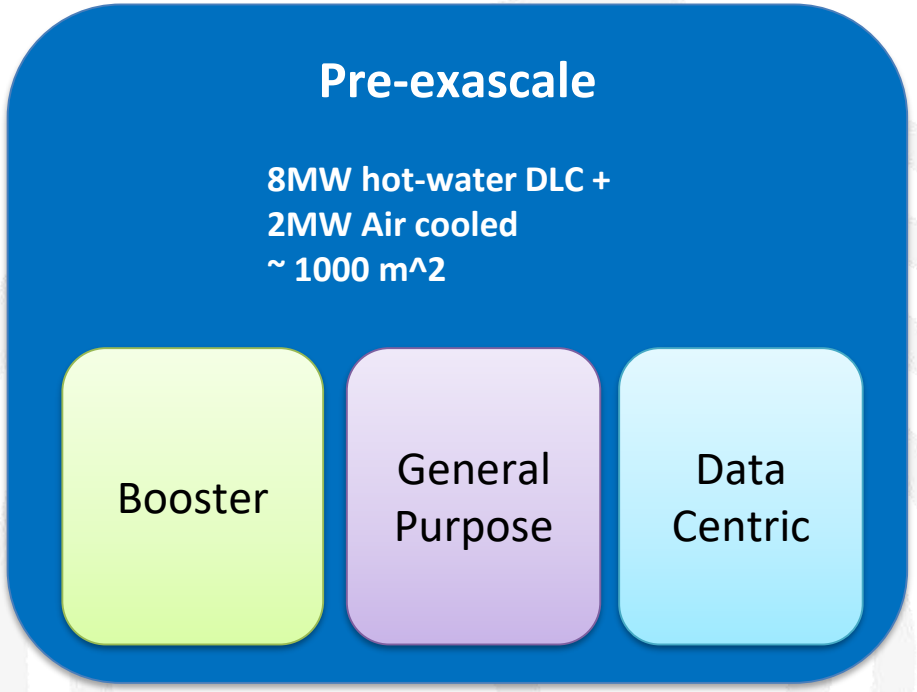


Balancing hardware needs with cost-efficiency & usability



Proposed Systems

System name	Leonardo
Modules	3 (booster, general purpose, data centric)
Number of computing nodes	????
Storage	Capacity: 150 PB, bandwidth: 1 TB/s
HPL Targeted Performance (peak)	150-180 PFlops (210-250 PFlops); Top 3
HPCG Targeted Performance	2.8-3.3 PFlops; Top 3
I/O	≥ 150 PB
Interconnection Bandwidth	≥ 200 Gb/s
Estimated Power consumption (after PUE)	8-9 MW (8.8-9.9 MW)



Use cases:

- 10x computing capability in a large set of key applications for science, industry and society (CoEs, HEP, Pharma, Oil&GAS), and keep the European leadership.
- gain sovereignty on strategic technologies for the European economic wealth, like Artificial Intelligence, Cybersecurity and Internet of Thing,
- tackle relevant and urgent societal challenges.

Experimental platform

	Phase 1 partition	Phase 2 partition
Installation date	Q1 2021	Q2-Q3 2022
Node architecture	Dual socket best in class ARM processor available beginning 2021.	Dual socket ARM processor with SVE extension, use of EPI GPP, if available.
Accelerator	2 PCI express FPGA card per node.	2 PCI express FPGA card per node, use of PCI card designed in Europe, if available
Memory per node	>= 256GByte DDR4	>= 256GByte DDR4 + HBM if available
Number of nodes	64	64
Price estimate (including non recoverable engineering costs)	800 k€	1'200 k€
Co-design and SW development effort	120 Person Months	180 Person Months

Objectives:

O1. To accelerate the development and the experimentation of novel software toolchains and solutions for DL specific DSA across the whole stack form hardware IP to high-level programming tools.

O2. To develop a novel "custom network of hardware accelerators as-a-Service" (DSAaaS) provisioning model able to support dynamic partitioning of heterogenous platforms accelerated with GPU, FPGA and DSA conceptually extending from single node (such as AWS EC2 F1) to cluster the provisioning of custom hardware accelerators.

O3. To support experimentation of novel DSA IP for HPC (e.g. for DL training) and the simulation of complex edge system composed of many low-power DSA IP (e.g. for DL inference and federated learning).

DEEP Projects

EPI European Processor Initiative

EUROEXA

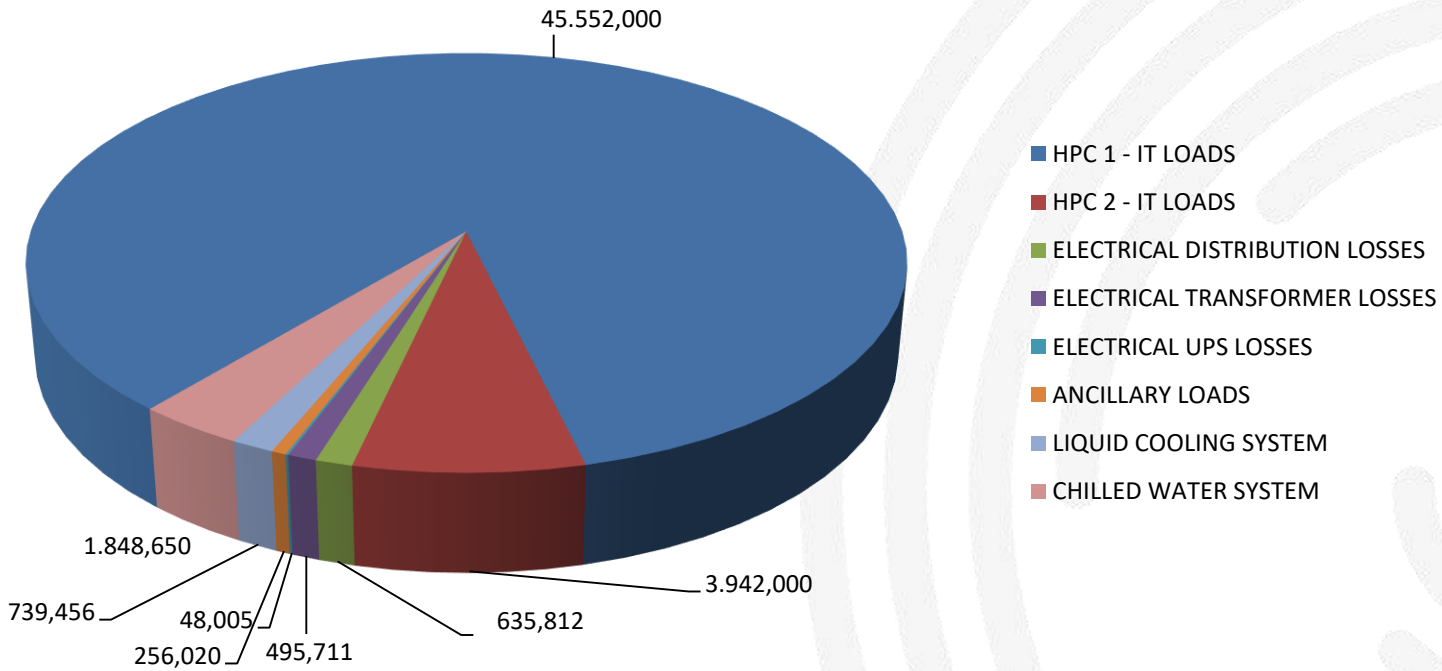
ExaNoDe

Leonardo "Module"
Experimental platform
ARM + FPGA (phase I)
EPI + FPGA (phase II)

EuroHPC exascale pilot

PUE

ANNUAL ENERGIES - PUE = 1,081296



Some considerations about the new infrastucture

- Currently we don’t know the architecture to be installed in the new machines
- But clearly:
  - The architecture of the new machine will be at the state of the art when it will start the production
  - Large part of the user’s codes need to be re-writen to be ready for the booster/accelerated partition
  - A smoot transition will be ensured by the general purposes partition
  - New possibilities will be ensured by the data centric partition (ML, AI, NVM,…)
  - Some experiments on new promising architecture will be possible thanks to the Experimental platform

Competitive calls

Thank You

m.guarrasi@cineca.it