European Strategy 2025-2026
- Computing -

G. Carlino
per il C3SN

Milano Bicocca - 4/2/25

Role of Computing in the Particle Physics Strategy

Why discussing Computing in the Particle Physics Strategy ?

The «HL-LHC era» experiments will produce a huge amount of data with increasing complexity whose management (production, simulation, reconstruction, analysis ...) can't be addressed without substantial changes in

- infrastructure and technology: data centres and IT resources
- software and computing models of the experiments

to keep costs to an acceptable level (flat budget model)

Hence Computing must play a strategic role in HEP

Role of Computing in the Particle Physics Strategy

The Computing Infrastructure

Evolution towards a sustainable, heterogeneous and multidisciplinary infrastructure

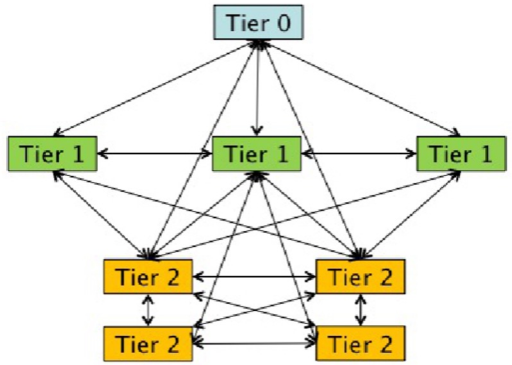
Hot topics in HEP
INFN is developing a strategy for the coming years on

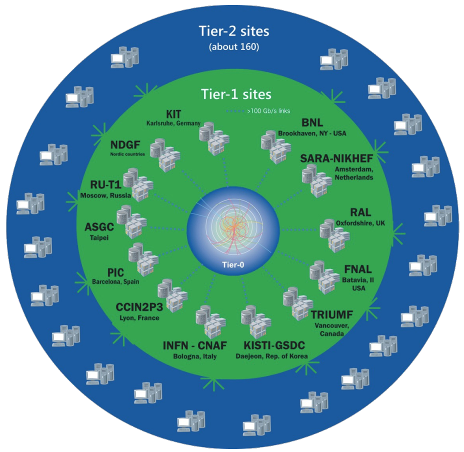
Artificial Intelligence and Quantum Computing

The WLCG Infrastructure

The current computing WLCG infrastructure is based on a grid of Tier1 and Tier2 centres (HTL centres) dating back about 25 years ago to the Monarc project

- 170 centres in more than 40 countries to provide computing and storage resources in a single infrastructure, to distribute and analyse LHC and simulated data
- centres with specific functions connected by a high bandwidth geographical network connections





INFN Scientific Computing Infrastructure

In the last two decades, INFN developed a big scientific computing infrastructure for HEP, part of the WLCG and EGI infrastructures , composed of 10 centres:

- 1 Tier1: CNAF @ BO
- 9 Tier2: BA – CT – LNF - LNL/PD - MI – NA – PI - RM1 - TO

connected with dedicated links at 240 Gbps (Tier1) and at 100 Gbps (Tier2) to the GARR backbone

Total compute and storage resources:

- CPU: ˜ 200.000 cores
- Storage: ˜ 140 PB disk and 250 PB tape

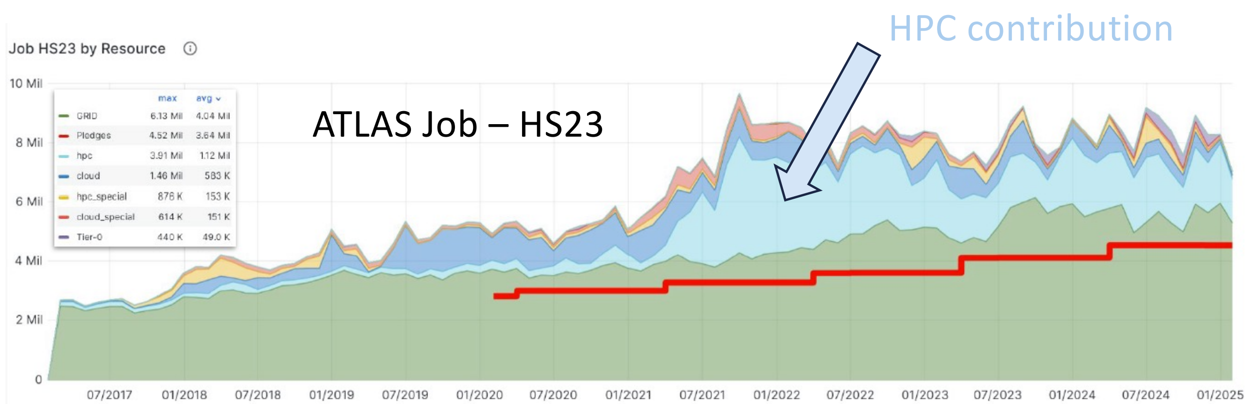
Key element facilitating interoperability between heterogeneous centres: development and adoption of a standardised set of protocols and tools for authorisation, authentication, job management and data access

The WLCG Infrastructure

Over the years

- Not only LHC, now extended to other partners: DuNE, Belle2, JUNO, Virgo ….
- Not only HTC centres, now using HPC and cloud centres mainly for non pledge resources
- Not only standard x86 processors, also GPUs for parallel applications or ARM for energetic sustainability

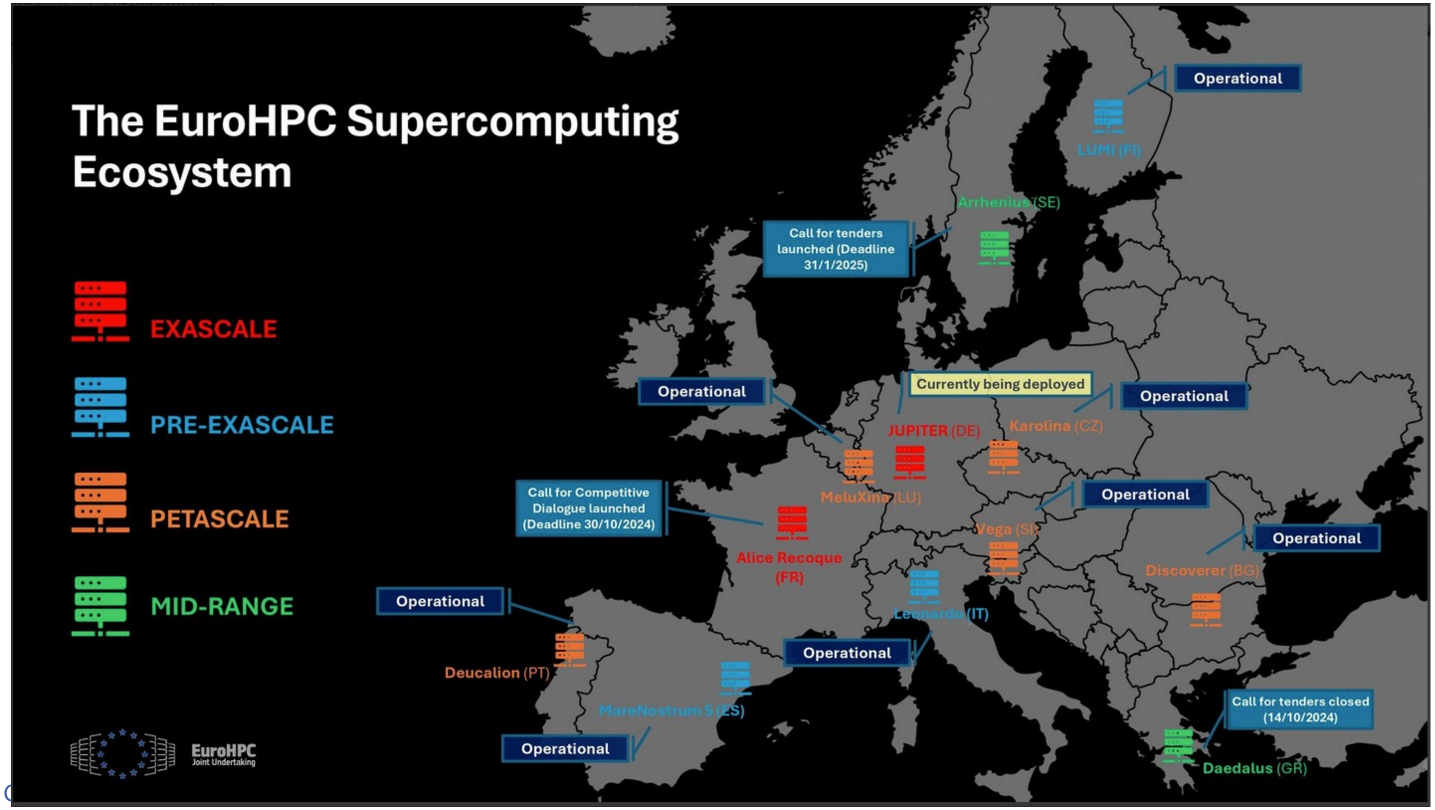
The WLCG infrastructure performed very well up to LHC Run3 but is no longer economically sustainable and FA are pushing towards interdisciplinary infrastructures, shared with other sciences



EuroHPC

High-Performance Computing is a very important area of interest for Europe, particularly as part of its wider strategy to improve digital sovereignty, innovation, and scientific research capabilities

The most recent initiatives is the European High-Performance Computing Joint Undertaking (EuroHPC JU): a flagship initiative aimed at developing a world-class supercomputing infrastructure at pre-exa- and exa-scale levels to support research and innovation in HPC technologies, software and applications.



HPC centres

HPC centres ideal for compute-intensive applications, as for example Lattice QCD, and less for HEP data-intensive applications

HEP experiments are using HPC centres to exploit the high computing power offered by CPU partitions

- In Italy Leonardo pre-exascale EuroHPC centre at CINECA tightly coupled with the INFN Tier-1 resources and transparently accessible from the Tier1 grid interfaces

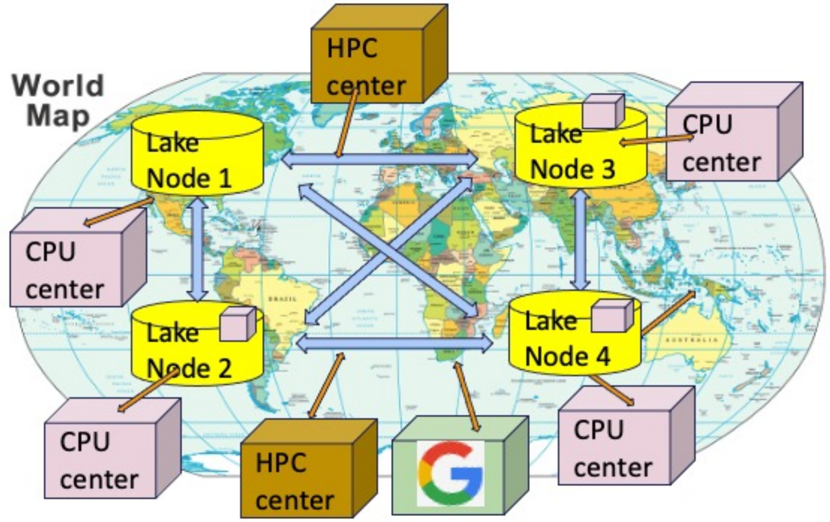
Nevertheless, **extensive use of HPC for HEP not trivial**: big challenge

- Future HPC centres, designed for AI and ML workloads, will adopt architectures largely based on GPU
  - HEP experiments must evolve the computing models increasing the GPU based workflow
- Technical issues: AAI systems, high bandwidth geographical network connections, access to remote data
- “Political” issues: co-management of the centres (e.g. different SLA)

A combined HTC and HPC infrastructure

INFN

Data Lake Model



DATA

few WLCG big centres to store an manage data
- data ownership -
data are the real treasure

CPU

CPU power provided by internal and external centres:
HTC - HPC – Cloud
HPC centres key element of the infrastructure
- HEP community, experimental and theoretical, should become a foundational part of the HPC strategy to have a role in the future centres technical designs

MIDDLEWARE

Cloud approach: PaaS orchestrator and workload offloading
- high level services built by communities
- central services managed by infrastructure

ICSC - an example in Italy

ICSC - Centro Nazionale di Ricerca in HPC, Big Data e Quantum Computing

ICSC - multidisciplinary

Public Research Institutions Founding Members: a pervasive initiative throughout Italy

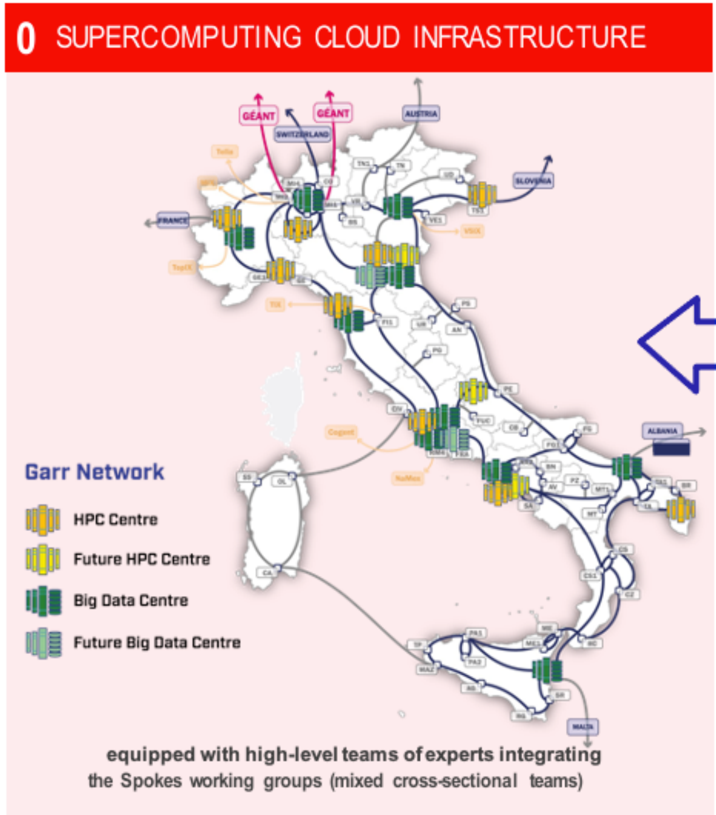
National Institutions

INFN, CINECA, ENEA, Consiglio Nazionale delle Ricerche, INGV, Consortium GARR, INAF

Hub Only

UNIMORE, UNIVERSITÀ DI PARMA, OGS

ICSC – HTC & HPC infrastructure supporting many domains



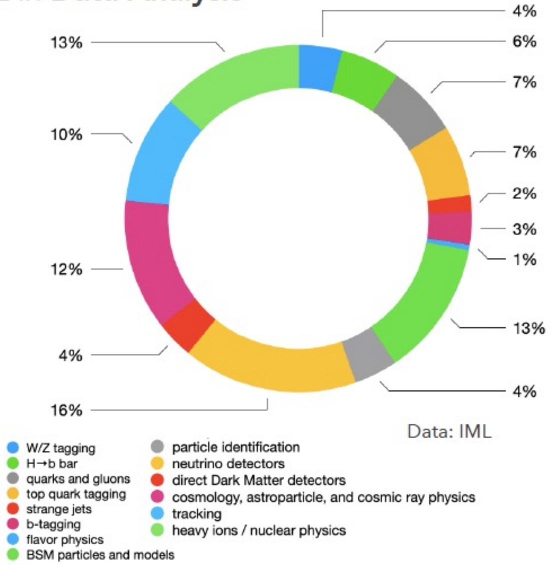
AI in HEP – to fully exploit the physics potential of data

AI has long been a part of HEP, with ML techniques first emerging over two decades ago as tools for data exploration

ML in **data acquisition** and **trigger**

- Bkg and trigger rate reduction
- Signal specific trigger paths
- Anomaly detection in data taking
- Unsupervised new physics mining

ML in **Data Analysis**



ML in Event **Simulation**

The production of simulated events (full/fast simulation) is extremely intense from the computation standpoint (up to the point it might impact the physics reach of the experiments). ML can help to reduce such load

.)

- Calorimeter shower surrogate simulator
- Analysis level simulator
- Pile-up overlay generator
- Monte Carlo integration
- ML-enabled fast-simulation
- Invertible full-simulation (probabilistic programming, …)

D. Bonaccorsi – Piano Triennale INFN 2024

ML in Event **Reconstruction**

Online/offline reconstruction might be partially replaced by surrogate models (approximate → faster) or by new algorithms (that might offer unprecedented performances)

- Charged particle tracking (GraphNN, vertexing, …)
- Calorimeter reconstruction (local, clustering, …)
- Particle flow (GraphNN, …)
- Particle identification (boosted jets, isolation, …)
- Pileup mitigation
- Energy regression (end-2-end, …)

ML in **Computing Operations**

Application of ML to **non-collision (meta-)data** might help to increase efficiency and reduce the need omg personpower in Ops, e.g. automating specific tasks, developing intelligent/adaptive systems, ultimately acting on the full chain - from data collection to data analysis - and make it more agile

- Detector control
- Data quality monitoring
- Operational intelligence
- Predictive maintenance

The INFN AI strategy

AI applications require very large computing power, offered by CPU and mainly GPU or accelerators like TPU and FPGA.

HPC centres are the ideal structures to handle such massive computational demands

INFN supports the creation of a pan-European initiative for research, society and productive systems, aimed at promoting and supporting the development of an AI ecosystem at continental level, both at fundamental and applicative levels:

- Distributed hybrid Infrastructure: leveraging the existing HPC infrastructure coordinated by EuroHPC to be preferred to a very expensive centralised one
- Institute for AI: funding a federated virtual institute to aggregate national initiatives, to be preferred to a physical institute, to coordinate the high-level activities of the research institutes and companies involved and to support funding opportunities for researchers' training and mobility

The INFN AI strategy

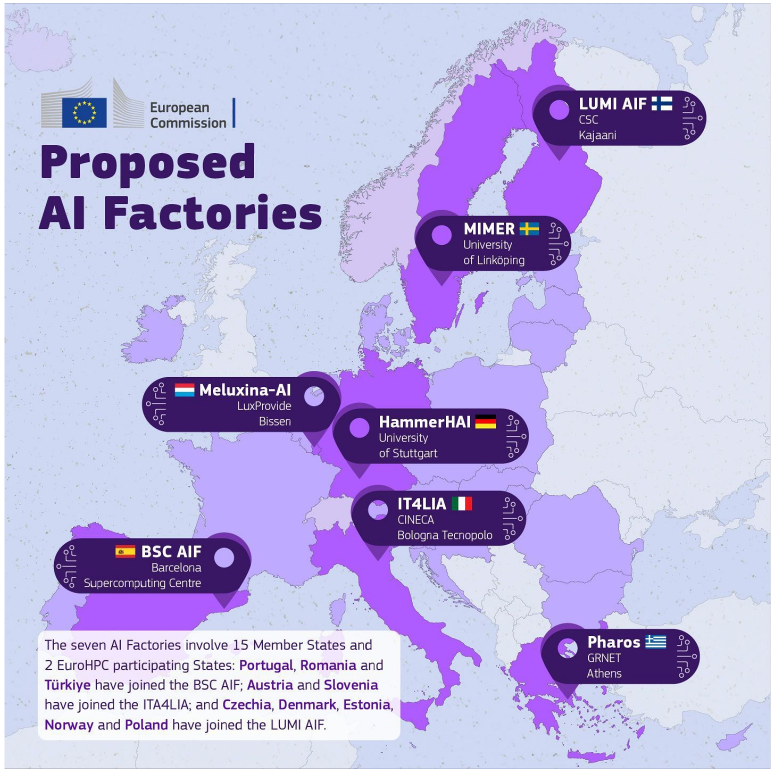
Strategic research focus

- develop the datalake model: middleware to operate a distributed computing infrastructure with standard and AI-specialised hardware and large data repository
- incorporate facilities dedicated to handling sensitive data, in compliance with ISO standard

- progress on foundational AI algorithms tailored for scientific research: Quantum AI, Explainable AI, Physics-Informed ML, generative models for detector simulation ….

Again, the role of WLCG in the EuroHPC coordination will be crucial

AI – European map



Network of **7 supercomputers** for AI coordinated by **EuroHPC** JU.
Funding 1,5 G€

- 6 new generation supercomputers: Italy, Finland, Spain (3 HPC pre-exascale centres), Luxemburg, Sweden and Greece

Italy – IT4LIA AI Factory Project

- Proposed by Italy with CINECA hosting entity, the participation of Austria and Slovenia and involving ICSC, INFN and other Italian institutions
- Location: Bologna Tecnopolo, together with Leonardo's evolution Lisa
- Funding: 430 M€ - joint funding between MUR e Agenzia Cybersicurezza Nazionale, CINECA, INFN, Regione Emilia-Romagna, Agenzia ItaliaMeteo, Fondazione per l'IA e Fondazione Bruno Kessler.

Quantum Computing

- Quantum computers hold exceptional promise to transform computation
  - N "perfect" qubits describe 2^N states at the same time
  - exponential increase of theoretical computing power and information storage with increasing N
- The realization of fault tolerant QC requires overcoming many significant challenges
  - current NISQ devices can be exploited, by utilising quantum error correction and mitigation techniques, and offer already today advantages in terms of data processing speed and simulations of complex quantum phenomena
- Developing quantum algorithms is very challenging
- Active involvement of HEP theory and experiments

https://arxiv.org/pdf/2307.03236

  - Theory: scattering phenomena, string breaking, quenching, phase transitions, and quantum simulations of (simplified) lattice gauge theories
  - Experiment: QML techniques can be exploited for signal processing, detector reconstruction and simulation

Quantum Computing: actions in HEP

Increase the effort on Quantum Computing

- **Hardware**: national (e.g. CSN5) and international (e.g. DRD5) initiatives
- **Software**: increase the existing R&D efforts
  - expanded use cases, e.g. charged particle tracking, jet tagging, theory simulations, on the available hardware and classical emulators, with the goal of gaining experience and identifying the most promising avenues
  - target NISQ devices on a shorter-medium term. Error mitigation and correction techniques are essential
- **Middleware**: components in rapid development
  - quantum control firmware, operating systems, compilers, software development kits (SDKs), programming languages, runtime systems, workflow scheduling, high-level libraries, domain-specific toolkits
  - existing INFN efforts in e.g. tensor networks and their application in Level-1 triggers, the QIBO library and contributions to open-source tools

Coordinate with, and contribute to strategic initiatives

- e.g. working group on the Italian National Strategy for Quantum Technologies

Quantum Computing: actions in HEP

Streamline resource acquisition and access

- Cloud access to existing devices (IBM, QuEra, IQM, D-Wave, …)
- Seamless integration to research infrastructures (e.g. ICSC, CERN,…)
- Involvement in the design and federation of future facilities (e.g. EuroQCS)
- Large-scale access to emulators on classical hardware (e.g. ICSC, EuroHPC, …)

Training, support and career recognition

- Quantum Computing involves many scientific domains
- Paradigm shift in HEP algorithms, integration into HEP infrastructures
  - Training through e.g. hackathons, courses, seminar, workshops
- Recognise QC careers in HEP at various levels
  - PhD, postdoc, researcher, technician, engineer…

Recommendations

The «HL-LHC era» is posing very interesting and complicated challenges to the HEP community

Experiments must continue the efforts to reduce the resources needs and the energy consumption:

- higher concurrency increasing the GPU-based workflow
- algorithms optimization to reduce reconstruction time, with increased pile-up, and optimization of full sim, increase use of fast sim and AI-based sim
- use of non-x86 architectures
- adopt AI techniques that can be more efficient and less resource-intensive

Infrastructure

- develop a robust Data Lake model at European Level including external HPC centres
- tight collaboration between WLCG and EuroHPC communities to create and fully exploit a combined multi-science infrastructure compliant with HEP technical needs
    - Involvement in the design and federation of future facilities AI-specialized

Emerging (or emerged) technologies

- Increase effort at national and European levels on AI and QC activities