

# SOSC22 – Introduction to INFN and to its Vision for Computing

Perugia, 28/11/2022

Davide Salomoni ( <u>davide@infn.it</u> )

#### INFŃ Istituto Nazionale di Fisica Nucleare

The National Institute for Nuclear Physics (INFN) is the Italian research agency dedicated to the study of the fundamental constituents of matter and the laws that govern them.

## The **5 research lines** and the INFN National Scientific Committees



# The INFN Facilities



## Computing @ INFN



In the 2000's, ten international centers were selected to host the Worldwide LHC Computing Grid (WLCG):

- In Italy, this was the Tier-1 at CNAF (red in the picture)
- 9 additional "Tier-2" centers were then added, at LNL, LNF, Turin, Milan, Pisa, Rome, Naples, Bari, Catania (yellow in the picture).
- Then came the GRID, the Cloud, and other technological developments.
- All these centers are still operational, even if their size has increased ~100x since then, and their interconnectivity now reaches multiples of 100 Gbps, thanks to the GARR-X network.
- Collectively, our distributed infrastructure currently offers about 140,000 CPU cores, 120PB of enterprise-level disk space, 100PB of tape storage.



#### INFN **HPC** World center Map CPU Lake Lake center Node 3 Node 1 CPU center amazon Lake Lake Node 4 Node 2 CPU CPU **HPC** center center center

# The Cloud / Data Lake Model

- A blueprint defined by international physics collaborations to:
  - Keep the real value from the 1. experiments safe, with:
    - (RAW) data and a solid baseline of CPU in owned and stable sites
      - Allow multiple CPU resources to • join, even temporarily, possibly choosing the cheapest at any given moment
    - Solid networking
      - Using caches or streaming to • access data
  - Reduce requirements for 2. computing resources, through:
    - Commercial Clouds
    - Other sciences' resources
      - SKA, CTA, Dune, Genomics, ...
    - HPC systems *if possible*

The main drivers for the next 10 years of scientific computing for INFN



### 1. Infrastructure

- Renew infrastructures to be ready for the High Luminosity-LHC (HL-LHC) era, to ~2035 and beyond
- Use more compact computing (from today's ~20 kW/rack to 80 or more)
- Lower the PUE (power usage effectiveness), be greener
- Extend and expand networking for a future-proof infrastructure

#### SOSC22 covers some of these points

2. Hardware, Software, Services

- Foster and simplify the utilization of more viable technologies (€/task or J/task), like GPUs, FPGA, down to Quantum when available
- Be more efficient, elastic and resilient:
  - Pervasive use of geographically distributed storage ("the Datalake")
  - Abstraction from physical machines, with a national pool of resources and high-level services ("the Cloud")
  - Elastic extension to external providers, such as traditional HPC centers or to other cloud providers ("dynamic federations")
  - Evolution and exploitation of AI in multiple fields and ways.

### **Digital Compass:** The European way for the Digital Decade



#### SKILLS

ICT Specialists: 20 million + Gender convergence Basic Digital Skills: min 80% of population

### INFRASTRUCTURES

**Connectivity:** Gigabit for everyone, 5G everywhere **Cutting edge semiconductors:** double EU share in global production **Data – Edge & Cloud:** 10,000 climate neutral highly secure edge nodes **Computing:** first computer with quantum acceleration

#### **BUSINESS**

**Tech up-take**: 75% of EU companies using Cloud/AI/Big Data **Innovators:** grow scale ups & finance to double EU Unicorns

Late adopters: more than 90% European SMEs reach at least a basic level of digital intensity



Davide Salomoni

SOSC22 - Perugia, 28/11/2022

### The European Data Strategy Shaping Europe's digital future (1/2)





SHAPING EUROPE'S DIGITAL FUTURE

#### Global Data Volume will grow:

#### Value of data economy (EU27)

#### Number of data professionals (EU27)



### **The European Data Strategy** Shaping Europe's digital future (2/2)





SHAPING EUROPE'S DIGITAL FUTURE

#### Creating a single market for data will make EU more competitive globally and will lead to innovative processes, products and services

Industrial and commercial data are key drivers of digital economy. The European Data Strategy will make more data available for use in economy and society, while keeping the data control to those who generate it.

#### Examples of industrial and commercial data use



Jet engines filled with thousands of sensors collect and transmit data back to ensure efficient operation



Real-time traffic avoidance navigation can save up to **730 million hours**. This represents up to **€20 billion** in labor cost



Real-time notification of delayed trains can save **27 million working hours**. This amounts to **€740 million** in labor costs



Wind farms industrial data to reduce visual impact and optimize wind power



Better allocation of resources to fight malaria could save up to **€5 billion** in health care cost globally.

Davide Salomoni

# INFN Cloud, <u>https://www.cloud.infn.it/</u>



- In production since March 2021.
- The initial seed of a National Datalake for research and beyond, building on (existing | renewed | new) e-Infrastructures.
- The **base of the evolution** of the INFN Distributed Computing vision.
- Built on a **thin middleware layer** running on top of *federated clouds*, decoupling physical and logical views via a **service composition** mechanism.
- The **INFN foundation** for all the NRRP computing-related initiatives (more on this later).
- A **multi-site**, **federated Cloud** infrastructure integrating HPC and HTC resources.

INFN

# The "Big Data Valley"

- In 2017, Bologna won a bid to host the "European Centre for Medium-Range Weather Forecasts" (ECMWF)
- The Emilia-Romagna region decided to repurpose the *"Manifattura Tabacchi \*"* area in Bologna to host a technology district, for ECMWF <u>and more</u>: the Tecnopolo

(\*a former tobacco factory)

How it will be



# INFN & CINECA moving to the Tecnopolo



• ECMWF is up and running since September 2021











- The CINECA ("C2") and INFN ("B5") barrels are expected to be ready by:
- November 2022 (CINECA)
  - ~Mid 2023 (INFN)
- Two phases are expected:
  - Phase-1 (2023-2025): Leonardo + CNAF data center relocated. Total 13 MW.
  - Phase-2 (2025+): infrastructure up to 23 MW ready for post-exascale HPC and for the next generation of scientific experiments.

# LEONARDO is inaugurated: Europe welcomes a new world-leading supercomputer

Today, the European High Performance Computing Joint Undertaking (EuroHPC JU) and CINECA officially inaugurated LEONARDO, the newest EuroHPC pre-exascale supercomputer ranked as the 4th fastest in the world, at the Bologna Technopole in Italy.



### Bologna, 24/11/2022





# The Italian National Recovery and Resilience Plan (NRRP)







#### Federazione Cloud

Modello: integrazione «debole» di cloud indipendenti

In uso sulla cloud INFN

Tutti i centri INFN adotteranno il modello nell'arco del progetto ICSC

Estensione a CINECA e ad altri provider del centro nazionale (CMCC, ...) Attività già iniziate nell'ambito del POR-FESR SUPER dell'Emilia Romagna Federazione integrata dei sistemi Tier1 e dei sistemi Europei di EuroHPC mediante rete ad iperconnetività



Italian Research Center on High-Performance Computing, Big Data and Quantum Computing

Proof of concept previsto a M8

ICSC

Centro Nazionale HPC, Big Data e Quantum Computing



Cloud national infrastructure for supercomputing. Hub & Spoke organization:

10 vertical spokes for technology developments and software applications



### ICSC Kick-off Meeting Bologna, 25-26/11/2022



### The Big Data Technopole, Bologna

ECMWF Data Center

CINECA & INFN Exascale Supercomputer Center

Civic Protection and agency for development and innovation

Meteo National Agency

Bologna

IFAB - International foundation Big data

ICSC

University Center

Biobank and Life Science

SENEAPEester8/11,



Big Data Association and Foundation

> Conference and Educational Center

**Innovation Center** 

Competence Center Industry 4.0







by the European Union

### **The Scenario** Why a National research Center on HPC, BD & QC?



With the current Data explosion...

- An unprecedented amount of data is going to be produced
- The real competitiveness challenge is extracting value from data
- Supercomputing, simulation, AI, high-performance data analytics and Big Data are essential for innovation and growth in a datadriven society

... need for an ambitious Italian strategy ...



- Europe has a clear strategy (e.g. EuroHPC, EOSC, EPI, Chip Act, Quantum Flagship) - European Data Strategy
- People, businesses and organisations should be empowered to make better decisions based on insights from data

... to "close the gap" with best in class

- First actions from 2015: Bologna's Technopole, ECMWF Data Centre, Leonardo preexascale supercomputer
- A step forward based on 5 pillars

### The ICSC aim and objectives



Create the **national digital infrastructure** for research and innovation, starting from the existing HPC, HTC and Big Data infrastructures ...

... evolving towards a **cloud datalake** model accessible by the scientific and industrial communities through flexible and uniform cloud web interfaces, relying on a high-level support team ...

... form a globally attractive **ecosystem based on strategic public-private partnerships** to fully exploit top level digital infrastructure for scientific and technical computing and promote the development of new computing technologies



# The ICSC will include ten **thematic Spokes** and one **Infrastructure spoke**





10



# How all this fits with SOSC22

- At SOSC22, you will work with some of the state-of-the-art technologies that form the core of the INFN vision on computing.
- INFN has the ambition to create, evolve and operate a vendorneutral, open, scalable and flexible "data lake" that serves much more than just INFN users and experiments.
  - This will become a key asset for fundamental, applied and industrial research in Italy and beyond and will be based on continuing international cooperation.





# Have fun at SOSC22!