Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

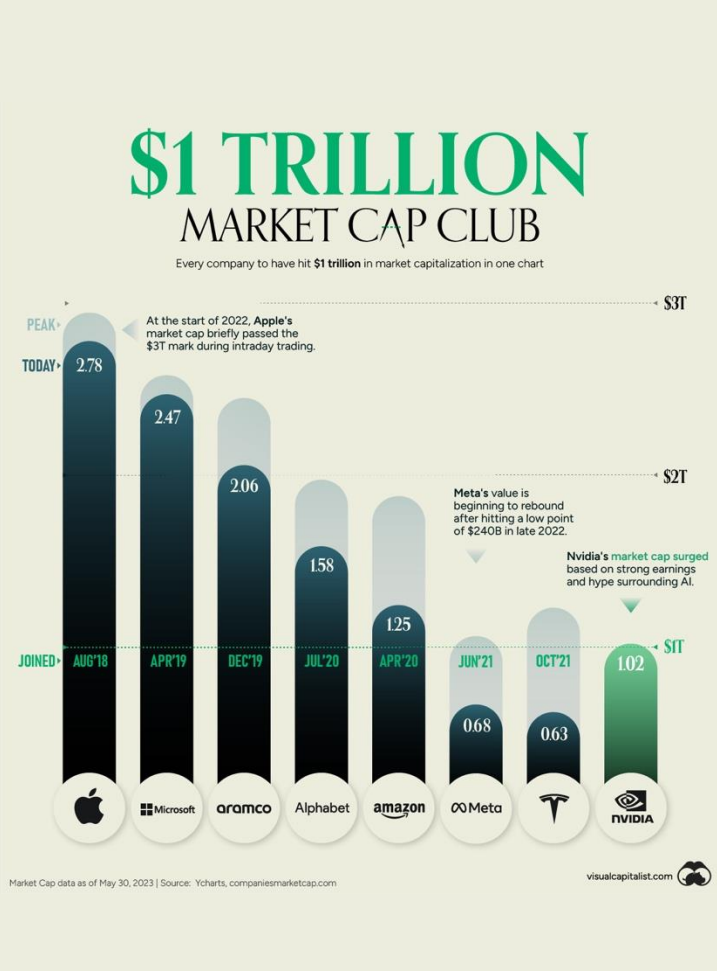
ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Why the SOSC, after all?

Davide Salomoni, ICSC Innovation Manager – davide@supercomputing-icsc.it

SOSC 2024, Bologna 2-6 December 2024

May 23



June 24

August 24

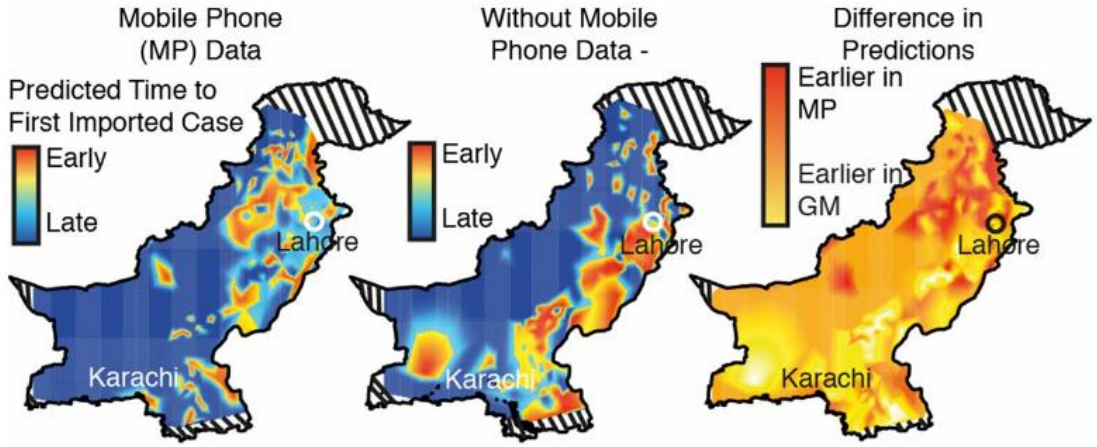
Why do we need all this?

- Because we want to **collect, analyze and manage [big] data** and **extract value from them**.
  - Right, but what do we mean with this *big data* thing?
- Let's start with an example from some years ago (2013-2015).
  - "Impact of human mobility on the emergence of dengue epidemics in Pakistan" – see https://www.pnas.org/doi/10.1073/pnas.1504964112

Fig. 1.

Human mobility dynamics in Pakistan. (A) Population density (red, high density; yellow, low density) and mobile phone tower coverage from the mobile phone operator in Pakistan (colored in gray) per tehsil. (B) The top routes of travel between pairs of tehsils in Pakistan. A line is drawn if at least 20,000 trips occurred between the origin and destination between June and December 2013. The top routes occur between Karachi and cities in northern Punjab province, particularly Lahore tehsil. (C) Relative direction and volume of travel. For each trip, we calculated the distance traveled from the origin and the destination. The origin location was centered at 0,0 and the longitude distance and latitude distance to the destination are shown. Although many trips occurred over short distances, a substantial amount of travel occurred between the southeastern and northern parts of the country, reflecting the geography and population distribution of Pakistan.

Fig. 4.



The estimated spatial spread of imported dengue. Using the modeled dengue dynamics in Karachi and mobility measured from the mobile phone data or a diffusion model, we estimated the time of the first introduced case to the rest of the country. The mobile phone data predict the earliest introductions in eastern Pakistan near Lahore and inland toward Swat Valley (Mingora). In comparison, the mobility model predicts early introductions in southern Pakistan with few introductions in Mingora. These differences are highlighted in the difference in predictions plot—the number of days earlier (red) from the mobile phone predictions or earlier (yellow) from the diffusion model (without the mobile phone data).

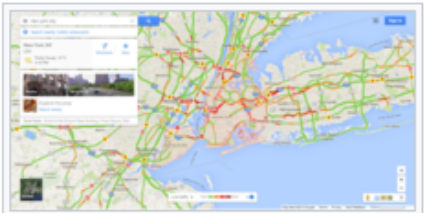
Mobile phone data provide dynamic population mobility estimates that can be combined with infectious disease surveillance data and seasonally varying environmental data to map these changing patterns of vulnerability in a country where dengue outbreaks are emerging and irregular in many regions. Because these data are continuously being collected by mobile phone operators, these methods could be integrated into national control programs in near real time.

Another example: traffic maps

- Google traffic maps work by analyzing the GPS-determined locations transmitted to Google by a **large number** of **mobile phone users**
- By calculating the speed of users along a length of road, Google can **generate a live traffic map**
- Google excludes **anomalies such as postal vehicles** which make frequent stops

Traffic conditions [edit]

In 2007, Google began offering traffic data as a colored overlay on top of roads and motorways to represent the speed of vehicles on particular roads. Crowdsourcing is used to obtain the GPS-determined locations of a large number of cellphone users, from which live traffic maps are produced.[64][65][66]



Screenshot of Google Maps with traffic option enabled

Google has stated that the speed and location information it collects to calculate traffic conditions is anonymous.[67] Options available in each phone's settings allow users not to share information about their location with Google Maps.[68] Google stated, "Once you disable or opt out of My Location, Maps will not continue to send radio information back to Google servers to determine your handset's approximate location".[69][*failed verification*]

All very nice, however…

An Artist Used 99 Phones to Fake a Google Maps Traffic Jam

What do these two examples have in common?

There is a lot of data, coming from multiple sources

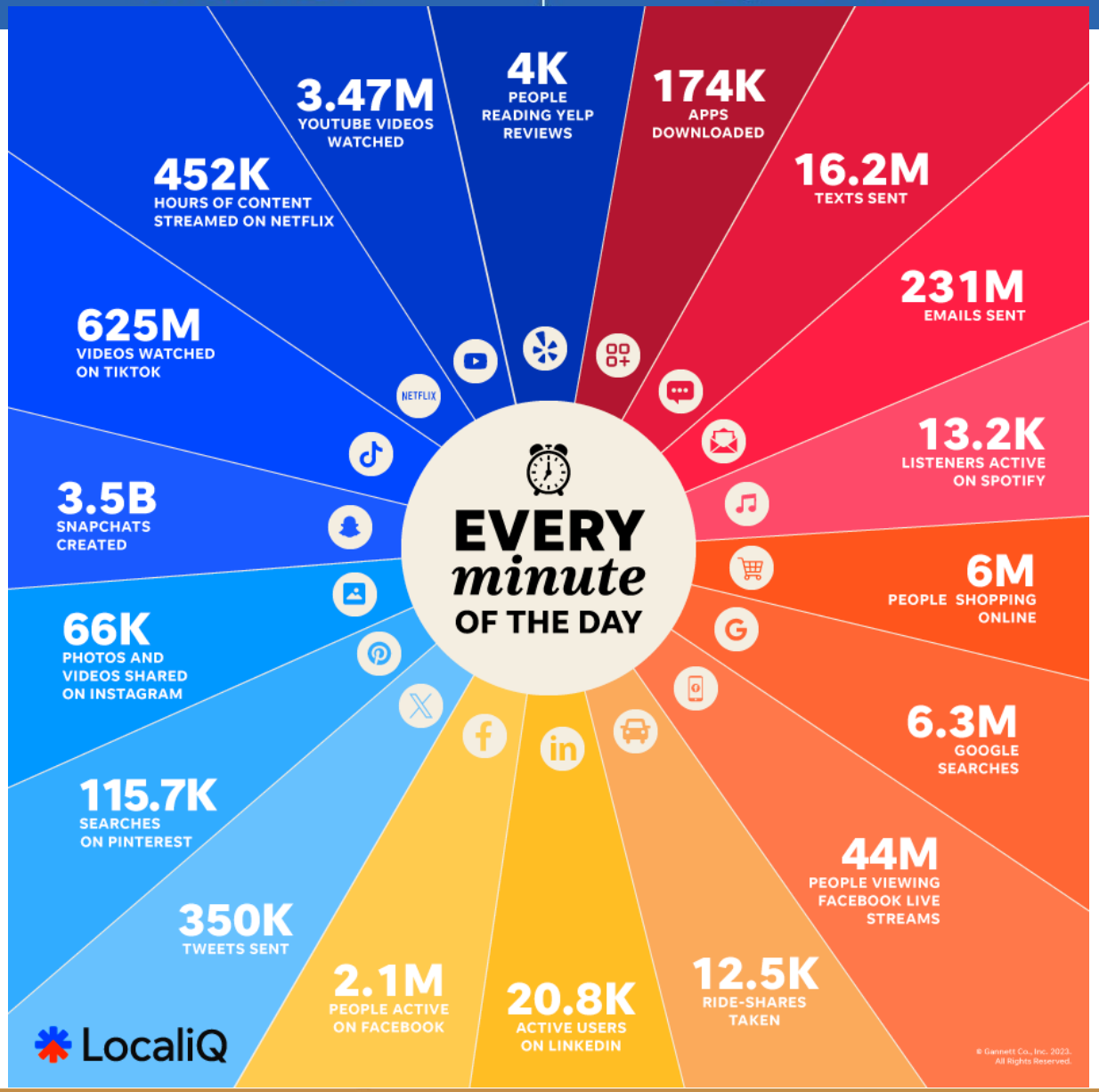
Data flows without a pre-determined schedule (=> highly *dynamic* environments)

We need a technology capable of processing data so that strategic decisions can be swiftly taken

We need ways to ensure data reliability

Data are based only on 5 numbers: 3 geographical coordinates, time and an identifier of the source (this is a simple case, we often have many more coordinates/metrics)

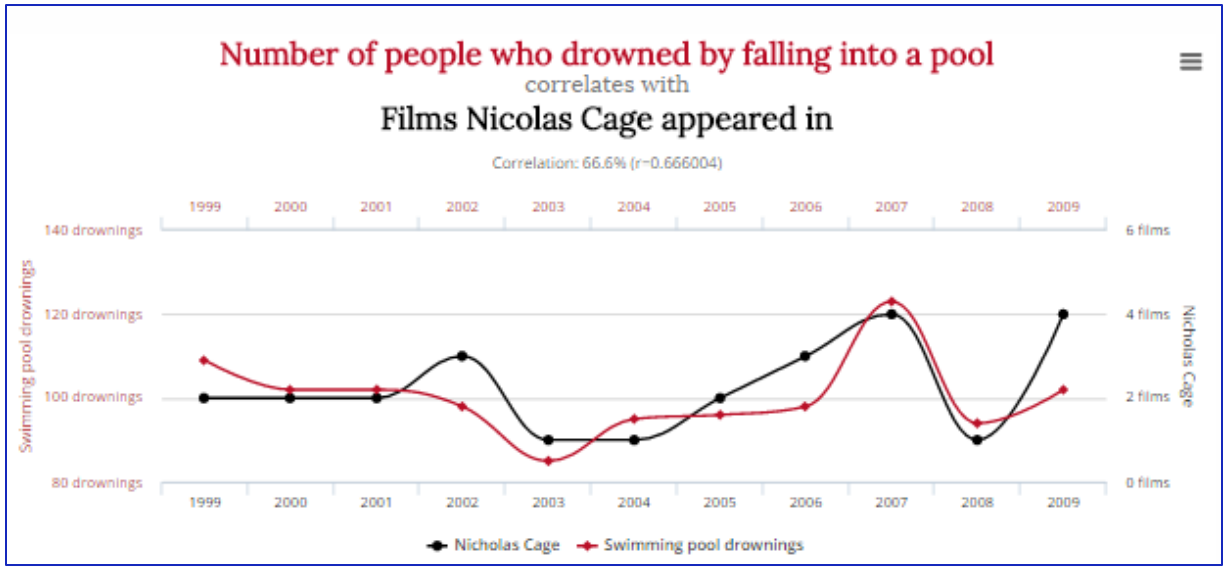
We can extract information from data, that may have a significant impact on our lives

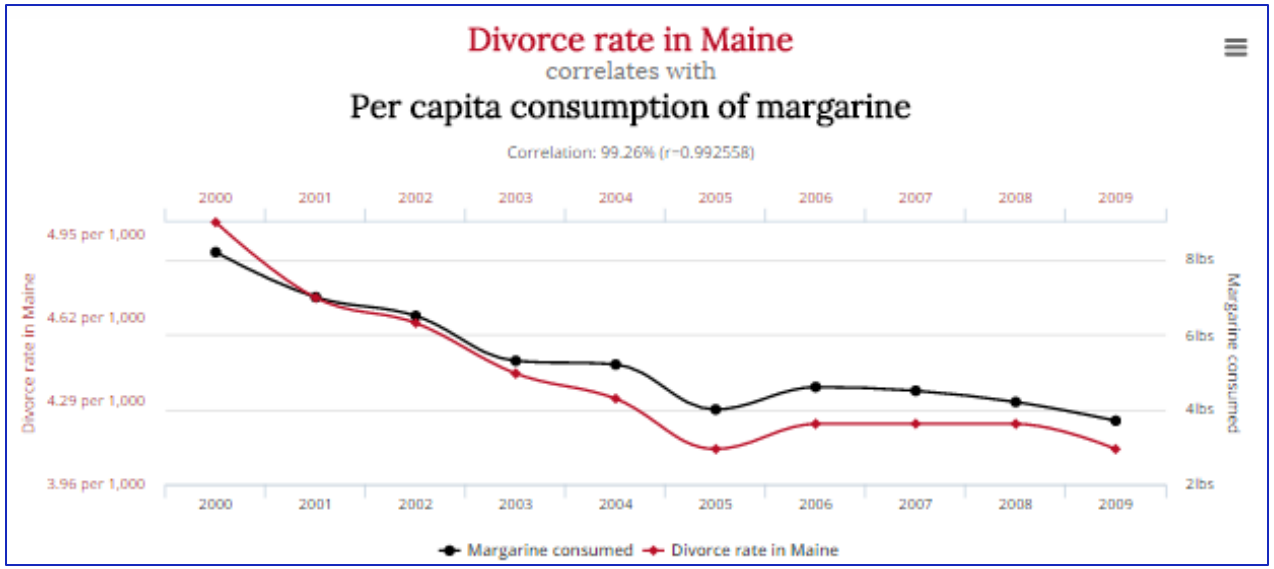

EVERY minute OF THE DAY
3.47M YOUTUBE VIDEOS WATCHED
4K PEOPLE READING YELP REVIEWS
174K APPS DOWNLOADED
16.2M TEXTS SENT
231M EMAILS SENT
13.2K LISTENERS ACTIVE ON SPOTIFY
6M PEOPLE SHOPPING ONLINE
6.3M GOOGLE SEARCHES
44M PEOPLE VIEWING FACEBOOK LIVE STREAMS
12.5K RIDE-SHARES TAKEN
20.8K ACTIVE USERS ON LINKEDIN
2.1M PEOPLE ACTIVE ON FACEBOOK
350K TWEETS SENT
115.7K SEARCHES ON PINTEREST
66K PHOTOS AND VIDEOS SHARED ON INSTAGRAM
3.5B SNAPCHATS CREATED
625M VIDEOS WATCHED ON TIKTOK
452K HOURS OF CONTENT STREAMED ON NETFLIX
NETFLIX
LocaliQ
© Gannett Co., Inc. 2023. All Rights Reserved.

The 5 Big Data "V" keywords

Volume	• We need ways (and premises) to transfer and store a lot of data
Velocity	• Rapid transactions, real-time streaming (Netflix, YouTube, in general any device collecting data)
Variety	• Many heterogeneous sources, data with different formats (texts, images, video, audio, other types od data), with or without a pre-defined structure
Veracity	• Data must be accurate and reliable
Value	• The transformation of a "data tsunami" into useful, valuable information

Value: be wary of false correlations ("correlation is not causation")





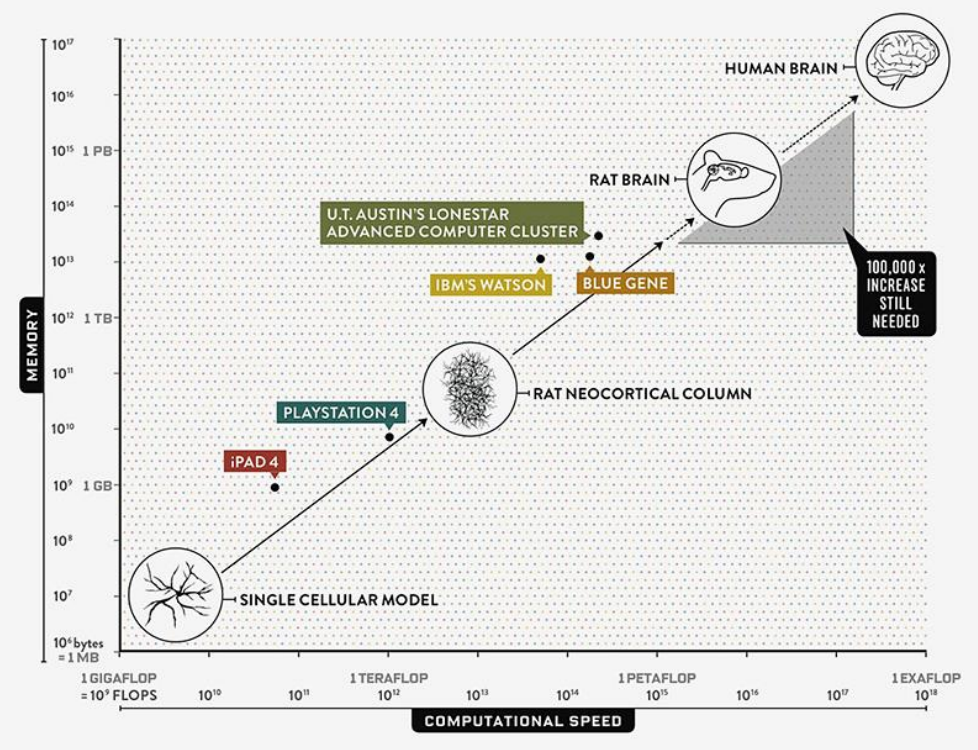
OK, but what do we do with all this data now?

- We have somehow collected and located it “somewhere” (in *data centers*); we should now analyze and extract some value from it.

- To do this, we need **computing power**.
  - Plus, ways and know-how to use it…

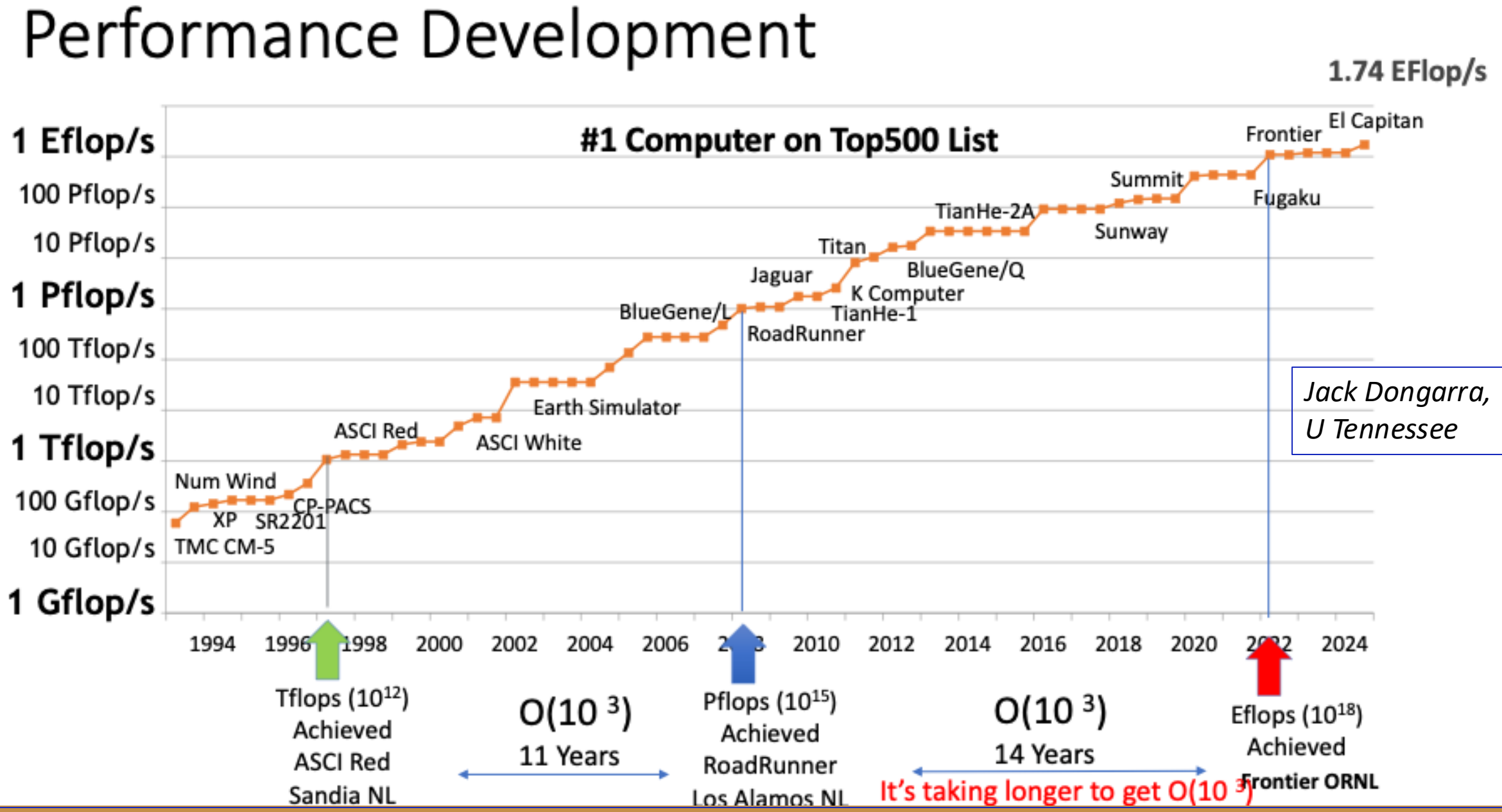
How powerful or "fast' is a computer?

- There are several ways to indicate the "computing power" or "computational speed" of a computer
  - An often-used metric is called FLOPS = **Floating point operations per second**. It measures a computer "speed" in terms of how many *real* (i.e., not simply integer) operations a computer can perform in one second.
- On the right, see a figure from a few years ago, where 1 Exaflop (10^{18} flops) is about equivalent to the computing power of a single human brain.
- The **Leonardo supercomputer** installed at the Bologna Technopole currently has a speed of ~0.25 Exaflops.
  - A lot? Not so much? (that's 250 million billion operations per second)
- The most powerful supercomputing in the world (El Capitan, USA, 2024) has a speed of 1.7 Exaflops (about 7 times Leonardo's speed).
- However: the Apollo 11 on-board computer (Moon, 1968) had a computing speed about *120 million times less* then an iPhone.
  - On the other hand, an «old» iPhone 12 has a speed of about 11 Teraflops. This is ~5000 times more than the speed of supercomputers operating in the 80s (which occupied about 1.5 m^2 and weighted about 2.5 tonnes).



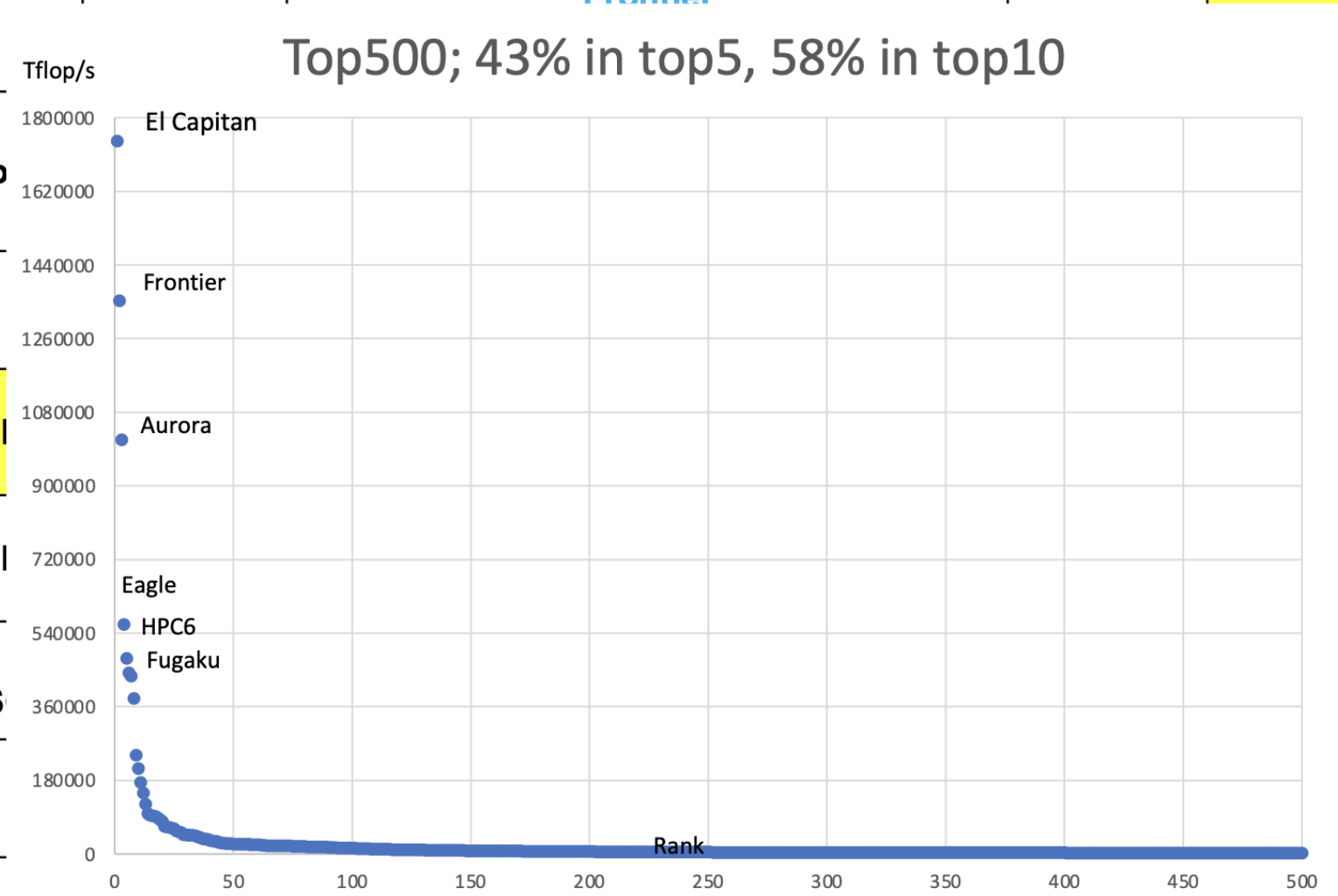
https://www.visualcapitalist.com/visualizing-trillion-fold-increase-computing-power/

High-Performance Computing (HPC)



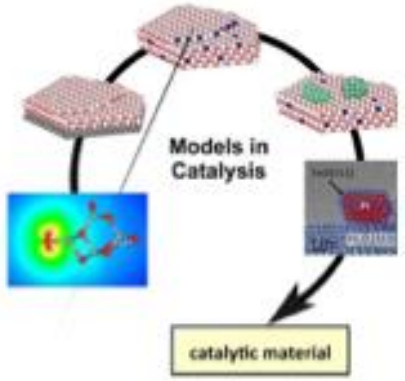
Jack Dongarra, U Tennessee

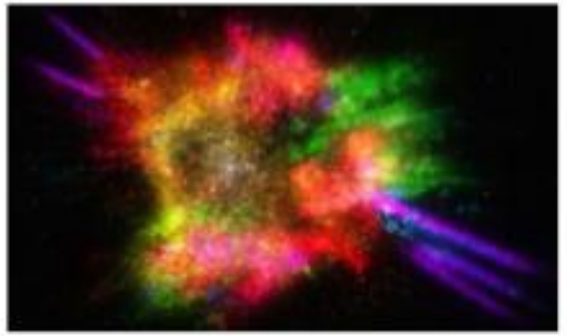
#	Site	Manufacturer	TOP10 Computer of the TOP500	Country	Cores	Rmax [Pflops]	Power [MW]
1	Lawrence Livermore National Laboratory	HPE	El Capitan HPE Cray EX255a, AMD EPYC 24C 1.8GHz, Instinct MI300A, Slingshot-11	USA	11,039,616	1,742	29.6
2	Oak Ridge National Laboratory		Frontier		,176	1,353	24.6
3	Argonne National Laborato				,808	1,012	38.7
4	Microsoft Azure				,200	561.2	
5	Eni S.p.A. Center for Computational Science				,520	477.9	8.5
6	RIKEN Center for Computational Science				,848	442.0	29.9
7	Swiss National Supercomputing Centre (CS				,600	434.9	7.1
8	EuroHPC / CSC				,704	379.7	7.1
9	EuroHPC / CINECA	EVIDEN	Atos BullSequana XH2000, Xeon 32C 2.6GHz, NVIDIA A100, HDR Infiniband	Italy	1,824,768	241.2	7.5
10	Lawrence Livermore National Laboratory	HPE	Tuolumne HPE Cray EX255a,	USA	1,161,216	208.1	3.4

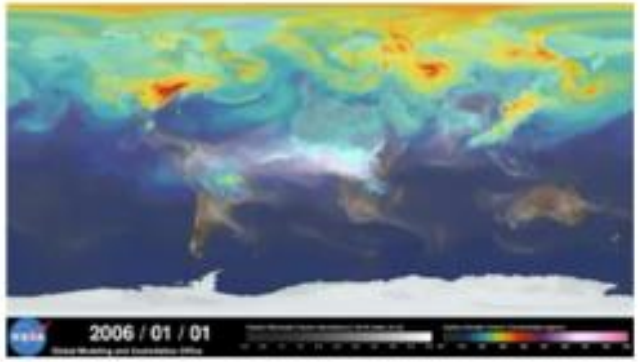


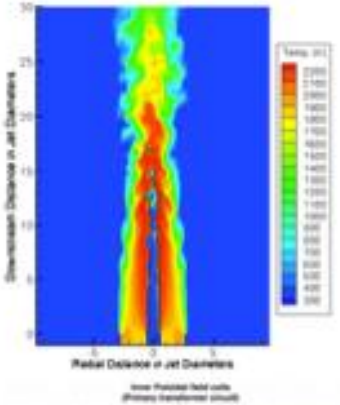
Today's Top HPC Systems Used to do Simulations

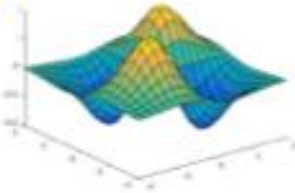
- *Climate*
- *Combustion*
- *Nuclear Reactors*
- *Catalysis*
- *Electric Grid*
- *Fusion*
- *Stockpile*
- *Supernovae*
- *Materials*
- *Digital Twins*
- *Accelerators*
- …

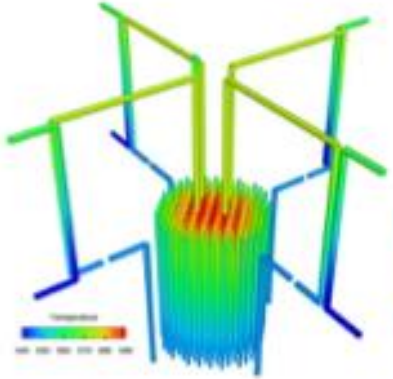


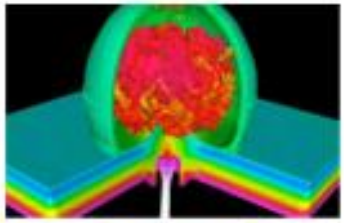


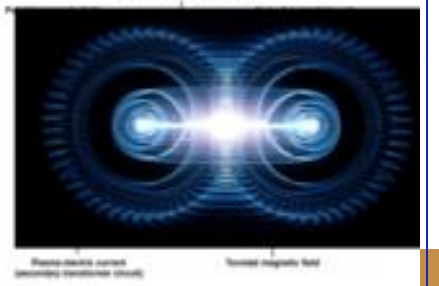












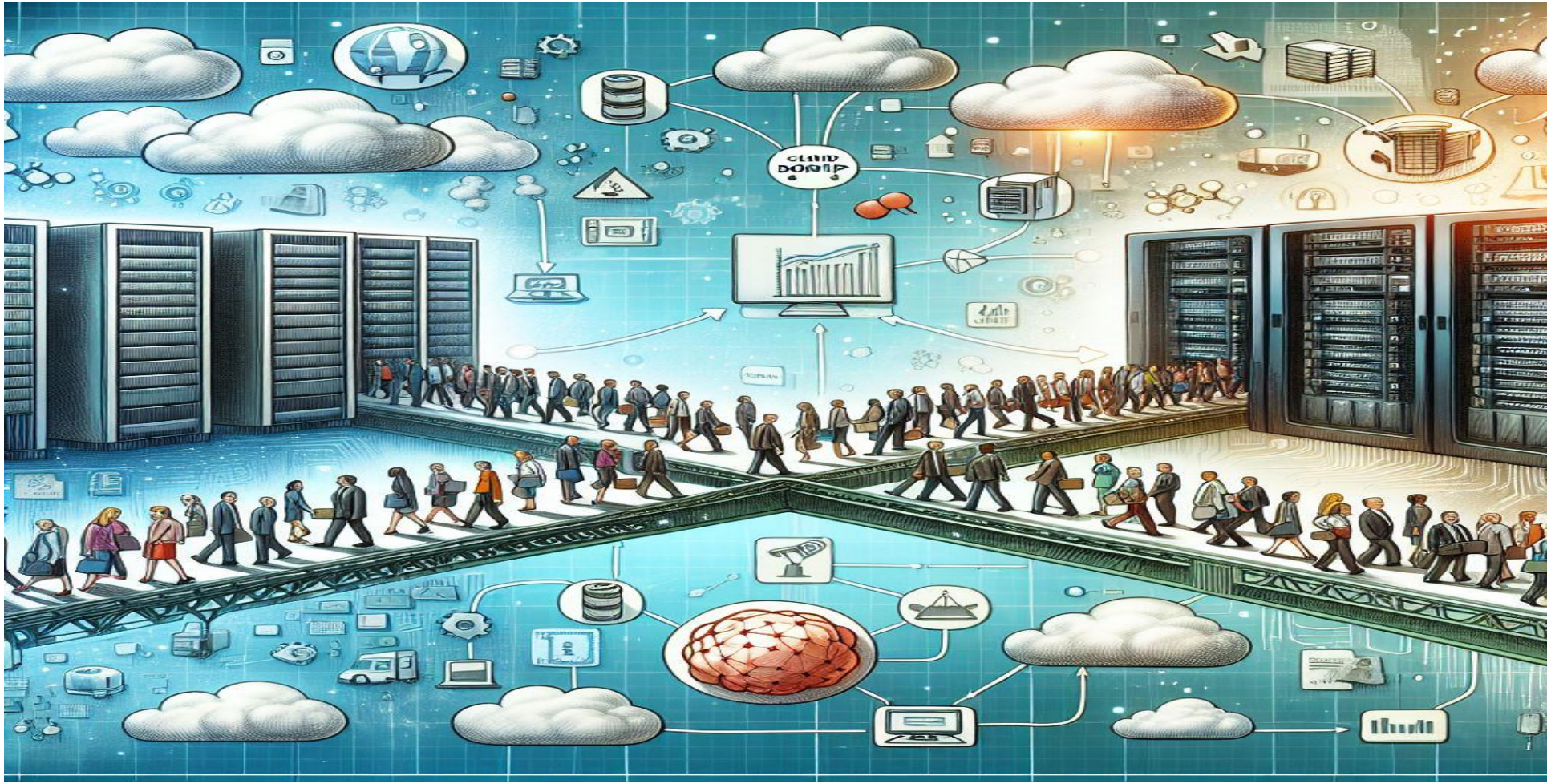
- Usually 3-D PDE's
  - Sparse matrix computations, not dense

Meanwhile… The Five Epochs of Distributed Computing

according to https://cloud.google.com/blog/topics/systems/the-fifth-epoch-of-distributed-computing

1. **1970-1985**: *Making Contact* → Asynchronous tasks (with SQL, FTP, Telnet)… and the Personal Computer
2. **1985-1995**: *Computer to Computer Communication* → RPCs, LANs, shared resources… and the World Wide Web
3. **1995-2005**: *Scale-out Global Computing* → multiple servers, web search, widespread email… and Cluster-based Internet Services
4. **2005-2020**: *Ubiquitous Information Access* → global cellular data coverage, ubiquitous video… and Cloud Computing
5. **2020-?**: *From Information to Insights* → Machine learning, Generative AI, privacy, societal infrastructure… and ???

Bridging the Centralized and Distributed Worlds



Where are we going? Here are some trends

- Declarative programming models
  - Declarative, i.e., not imperative: describe *what* should happen, not *how*
- Hardware specialization / segmentation
  - Different tasks may have widely different requirements (take AI Training vs. Inferencing, for example)
- Software-defined infrastructure
  - Programmatically declare infrastructural requirements for applications
- Security and reliability
  - Privacy and confidentiality, data sovereignty
- Sustainability
  - Look at the power consumption of the large HPC infrastructures previously mentioned. A new metric we must consider is *power-efficiency*, certainly for hardware & facilities, but also for software.
- Algorithmic innovation
  - Gone are the days when we just improved the performance of our applications because the next computer would have a faster clock. There are ample opportunities to remove software inefficiencies. Maybe in some cases shifting also to something else, such as Quantum-based algorithms.

What does this tell us?

- First, we need to be aware of these trends – **and not just in theory, but in practice**. The SOSC attempts to move in this direction, and you will **act** on some of these points throughout the week.

- Second, we must have a vision that allows us to govern and shape the trends, to avoid completely delegating our tasks – and our future – to external entities. This is **essential for science, but also for society and industry**. This point will be explored in the next part of this presentation.

The Italian Vision for HPC & Data

- As part of the **Italian National Recovery and Resilience Plan**, the Ministry of University and Research funded the constitution of the

National Research Center on HPC, Big Data and Quantum Computing (in short, **ICSC**)

- The initial **funding** to ICSC is €320M ($337M) in 4 years (2022-2025)
- The Ministerial main goal is to sustain research so that innovative, low TRL research outputs may be brought to a ready-to-market state, through direct involvement of public and private institutions:

From Research to Business

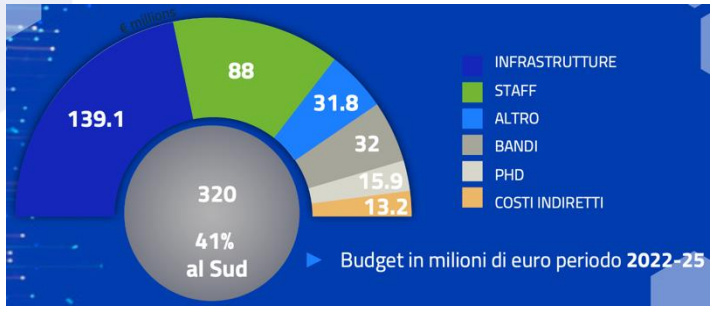
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca

The 2 ICSC Main Objectives

1. To **create an integrated national computing infrastructure for Research and Innovation**, boosting and federating the existing HPC, Big Data and network infrastructures and adding new targeted resources (e.g., Quantum).

2. To **create an attractive ecosystem around the infrastructure for Italy and beyond**, supporting academia and enterprises. A key goal is to simplify and foster the exploitation of computing resources and the development of new computing technologies.

12 Research Institutes



25 Universities

Istituti Nazionali

INFN Istituto Nazionale di Fisica Nucleare · CINECA · ENEA · Consiglio Nazionale delle Ricerche · INAF · Istituto Nazionale di Geofisica e Vulcanologia · Consortium GARR

HUBs

UNIMORE Università degli Studi di Modena e Reggio Emilia · Università di Parma · OGS

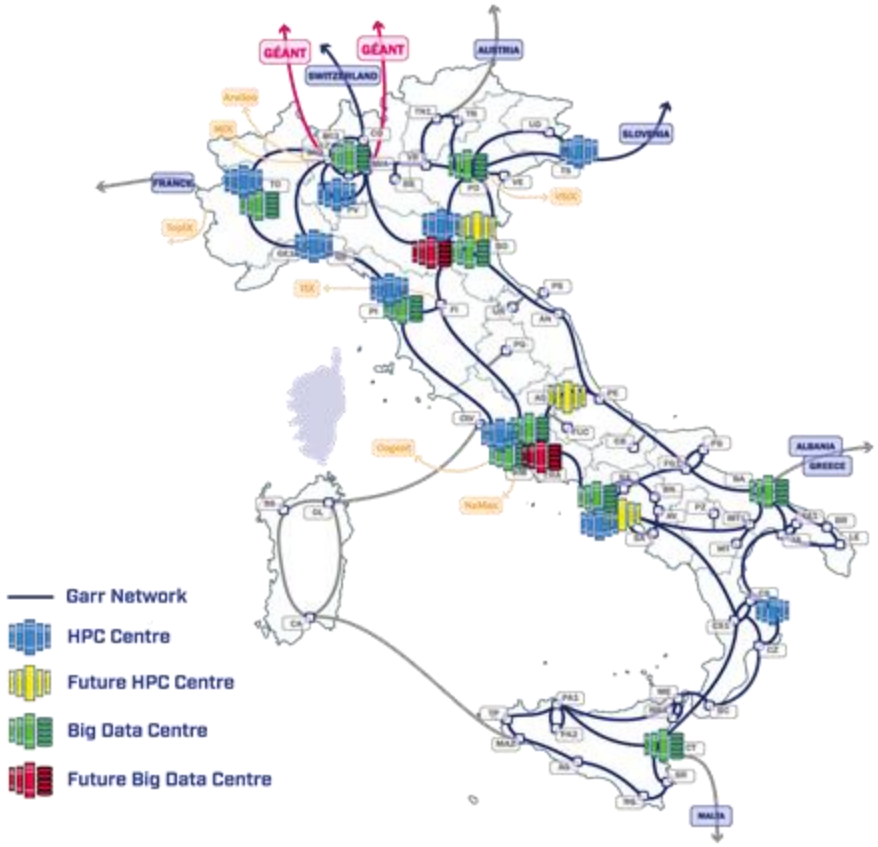
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

14 Private Enterprises

Engineering · eni · Ferrovie dello Stato Italiane · Fincantieri · fondazione innovazione urbana · autostrade per l'italia · Humanitas Research Hospital · iFAB · Intesa Sanpaolo · Leonardo · Sogei · ThalesAlenia Space · Terna Driving Energy · UnipolSai Assicurazioni

Organization

0 SUPERCOMPUTING CLOUD INFRASTRUCTURE



1 FUTURE HPC & BIG DATA

2 FUNDAMENTAL RESEARCH & SPACE ECONOMY

3 ASTROPHYSICS & COSMOS OBSERVATIONS

4 EARTH & CLIMATE

5 ENVIRONMENT & NATURAL DISASTERS

6 MULTISCALE MODELING & ENGINEERING APPLICATIONS

7 MATERIALS & MOLECULAR SCIENCES

8 IN-SILICO MEDICINE & OMICS DATA

9 DIGITAL SOCIETY & SMART CITIES

10 QUANTUM COMPUTING

Overall management by the **ICSC Foundation** (the "**Hub**")

Scientific, Technological, Industry-related Activities spread across 11 large sub-projects (the "**Spokes**")

A Transversal Group on **impact and society**, an Advisory Board **on Ethics**

SII TRANSVERSAL RESEARCH GROUP on SOCIETAL IMPLICATIONS AND IMPACT

The #Headquarters

Bologna Technopole

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

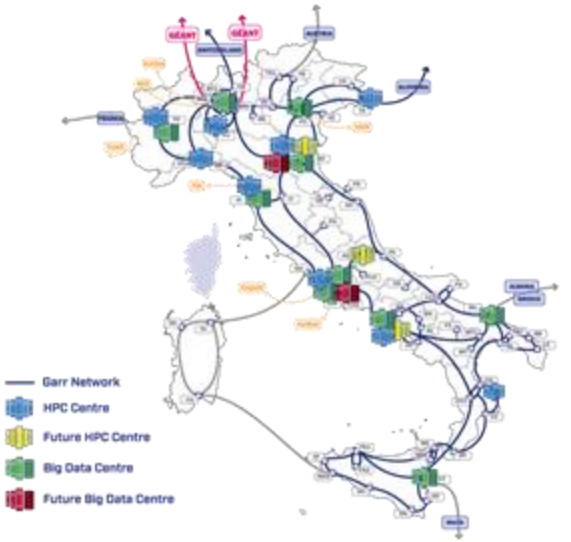
Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

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

What types of computational resources does ICSC offer? Where are they?



- Currently:
  - HPC CPU-hours from the **CPU-based Leonardo General Partition (CINECA)**
  - HPC GPU-hours from the **GPU-based Leonardo Booster Partition (CINECA)**
  - CPU Cores from the **INFN Grid and Cloud distributed infrastructure (INFN)**
  - HPC-optimized cloud-accessible GPUs or FPGA from the **INFN distributed HPC Bubbles**
  - Virtual Machines, PaaS or SaaS services from **CINECA or INFN**
  - Disk space from **CINECA or INFN**
  - Tape space (for backup) from **INFN**
  - **(soon)** Quantum resources

The ICSC resources, distributed over a national publicly-funded infrastructure (**no vendor lock-in!**), may also be provisioned in **ISO-27001 certified data centers**, for instance to address use cases with strong requirement on **security, high availability or sensitive data handling** (with GDPR compliance)

At the Bologna Technopole

The INFN Tier1
Inaugurated May 10, 2024

INFN CNAF
Istituto Nazionale di Fisica Nucleare

Leonardo
n. 9 Top500 supercomputer in the world

CINECA

PRESS RELEASE | 19 September 2024 | European High-Performance Computing Joint Undertaking | 2 min read

New Procurement Call to Upgrade LEONARDO, the EuroHPC supercomputer located in Italy

The European High Performance Computing Joint Undertaking (EuroHPC JU) has launched a new procurement call for the acquisition, delivery, installation and maintenance of the hardware and software of LISA, the upgrade to LEONARDO Supercomputer.

HPC and the "HPC Bubbles"

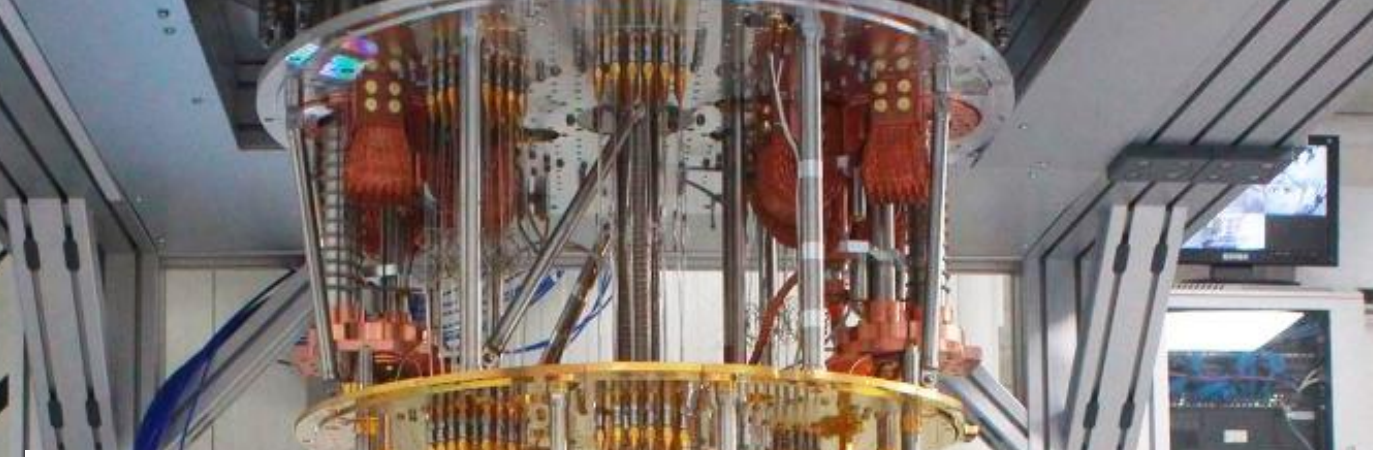
- To complement big traditional HPC systems, such as Leonardo and PRACE-Italy, **ICSC is implementing an "HPC at all scales" concept** through its distributed "**HPC Bubbles**". With the HPC Bubbles, ICSC provides:
  - Distributed, Cloud-native, scalable HPC resources and services at the IaaS, PaaS and SaaS levels, with optional instantiation in ISO 27001-certified data centers for proper handling of sensitive data.
  - Strong integration between network, big data, cloud and HPC resources.
  - Communication and federation between the HPC Bubbles and other HPC infrastructures.

Technically, what are the HPC Bubbles?

1. Modular AI-specialized HPC Clusters (8-16 nodes with 4 GPU NVIDIA H100 each)
2. Modular general purpose HPC Clusters (8-16 nodes with 192 CPU cores and 1.5 TB RAM each)
3. Modular FPGA-capable HPC Clusters (nodes with 32 CPU cores and 4 FPGA each)

All the HPC Bubbles nodes have SSD drives and InfiniBand and Ethernet network interfaces

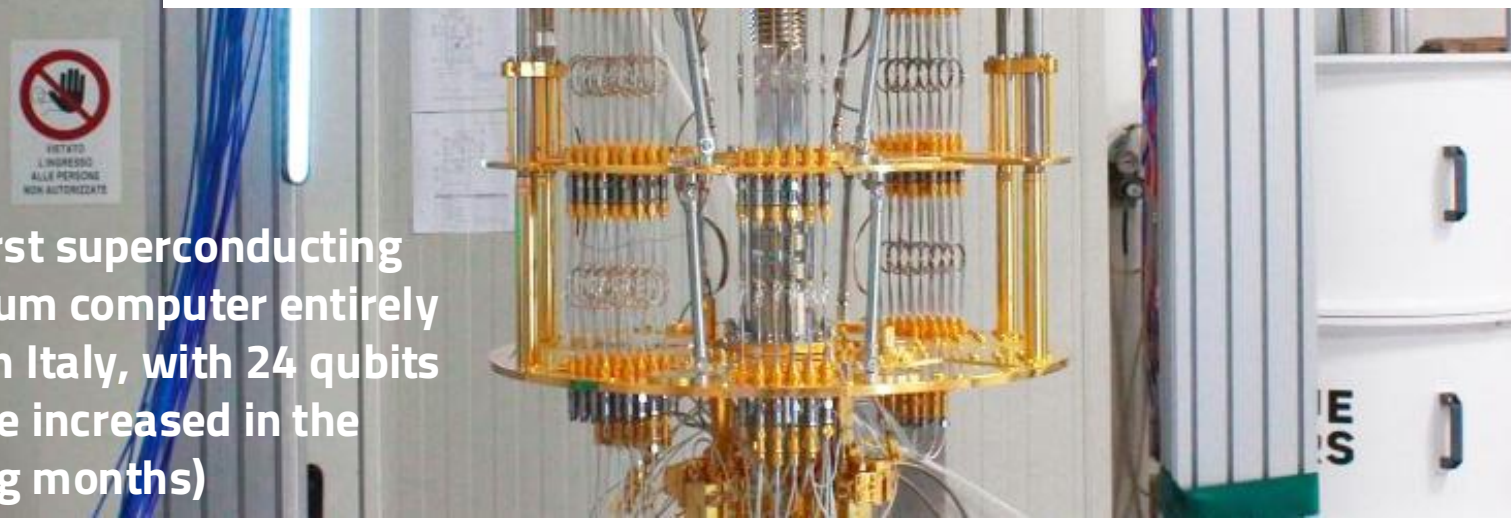
Napoli 29-30/5/2024
Inauguration of the UNINA Superconducting Quantum Computing Center



PRESS RELEASE | 1 August 2024 | European High-Performance Computing Joint Undertaking | 3 min read

EuroHPC JU Launches the Procurement for a New Quantum Computer in Italy

The European High Performance Computing Joint Undertaking (EuroHPC JU) has launched a call for tender for the installation of EuroQCS-Italy, a new EuroHPC quantum computer to be integrated into the EuroHPC pre-exascale system Leonardo.



The first superconducting quantum computer entirely built in Italy, with 24 qubits (will be increased in the coming months)

Ethics and Data Governance

Coordination of the ICSC **data management activities** and of the ICSC overall **Data Management Plan**

Definition and enforcement of **Data Governance Policies**

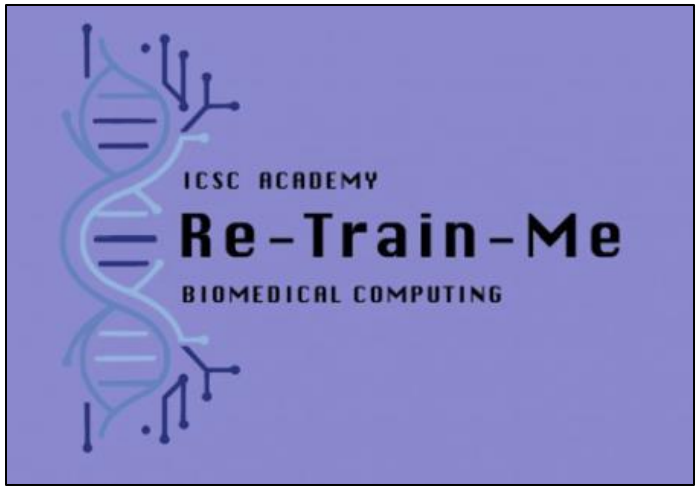
Analysis of the **ethical aspects associated to Artificial Intelligence (AI) activities**, with a special focus on the EU AI Act

New: ICSC candidacy to become an **EOSC (European Open Science Cloud) national node**, based on the FAIR principles

Ethical review of all the ICSC activities, through consultation and advice from the ICSC Ethics and Data Governance Board

Education

- Overall objective: **bridge the gap** between professionals with strong domain skills and those with IT skills in HPC, Big Data, Cloud and Quantum Computing fields.
- Train new professionals in areas in scope with ICSC interests, esp. where demand exceeds supply.
- Organize activities with and for start-ups and SMEs, and to attract young people to ICSC areas.
- **Define recognized, professional profiles** for supercomputing and data management professionals; **deliver appropriate qualification paths, managed and issued by ICSC**.



SOSC 2024 Sixth International School on Open Science Cloud

The theme of the sixth International School on Open Science Cloud is **"Computing Models for Scientific Experiments"**

The SOSC24 is organized by

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing
unipg UNIVERSITÀ DEGLI STUDI DI PERUGIA
INFN Istituto Nazionale di Fisica Nucleare

The ICSC Observatory

OS
Osservatorio sulle tendenze
e le applicazioni del Supercalcolo

Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

The ICSC Observatory on Supercomputing trends and applications addresses the **need to keep up with the data revolution**, observing and understanding technological transformations, and transmitting its findings to professionals and society. Among other activities, it produces free white papers on technology trends.

https://osservatorio.supercomputing-icsc.it/

Toward the Future…

(not now – we'll discuss this on Friday)


ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing
Supercomputing shaping the future
https://www.supercomputing-icsc.it/
e-mail: info@supercomputing-icsc.it