

# Le attività di Calcolo INFN

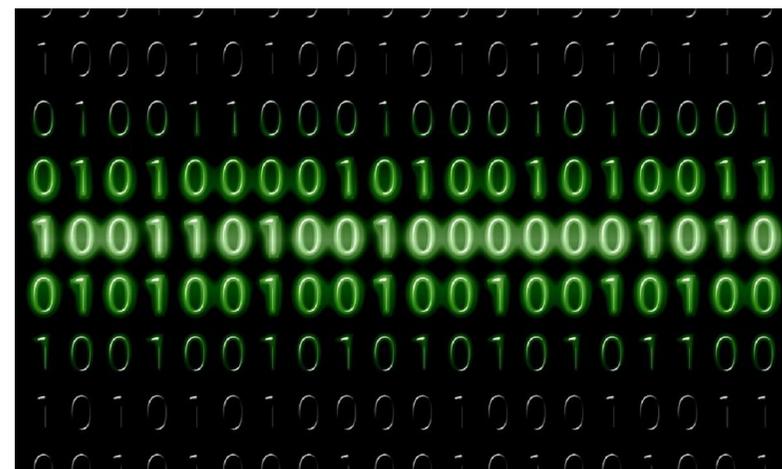
*G. Carlino – INFN Napoli*



# Il Calcolo nell'INFN

Il Calcolo, scientifico e non, ha un ruolo molto importante nell'INFN, trasversale alle Commissioni Scientifiche, necessario per raggiungere l'obiettivo finale: lo studio dei costituenti fondamentali della materia e le leggi che li governano.

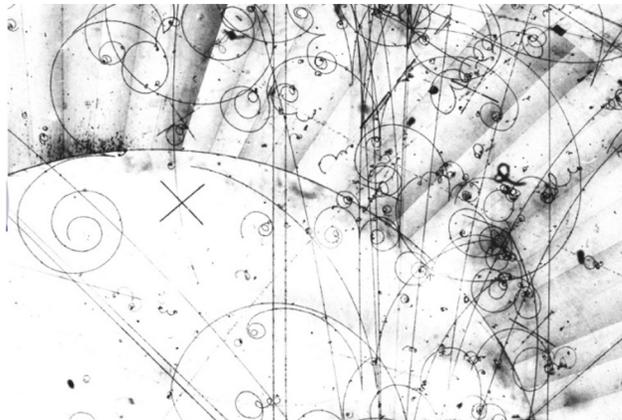
Tutte le attività, sia scientifiche che amministrative, richiedono un'infrastruttura di calcolo e un insieme di servizi altamente efficienti e personale altamente qualificato e l'INFN se ne è dotato nei decenni di attività.



# Un pò di storia del calcolo scientifico @INFN – Anni '60

Il CNAF (Centro Nazionale Analisi Fotogrammi) a Bologna è stato fondato nel 1962 e dedicato all'analisi delle immagini delle camere a bolle.

Sfida tecnologica molto importante all'epoca che richiedeva l'uso sistematico di computer



IBM 7094 operator's console showing additional index register displays in a distinctive extra box on top. Note "Multiple Tag Mode" light in the top center.

IBM System/360 Model 44

**IBM**



System/360 Model 44 front panel

**Manufacturer** International Business Machines Corporation (IBM)

**Product family** System/360

**Release date** August 16, 1965

**Discontinued** September 23, 1973

**Memory** 32–256 KB Core

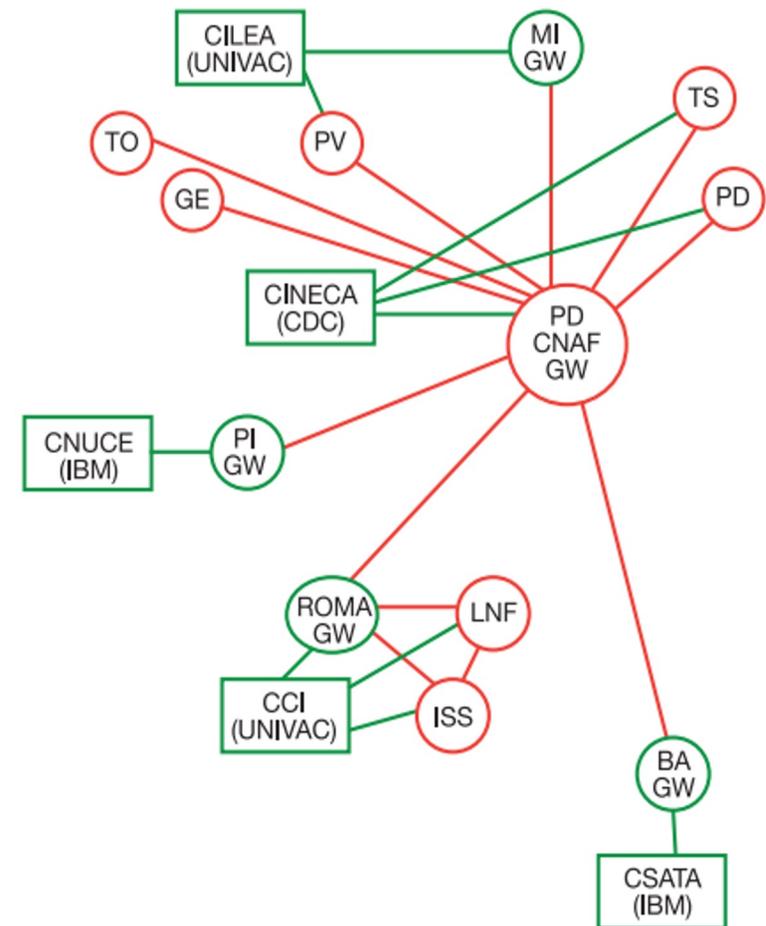
# Un pò di storia del calcolo scientifico @INFN – Anni '70/80

L'uso dei computer diventa sempre più popolare e necessario per i fisici e si crea la prima «infrastruttura distribuita» nelle sedi INFN.

I computer erano gestiti indipendentemente ma era necessaria una connessione tra le sedi. Nasce il progetto INFN-net e il CNAF assume un ruolo centrale.

All'inizio degli anni 80 si crea una connessione con il CERN (CERNnet) e poi con FNAL.

Alla fine degli anni 80, INFNnet raggiunge i 64 kbit/s



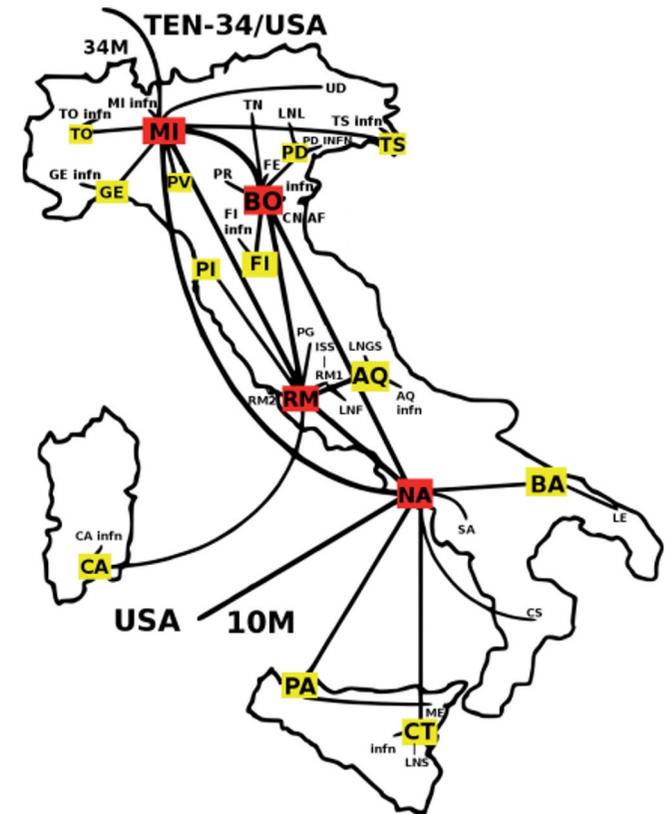
# Un pò di storia del calcolo scientifico @INFN – Anni '90

Si potenzia l'infrastruttura INFN al servizio degli esperimenti al LEP al CERN e Tevatron a FNAL.

Il GARR garantisce la connessione a 2 Mbps tra CNAF e CERN costituendo il backbone della ricerca connettendo tutti i centri INFN fino a 34 Mbps entro il 1995

Si prepara l'era LHC iniziando gli studi sul calcolo distribuito e la GRID. È evidente che il calcolo avrà un ruolo determinante per il successo degli esperimenti

Si crea un'infrastruttura distribuita a livello mondiale



# Il calcolo di LHC

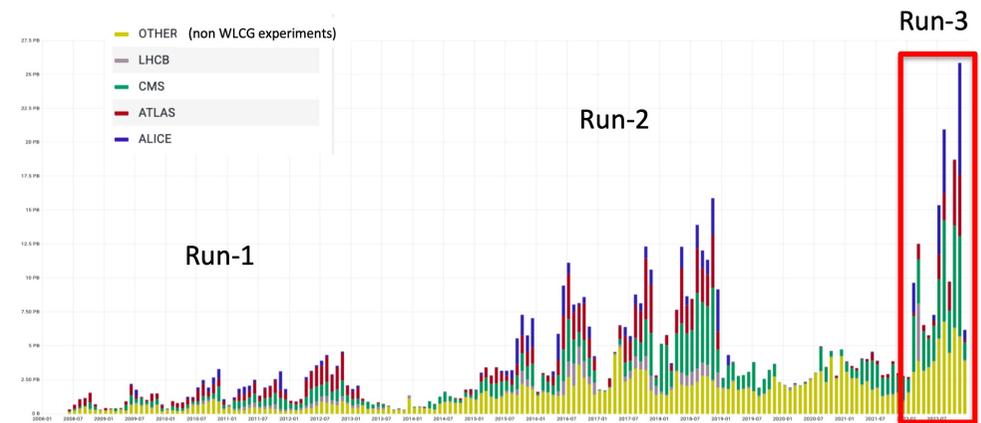
- A LHC collisioni protone protone a 40 MHz
- Ogni rivelatore (4 i principali) ha ~ 100Milioni di canali di lettura
  - A 1 byte/canale, farebbero 4x4 PB/s
  - Dopo vari sistemi di riduzione (trigger) si scende a ATLAS/CMS < 10 GB/s (gestibili)
- Corrisponde a ~ 30 PB/anno + stesse quantità per dati ricostruiti/derivati e simulati + repliche => più di 100 PB/anno/esperimento

Serve un'infrastruttura di calcolo e software molto efficaci per gestire questa quantità di dati

## Totale LHC 2023:

- 1 EB di Disco
- 1.5 EB di Tape
- ~1 Milione di CPU cores

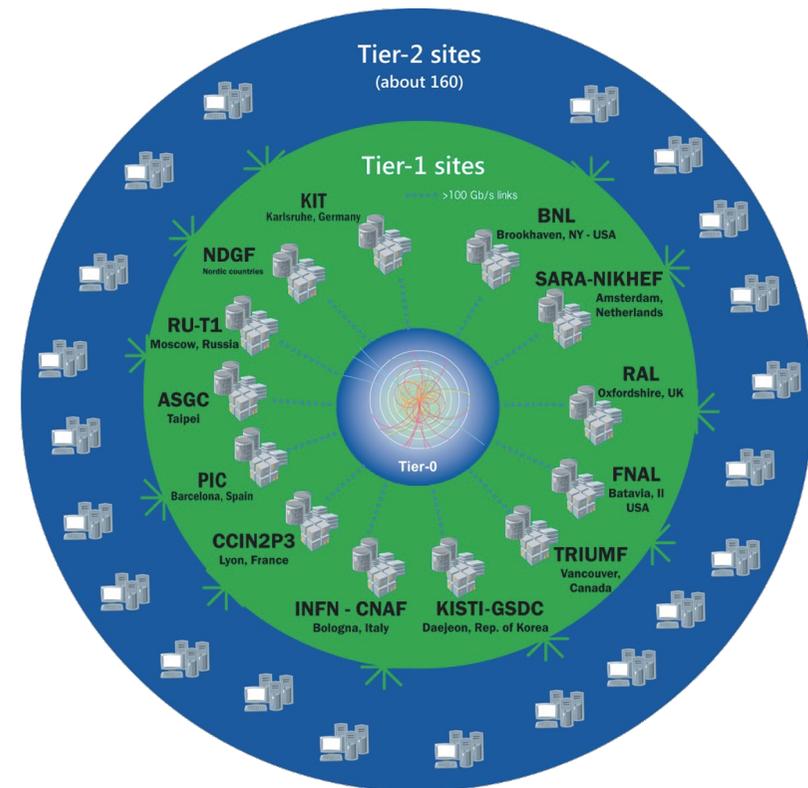
Data written in the CERN tape storage per month (since 2008)



26 PB of data written on tape in November 2022

# WLCG – Il Calcolo Distribuito

- Il progetto WLCG (Worldwide LHC Computing Grid) è una collaborazione internazionale di circa 170 centri in più di 40 paesi che nasce con la missione di fornire risorse di calcolo e storage in una singola infrastruttura, per distribuire e analizzare i dati raccolti da LHC
- Evoluzione di WLCG per includere esperimenti non LHC o non HEP (astroparticles)



# L'Infrastruttura di Calcolo Scientifico INFN – anni 2000

L'Infrastruttura di Calcolo Scientifico INFN è parte dell'infrastruttura mondiale di calcolo distribuito gestita da WLCG in attività dall'inizio degli anni 2000

L'Infrastruttura di Calcolo INFN, in attività da più di 20 anni, è costituita da 10 centri:

- 1 Tier1: CNAF @ BO
  - 9 Tier2: BA – CT – LNF - LNL/PD - MI – NA – PI - RM1 - TO
- connessi con link dedicati a 240 Gbps (Tier1) e a 10 / 100 Gbps (Tier2) al backbone del GARR

Non solo centri INFN

- HPC @CINECA: da anni l'INFN utilizza i sistemi di calcolo ad alte prestazioni sia per l'HTC che per attività che richiedono alta parallelizzazione

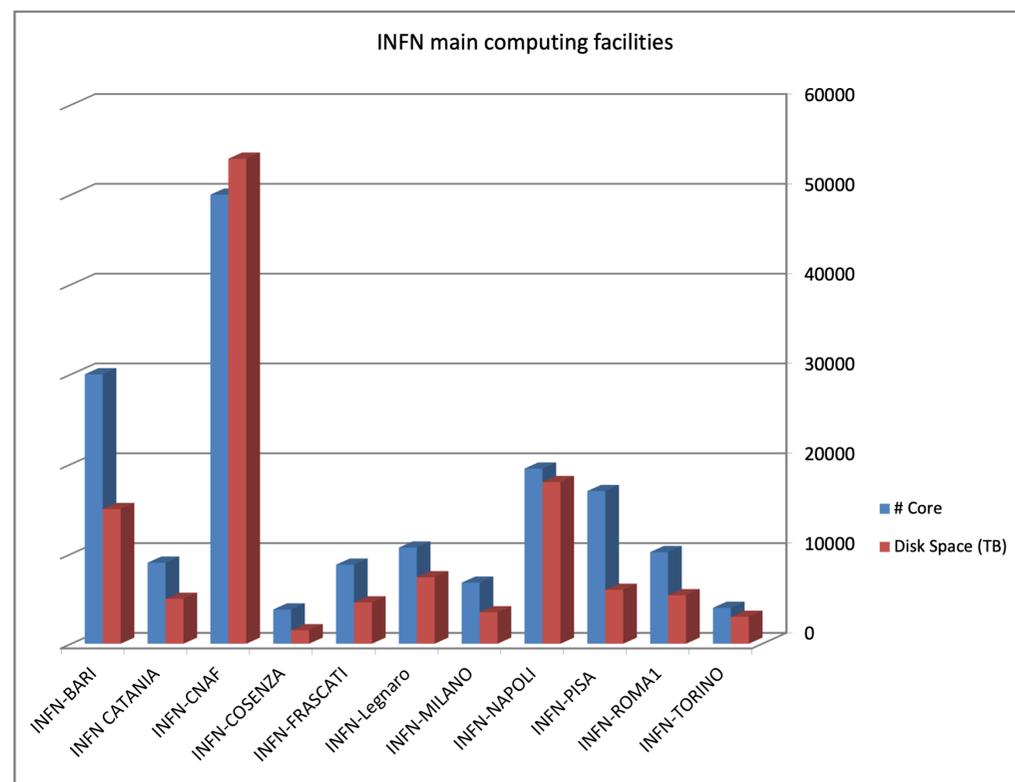


# L'Infrastruttura di Calcolo Scientifico INFN

Risorse di calcolo e storage nei centri:

- Calcolo
  - 170.000 cores
  - 2.200.000 HepSpec06
- Storage
  - 125 PB Disco
  - 220 PB Nastro

L'infrastruttura si sta arricchendo di architetture alternative (GPU, FPGA, ARM) per supportare applicazioni parallele e studiare come aumentare l'efficienza energetica dei centri



# Le sfide del Calcolo Scientifico

La prossima generazione di esperimenti HEP (HL-LHC) richiederanno enormi risorse di calcolo e storage (e rete) per la crescente complessità degli eventi (track multiplicity) e la quantità di dati raccolti (quasi un ordine di grandezza maggiore a LHC)

- Risorse di calcolo necessarie x20 rispetto a oggi ma il budget a disposizione resterà sostanzialmente costante (flat budget) e non c'è sempre un PNRR
- Nuovi esperimenti richiederanno risorse paragonabili a HL-LHC

| Parameter             | unit            | 2023 Run3 | 2029 Run 4 | 2033 Run5 |
|-----------------------|-----------------|-----------|------------|-----------|
| Interaction/crossing  | max $\mu$       | 55        | 140        | 200       |
| Integrated luminosity | $fb^{-1}y^{-1}$ | 100       | 300        | 450       |
| LHC ready for physics | $10^6$ s        | 7         | 7          | 7         |
| Rate                  | kHz             | 1.4       | 10         | 10        |
| Recorded events       | $10^9$          | 10        | 70         | 70        |

L'evoluzione della tecnologia e delle infrastrutture «standard» usate fino al Run2 di LHC non aiuta a ridurre i costi!

È necessario apportare innovazioni in

- Tecnologia
  - Parallelizzazione (HP06/\$\$), eterogeneità (HP06/Watt)
- Infrastruttura
  - Data Lake: consolidamento dello storage in pochi centri, CPU da centri HPC o cloud commerciali, riduzione costi operations (QoS)
- Computing e Analysis Model degli esperimenti

**Sfida principale per il calcolo scientifico**

# Il Centro Nazionale ICSC - oggi

## La sfida del PNRR

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

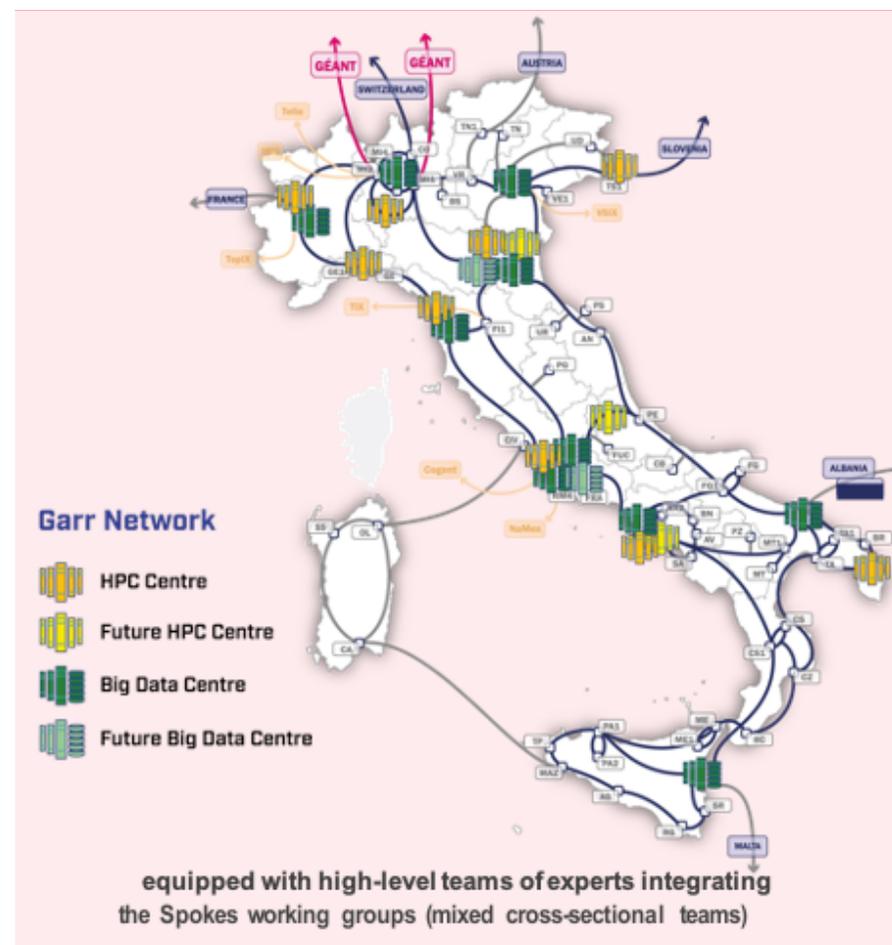
L'intera infrastruttura di calcolo INFN è inserita nel Centro Nazionale

Il Tier1 e tutti i 9 Tier2 vengono potenziati in termini di impianti, rete e risorse

→ I Tier2 dovranno essere tutti attivi per almeno i prossimi dieci anni. Non si prevede la dismissione di nessun centro. Importante

l'immissione e la formazione di personale nuovo negli anni

→ sarà comunque possibile differenziare i ruoli



# Calcolo Scientifico al Tecnopolo





# Private Founding Members: strategic players for digital transformation



Highly-qualified group of large leading companies covering most of the strategic industrial sectors involved by digital transformation at the national level

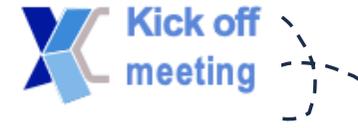


Strategic partner to implement and develop the digital twin pilot case of an urban complex system

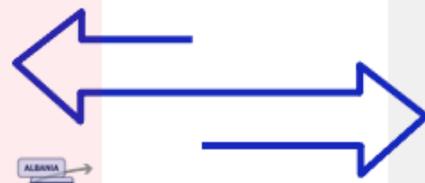
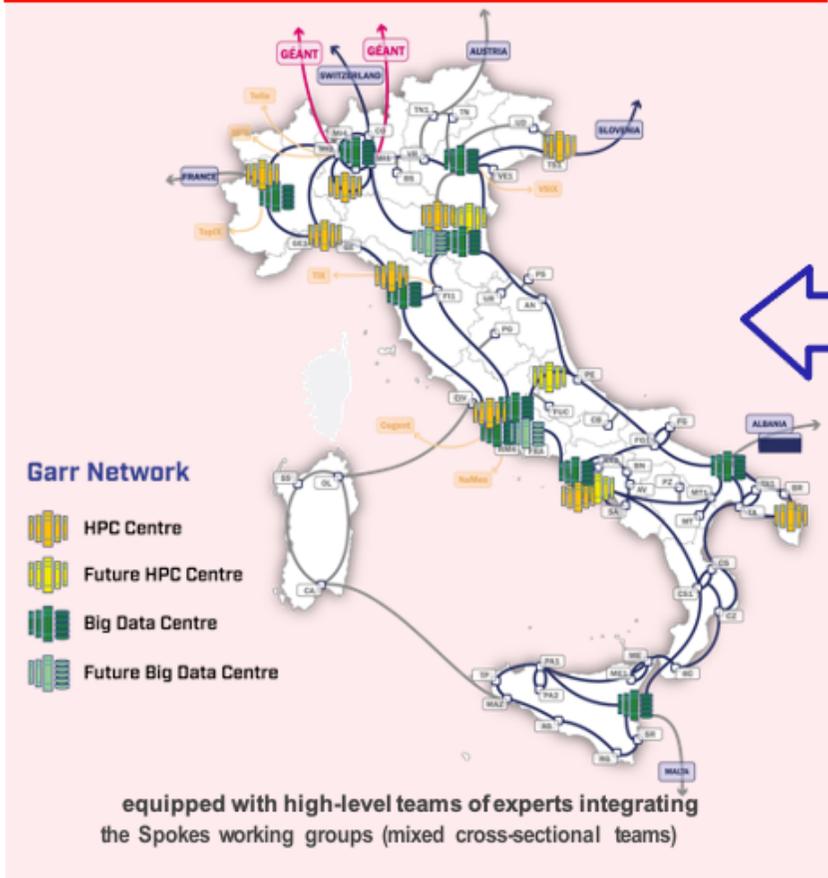


Industry-driven not-for-profit international organization aimed at: (1) aggregating companies, including SMEs, to engage with ICSC through a structured partnership, (2) funding research and innovation projects, (3) promoting the Big Data Technopole

# The ICSC will include 10 thematic Spokes and 1 Infrastructure spoke



## 0 SUPERCOMPUTING CLOUD INFRASTRUCTURE



EDUCATION & TRAINING, ENTREPRENEURSHIP, KNOWLEDGE TRANSFER, POLICY, OUTREACH

|   |   |
|---|---|
| <b>1</b><br><b>FUTURE HPC &amp; BIG DATA</b>              | <b>2</b><br><b>FUNDAMENTAL RESEARCH &amp; SPACE ECONOMY</b>           |
| <b>3</b><br><b>ASTROPHYSICS &amp; COSMOS OBSERVATIONS</b> | <b>4</b><br><b>EARTH &amp; CLIMATE</b>                                |
| <b>5</b><br><b>ENVIRONMENT &amp; NATURAL DISASTERS</b>    | <b>6</b><br><b>MULTISCALE MODELING &amp; ENGINEERING APPLICATIONS</b> |
| <b>7</b><br><b>MATERIALS &amp; MOLECULAR SCIENCES</b>     | <b>8</b><br><b>IN-SILICO MEDICINE &amp; OMICS DATA</b>                |
| <b>9</b><br><b>DIGITAL SOCIETY &amp; SMART CITIES</b>     | <b>10</b><br><b>QUANTUM COMPUTING</b>                                 |

# Il calcolo Teorico

Gruppi della CSN4 effettuano calcoli avanzati per studi di Lattice QCD e altri su infrastrutture HPC ad alta parallelizzazione (molti nodi di calcolo partecipano al processing, mediante reti veloci) molto diverse da quelle usate per LHC. Inizialmente computer sviluppati in casa (APE) poi utilizzo dei supercomputer del CINECA

| name         | year | authors         | peak speed   |
|--------------|------|-----------------|--------------|
| Columbia     | 1984 | Christ-Terrano  | –            |
| Columbia-16  | 1985 | Christ et al    | 0.25 GFlop/s |
| ➔ APE1       | 1988 | Cabibbo-Parisi  | 1 GFlop/s    |
| Columbia-64  | 1987 | Christ et al    | 1 GFlop/s    |
| Columbia-256 | 1989 | Christ et al    | 16 GFlop/s   |
| ACPMAPS      | 1991 | Mackenzie et al | 5 GFlop/s    |
| QCDPAX       | 1991 | Iwasaki-Hoshino | 14 GFlop/s   |
| GF11         | 1992 | Weingarten      | 11 GFlop/s   |
| ➔ APE100     | 1994 | APE Collab.     | 0.1 TFlop/s  |
| CP-PACS      | 1996 | Iwasaki et al   | 0.6 TFlop/s  |
| QCDSP        | 1998 | Christ et al    | 0.6 TFlop/s  |
| ➔ APEmille   | 2000 | APE Collab.     | 0.8 TFlop/s  |
| ➔ apeNEXT    | 2004 | APE Collab.     | 10 TFlop/s   |
| QCDOC        | 2005 | Christ et al    | 10 TFlop/s   |
| PACS-CS      | 2006 | Ukawa et al     | 14 TFlop/s   |
| QCDCQ        | 2011 | Christ et al    | 500 TFlop/s  |
| QPACE        | 2012 | Wettig et al    | 200 TFlop/s  |

## Computational Theoretical Physics @ INFN

Many research groups are involved in the activities of ICSC (Centro Nazionale di ricerca in HPC, Big Data and Quantum Computing)

### Cosmology and Astroparticle Physics

INDARK  
NEUMATT  
TEONGRAV

### Lattice QCD

LQCD123  
NPQCD  
SFT  
GAGRA  
QC DLAT  
SIM

### Physics of Complex Systems

BIOPHYS  
ENESMA  
FIELDTURB

### Standard Model Phenomenology

QFTATCOL

### Quantum Information

QUANTUM

### Nuclear Physics

MONSTRE  
NUCSYS

### Condensed Matter

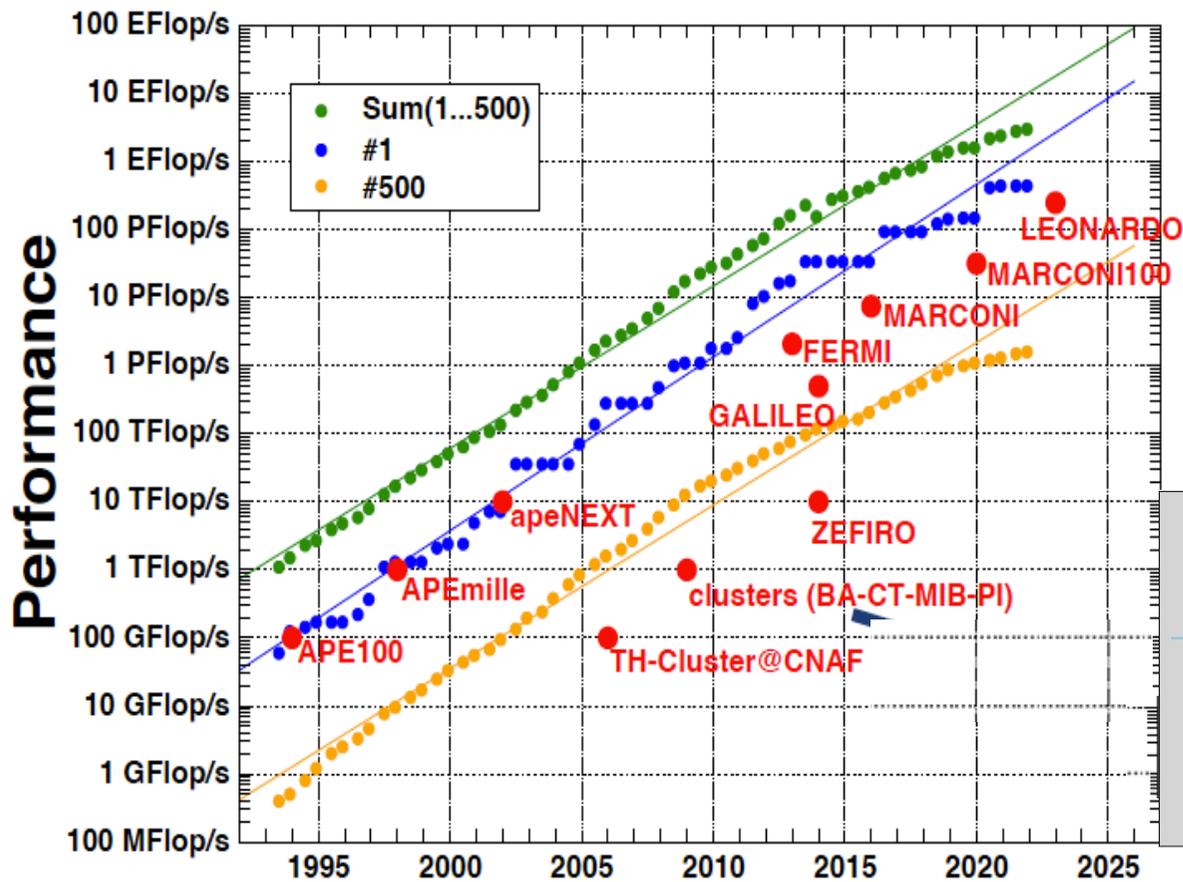
NEMESYS

Leonardo Cosmai - INFN Sezione di Bari

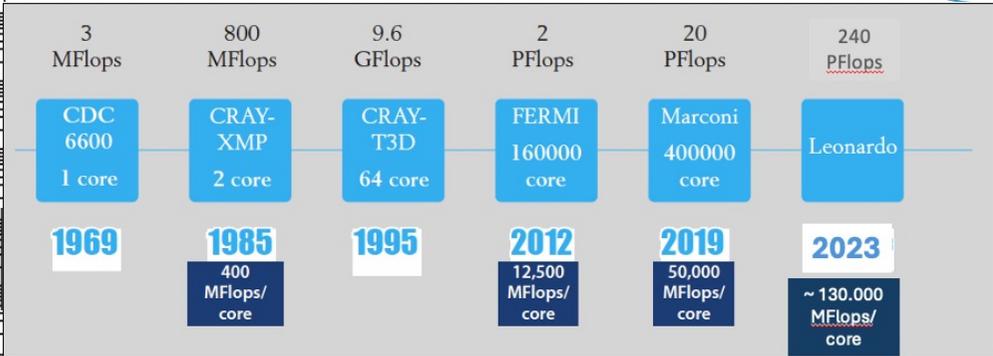
Leonardo Cosmai - INFN Sezione di Bari

# Supercomputer INFN e CINECA

Top 500 - Giugno 2023



| Rank (previous) | Rmax Rpeak (PetaFLOPs) | Name     | Model                  | CPU cores   | Accelerator (e.g. GPU) cores     | Interconnect             | Manufacturer | Site country                                    | Year |
|-----------------|------------------------|----------|------------------------|---|----------------------------------|--------------------------|--------------|---|------|
| 1               | 1,194.00<br>1,679.82   | Frontier | HPE Cray EX235a        | 561,664<br>(8,776 × 64-core Optimized 3rd Generation EPYC 64C @2.0 GHz) | 36,992 × 220 AMD Instinct MI250X | Slingshot-11             | HPE          | Oak Ridge National Laboratory<br>United States  | 2023 |
| 2               | 442,010<br>537,212     | Fugaku   | Supercomputer Fugaku   | 7,630,848<br>(150,976 × 48-core Fujitsu A64FX @2.2 GHz)                 | 158,976 × Fujitsu A64FX          | Tofu interconnect D      | Fujitsu      | RIKEN Center for Computational Science<br>Japan | 2020 |
| 3               | 309.10<br>428.70       | LUMI     | HPE Cray EX235a        | 150,528<br>(2,352 × 64-core Optimized 3rd Generation EPYC 64C @2.0 GHz) | 9,408 × 220 AMD Instinct MI250X  | Slingshot-11             | HPE          | EuroHPC JU<br>European Union, Kajaani, Finland  | 2022 |
| 4               | 238.70<br>304.47       | Leonardo | BullSequana XH2000     | 110,592<br>(3,456 × 32-core Xeon Platinum 8368 @2.6 GHz)                | 15,872 × 108 Nvidia Ampere A100  | Nvidia HDR100 Infiniband | Atos         | EuroHPC JU<br>European Union, Bologna, Italy    | 2023 |
| 5               | 148.600<br>200.795     | Summit   | IBM Power System AC922 | 202,752<br>(9,216 × 22-core IBM POWER9 @3.07 GHz)                       | 27,648 × 80 Nvidia Tesla V100    | InfiniBand EDR           | IBM          | Oak Ridge National Laboratory<br>United States  | 2018 |



# Il calcolo INFN

## Come si organizza l'INFN?

Nel 2019 si è avviato un processo di analisi e revisione dell'organizzazione del calcolo INFN per razionalizzare il modello di gestione e nel 2022 è stato istituito il **CNC** (Coordinamento Nazionale del Calcolo)

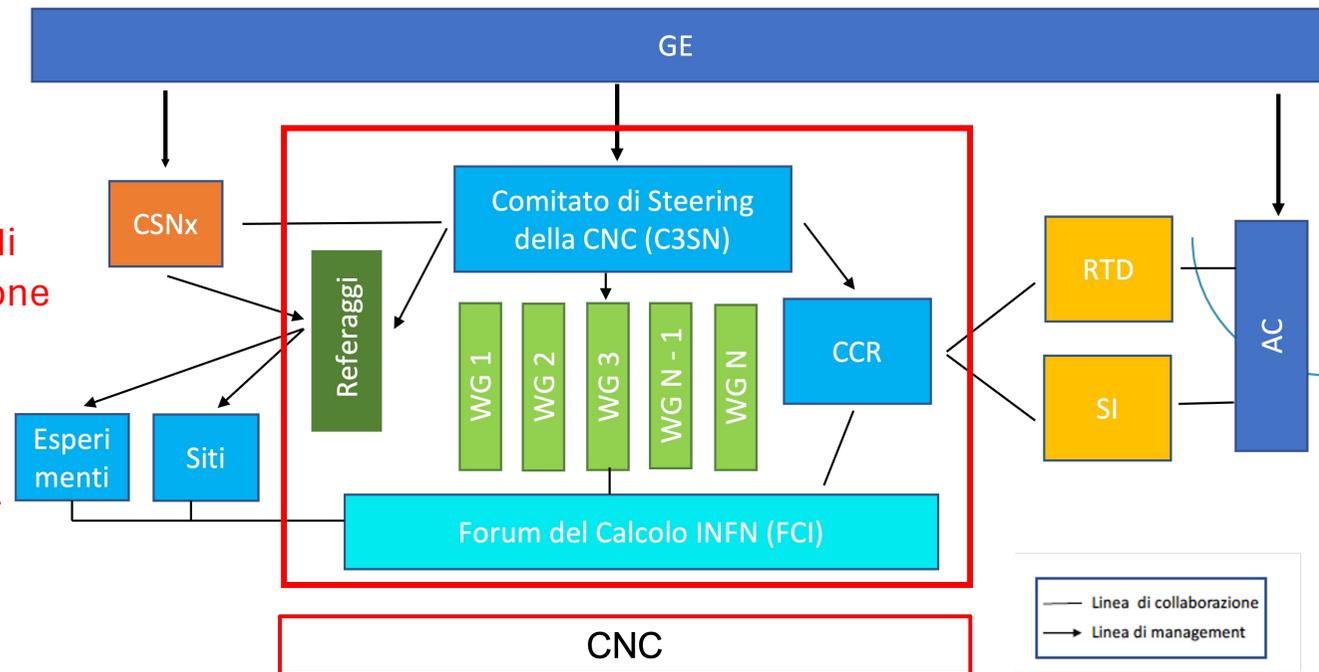
Per il calcolo scientifico, l'obiettivo è definire le strategie per affrontare le sfide proposte da HL-LHC e dai progetti emergenti (astroparticle, fisica nucleare, calcolo dei dati sensibili biomedicali):

- Evoluzione dell'infrastruttura INFN
  - Ruolo del Tier1 e dei Tier2
    - Nell'approccio Data Lake, l'infrastruttura attuale costituita da 9 Tier2 ha senso? I Tier2 vanno differenziati assumendo ruoli diversi? Sono tutti necessari?
  - Interazione con il CINECA
    - Il Tecnopolo a Bologna
    - Leonardo, uno dei due supercomputer pre-exascale Europei
  - Sviluppo del middleware per il Centro Nazionale e coordinamento dei progetti PNRR
  - Certificazione dei siti per la gestione dei dati sensibili

# Organizzazione del Calcolo INFN

## Il Coordinamento Nazionale del Calcolo (CNC)

- Strategia e Coordinamento: il **Comitato di Steering del CNC (C3SN)** e la **Commissione Calcolo e Reti (CCR)**
- Attività Operative: i **gruppi di lavoro del CNC (WG)**
- Discussione e Condivisione: il **Forum del Calcolo Italiano (FCI)**



## II C3SN

La gestione strategica e di coordinamento del calcolo INFN viene svolta in un comitato di steering ristretto, il Comitato di Steering del CNC (C3SN)

### Compiti:

- elabora e propone al management INFN la strategia del calcolo e l'evoluzione dell'infrastruttura
- definisce le linee guida e armonizza le attività che vengono svolte nei WG
- mantiene i contatti con le Commissioni Scientifiche, i gruppi di ricerca e i Fondi Esterni
- valuta e armonizza la partecipazione italiana a progetti internazionali e nazionali
- indirizza la rappresentanza dell'INFN nei confronti delle organizzazioni nazionali (CINECA, ICDI, GARR, EPR, ...) e internazionali (WLCG, EOSC .....
- gestisce i finanziamenti del CNC

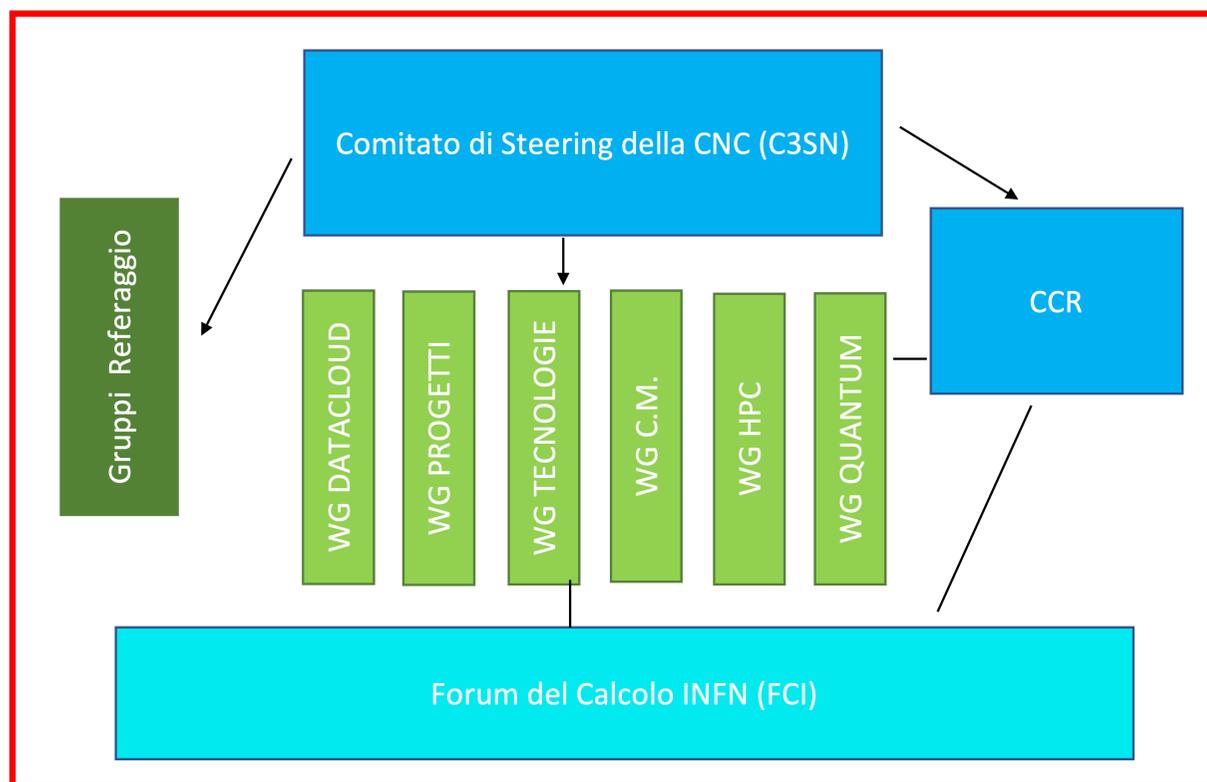
# La CCR

Coordina le attività che riguardano i Servizi di Calcolo e i Servizi Nazionali, l'infrastruttura di rete nazionale e le tematiche di sicurezza informatica

## Compiti:

- definizione e armonizzazione delle attività dei Servizi di Calcolo Locali
- Infrastruttura di Rete nazionale e rapporti tecnologici con il GARR
- Infrastruttura di Business Continuity e Disaster Recovery
- SSNN (Infrastruttura di Autenticazione e Autorizzazione (AAI), mailing, servizi multimediali, servizi web, strumenti collaborativi)
- Sicurezza informatica (di concerto con il RTD)
- Policy di accesso alle risorse informatiche
- Acquisti centralizzati di software e manutenzioni apparati di rete
- Supporto al SI

# I Gruppi di Lavoro



# I Gruppi di Lavoro

## **WG DataCloud**

Infrastruttura: sviluppo e gestione del modello architetturale di Data Lake e delle infrastrutture ISO, supporto alla gestione e all'operazione dei siti, agli esperimenti e agli utenti, sviluppo nuovi servizi

## **WG Progetti Esterni**

supporto, direttive e linee guida per organizzare le attività nei progetti esterni

## **WG Quantum Computing**

coordinamento delle attività di calcolo quantistico

## **WG Computing Model**

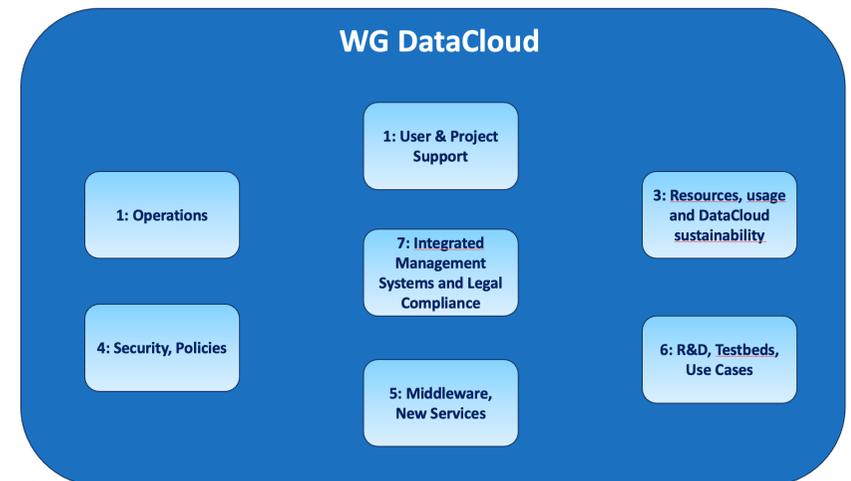
studio dell'evoluzione dei CM degli esperimenti in corso o futuri, ottimizzazione dei modelli per l'utilizzo efficace sull'infrastruttura INFN

## **WG Tecnologia Informatica**

studio dell'evoluzione della tecnologia informatica collaborando con istituzioni internazionali (Cern, Hepix), supporto tecnico agli esperimenti nella preparazione delle gare e al gruppo di referaggio, R&D

# Il WG DataCloud

- Ha il compito di **gestire ed evolvere l'infrastruttura per il calcolo scientifico dell'INFN**
  - sviluppo di un'architettura globale basata sulla federazione di risorse a diversi livelli (IaaS, PaaS, SaaS) dei centri di calcolo (INFN Cloud)
  - sviluppo e gestione del modello architetturale di datalake
  - sviluppo delle infrastrutture ISO-certified
  - supporto alla gestione e all'operazione dei siti, agli esperimenti, ai progetti e agli utenti
  - integrazione tra l'infrastruttura WLCG a Tier e il modello cloud
- Deve rispondere alle esigenze dei progetti delle CSN (progetti interni) e dei progetti derivanti da collaborazioni o call competitive nazionali ed internazionali (progetti esterni), che insistono sull'infrastruttura di calcolo scientifico dell'INFN
- Ha un ruolo essenziale nella partecipazione ai progetti PNRR sul calcolo che coinvolgono l'INFN, in particolare il Centro Nazionale ICSC che si basa sull'architettura DataCloud



# Conclusioni

Il calcolo ha un ruolo molto importante nell'INFN

- Il calcolo scientifico di supporto agli esperimenti e ai progetti esterni
- Il calcolo dei servizi di supporto alle attività amministrative e per garantire la sicurezza di tutte le attività
- Le sfide che abbiamo di fronte sono molto impegnative, ma l'infrastruttura realizzata, la partecipazione al centro nazionale e, soprattutto, il personale di grande esperienza e competenza sono una garanzia di successo
  - Sarà molto importante il contributo dei nuovi assunti su tutti questi aspetti