

# Il calcolo ai Laboratori del Gran Sasso!

# Perche' un servizio di Calcolo e Reti ai LNGS?

Come vedremo, gli esperimenti di fisica acquisiscono, elaborano e conservano grandi moli di dati sperimentali!

La raccolta e l'analisi dei dati richiedono **reti veloci** e grande **potenza di calcolo**; non solo: richiedono una conoscenza profonda dei sistemi informatici!

Sono necessari **specialisti delle reti di trasmissione dati, di calcolatori, di software e di tecnologie innovative** per dare supporto agli esperimenti!

Questo e' il compito principale dei servizi di calcolo nei laboratori e nelle sezioni INFN!

# I dati, così' preziosi!



**I dati sono la materia prima della ricerca scientifica!**



**Sono costosi e preziosi!**



Tutti gli esperimenti di fisica, per arrivare a trovare un risultato, si basano sull'analisi dei dati.



I dati un tempo erano scritti a mano con carta e penna!

# Carta, penna, carta millimetrata...

## West Point Ozone Project

Sheet 43 Operator J. Doe

Date	Time	Wind Dir	Wind Spd	Ozone
08/01/93	1000	359	2.68	53
08/01/93	1100	1	1.11	55
08/01/93	1200	103	2.46	59
08/01/93	1300	709	1.11	65
08/01/93	1400	307	1.78	74
08/01/93	1500	164	3.13	80
08/01/93	1600	167	4.47	80
08/01/93	500	1777	6.26	93,702
08/01/93	1800	177	6.48	100
08/01/93	1900	157	4.02	81
08/01/93	2000	158	1.78	61
08/01/93	2100	189	1.34	74
08/01/93	2200	348	0.67	37
08/01/93	2300	316	0.67	30
08/01/93	2400	310	0.44	26
08/02/93	100	290	0.22	24
08/03/93	200	316	0.22	21
08/04/93	300	15	0.22	20
08/02/93	400	5	4.47	15
08/02/93	400	16	0.44	11
08/02/93	600	49	0.44	5
08/02/93	700	137	0.44	8



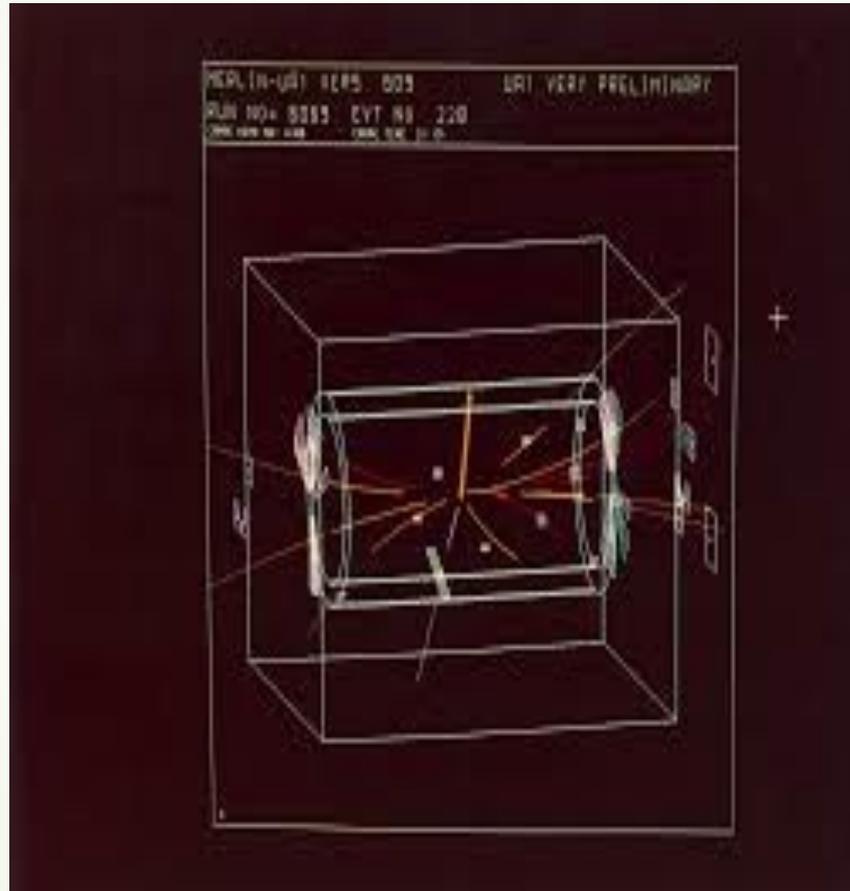
# I dati, i computer e le reti

- Fin dagli anni '60 in fisica delle particelle si sono usati i computer: dal sistema di **acquisizione** sull'apparato sperimentale, alla **rappresentazione grafica** dei dati, all'**analisi** complessiva dei dati acquisiti, alla scoperta scientifica e alla pubblicazione degli articoli su riviste specializzate:

**TUTTO QUESTO NON SAREBBE POSSIBILE SENZA I  
CALCOLATORI e le RETI!**

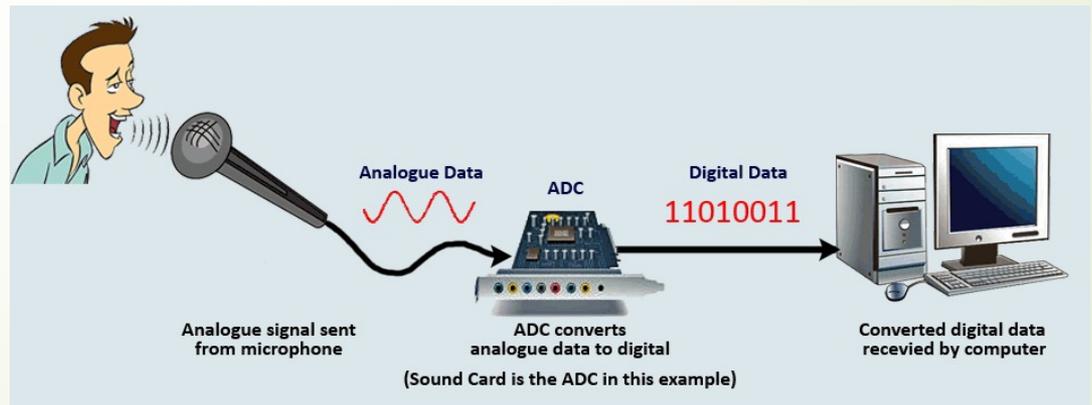
**La necessita' di potenza di calcolo per gli  
esperimenti ha portato l'INFN ad essere uno degli  
attori principali nel campo delle infrastrutture di  
calcolo e dell'evoluzione tecnologica!**

# 1981 – Esperimento UA1 al CERN

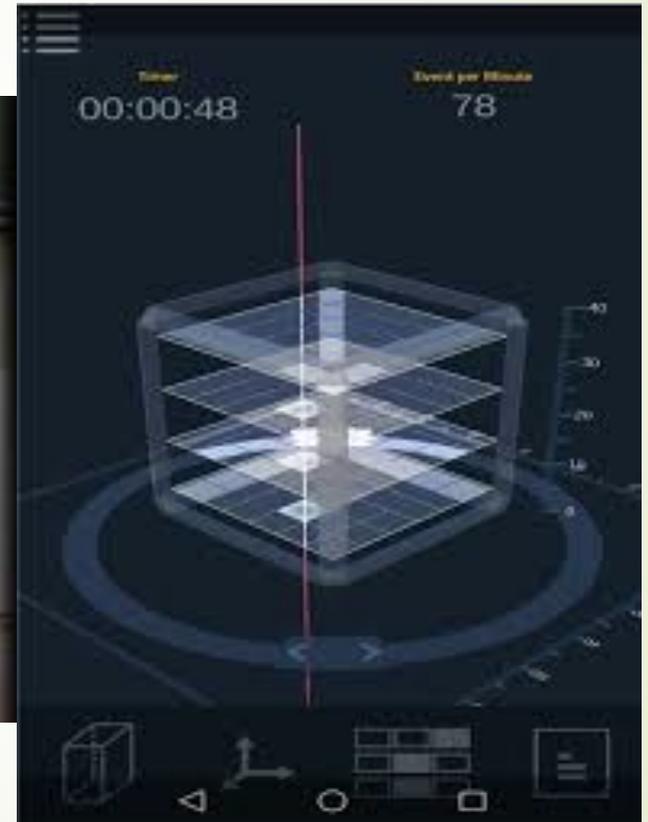


# I dati in formato digitale

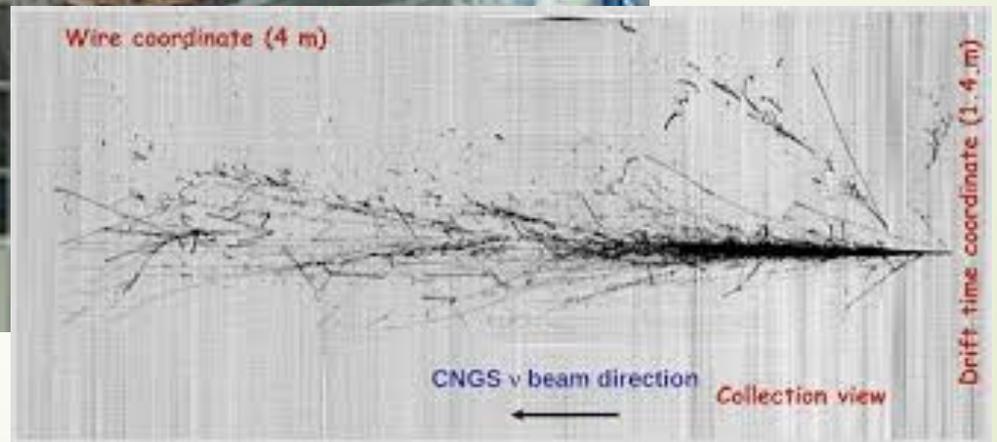
- Tutti i rivelatori di particelle quindi sono progettati per produrre un segnale in uscita che sia convertibile in un dato digitale.
- Generalmente un rivelatore converte il segnale lasciato da una particella (luce, temperatura,..) in un segnale elettrico analogico il quale viene successivamente trasformato in un segnale digitale pronto per essere salvato e analizzato con un computer



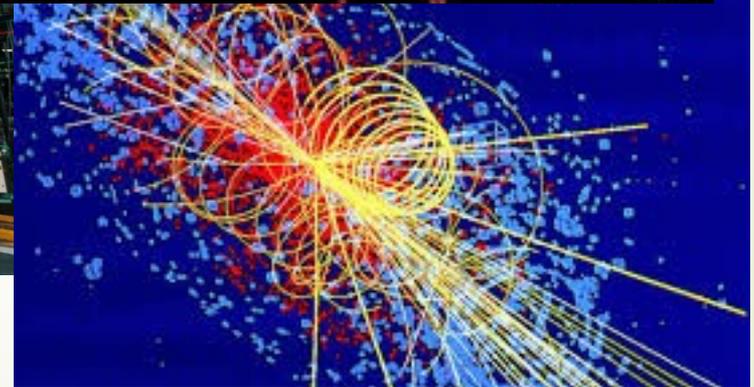
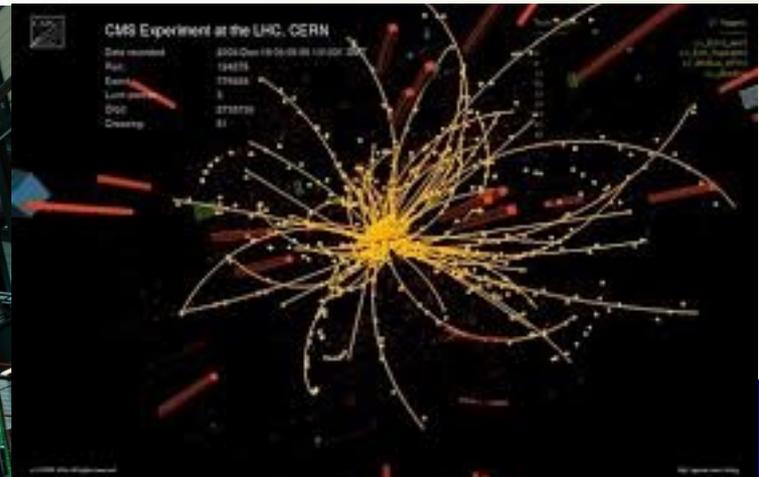
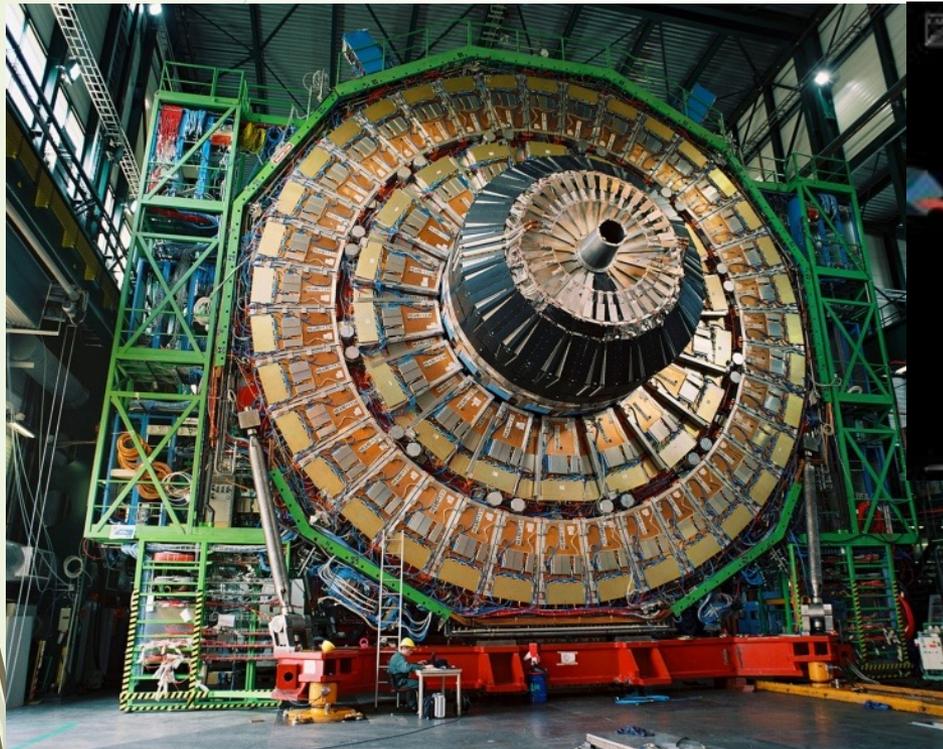
# Un semplice rivelatore di raggi cosmici



# ...un grande esperimento ai LNGS



# ..un esperimento complesso al CERN



# Tante fasi per arrivare al risultato!

Acquisizione dati e primo processamento



Memorizzazione su disco/nastro



Analisi dati



Visualizzazione degli dati



Risultato/scoperta



Pubblicazione

$m_1 m_2$

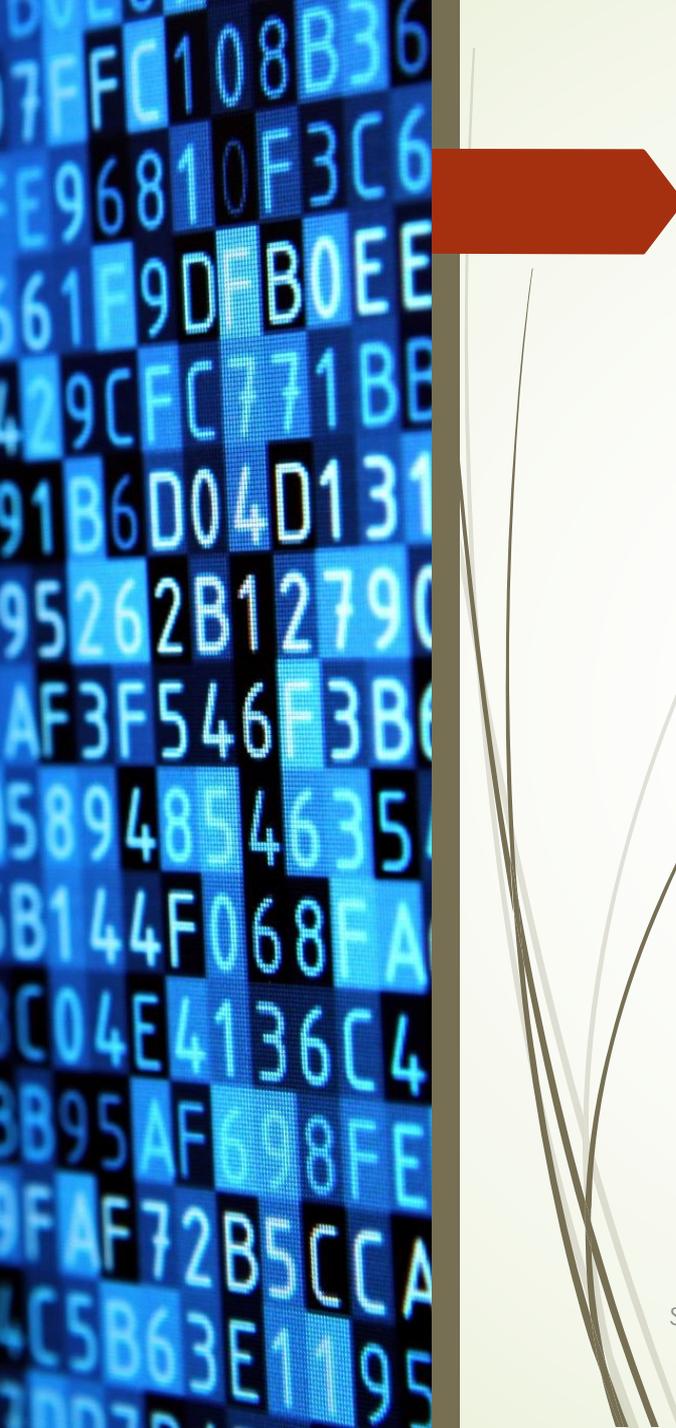
$d^2$

$$\hbar \frac{\partial \psi}{\partial t} = \hat{H} \psi$$

$E =$

# Ma quanti dati!!!!

- ▶ Un esperimento di fisica delle particelle può essere molto complesso e formato da migliaia di rivelatori indipendenti
- ▶ Facciamo un semplice esempio:
  - ▶ Ogni singolo rivelatore, un canale di elettronica, può generare un segnale da 1KByte
  - ▶ Ogni 'evento' può coinvolgere migliaia di canali 1MByte
  - ▶ La frequenza di eventi, *rate* di eventi, può essere di migliaia di eventi al secondo: **1GByte/s**
- ▶ Sono tantissimi dati, ma non tutti portano informazioni utili!



# (una parentesi per ricordare)

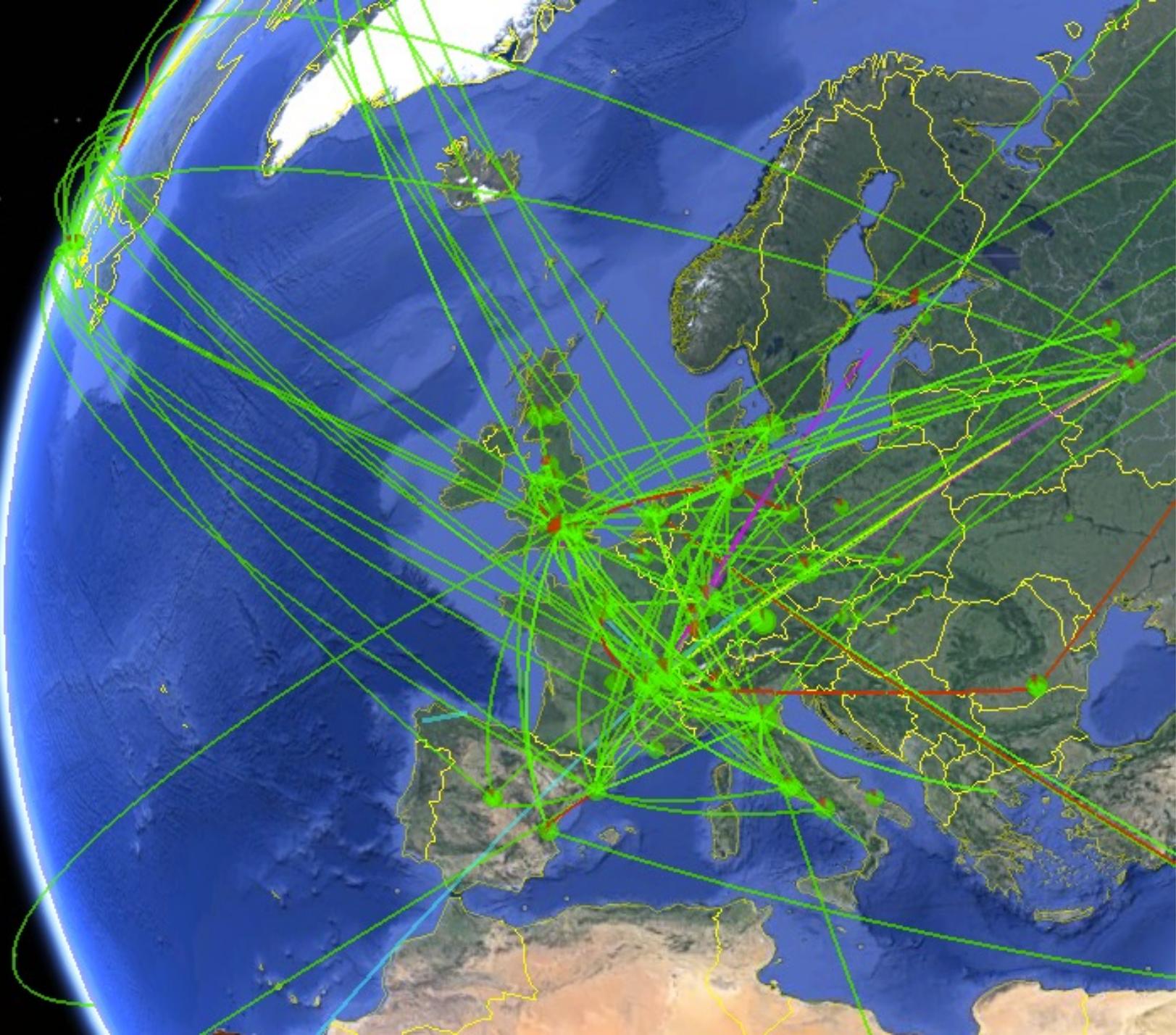
- Byte – 8 bit unità fondamentale di memorizzazione
- KB –  $10^3$  byte
- MB –  $10^6$  byte
- GB –  $10^9$  byte
- TB –  $10^{12}$  byte
- PB –  $10^{15}$  byte
- EB -  $10^{18}$  byte

# ... selezionare i dati interessanti

- ▶ L'elettronica di acquisizione raccoglie i dati provenienti da ogni singolo elemento di un rivelatore e li convoglia verso un punto di raccolta; in questa fase puo' essere operata una prima scrematura dei dati (**L1 trigger**), ad opera di hardware dedicato, a volte con fattori di riduzione drammatici
- ▶ Successivamente i dati sono indirizzati attraverso la rete (cavi in rame e fibre ottiche) a *farm* di calcolatori commerciali che immagazzinano temporaneamente i dati e operano una seconda scrematura (**L2 trigger**) per selezionare solo gli eventi interessanti; anche in questa fase la quantita' di dati e' ridotta di un fattore significativo

# Ai LNGS

- Gli esperimenti ai LNGS tipicamente studiano fenomeni prodotti naturalmente e hanno *rate* di eventi decisamente inferiori
- I dati complessivamente prodotti sono dell'ordine di **qualche PB/anno**
- ...vanno analizzati!
- In parte i dati sono analizzati ai LNGS su *farm* di server installate in sala calcolo e in parte prendono una strada differente..

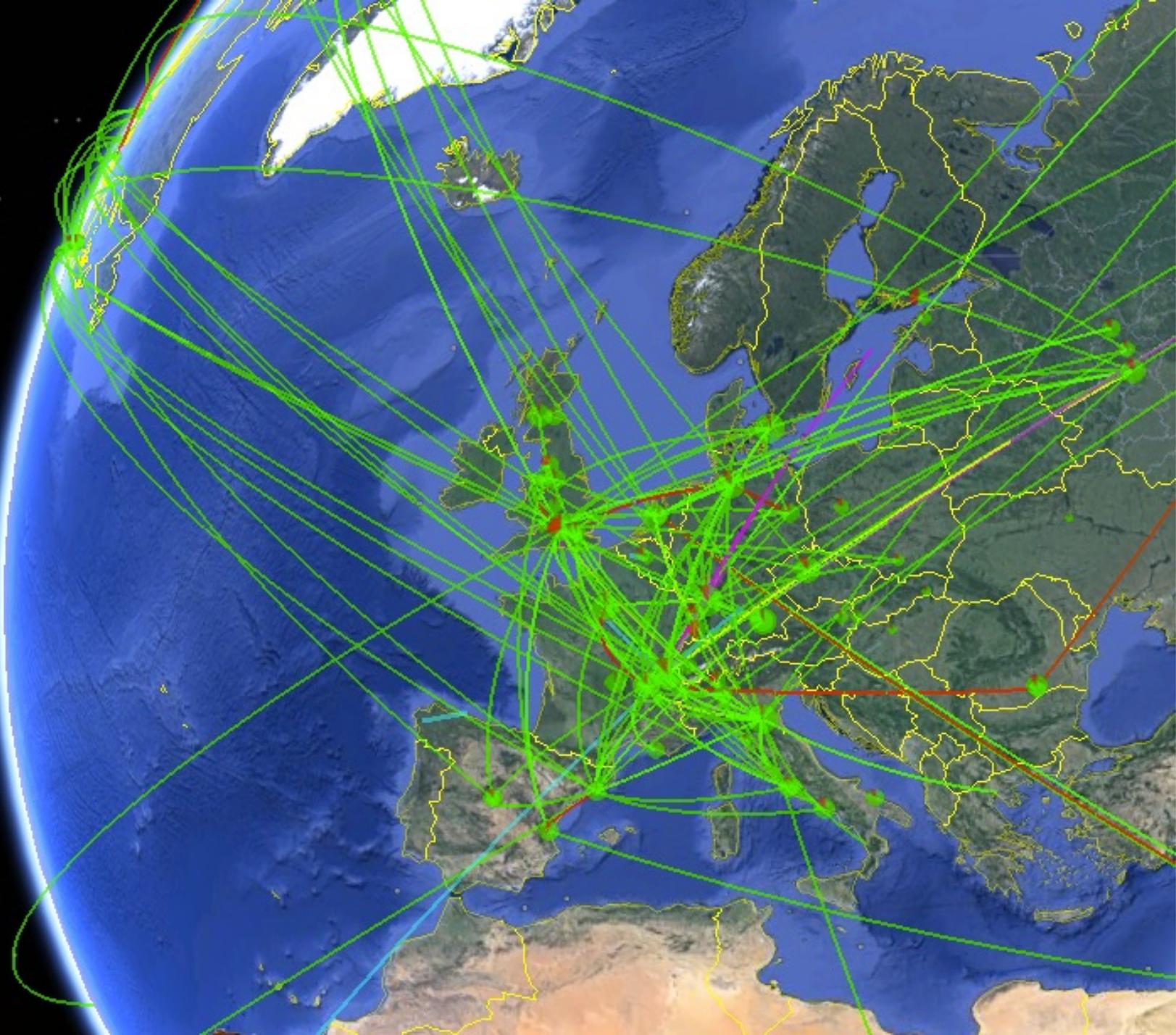


# I dati sono salvi...e ora?

- ▶ La potenza di calcolo offerta da un singolo computer o una singola sala calcolo non e' sufficiente ad analizzare i dati, si ricorre ad una **griglia di computer** collegati in rete tra di loro: la **GRID**
- ▶ La GRID e' costituita da data center che contengono **storage** per i dati e **server** per il loro processamento; i datacenter sono interconnessi da **reti** di trasmissione dati dedicate.
- ▶ Nei prossimi anni le risorse di calcolo distribuite saranno accessibili con interfacce **CLOUD** semplificate rispetto alla GRID.

- Collegamenti di rete ad alta velocità sono fondamentali per l'esecuzione dei job sulla GRID/cloud
- Il modello della GRID e del calcolo distribuito si è potuto sviluppare solo quando le reti di trasmissione dati hanno raggiunto la maturità
- Collegamenti attuali tra centri della GRID variano tra multipli di **10Gb/s** di centri Tier2 al Tier1 e multipli di **100Gb/s** per i Tier1
- La connessione dei LNGS alla rete GARR e alla GRID è attualmente di 10Gb/s.







I  
datacenter  
per il  
calcolo  
scientifico

---

Migliaia di **server** montati su armadi rack

---

Migliaia di dischi organizzati in sistemi di **storage**

---

Migliaia di Km cavi in **fibra ottica** per  
connettere server e storage tra loro

---

Apparati di **rete** per il trasporto dei dati

---

Librerie di nastri magnetici o **tape** per il backup  
e la conservazione a lungo termine dei dati

---

Sistemi di **raffreddamento**

---

Alimentazione elettrica!!! **IL COSTO E' IN  
CONTINUO AUMENTO!!!!**

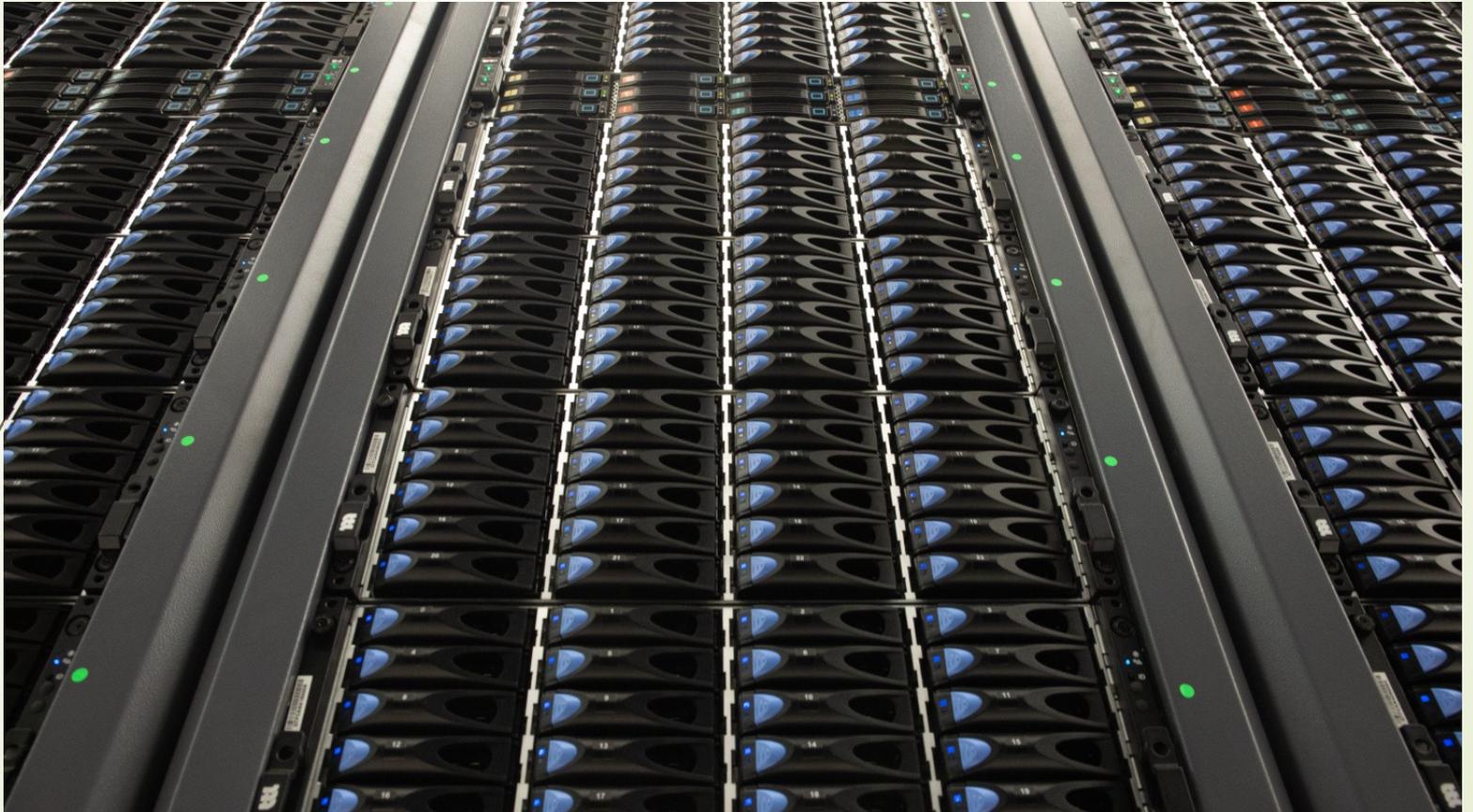


Sandra Parlati - PID LNGS 2022 - 17-21 ottobre 2022

# Server negli armadi rack



# Storage



# Nastri magnetici



Sandra Parlati - PID LNGS 2022 - 17-21 ottobre 2022

# Raffreddamento del data center



# Alimentazione elettrica





# Qualche numero sul data center LNGS

- ▶ Circa **2PB** di dati distribuiti tra vari sistemi di storage (i dischi piu' moderni acquistati ai LNGS hanno una capacita' di 16TB)
- ▶ Circa **1000 core** per l'analisi dati, tra server gestiti dal servizio calcolo e server degli esperimenti (i server piu' nuovi acquistati ai LNGS hanno doppio processore 2.7GHz, 16C/32T e 128 GB RAM)
- ▶ Collegamento a 10Gb/s con la rete GARR
- ▶ Circa **40KW di potenza assorbita** e protetta da un sistema di 3 UPS
- ▶ Sistema di raffreddamento ridondato

# ..le richieste di potenza di calcolo degli esperimenti esplodono!

L'evoluzione delle **CPU** dei calcolatori, impressionante fino a qualche anno fa, **ora segna il passo..**

Nuove soluzioni tecnologiche, idonee a soddisfare le richieste di potenza

Big player del mercato dell'hardware (INTEL, NVIDIA, AMD) stanno investendo miliardi di dollari su queste nuove tecnologie!!!

**HPC**

(calcolo ad alte prestazioni)

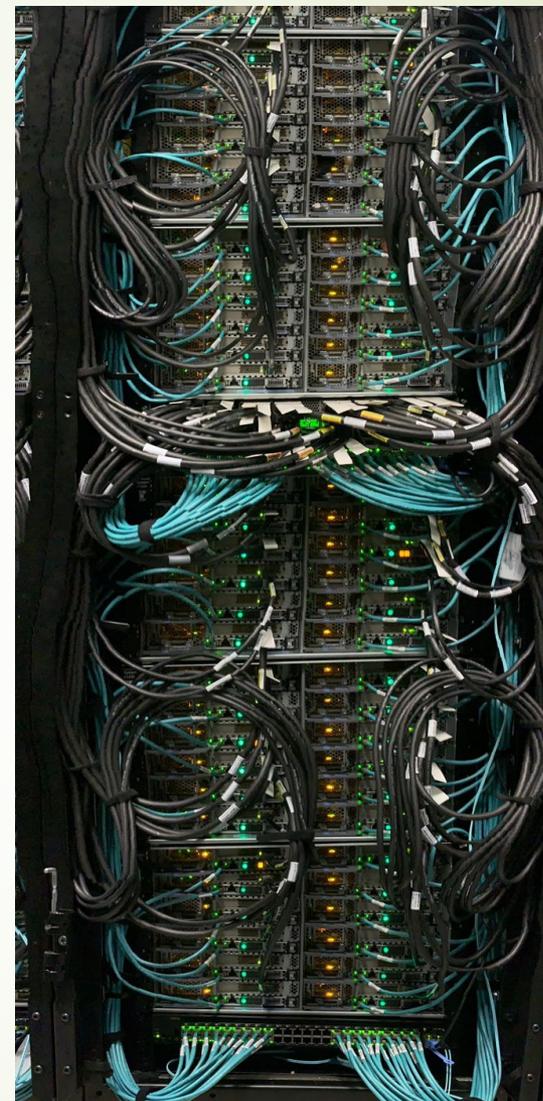
**GPU**

(processore grafico)

**FPGA**

# HPC@LNGS

Nel corso del 2022 i LNGS hanno acquisito dal CINECA, il più importante centro di supercalcolo in Italia, **400 server di calcolo HPC**, interconnessi da una rete ad alta velocità (100Gb/s), per il calcolo parallelo!!!



# HPC@LNGS

Stiamo installando le macchine, la rete, lo storage e stiamo implementando i servizi per iniziare nel 2023 le prime attività di test del sistema!!





# HPC4DR ai LNGS

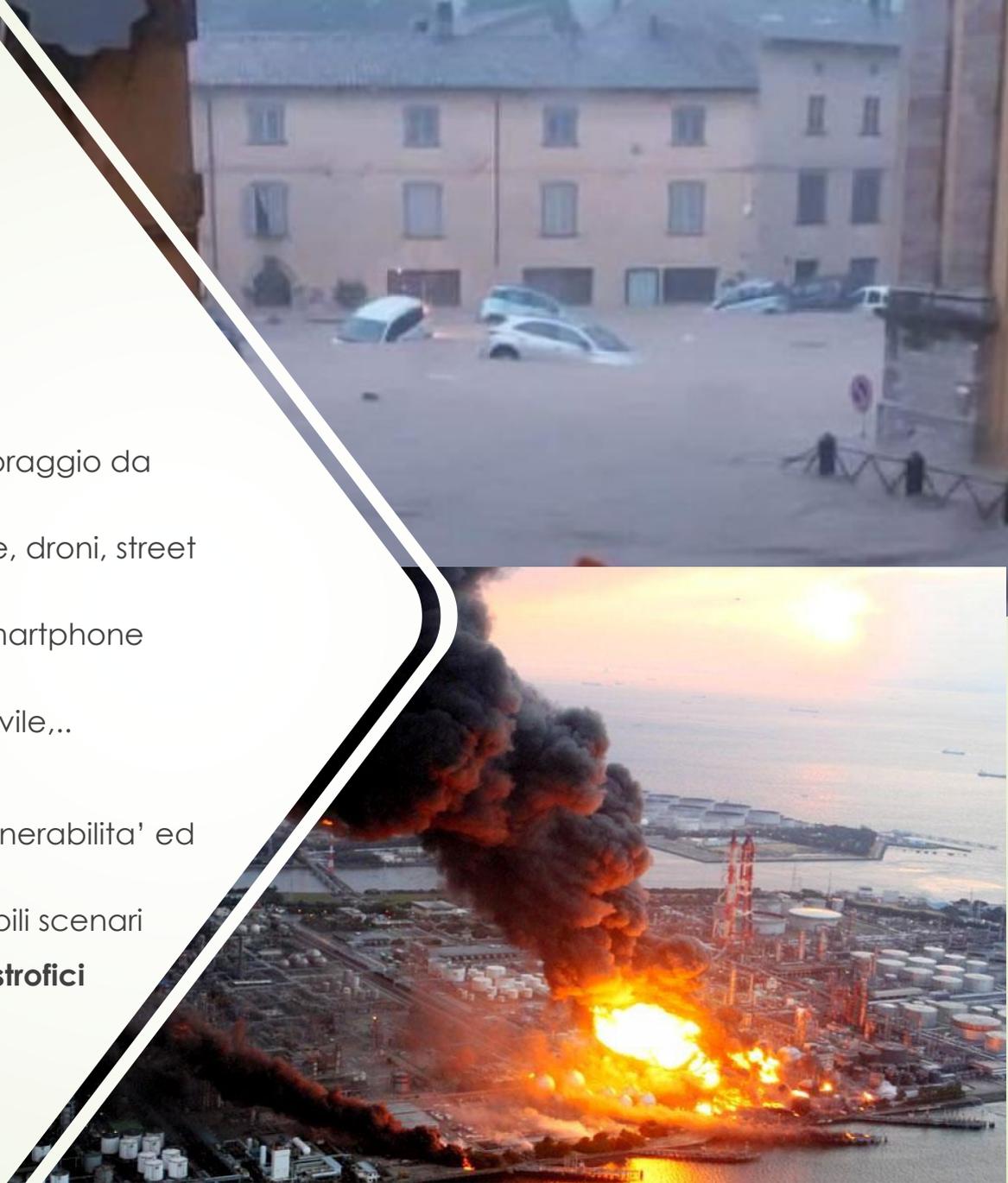
La comunità scientifica che utilizzerà l'HPC sarà quella degli esperimenti dei LNGS e anche quella di un consorzio, **HPC4DR**

Il progetto **HPC4DR** nasce nel 2020 da Università e centri di ricerca delle regioni Abruzzo, Marche e Molise, duramente colpite dai terremoti e da altri eventi catastrofici

L'idea è quella di "realizzare un **centro di competenze** per la riduzione dei rischi connessi ai disastri dovuti a fenomeni naturali e di origine umana, dotato di un'**infrastruttura tecnologica di calcolo ad alte prestazioni**

# HPC e AI per capire il rischio

- ▶ FAIR e Open BIG DATA
  - ▶ IoT per raccolta dati di monitoraggio da sensori
  - ▶ Raccolta immagini da satellite, droni, street level
  - ▶ Immagini e informazioni da smartphone applications
  - ▶ Dati su infrastrutture, edilizia civile,..
- ▶ ML, AI e HPC
  - ▶ Generare analisi su rischio, vulnerabilita' ed esposizione
  - ▶ Simulazioni per produrre possibili scenari
  - ▶ **Early warning per eventi catastrofici**
  - ▶ ...

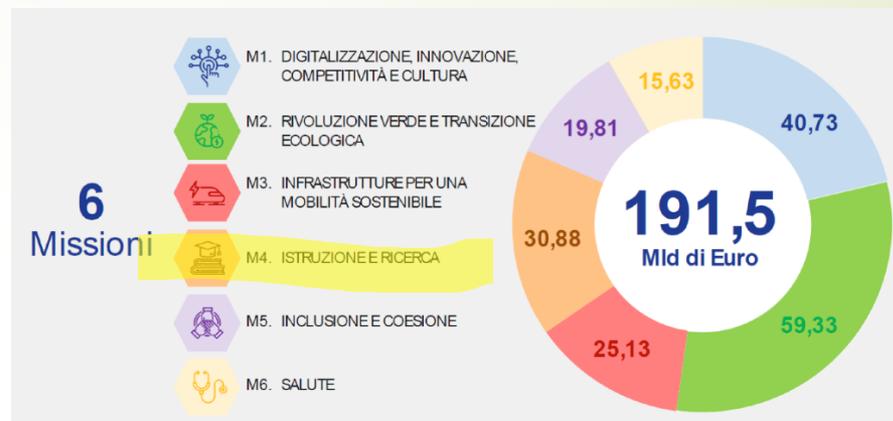


# HPC4DR ai LNGS

L'infrastruttura tecnologica di calcolo ad alte prestazioni ospitata ai LNGS richiede l'adeguamento del CED dei LNGS

- Potenza fino a 400KW
- Acquisizione di GPU e storage fino a 5PB

I lavori per l'adeguamento del CED e l'acquisizione di nuove risorse di calcolo sono inseriti nel **PNRR**



Le LINEE GUIDA

## Le LINEE GUIDA

Le Linee Guida contengono indicazioni chiave per 4 **Misure** della Componente 2 «Dalla ricerca all'impresa» della Missione 4 «Istruzione e ricerca» del PNRR:

INVESTIMENTO	DURATA	RISORSE IN MILIONI DI €	BENEFICIARI
1.3 Partenariati allargati estesi a Università, centri di ricerca, imprese e finanziamento progetti di ricerca di base	2022-2026	1610	Università, enti pubblici di ricerca, imprese
1.4 Potenziamento strutture di ricerca e creazione di «campioni nazionali di R&S» su alcune KET	2022-2026	1600	Università, enti pubblici di ricerca, imprese
1.5 Creazione e rafforzamento di «ecosistemi dell'innovazione», costruzione di «leader territoriali di R&S»	2022-2026	1300	Università, enti pubblici di ricerca, imprese
3.1 Fondo per la costruzione di un sistema integrato di infrastrutture di ricerca e di innovazione	2022-2026	1580	Università, enti pubblici di ricerca, imprese

PNEXTGENERATIONITALIA

Ministero dell'Università e della Ricerca

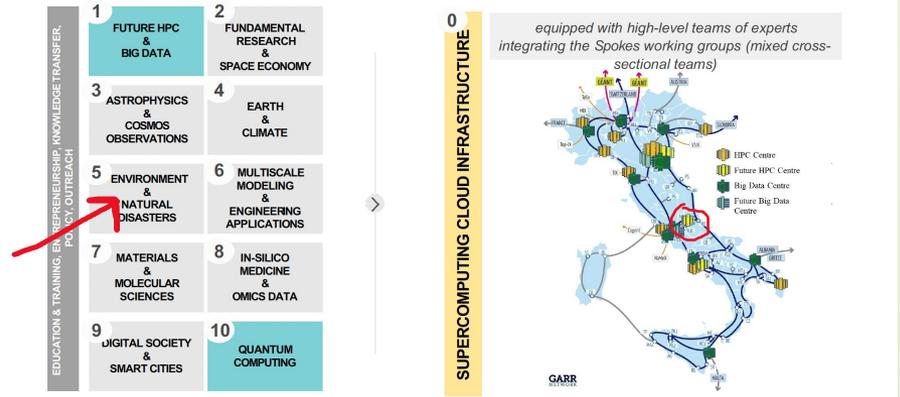
# HPC4DR ai LNGS

I **LNGS** sono stati **inclusi nel Centro Nazionale HPC, Big Data e Quantum Computing** finanziato dal **PNRR**

Oltre alle risorse finanziarie per il datacenter e per i calcolatori, il PNRR creerà opportunità di **lavoro per giovani diplomati e laureati!**



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



# PCTO e borse di studio ai LNGS

Il Servizio Calcolo e Reti dei LNGS ospita studenti delle scuole superiori della Regione Abruzzo per percorsi PCTO  
L'ultimo, il 12 settembre 2022.

Borse di studio per diplomati e laureati sono emesse regolarmente dai LNGS per la formazione di giovani!



# La fisica e' da sempre al passo con le nuove tecnologie..

- Web nato al CERN negli anni '90
- L'INFN aveva una propria rete dati gia' dalla fine degli anni '80 quando la rete non c'era!!!
- I Data Center di GRID/cloud sono molto diversi dai data center dei provider commerciali

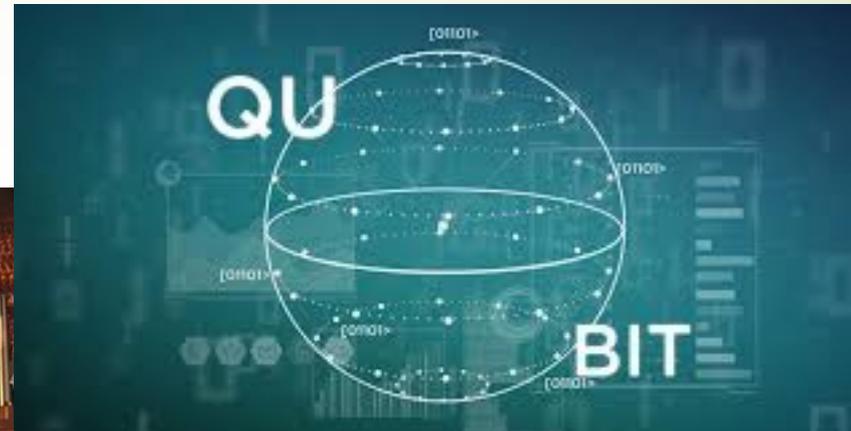


# Il futuro prossimo: AI

- ▶ L' **intelligenza artificiale, AI** e in particolare le tecniche di **Machine Learning** già' applicata dai provider commerciali per la ricerca di informazione all'interno dei Big Data, sarà applicata alla selezione degli eventi interessanti (L1 e L2 trigger) per aumentare la soppressione dei falsi-positivi e la conseguente riduzione dei dati da salvare e analizzare!



# Il futuro non così remoto: quantum computing



# La fisica: non solo fisici!!!!

- **Amministratori di sistemi** informatici: gestiscono server, storage, rete, sistemi operativi e applicazioni, sicurezza informatica
- **Tecnologi** che gestiscono Data center
- **Ricercatori/tecnologi** che sviluppano software online DAQ/Trigger 2L
- Ricercatori/tecnologi che sviluppano il modello di calcolo dell'esperimento
- Ricercatori/tecnologi che sviluppano software per l'analisi e la visualizzazione dei dati (qui le nostre esercitazioni)

**Lavorano dietro le quinte: non prenderanno mai un Nobel ma senza di essi non ci sarebbero risultati e scoperte scientifiche!**



..grazie!

Per ulteriori info..

[sandra.parlati@lngs.infn.it](mailto:sandra.parlati@lngs.infn.it)