



IL CALCOLO AI LABORATORI DEL GRAN SASSO

Miriam Olmi – PID LNGS 2024

PERCHÉ 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** ma 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 è il compito principale dei servizi di calcolo nei laboratori e nelle sezioni INFN!

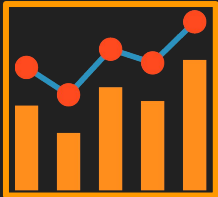
I DATI, COSÌ PREZIOSI!



I dati sono la materia prima della ricerca scientifica!



I dati sono costosi e preziosi!

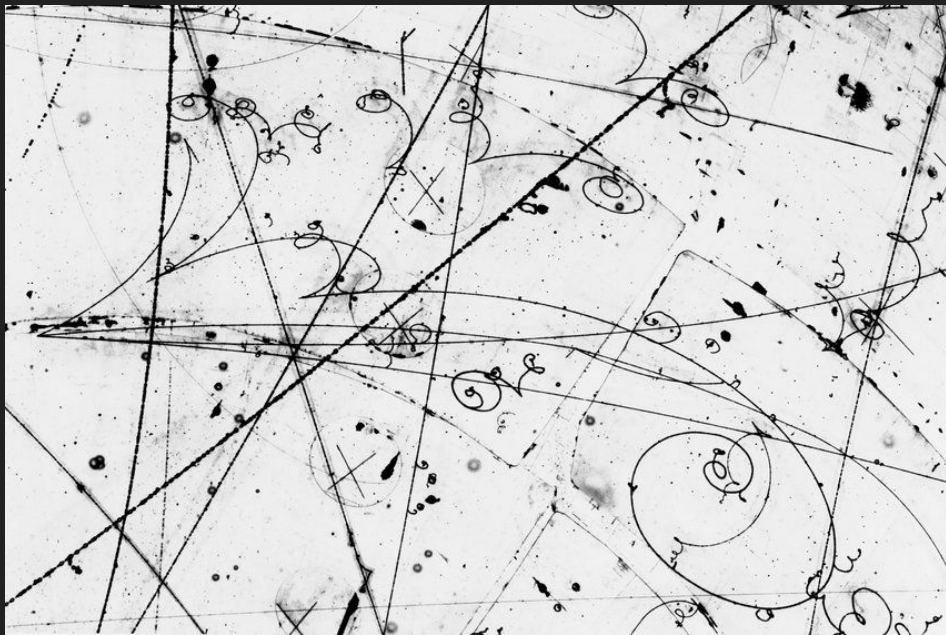


Tutti gli esperimenti di fisica, per arrivare a trovare un risultato, si basano sull'analisi dei dati.

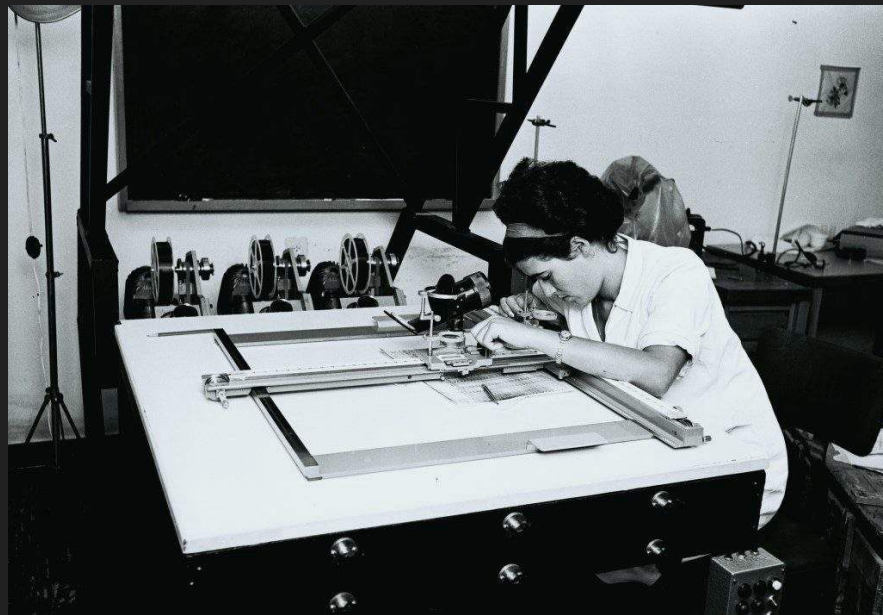


I dati un tempo erano scritti a mano con carta e penna!

CARTA, PENNA E TANTA PAZIENZA



An example of tracks captured by a bubble chamber. These images were typically manually reconstructed by trained human scanners. Illustration courtesy of Fermilab.



Scanning, circa 1963. Image: Weizmann Institute Archives



I DATI, I COMPUTER E LE RETI

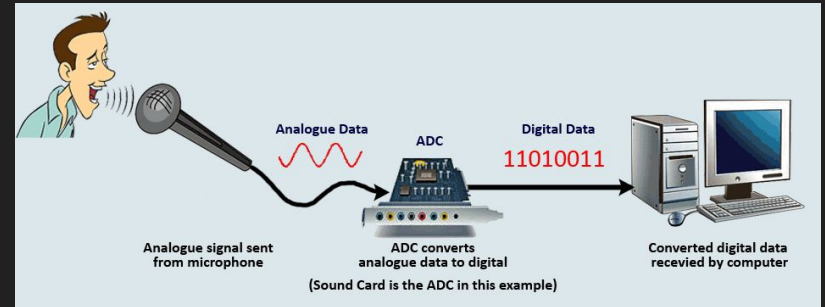
Fin dagli anni '60 in fisica delle particelle si sono usati i computer: dal sistema di **acquisizione** dell'apparato sperimentale, alla **rappresentazione grafica** dei dati, all'**analisi** complessiva dei dati acquisiti, alla scoperta scientifica e alla pubblicazione degli articoli su riviste specializzate:

TUTTO QUESTO NON SAREBBE POSSIBILE SENZA I CALCOLATORI E LE RETI!

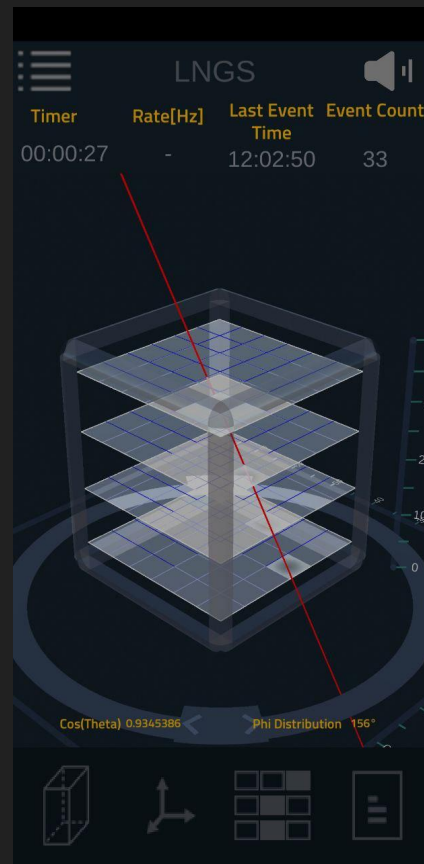
La necessità di potenza di calcolo per gli esperimenti ha portato l'INFN ad essere uno degli attori principali nel campo delle infrastrutture di calcolo e dell'evoluzione tecnologica!

I DATI IN FORMATO DIGITALE

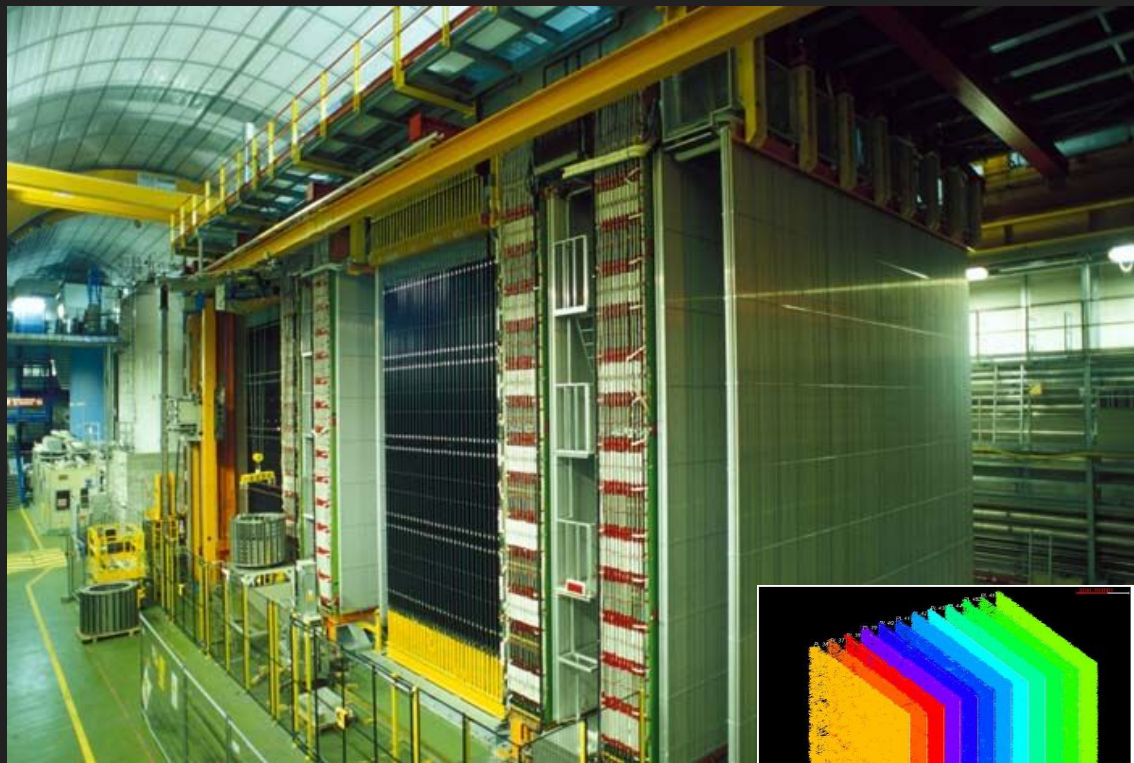
- Tutti i rivelatori di particelle quindi sono progettati per produrre un segnale in uscita che sia convertibile in un dato digitale.
- Generalmente un rivelatore converte il segnale lasciato da una particella (luce, temperatura,..) in un segnale elettrico analogico il quale viene successivamente trasformato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



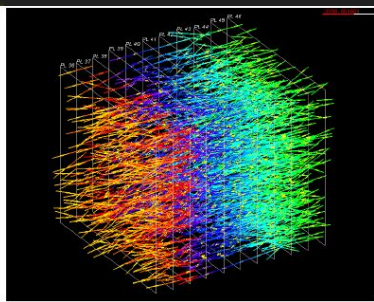
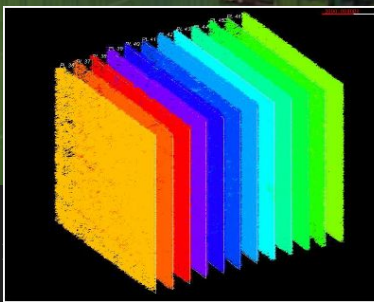
UN SEMPLICE RIVELATORE DI RAGGI COSMICI...



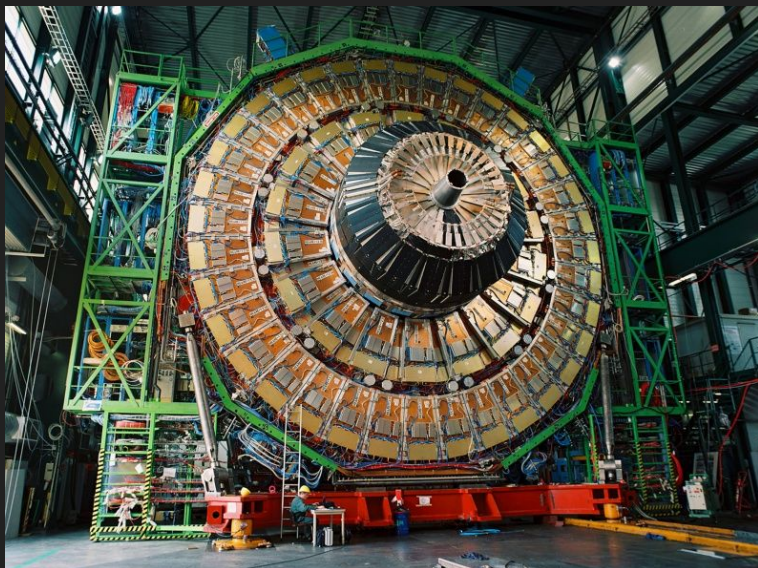
...UN GRANDE ESPERIMENTO AI LNGS...



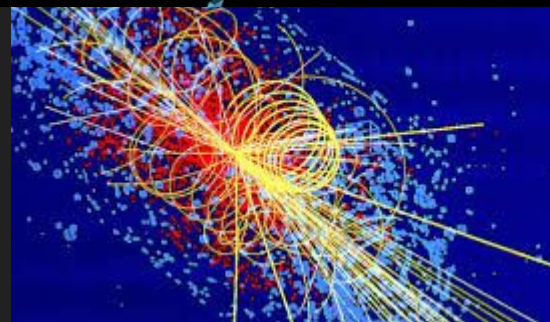
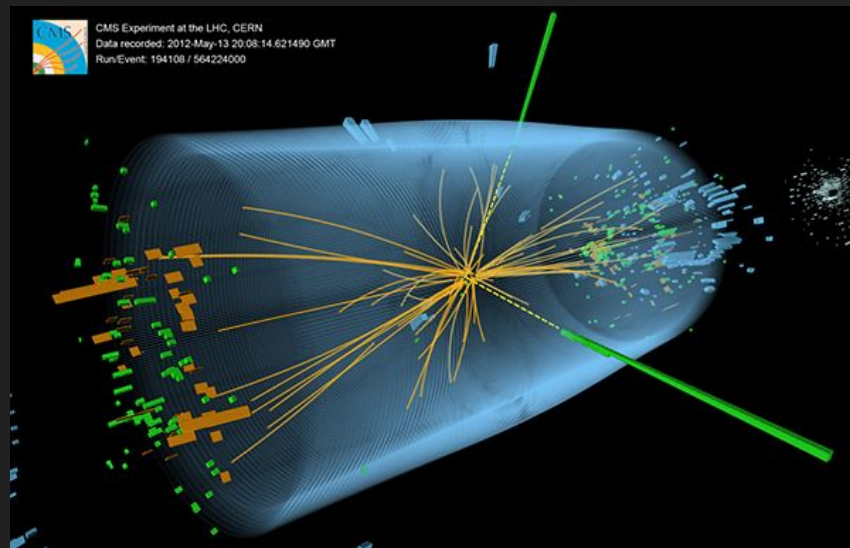
OPERA. Different steps of the emulsion data processing.
(Left) All base-tracks in 11 films of the volume under study are reconstructed; they participate in the alignment process from which tracks are reconstructed (middle); passing-through tracks are discarded and the interaction vertex is reconstructed (right).



...UN ESPERIMENTO COMPLESSO AL CERN

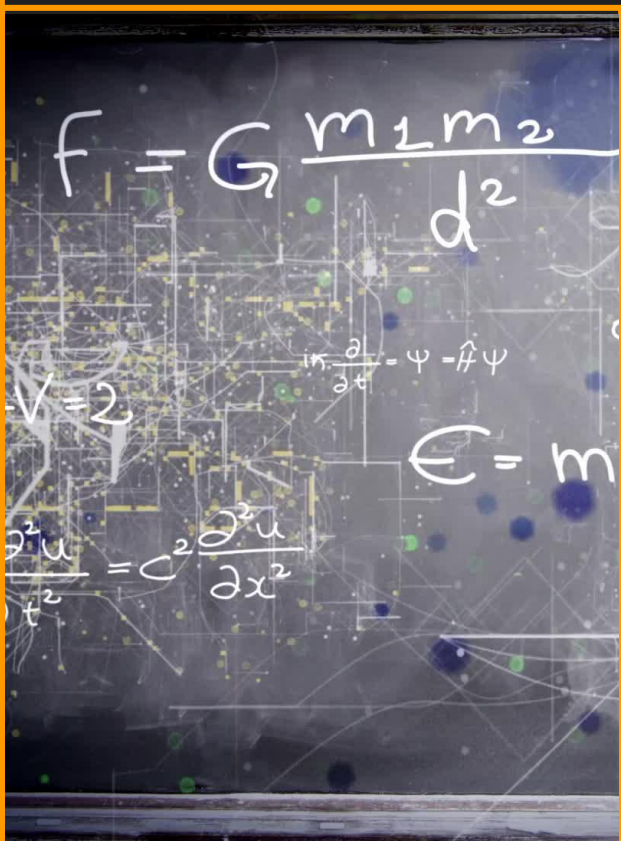


CMS experiment at CERN. Event reconstruction in CMQ viewer (top right) and picture of a Higgs event (bottom right)



TANTE FASI PER ARRIVARE AL RISULTATO





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**

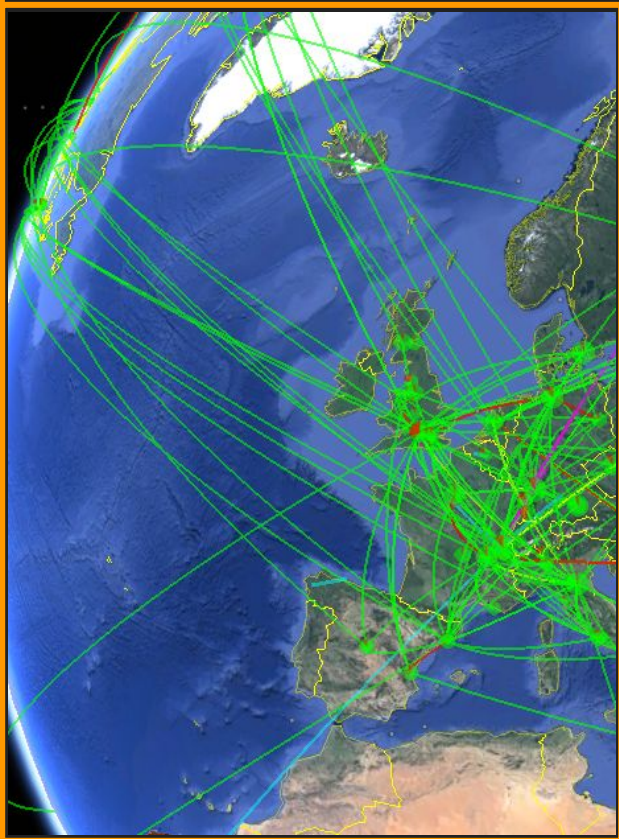
Sono tantissimi dati, ma non tutti portano informazioni utili!

SELEZIONARE EVENTI INTERESSANTI È FONDAMENTALE

- ❑ 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 può 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 quantità di dati è ridotta di un fattore significativo

QUAL È LA SITUAZIONE 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**
- ❑ In parte i dati sono analizzati ai LNGS su farm di server installate in sala calcolo e in parte prendono una strada differente..
- ❑ ...ma in ogni caso vanno analizzati!



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.

GARR NETWORK

- ❑ 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** per i centri Tier2 e multipli di **100Gb/s** per i Tier1
- ❑ La connessione dei LNGS alla rete GARR e alla GRID è attualmente di 10Gb/s.





I DATA-CENTER 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

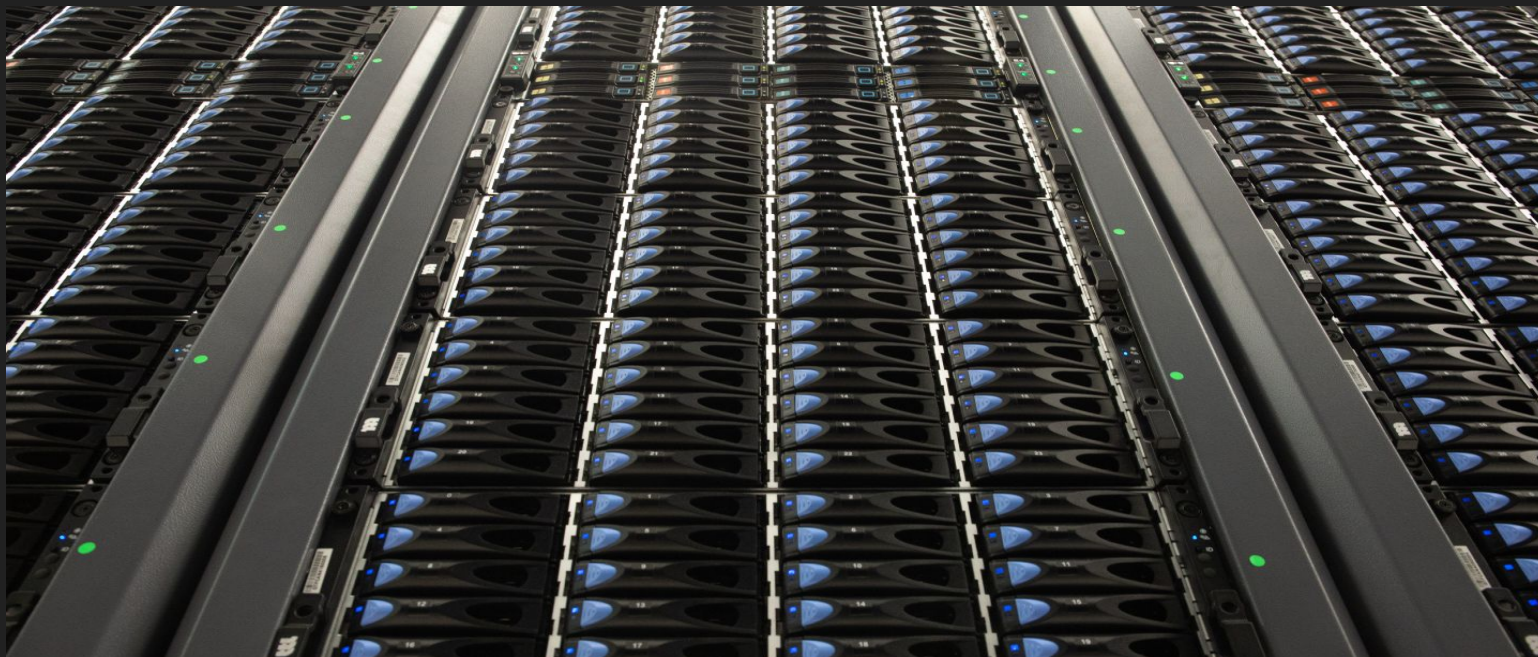
DATA-CENTER



SERVER NEGLI ARMADI RACK



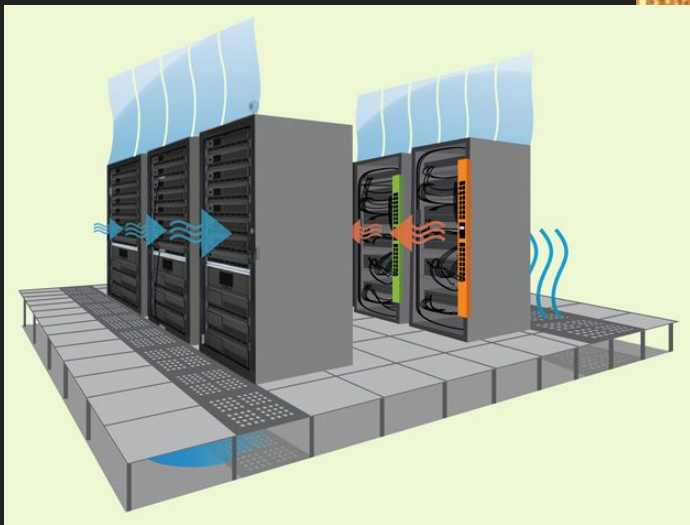
STORAGE



NASTRI MAGNETICI



RAFFREDDAMENTO DEL DATA-CENTER



ALIMENTAZIONE DEL DATA-CENTER





QUALCHE NUMERO SUL DATA-CENTER LNGS

- ❑ Circa **2PB** di dati distribuiti tra vari sistemi di storage (i dischi più moderni acquistati ai LNGS hanno una capacità di 16TB)
- ❑ Circa **1000 core** per l'analisi dati, tra server gestiti dal servizio calcolo e server degli esperimenti (i server più 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

MA 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 soluzione 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!!!

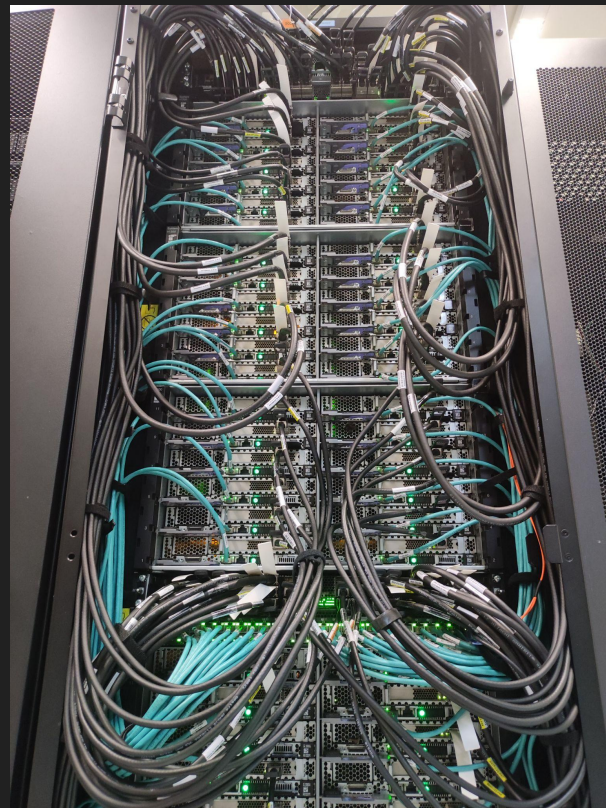
HPC
(calcolo ad alte prestazioni)

GPU
(processore grafico)

FPGA
(Field Programmable Gate Arrays)

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!!!

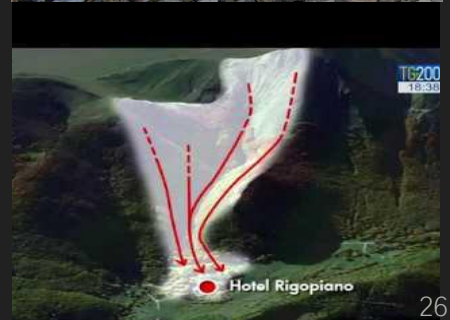


HPC4DR @ LNGS

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
- ❑ ...

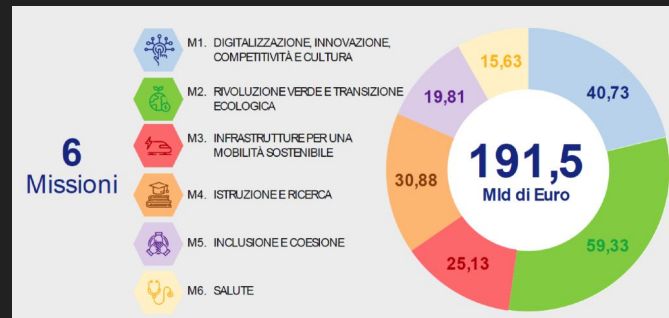


HPC4DR @ LNGS

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

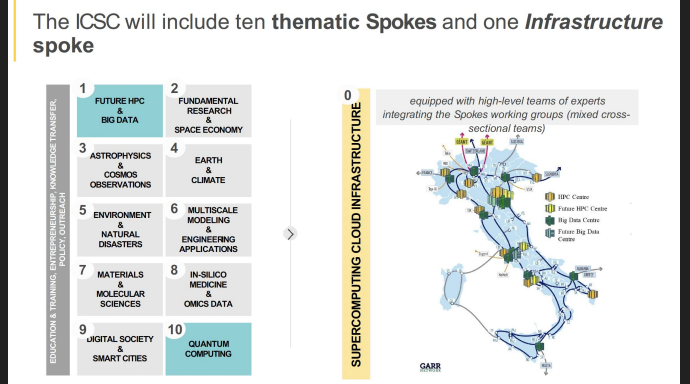
INVESTIMENTO	DURATA	RISORSE IN MILIONI DI €	BENEFICIARI
1.3 Partenariati allargati estesi a Università, centri di ricerca, imprese e finanziamento progetti di ricerca di base	2022-2026	1610	Università, enti pubblici di ricerca, imprese
1.4 Potenziamento strutture di ricerca e creazione di «campioni nazionali 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

INTEGRO NAZIONALE
Ministero dell'Università e della Ricerca

HPC4DR @ 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 ha creato opportunità di lavoro per giovani diplomati e laureati!



PCTO E BORSE DI STUDIO @ 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 È SEMPRE AL PASSO CON LE NUOVE TECNOLOGIE

- ❑ Web nato al CERN negli anni '90
- ❑ L'INFN aveva una propria rete dati già' 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

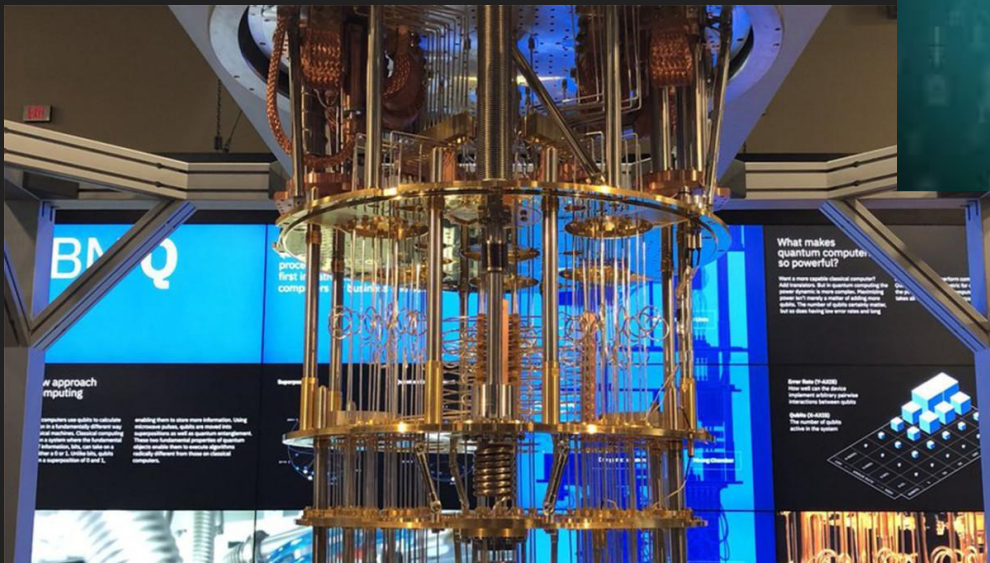


FUTURO PROSSIMO: AI

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



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)

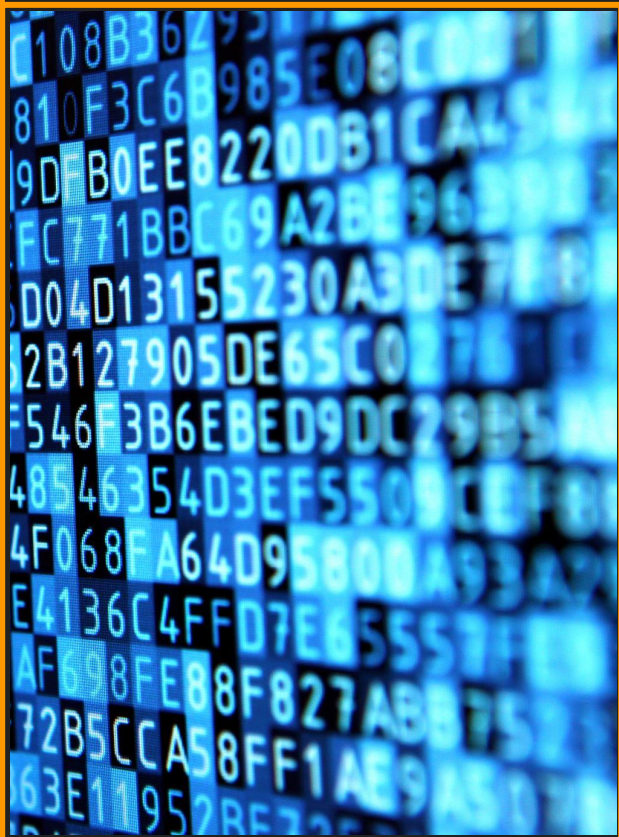
Lavorano dietro le quinte: non prenderanno mai un Nobel ma senza di essi non ci sarebbero risultati e scoperte scientifiche!



**GRAZIE PER
L'ATTENZIONE**

miriam.olmi@lngs.infn.it

SPARE



“RIPASSO”

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