

CINECA

**Big data e intelligenza artificiale:
il caso CoronaVirus**

David Vannozzi
CINECA – Consorzio Interuniversitario
Direttore Generale
Motore Sanità
Asiago 18 Settembre 2020

UN CONSORZIO IN CONTINUA ESPANSIONE

CINECA opera su tutto il territorio al servizio del **sistema accademico**, dell'**istruzione** e della **ricerca nazionale**.

69

ATENEI

11

**ISTITUZIONI PUBBLICHE
NAZIONALI**



**Ministero
dell'Istruzione**

**Ministero dell'Università
e della Ricerca**

50 ANNI DI STORIA

Al servizio del Paese e del suo Sistema Accademico



4 ATENEI FONDATORI

Bologna, Firenze, Padova, Venezia

1969

SUPERCOMPUTING



1980

MIUR



SANITA' PA INDUSTRIA



1990

**DIGITALIZZAZIONE
UNIVERSITA**



Futuro

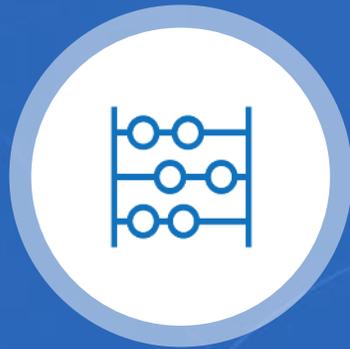
**CHATBOT
BLOCKCHAIN
BIG DATA
QUANTUM
COMPUTING**



200.000.000.000.000.000

LE ATTIVITÀ PRINCIPALI DEL CONSORZIO

CINECA offre supporto alle attività della comunità scientifica tramite il **supercalcolo** e le sue applicazioni, realizza **sistemi gestionali** per le amministrazioni universitarie e il MIUR, progetta e sviluppa **sistemi informativi** per pubblica amministrazione, sanità e imprese.



SUPERCALCOLO



**SUPPORTO
AL MIUR**



**SUPPORTO
AGLI ATENEI ED
AI
CONSORZIATI**



**TRASFERIMENTO
TECNOLOGICO**

RISORSE DI SUPERCALCOLO A SUPPORTO DELLA RICERCA ACCADEMICA NAZIONALE

MARCONI100: installato nel 2020, è al **9° posto** nella classifica dei supercalcolatori mondiali (20 milioni di miliardi di calcoli al secondo)



LEONARDO calcolatore **pre-exascale** finanziato da **EURO HPC** sarà ospitato da CINECA al **Tecnopolo di Bologna** (200 milioni di miliardi di calcoli al secondo)



DATACENTER A SUPPORTO DELLA PUBBLICA AMMINISTRAZIONE

2k+

MACCHINE
VIRTUALI

318.000

PROCESSORI

2
sedi

- BOLOGNA
- ROMA

24

SERVIZI
QUALIFICATI
SAAS

✓
Disaster
Recovery

✓
ISO9001

✓
ISO/IEC 27001
• + ISO/IEC 27017
• + ISO/IEC 27018

CINECA è **Cloud Service Provider**
per i propri consorziati e per gli Atenei italiani

CINECA INNOVATION LABS



BIG DATA

- ECMWF – Centro Meteo
- Analisi epidemiologiche



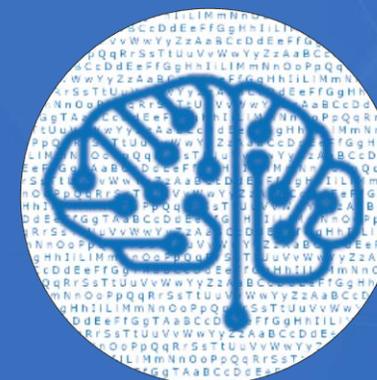
BLOCKCHAIN

- Embedded blockchain
- Blockcerts



MACHINE LEARNING

- Manutenzione predittiva
- Previsione abbandoni



NATURAL LANGUAGE PROCESSING

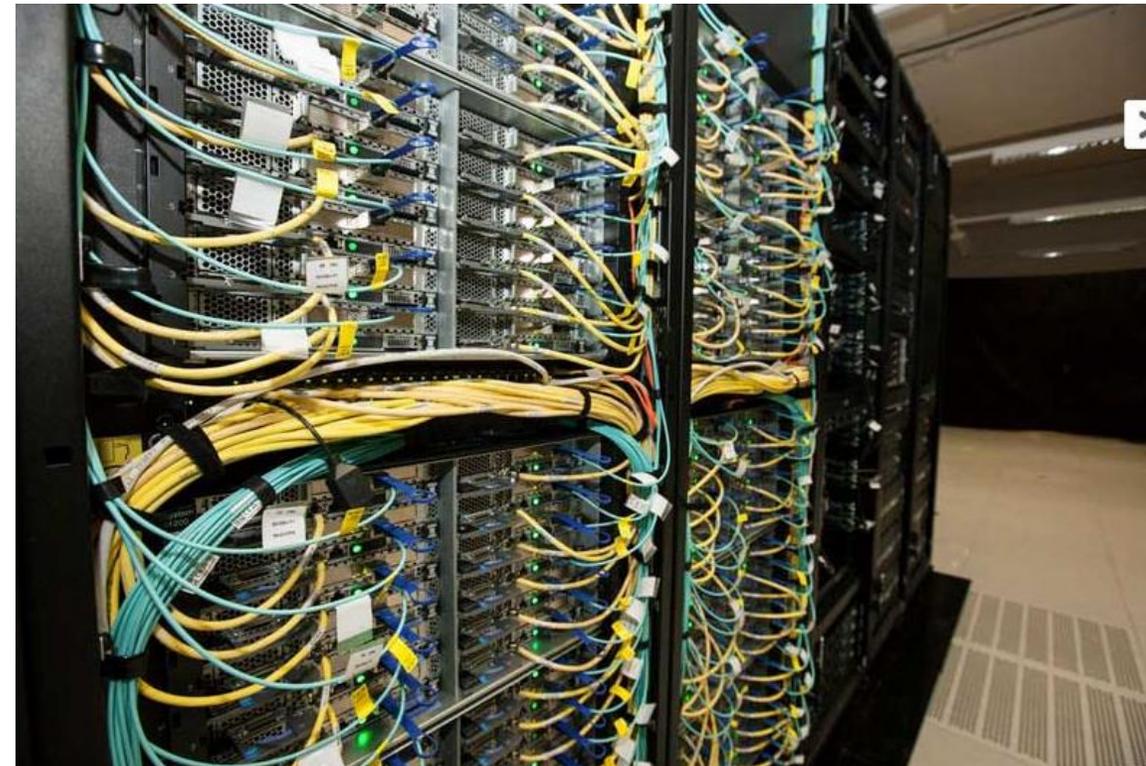
- Chatbot

COSA È UN SUPERCOMPUTER E COME È FATTO

La principale differenza tra un computer normale e un supercomputer è l'approccio con cui queste macchine vengono ingegnerizzate: i supercomputer, infatti, nascono per la cosiddetta "elaborazione parallela".

Detta in parole molto semplici, un supercomputer in realtà è formato da tanti computer normali che lavorano "in parallelo" sugli stessi dati. Quindi la potenza di ogni computer viene moltiplicata per il numero di computer installati dentro il supercomputer.

Da ciò deriva la particolarissima architettura di questi mostri da calcolo: delle lunghissime file di armadi, dentro i quali si trovano i computer veri e propri ma "nudi" (cioè senza il case che di norma ricopre i nostri desktop) e montati su dei cosiddetti rack. Sembrano dei vassoi, su cui sono montati scheda madre, processore, RAM, unità disco, alimentatore e tutte le altre componenti hardware. Ogni rack è collegato a tutti gli altri e la sua potenza si somma a quella degli altri "vassoi" creando, appunto, un unico supercomputer dalle prestazioni eccezionali.



QUANTO È POTENTE UN SUPERCOMPUTER

Le prestazioni di un supercomputer moderno si misurano in **Peta FLOPS**, dove FLOPS sta per *Floating Point Operations Per Second*, traducibile in italiano con "Calcoli a virgola mobile per secondo". Peta, invece, sta per 10 alla quindicesima, cioè un biliardo (equivalente al *quadrillion*, nei paesi anglosassoni), cioè 1.000.000.000.000.000.

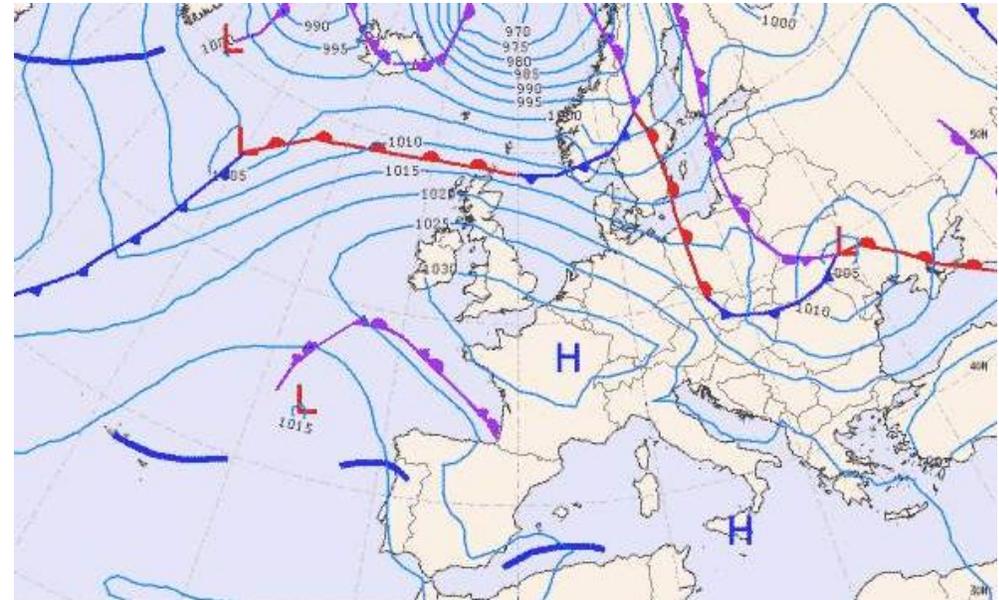
Il supercomputer del CINECA, attualmente quinto tra i più potenti al mondo, ha una capacità di calcolo pari a 50 Peta FLOPS. Per raggiungere questa potenza il Marconi 100, che è costato 25 milioni di Euro, occupa uno spazio di 500 **metri quadrati** e ha un peso totale di 100 **tonnellate**. È composto circa da 7.300.000 **computer** che lavorano in parallelo tramite 5.000 nodi , connessi da quasi 100 chilometri di cavi.



A COSA SERVONO I SUPERCOMPUTER

Vista la loro architettura parallela, i supercomputer sono estremamente efficienti nell'**eseguire calcoli ripetitivi su una enorme mole di dati**. Per questo sono adatti al calcolo puro a fini scientifici in ambiti come l'astronomia, la fisica, la chimica, ma l'utilizzo più famoso di queste macchine è per la **meteorologia**.

I supercomputer, infatti, vengono usati per calcolare i modelli climatici tramite i quali è possibile ipotizzare, con una buona approssimazione, che tempo farà nelle prossime ore, giorni, settimane. Le informazioni meteo, negli ultimi 30 anni, sono sempre più accurate proprio grazie all'aiuto dei supercomputer che elaborano le informazioni e riescono a combinare tra loro sempre più dati, creando modelli sempre più accurati. Tuttavia, è stato calcolato che per creare un modello meteo con affidabilità tendente al 100% è necessaria una potenza di calcolo nell'ordine dei 50/100 PetaFLOPS. Cioè almeno 50.000.000.000.000.000.000 di operazioni al secondo.



IL FUTURO, E I LIMITI, DEI SUPERCOMPUTER

I **supercomputer consumano moltissima energia elettrica** e per farli lavorare bisogna garantire una fornitura elettrica eccellente: abbondante, stabile, senza sbalzi e con un sistema di backup in caso venga a mancare l'alimentazione del tutto o in parte. Il Marconi 100, ad esempio, consuma 1,6 MW (MegaWatt) di energia elettrica. Una quantità pari a quella sufficiente per alimentare quasi 9000 utenze elettriche domestiche a pieno carico per un anno.

Come se non bastasse, all'energia consumata dal supercomputer bisogna aggiungere quella necessaria ad alimentare il sistema di raffreddamento dell'edificio che lo ospita: 7.300.000 computer che lavorano in parallelo scaldano e devono essere tenuti a temperatura controllata altrimenti avrebbero vita breve. Il futuro dei supercomputer, quindi, come è stato già in passato dipende dalle nuove tecnologie (in particolare dal miglioramento dei processi produttivi dei transistor) che permettono ai nuovi processori di consumare meno a parità di potenza, o, viceversa, di sviluppare più potenza a parità di consumi e di calore sviluppato.



IL PROGETTO EXSCALATE4COV

Tra i progetti di ricerca in corso spicca **Exscalate4CoV**, finanziato con tre milioni dalla Commissione Europea, che vede impegnato un folto gruppo di centri di ricerca del nostro paese nell'uso di tecnologie di supercalcolo e di urgent computing per contrastare la pandemia in corso.

Il progetto è condotto da un consorzio che aggrega 18 istituzioni di 7 nazioni europee, di cui 9 italiane, fra le quali il Consorzio Universitario Cineca.

Exscalate4CoV sta usando il supercomputer Marconi100 del Cineca per passare in rassegna una vasta collezione di molecole: si tratta di farmaci già approvati per l'uso nell'uomo e pronti per nuovi studi clinici, ma anche di principi attivi nuovi non ancora testati sulla nostra specie.

Scopo della valutazione virtuale di queste molecole è verificare le loro potenziali capacità di contrastare il virus in modo efficace. Attraverso lo screening virtuale di farmaci già approvati per uso umano, anche se non differenti indicazioni terapeutiche, è possibile selezionare un numero abbastanza ristretto di molecole da testare in tempi brevissimi sull'uomo, consentendo una risposta rapida all'emergenza generata dalla pandemia.

Il punto di partenza è la piattaforma ExScalate, che raccoglie una libreria digitale di 500 miliardi di molecole di facile sintesi che, quindi, possono essere trasformate rapidamente in farmaci. Questo enorme database, sviluppato negli ultimi 15 anni e messo a disposizione della comunità scientifica, consente di prevedere con tecniche computazionali l'attività farmacologica di un enorme numero di molecole.

L'ESPERIENZA MATURATA SUL VIRUS ZIKA

Le potenzialità di **ExScalate** nell'individuare in tempi rapidi farmaci contro possibili epidemie virali o batteriche sono già state testate in passato nell'ambito del progetto europeo **Antarex**. Nel corso di Antarex abbiamo simulato con il supercomputer Marconi un caso applicativo per individuare possibili molecole candidate alla cura del virus Zika, che nel 2016 arrivò a minacciare lo svolgimento delle Olimpiadi di Rio de Janeiro. Si è trattato del più grande esperimento di *virtual screening* effettuato con un software di simulazione fatto girare su *thread* paralleli, con una potenza di calcolo pari a 10 petaflop, cioè 10 milioni di miliardi di operazioni al secondo.

I risultati sul virus Zika sono stati sottoposti a una procedura di valutazione e validazione che oggi può essere messa a frutto anche dal progetto ExScalate4CoV.

Il progetto Antarex è stato un antesignano di ExScalate4CoV e ha consentito di definire procedure e di accumulare esperienze oggi molto utili sia per la ricerca su SARS-CoV-2, sia ottimizzare le operazioni di supercalcolo.

Oggi il supercomputer Marconi100 sta lavorando a una velocità da 50 petaflop.

E grazie agli investimenti deciso dall'Italia e dall'Europa è in corso un ulteriore upgrade che lo porterà a 200 petaflop, che potrebbe collocarlo tra i 5 supercomputer più potenti del mondo.

Non appena, il 17 gennaio scorso, è stato reso noto il genoma del virus, il gruppo di lavoro è partito con una prima valutazione di tutte le molecole già disponibili per l'uso umano.

Questa prima fase del progetto **ExScalate** ha portato ad isolare un ristretto numero di farmaci potenzialmente attivi contro il coronavirus SARS-CoV-2 e di pronta produzione, nell'arco di poco tempo. I candidati identificati saranno sottoposti ad altre valutazioni, per arrivare a strutturare insieme con la European Medicine Agency (EMA), un modello di sperimentazione efficace delle molecole per velocizzarne l'uso terapeutico.

La ricerca si sta estendendo poi a tutto il resto della piattaforma **ExScalate** e, quindi, in prospettiva a tutti i 500 miliardi di molecole del database.

La potenza di calcolo richiesta dalla simulazione di un numero così elevato di molecole è enorme.

Le proteine sono entità dinamiche, che si muovono nell'ambiente biologico in cui si trovano. Perché la simulazione sia accurata occorre applicare particolari tecniche di «dinamica molecolare», che diano al modello virtuale la dinamicità e plasticità delle molecole reali, così da valutare la loro effettiva capacità di legare le molecole testate. I calcoli necessari a questo tipo di simulazione sono molto lunghi e per questo Cineca ha messo a disposizione tutta la sua potenza di calcolo.

Il progetto **ExScalate2CoV** ha destato grande interesse in tutta la comunità scientifica e sono giunte subito proposte di collaborazione da tutto il mondo, sia da centri di ricerca che da aziende private.

Quando avremo un risultato positivo dagli screening delle 50 molecole già individuate si porrà infatti il problema di come procurarsi o sintetizzare velocemente il farmaco per i test dell'uomo.

Se c'è una lezione da imparare da Covid-19 è proprio che certe minacce vanno affrontate tutti insieme, e non in ordine sparso.



CINECA

Consorzio interuniversitario per il calcolo automatico