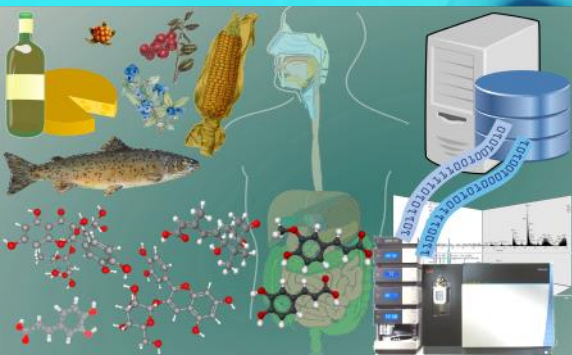


Festival della Scienza

Genova, 21 ottobre _ 5 novembre 2021

Mappe



Dalla scheda madre alla “Madre Terra” *L'interdisciplinarietà nella ricerca agroalimentare, nutrizionistica e sulla sicurezza alimentare*

2 - 5 novembre 2021 ore 10.00 – 11.00

- **Matematica e (super)calcolo per comprendere il presente, descrivere il passato, prevedere il futuro – Luisa Carracciuolo (IPCB)**
- **TITOLO – NOME Cognome relatore (ISTITUTO)**
- **TITOLO – NOME Cognome relatore (ISTITUTO)**

DigiLab: Estrazione DNA - (IGB, ISA e ISPAAM)

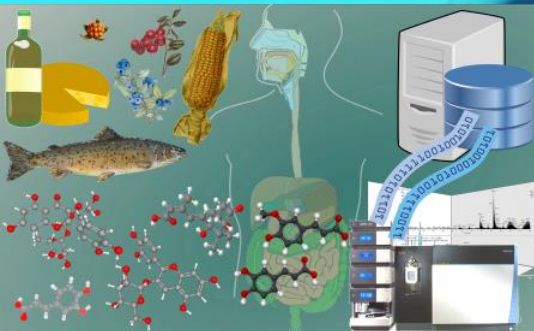
Regia di **Maria Staiano (ISA), Rosarita Tatè (IGB) e Pietro Amodeo (ICB)**



Consiglio Nazionale delle Ricerche



Mappe



2 - 5 novembre 2021 ore 10.00

**Matematica e (super)calcolo per
comprendere il presente,
descrivere il passato, prevedere il
futuro**

Luisa Carracciuolo (IPCB)



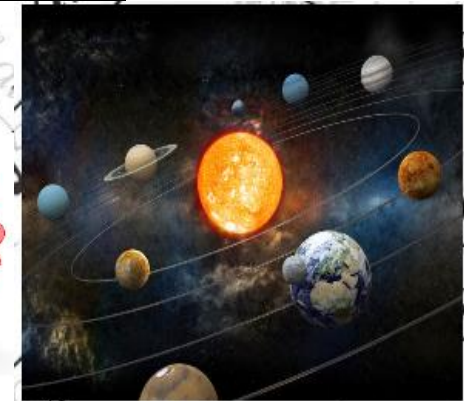
Festival della Scienza

Genova, 21 ottobre _ 5 novembre 2021

Mappe



Da sempre l'uomo si è servito, e si serve, degli strumenti della **matematica** e del **calcolo** per comprendere, e gestire, la realtà nella quale vive



Consiglio Nazionale delle Ricerche

iPCB ISTITUTO PER I
POLIMERI
COMPOSITI E
BIOMATERIALI

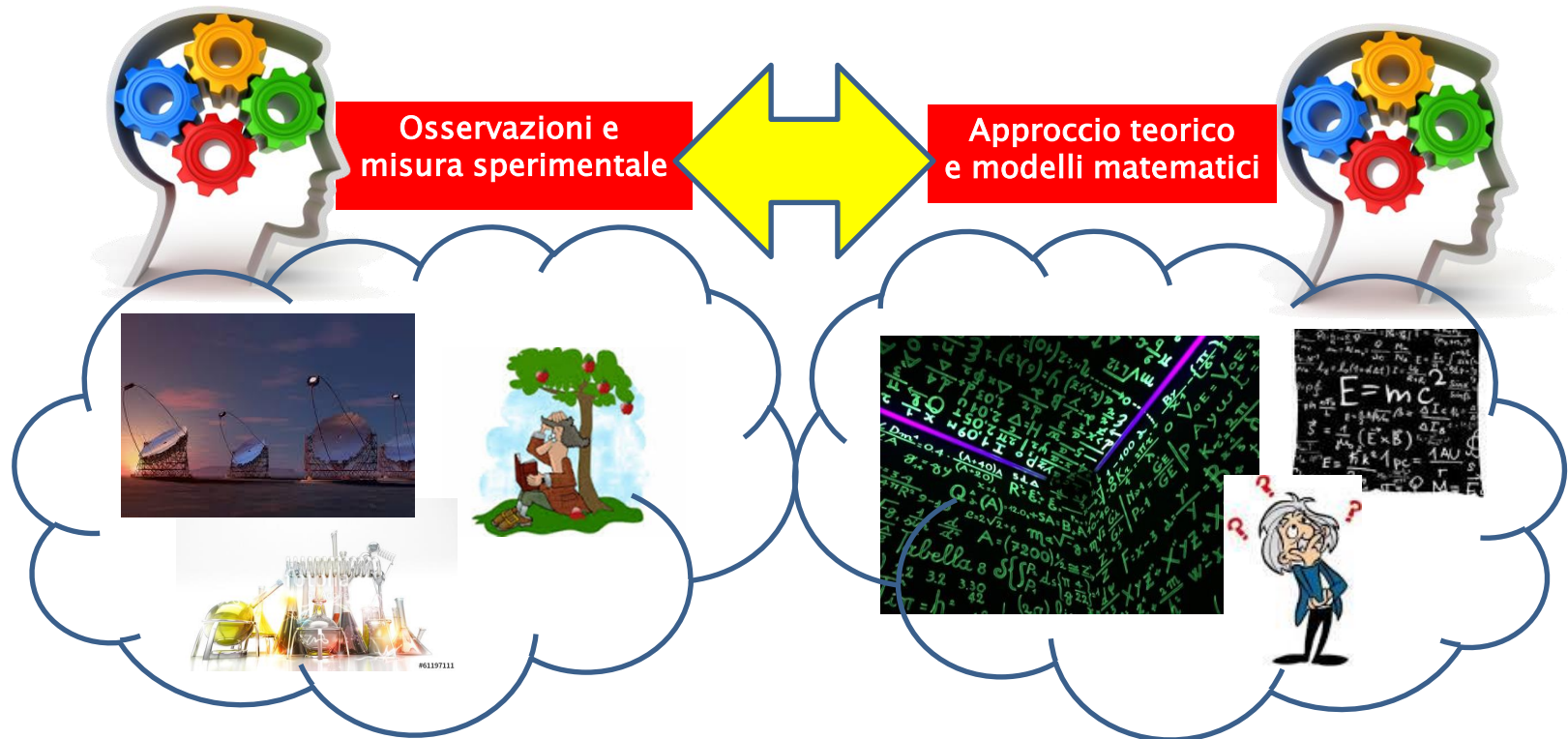
Festival della Scienza

Genova, 21 ottobre _ 5 novembre 2021

Mappe



Quindi l'osservazione e l'approccio teorico possono essere considerati i due paradigmi tradizionali sul quale si basa la conoscenza scientifica.



Consiglio Nazionale delle Ricerche

iPCB ISTITUTO PER I
POLIMERI
COMPOSITI E
BIOMATERIALI

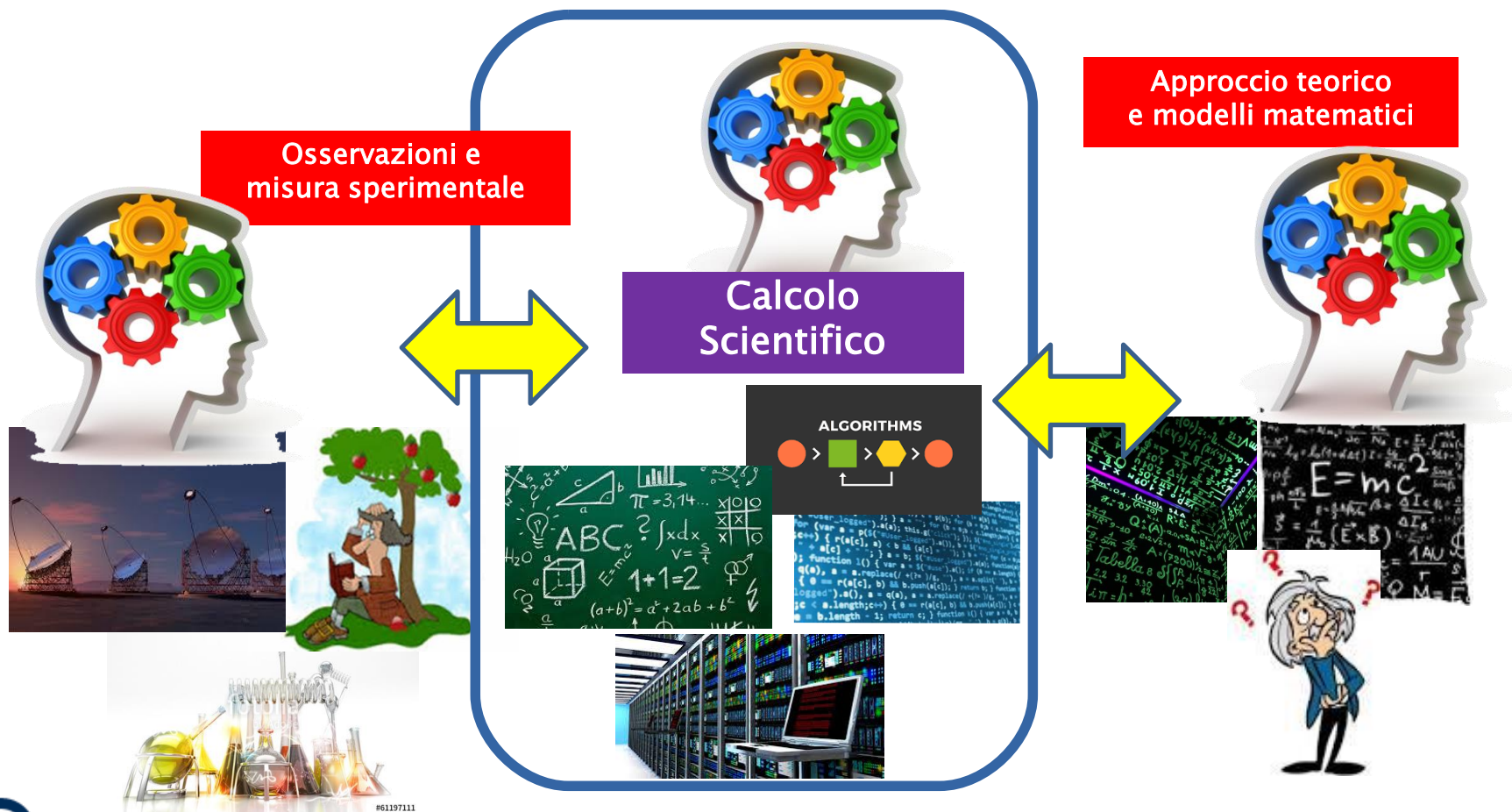
Festival della Scienza

Genova, 21 ottobre _ 5 novembre 2021

Mappe



L'uso del calcolo scientifico è ormai universalmente considerato essere il terzo paradigma nel processo di conoscenza



Consiglio Nazionale delle Ricerche

iPCB ISTITUTO PER I
POLIMERI
COMPOSITI E
BIOMATERIALI

Festival della Scienza

Genova, 21 ottobre _ 5 novembre 2021

Mappe



L'Era della "Data-Driven Science and Engineering"

*"Originally, there was just **experimental science**, and then there was **theoretical science**, with Kepler's Laws, Newton's Laws of Motion, Maxwell's equations, and so on.*

*Then, for many problems, the theoretical models grew too complicated to solve analytically, and people had to **start simulating**. These simulations have carried us through much of the last half of the last millennium.*

At this point, these simulations are generating a whole lot of data, along with a huge increase in data from the experimental sciences. People now do not actually look through telescopes. Instead, they are "looking" through large-scale, complex instruments which relay data to datacenters, and only then do they look at the information on their computers. The world of science has changed, and there is no question about this.

The new model is for the data to be captured by instruments or generated by simulations before being processed by software and for the resulting information or knowledge to be stored in computers. Scientists only get to look at their data fairly late in this pipeline.

*The techniques and technologies for such data-intensive science are so different that it is worth distinguishing data-intensive science from computational science as a new, **fourth paradigm for scientific exploration**"*

G. Bell, T. Hey, and A. Szalay, "Beyond the Data Deluge," Science, vol. 323, no. 5919, pp. 1297-1298, 2009, doi: 10.1126/science.1170411.

La nuova, grande, abbondanza di dati, annuncia l'era di un quarto paradigma per la conoscenza?



... ma come si fa ad ottenere conoscenza dai dati? ...



Consiglio Nazionale delle Ricerche

iPCB ISTITUTO PER I
POLIMERI
COMPOSITI E
BIOMATERIALI

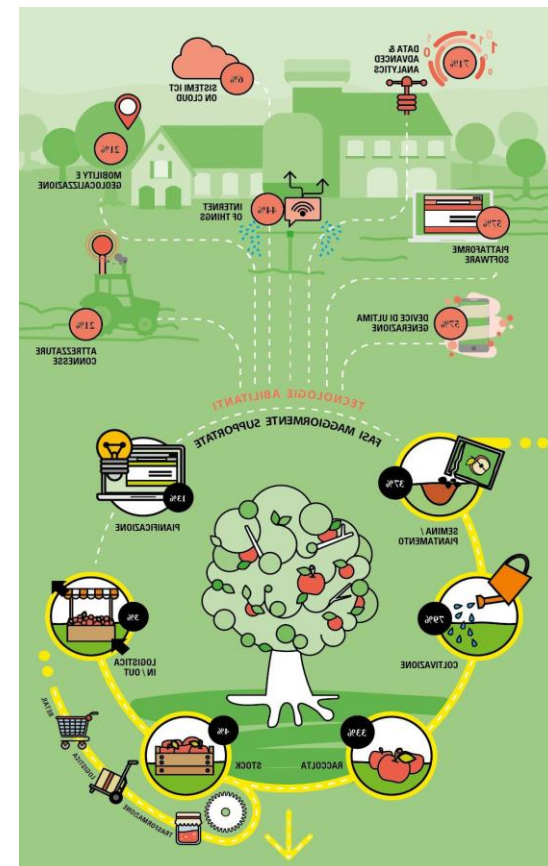
L'industria alimentare e l'apprendimento dai dati

Le filiere alimentari sono potenzialmente grandi produttori di dati grazie ad esempio ai sensori che monitorano ...

- lo stato dei componenti alimentari durante il processo produttivo e di trasformazione
- lo stato dell'approvvigionamento nei punti di distribuzione e vendita degli alimenti

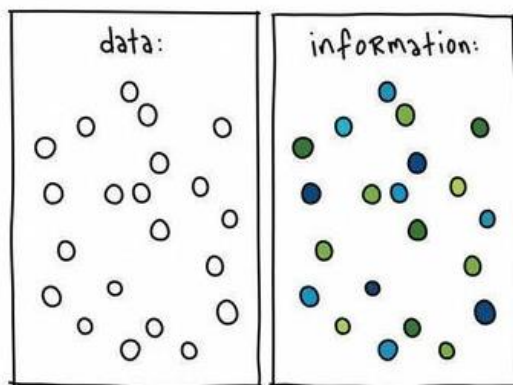
Dedurre conoscenza da tali dati in maniera automatica può migliorare la gestione di questioni significative quali ad esempio

- la sicurezza alimentare
- il contenimento degli sprechi.



Dai dati alla conoscenza

La riflessione sulla conoscenza è stata al centro dell'attenzione di filosofi e scienziati fin dai tempi di Platone e Aristotele, prima di diventare oggetto degli informatici e tecnologi della seconda metà del '900.

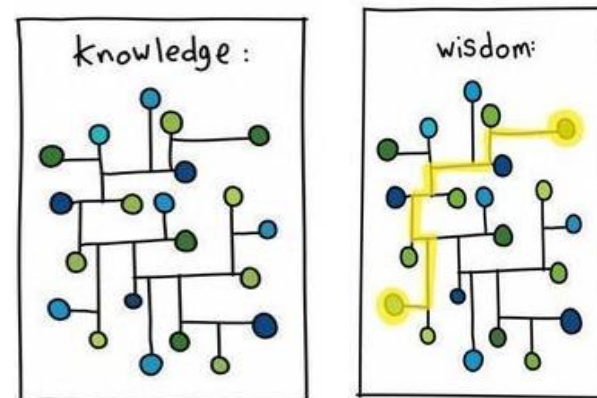


Dati: i dati sono una raccolta di numeri, eventi o simboli. Nella loro iniziale forma grezza non sono immediatamente utili (un esempio: i dati raccolti dai telescopi di tutto il mondo).

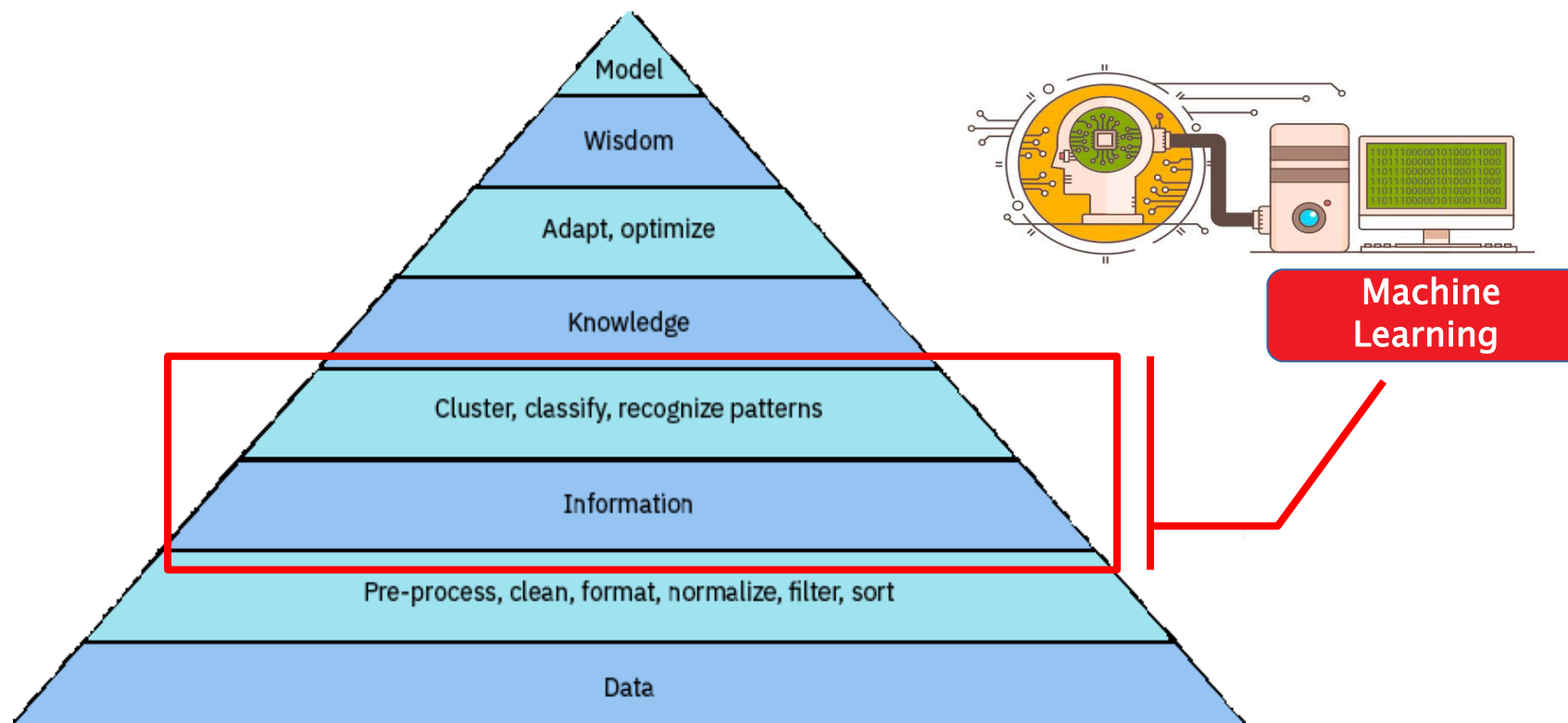
Informazione: l'informazione è una **raccolta di dati organizzata e ordinata in modo coerente**. I dati, organizzati in informazioni, guadagnano in utilità perché la loro fruibilità è semplificata (esempio: i dati raccolti dai telescopi organizzati per tipo di acquisizione e per data di acquisizione)

Conoscenza: La conoscenza è l'impiego, da parte di uno specifico individuo, di un certo numero di informazioni per attuare "azioni" e vivere delle esperienze nel suo ambiente (ad esempio saper riconoscere un pianeta da una stella).

Saggezza: la saggezza è la capacità di selezionare il modo migliore per raggiungere il risultato desiderato basato sulla conoscenza.



Come i dati diventano conoscenza nell'ambito della "Data-Driven Science and Engineering"



Vinay R. Rao, *How data becomes knowledge, Part 1 - From data to knowledge*, March 2018, IBM developerWorks®

Festival della Scienza

Genova, 21 ottobre _ 5 novembre 2021

Mappe



Come i dati diventano conoscenza nell'ambito della "Data-Driven Science and Engineering"

Supervised Machine Learning

Alcune informazioni accompagnano i dati

Input data



Annotations

These are apples



Model



Prediction

Its an apple!

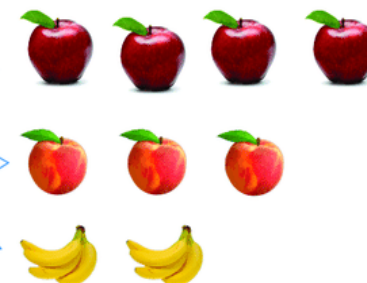
Unsupervised Machine Learning

Nessuna informazione accompagna i dati

Input data



Model



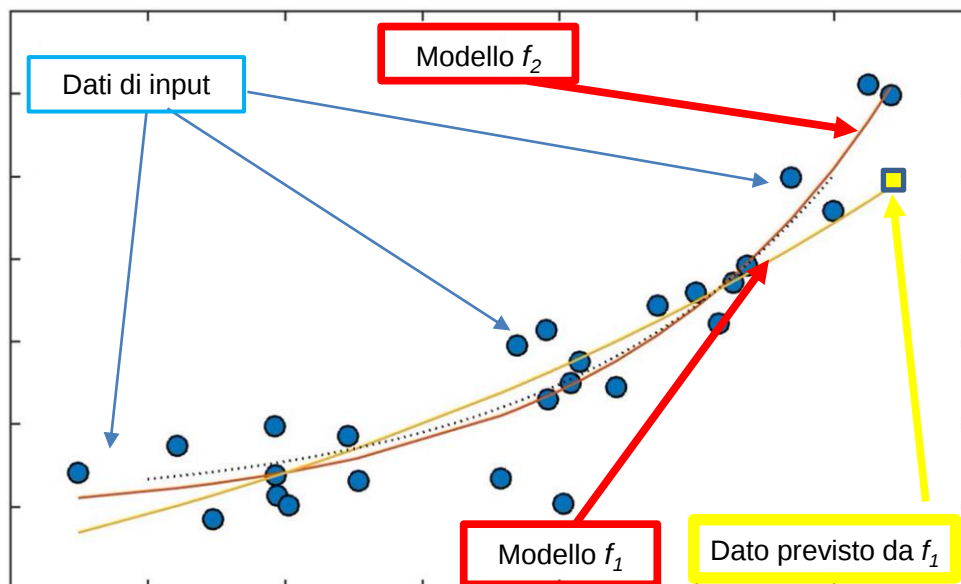
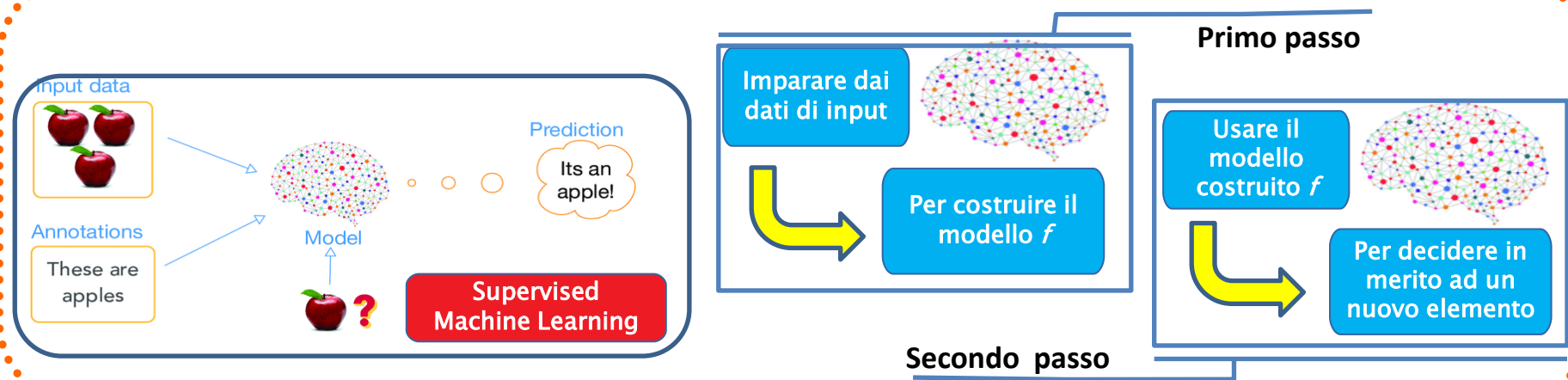
Festival della Scienza

Genova, 21 ottobre _ 5 novembre 2021

Mappe



Come i dati diventano conoscenza nell'ambito della "Data-Driven Science and Engineering"



- Il modello (una “funzione”) è una curva che cerca di «avvicinarsi» il più possibile ai dati
- Più modelli per gli stessi dati sono possibili: modelli diversi possono essere più o meno accurati nel predire i nuovi dati

Come sono evolute le machine calcolatrici



Il calcolatore Colossus

Dicembre 1943: diviene operativo il primo calcolatore elettronico, denominato «Colossus» nato dalla collaborazione di Alan Turing (matematico), Tom Flowers (matematico) e Max Newman (ingegnere)

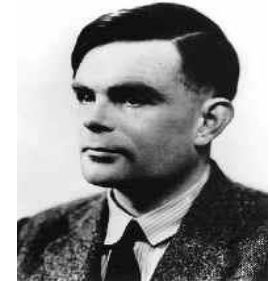
Colossus fu in grado di forzare i codici sviluppati dalla cifratrice Lorenz SZ 40/42 (Enigma) usata dai tedeschi per proteggere la corrispondenza fra Adolf Hitler e i suoi capi di stato maggiore

Colossus fu sviluppato per decifrare i messaggi codificati. La macchina, programmabile, comparava due flussi di dati: Un flusso era il messaggio da decifrare, mentre il secondo era generato dalla macchina che cercava di individuare la chiave di codifica effettuando delle prove e confrontando i risultati.

Colossus era un computer
special purpose

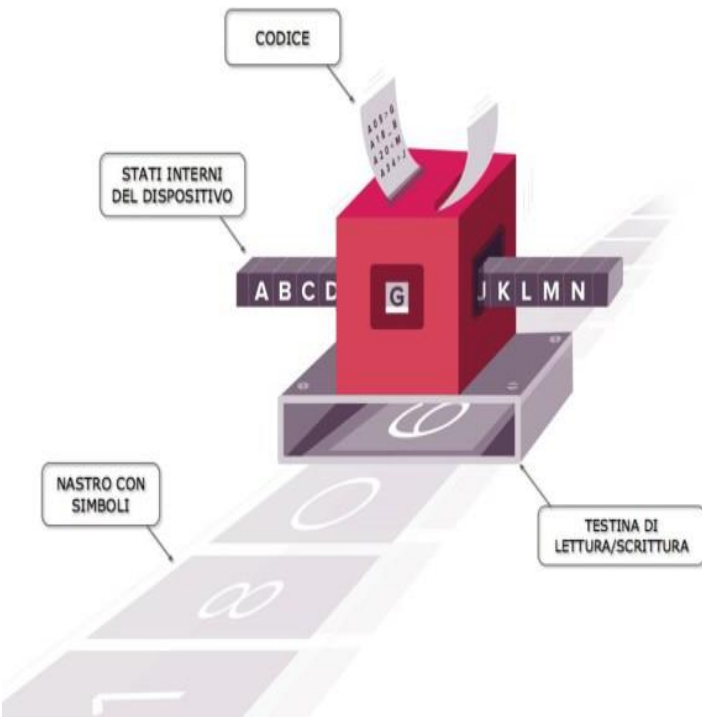
Come sono evolute le machine calcolatrici

Alan Turing



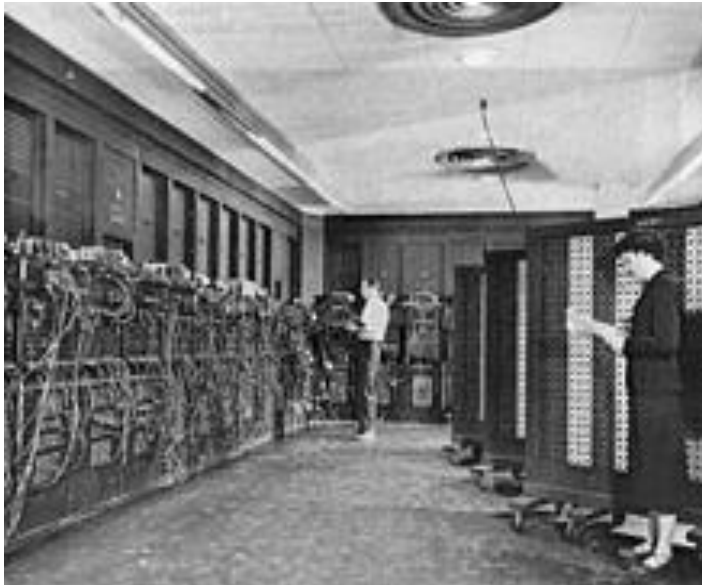
Nel 1936 il logico Alan Turing (1912-1954) enuncia il modello del calcolatore moderno, la cosiddetta "Macchina di Turing":

- è in grado di eseguire «atti primitivi» (non solo conti!) secondo uno schema che consente di risolvere ogni tipo di problema di logica simbolica in un numero finito di passi
- è un modello astratto che definisce una macchina in grado di eseguire algoritmi e dotata di un nastro potenzialmente infinito su cui può leggere e/o scrivere dei simboli



Come sono evolute le machine calcolatrici

L'Electronic Numerical Integrator and Computer (ENIAC), 1946, è il primo computer elettronico general purpose della storia.



ENIAC è l'antenato dei supercalcolatori moderni

- inizialmente realizzato ed utilizzato per scopi militari (calcolo delle traiettorie balistiche), era in grado di risolvere molti altri problemi basati sulla risoluzione di alcune equazioni differenziali (da ciò deriva il nome di integratore numerico)
- venne usato anche:
 - per scopi civili, come l'analisi dei dati dei censimenti
 - per **applicazioni scientifiche**, ad esempio John von Neumann se ne servì per realizzare la **prima previsione meteorologica al computer** (ENIAC fu in grado di processare **250000 operazioni in virgola mobile in circa 24 ore** e realizzare una previsione di 24 ore)

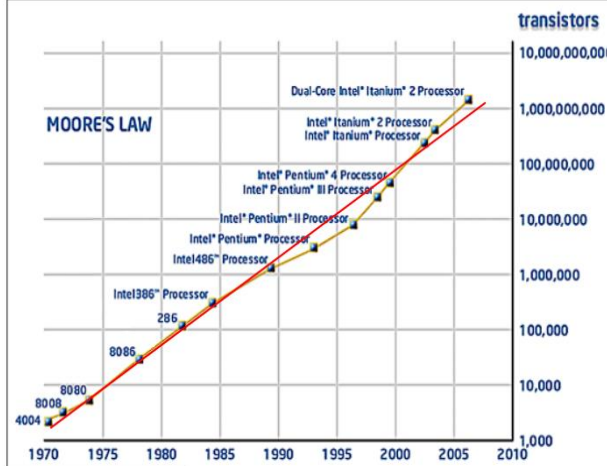
ENIAC era un computer *general purpose*

Era in grado di eseguire
 $250000/86400 \sim 3$ operazioni
floating point al secondo
(flops)

I calcolatori moderni

Nel 1965 Gordon Moore, uno dei fondatori di INTEL e dei pionieri della microelettronica, osservando l'evoluzione dell'hardware dei calcolatori, enunciò quella che va sotto il nome di 1° legge di Moore

Increasing Transistor Density -- Moore's Law



"The number of transistors in a dense integrated circuit doubles approximately every two years"



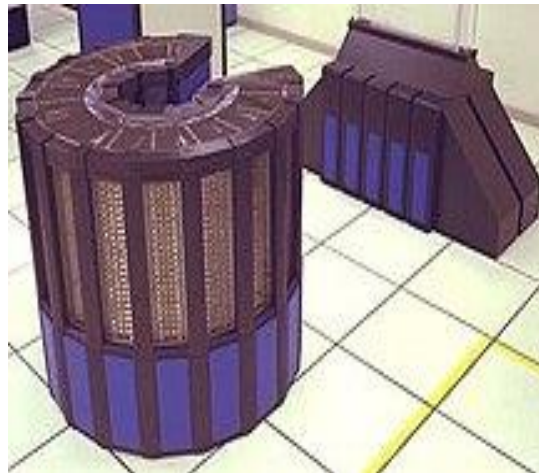
Per circa 20 anni questa legge empirica ha saputo descrivere abbastanza correttamente la realtà, ma qualcosa, ad un certo punto, l'ha resa inadatta a dare indicazioni sulla futura potenza computazionale dei processori

la riduzione delle dimensioni dei transistor, e quindi della scala di integrazione, esponeva i circuiti ad effetti fisici di natura fortemente non deterministica minando alla base l'affidabilità dei processori con questi realizzati

Un limite fisico era stato raggiunto!

Nasce l'idea di “parallelismo”

Se la potenza computazionale di un unico processore non può più crescere, si può pensare di realizzare maggiore potenza di calcolo “aggregando” le capacità elaborative di più processori (o “unità processanti”)



Il supercomputer Cray-2

- tra i primi esempi di calcolatore parallelo commerciale
- montava 4 cpu vettoriali ed è stato prodotto da Cray Research nel 1985
- era in grado di eseguire circa **2 GigaFlops (circa $2 \cdot 10^9$ operazioni al secondo)**

Festival della Scienza

Genova, 21 ottobre _ 5 novembre 2021

Mappe

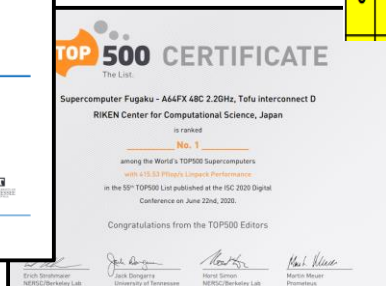


L'Era dello "EXASCALE COMPUTING"

... quando un calcolatore sarà capace di effettuare 10,000,000,000,000,000,000 (10¹⁸) operazioni floating point al secondo ...



#	Site	Manufacturer	Computer	Country	Cores	Rmax [P/ops]	Power [MW]
1	RIKEN Center for Computational Science	Fujitsu	Fugaku Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D	Japan	7,299,072	415.5	28.3
2	Oak Ridge National Laboratory	IBM	Summit IBM Power System, P9 22C 3.07GHz, Mellanox EDR, NVIDIA GV100	USA	2,414,592	148.6	10.1
3	Lawrence Livermore National Laboratory	IBM	Sierra IBM Power System, P9 22C 3.1GHz, Mellanox EDR, NVIDIA GV100	USA	1,572,480	94.6	7.4
4	National Supercomputing Center in Wuxi	NRPC	Sunway TaihuLight NRPC Sunway SW26010, 260C 1.45GHz	China	10,649,600	93.0	15.4
5	National University of Defense Technology	NUDT	Tianhe-2A ANUDT TH-IVB-FEP, Xeon 12C 2.2GHz, Matrix-2000	China	4,981,760	61.4	18.5
6	Eni S.p.A	Dell EMC	HPC5 PowerEdge C4140, Xeon 24C 2.1GHz, NVIDIA T. V100, Mellanox HDR	Italy	669,760	35.5	2.25



Le prime 6 posizioni della 55-sima TOP500 List
<https://www.top500.org>



Consiglio Nazionale delle Ricerche

iPCB ISTITUTO PER I
POLIMERI
COMPOSITI E
BIOMATERIALI

Festival della Scienza

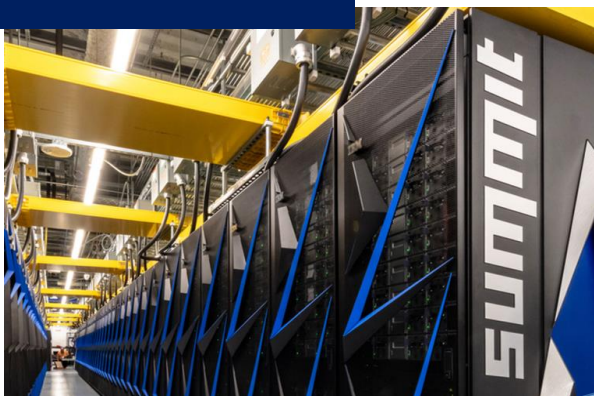
Genova, 21 ottobre _ 5 novembre 2021

Mappe



Che aspetto hanno i supercalcolatori più potenti del mondo?

Summit (circa 2.5 milioni di core)
DOE/SC/Oak Ridge National Laboratory
United States



Sunway TaihuLight (circa 10,5 milioni di core)
National Supercomputing Center in
Wuxi
China



Ciascuno di questi
calcolatori è in grado di
eseguire centinaia di
PetaFlops
(circa 10^{17} operazioni
al secondo)



Tianhe-2° (circa 5 milioni di core)
National Super Computer Center in
Guangzhou
China



Sierra circa (1,5 milioni di core)
DOE/NNSA/LLNL
United States



Consiglio Nazionale delle Ricerche

iPCB ISTITUTO PER I
POLIMERI
COMPOSITI E
BIOMATERIALI

Festival della Scienza

Genova, 21 ottobre _ 5 novembre 2021

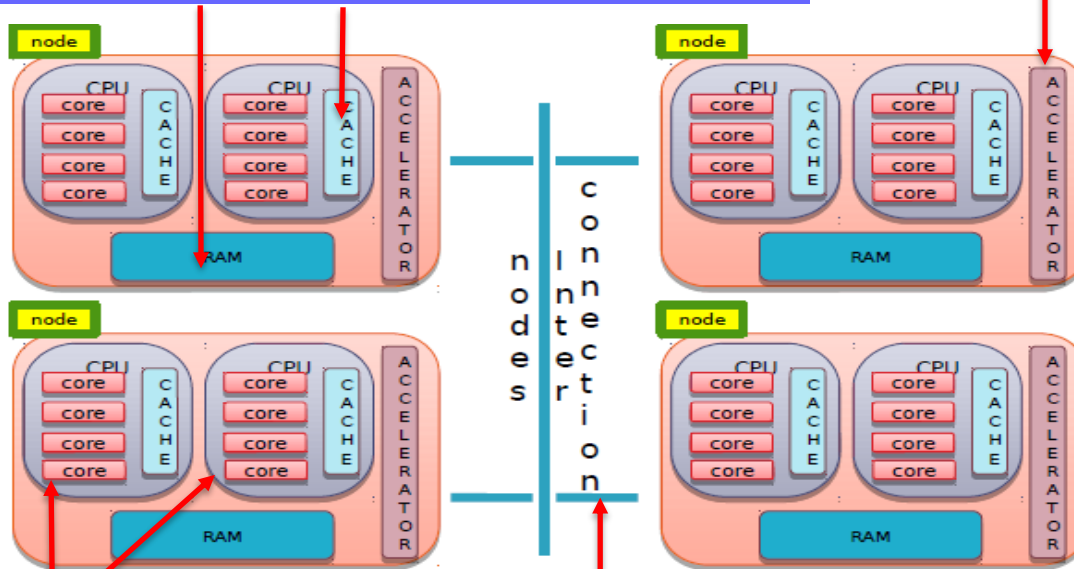
Mappe



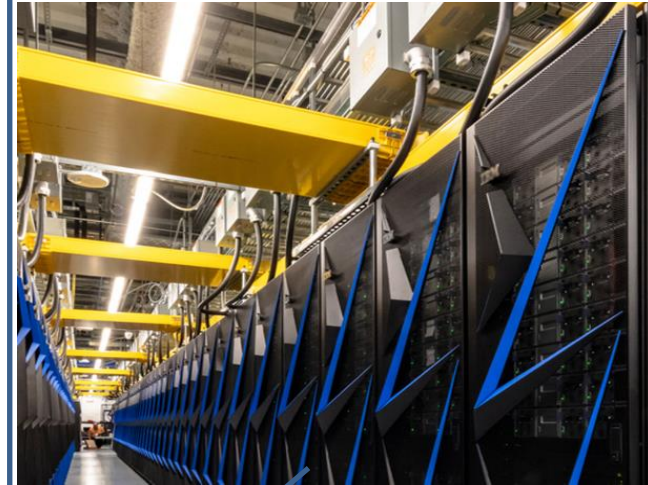
A very particular type of processing unit: the accelerators which mount thousands of cores

Memory is organized into levels

- Speed is inversely proportional to size
- The memory closest to the processing units is faster



- Several processing units (core/cpu) aggregate in a cpu/node sharing some memory devices
- Access to "remote" memory devices (on other nodes) is called "*communication*" and takes place thanks to an interconnection network



Che aspetto hanno i supercalcolatori più potenti del mondo "visti dall'interno"?

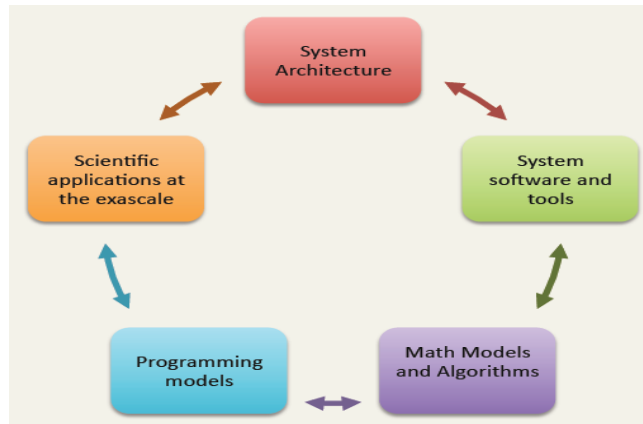
Festival della Scienza

Genova, 21 ottobre _ 5 novembre 2021

Mappe



Per supportare l'uomo nell'affrontare le grandi sfide della scienza e della tecnologia, le risorse computazionali dovranno essere sempre più potenti, efficaci ed efficienti ...



... ciò sarà possibile se tutte le forze coinvolte nella progettazione e realizzazione di tali risorse, da chi formula i modelli a chi progetta gli algoritmi, da chi li implementa in un software a chi costruisce l'hardware, sapranno coordinarsi in un approccio **multidisciplinare e interdisciplinare**

Al momento, l'unico approccio possibile, per vincere le sfide del presente e del futuro, sembra essere quello basato sul principio che **“L'unione fa la forza!”** soprattutto se l'unione è quella di intelligenze e competenze differenti



