Introduzione ai Big Data ed esempi applicativi

Prato
21 settembre 2023

Marco Bianchi

FUB
Fondazione Ugo Bordoni
Ricerca e Innovazione

AGENDA

- Introduzione ai problemi “Big Data”
  - Caratteristiche di un problema big data ed esempi applicativi
  - Scalabilità verticale vs. scalabilità orizzontale
  - Map-reduce, architetture Big data, “sistemi operativi big data”, figure professionali
- Big data in relazione a AI, IoT, Industrial IoT, edge computing
- Casi d’uso dei Big data in alcuni comparti industriali
  - settore agricolo
  - manifattura
  - logistica
- Mercato dei data Big data, *data strategy* e Big data nelle PMI

Una definizione di Big Data

Big data is high-volume, high-velocity and/or high-variety information assets that demand cost effective innovative forms of information processing that enable enhanced insight, decision making, and process automation. https://www.gartner.com/it-glossary/big-data/

I big data sono risorse informative ad alto volume, ad alta velocità e / o ad alta varietà che richiedono forme di elaborazione delle informazioni innovative ed economicamente sostenibili e che favoriscono le intuizioni, supportano i processi decisionali e l'automazione dei processi

Image Credit: arka38/Shutterstock

Dimensione # 1: Volume

Quanti dati devono essere memorizzati ed elaborati?

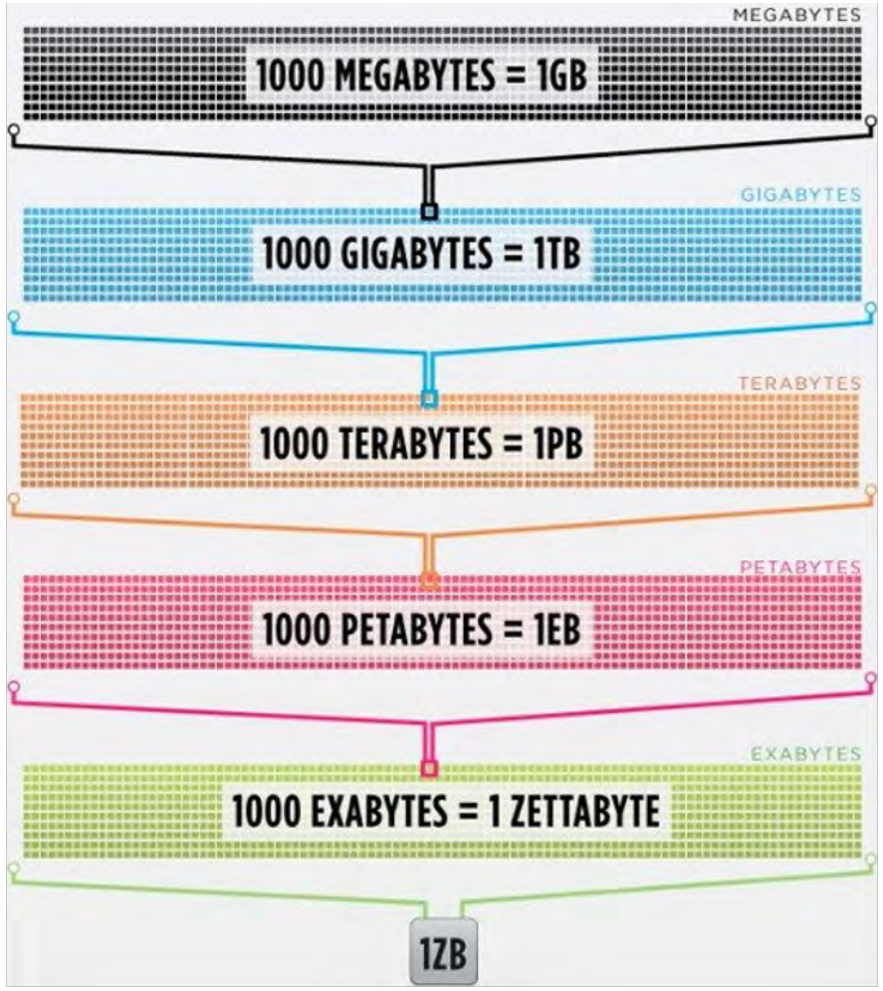


Image Credit: https://blogs.cisco.com/news/the-dawn-of-the-zettabyte-era-infographic

4.7 GB ~ Dimensione di un DVD-R standard

1 TB ~ HDD di un moderno computer portatile (circa 210 DVD)

1.5 PETABYTE ~ tutte le foto su Facebook

1 EXABYTE ~ L'intero catalogo di Netflix mandato in streaming più di 1000 volte

1.3 ZB ~ Traffico della rete Internet al 2016

Dimensione #1: Volume

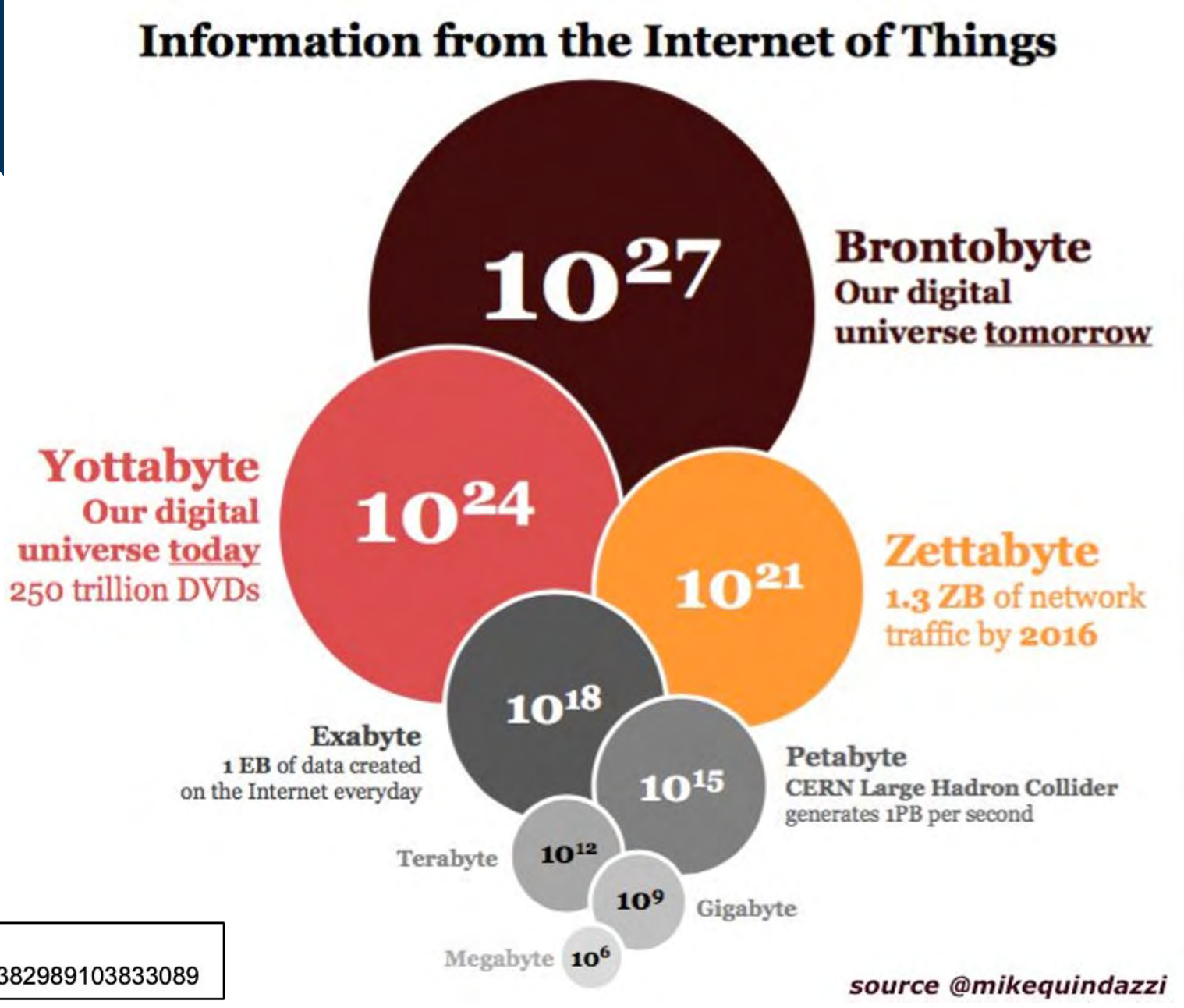


Image credit:
https://twitter.com/MikeQuindazzi/status/831382989103833089

Dimensione #2: Velocità

Quanto velocemente sono generati ed elaborati i dati?

COSA ACCADE SU INTERNET IN UN MINUTO…

- 5,9 milioni di ricerche effettuate su Google
- 1,7 milioni di contenuti condivisi su Facebook
- 66.000 foto condivise su Instagram
- 347.000 tweet pubblicati su Twitter
- 2,4 milioni di snap inviati su Snapchat
- 500 ore di video caricate su YouTube
- $ 443.000 spesi su Amazon
- 16 milioni di SMS inviati
- 231 milioni di email inviate
- $ 90 milioni spesi in criptovalute

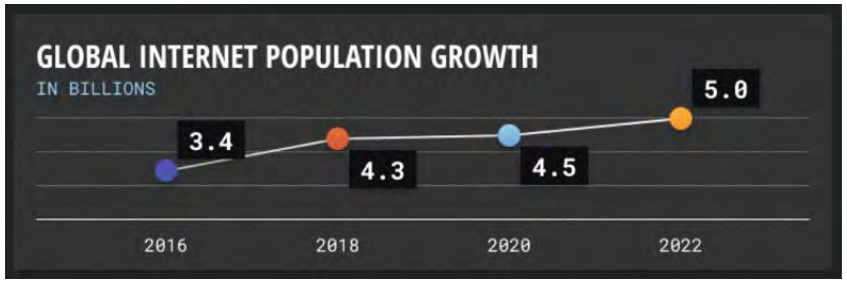
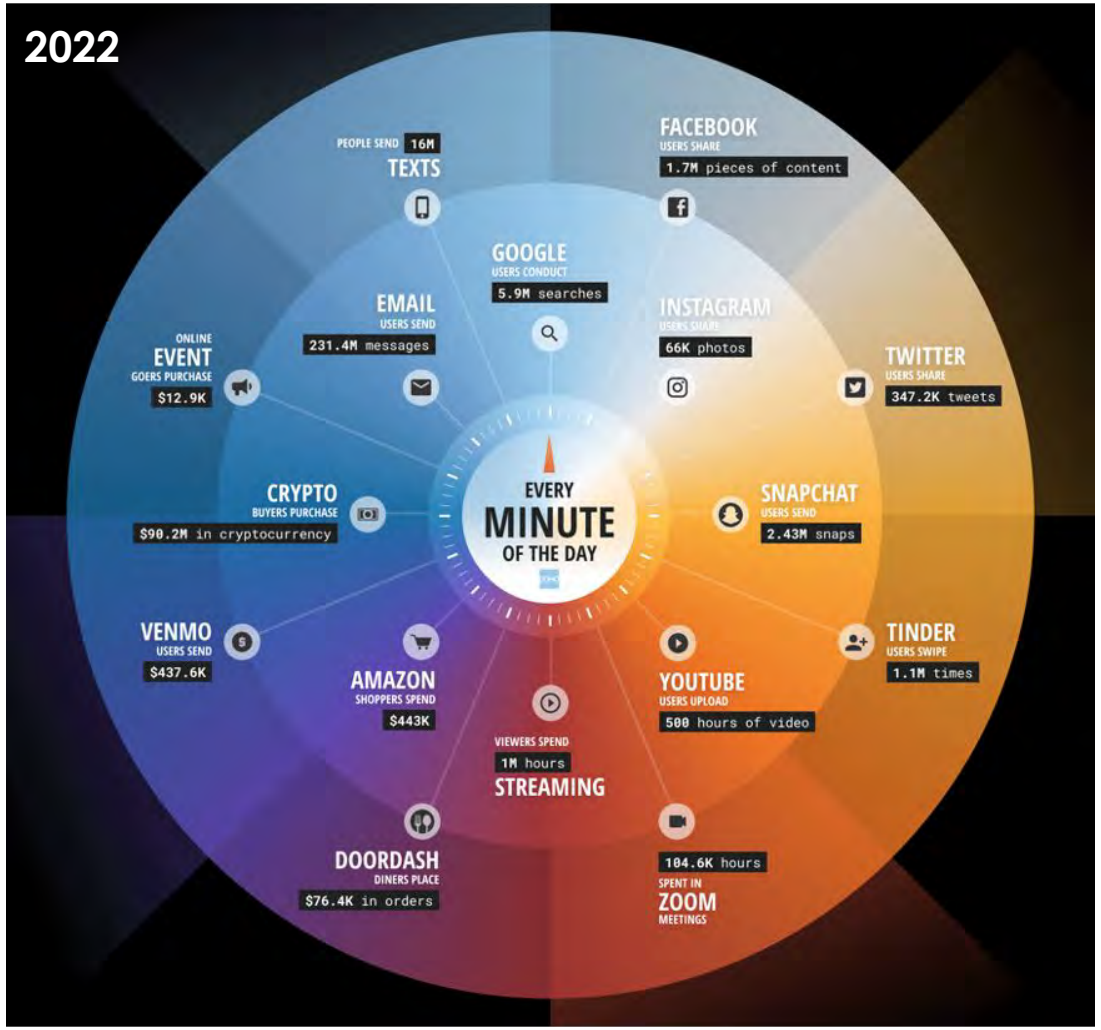


Image credit: https://www.domo.com/data-never-sleeps



Modalità di generazione ed elaborazione

La velocità dei dati è una caratteristica che riguarda la:

- **modalità di <u>generazione</u> dei dati → dipende dalla sorgente**
  Con quale frequenza i dati vengono generati?

- **modalità di <u>elaborazione</u> dei dati → dipende dal sistema che elabora**
  Con quale frequenza i dati vengono elaborati?

Tipologia delle sorgenti dati

- **Sorgenti streaming**:

  sorgenti sempre attive che producono dati continuamente

  - **Esempi: social networks, dati finanziari (es. borsa), dati da sensori (IoT)**
  - **Progettate per produrre e diffondere flussi di dati verso altri sistemi**
  - **Possono avere picchi di velocità e volume**

- **Sorgenti "passive"**:

  sorgenti che vengono visitate con una determinata frequenze

  - **Esempi: pagine Web, RSS, ecc.**
  - **Progettate prevalentemente per "pubblico umano"**
  - **La freschezza delle informazioni osservate dipende dalla frequenza di visita**

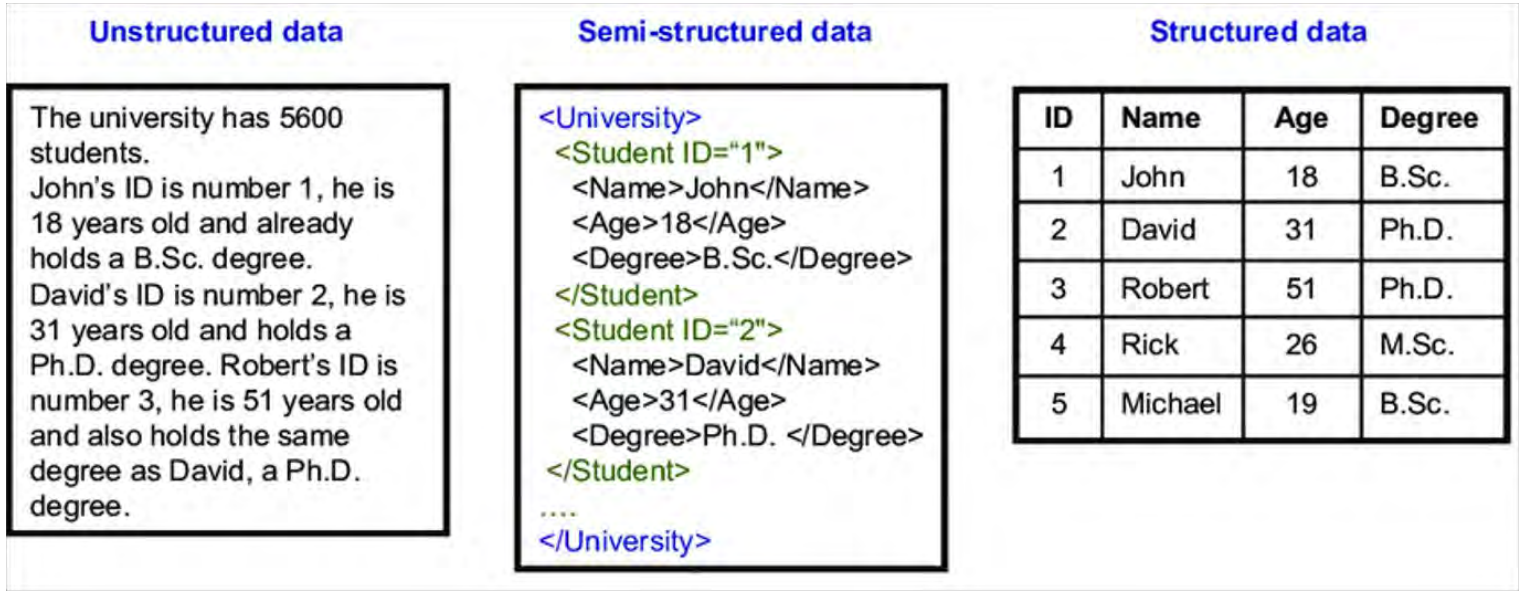
Modalità di elaborazione

- **Elaborazione batch**
  - Eseguita una-tantum, dopo aver raccolto tutti i dati necessari
  - Es. Quanto si è “parlato” dei musei italiani sui social network nel 2018?
- **Elaborazione periodica**
  - Tipicamente elaborazioni batch ripetute nel tempo
  - Es. Qual’è classifica settimanale dei film più apprezzati dagli utenti del Web?
- **Elaborazione near-real-time**
  - Dati processati appena arrivano sul sistema
  - Es. Offerta di biglietti scontati basati su sistemi di geolocalizzazione
- **Elaborazione real-time**
  - Dati processati appena arrivano sul sistema con alta criticità di latenza
  - Es. Advertising on-line

Dimensione #3: Varietà

Quanti tipi e formati di dati sono trattati?

- **Molti tipi e formati di dati**
  - **Testi, Immagini, Video, Tweet, Post Facebook, Email, Dati di sensori, …**
- **Molteplici strutture di dati**



Unstructured data

The university has 5600 students.
John's ID is number 1, he is 18 years old and already holds a B.Sc. degree.
David's ID is number 2, he is 31 years old and holds a Ph.D. degree. Robert's ID is number 3, he is 51 years old and also holds the same degree as David, a Ph.D. degree.

Semi-structured data

```
<University>
 <Student ID="1">
  <Name>John</Name>
  <Age>18</Age>
  <Degree>B.Sc.</Degree>
 </Student>
 <Student ID="2">
  <Name>David</Name>
  <Age>31</Age>
  <Degree>Ph.D. </Degree>
 </Student>
 ....
</University>
```

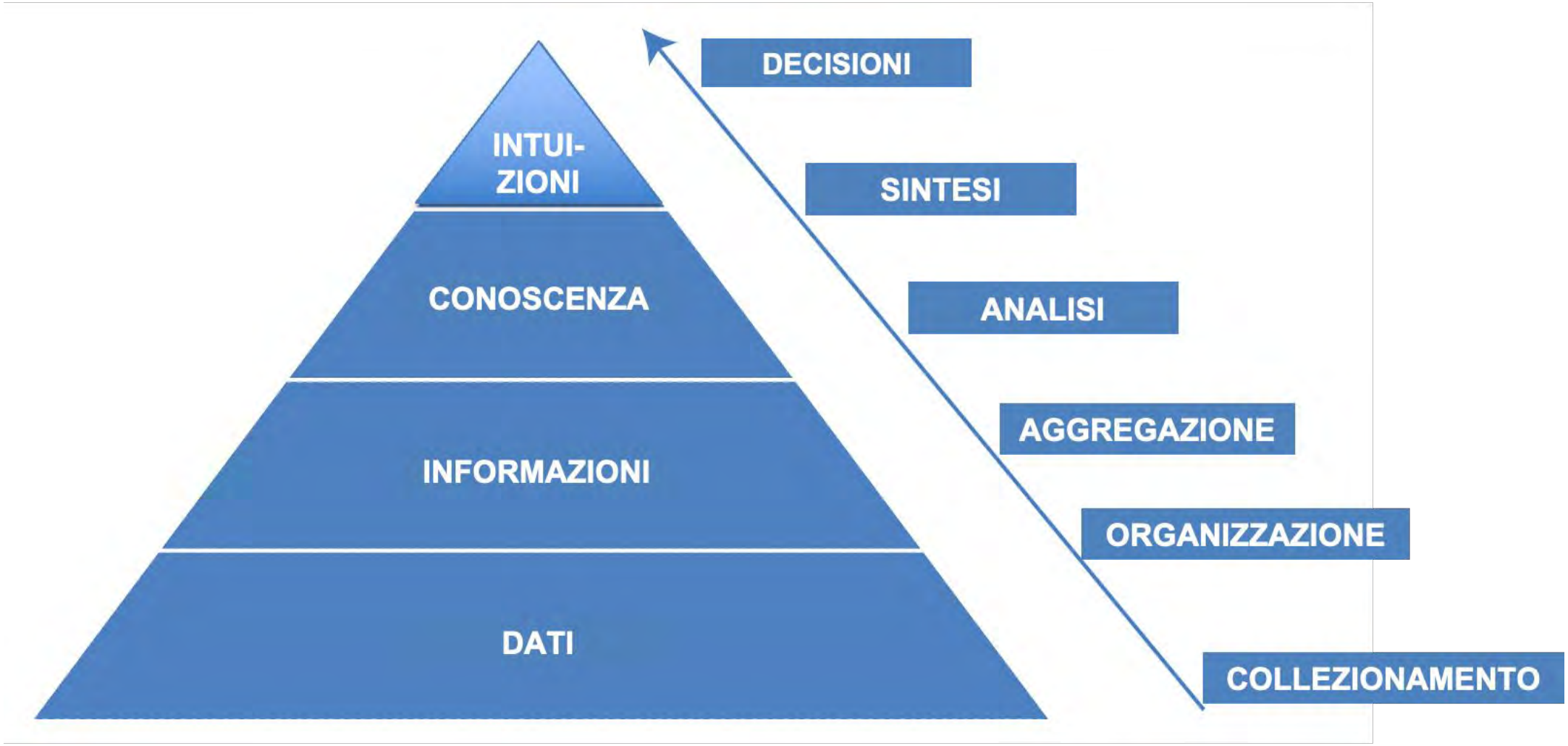
Structured data

ID	Name	Age	Degree
1	John	18	B.Sc.
2	David	31	Ph.D.
3	Robert	51	Ph.D.
4	Rick	26	M.Sc.
5	Michael	19	B.Sc.

Image credit:
Cardoso, Jorge. (2005). Developing Dynamic Packaging Applications using Semantic Web based Integration. 10.4018/978-1-59904-192-6.ch001.

La dimensione del Valore

Quanto vale
la conoscenza estratta?



Tipi di analitiche di Big Data

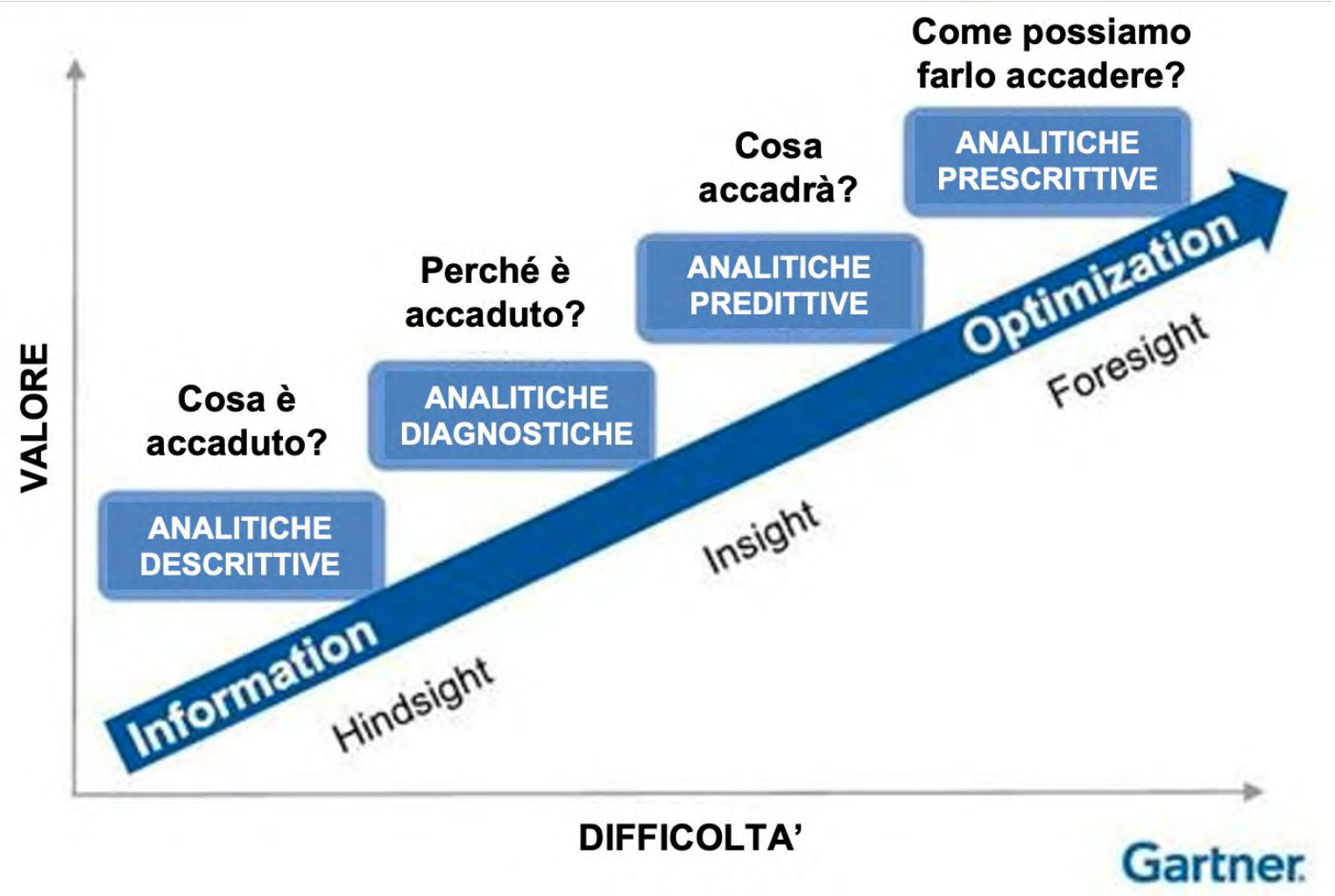
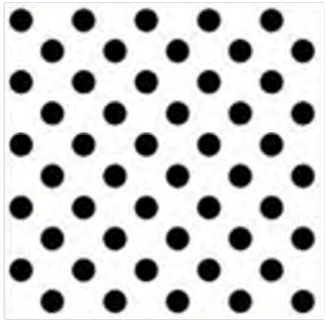
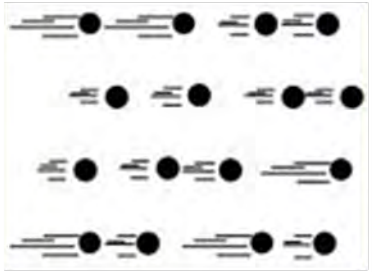
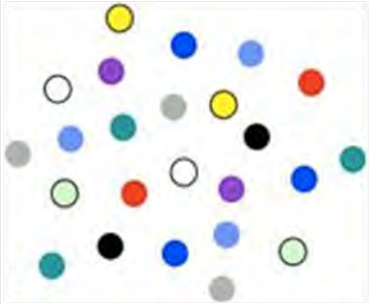


Image credit: https://www.pinterest.it/pin/272327108689539151/

Riassumendo

Volume	Velocity	Variety	Value
			
DATI "A RIPOSO"	**DATI "IN MOVIMENTO"**	**MOLTI TIPI DI DATI MOLTI FORMATI**	**VALORE DELLA CONOSCENZA**
Terabytes to Exabytes of existing data to process	Streaming data, requiring milliseconds to seconds to respond	Structured, unstructured, text, multimedia,...	Business models can be associated to the data

Adapted by a post of Michael Walker on 28 November 2012

Altre V…

- **Veridicità**

  la veridicità implica che i dati siano verificabili e veritieri. La veridicità riguarda l'affidabilità degli strumenti di misurazione, dei sensori, delle sincronizzazione di tempi, di eventuali dati fake inseriti da bot, ecc.

- **"Visibilità"**

  I silos informativi sono sempre esistiti all'interno delle aziende e sono stati uno dei principali ostacoli nel tentativo di estrarre valore dai dati. Le informazioni rilevanti non dovrebbero solo esistere, ma dovrebbero anche essere visibili alla persona giusta al momento giusto. I dati utilizzabili devono essere visibili superando i confini delle funzioni, dei dipartimenti e persino delle organizzazioni, per sbloccare il valore.

- **"Visualizzazione"**

  In un contesto di business, un'appropriata la presenza di dashboard in grado di visualizzare i dati e sintetizzare chiaramente i risultati delle analisi in modo facilmente comprensibile al management è un aspetto critico.

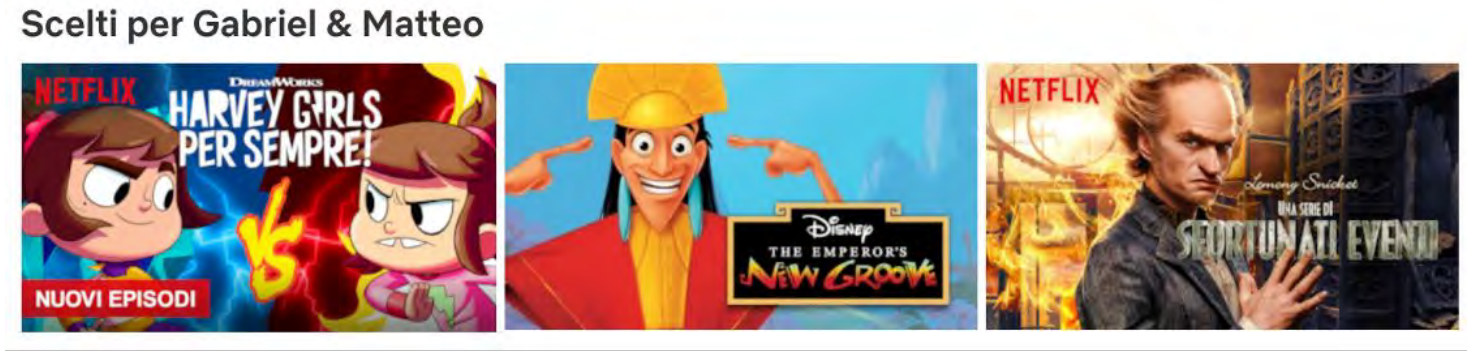
Esempi d'uso dei Big data

Image credit: https://www.artslife.com/wp-content/uploads/2018/02/yourstory_BigData-590x295.jpg

Alcuni esempi dalla vita di tutti i giorni…

- **Motori di ricerca**

- **Pubblicità on-line**
- **Metodi di raccomandazione (es. Amazon, Netflix, Spotify,…)**



Esempi meno conosciuti…

- **Macy's: azienda fondata a New York nel 1858, utilizza una tecnologia Big Data per cambiare, quasi in tempo reale, i prezzi dei circa 73 milioni di oggetti (prodotti di abbigliamento, calzature, mobili, gioielli, cosmetici e articoli per la casa) presenti nei suoi numerosi punti vendita. La variazione di prezzo avviene in base alla domanda e alla quantità di prodotti presenti in magazzino, in modo da ottimizzare sempre i costi e i guadagni.**

- **Predpol: La polizia di Los Angeles usa un sistema Big Data in grado di prevedere dove è più probabile che vengano commessi i crimini, con una precisione di circa 50 metri quadrati. Nelle zone della città dove la polizia utilizza questo sistema, c'è stata una riduzione dei furti del 33% e del 21% per i crimini violenti.**

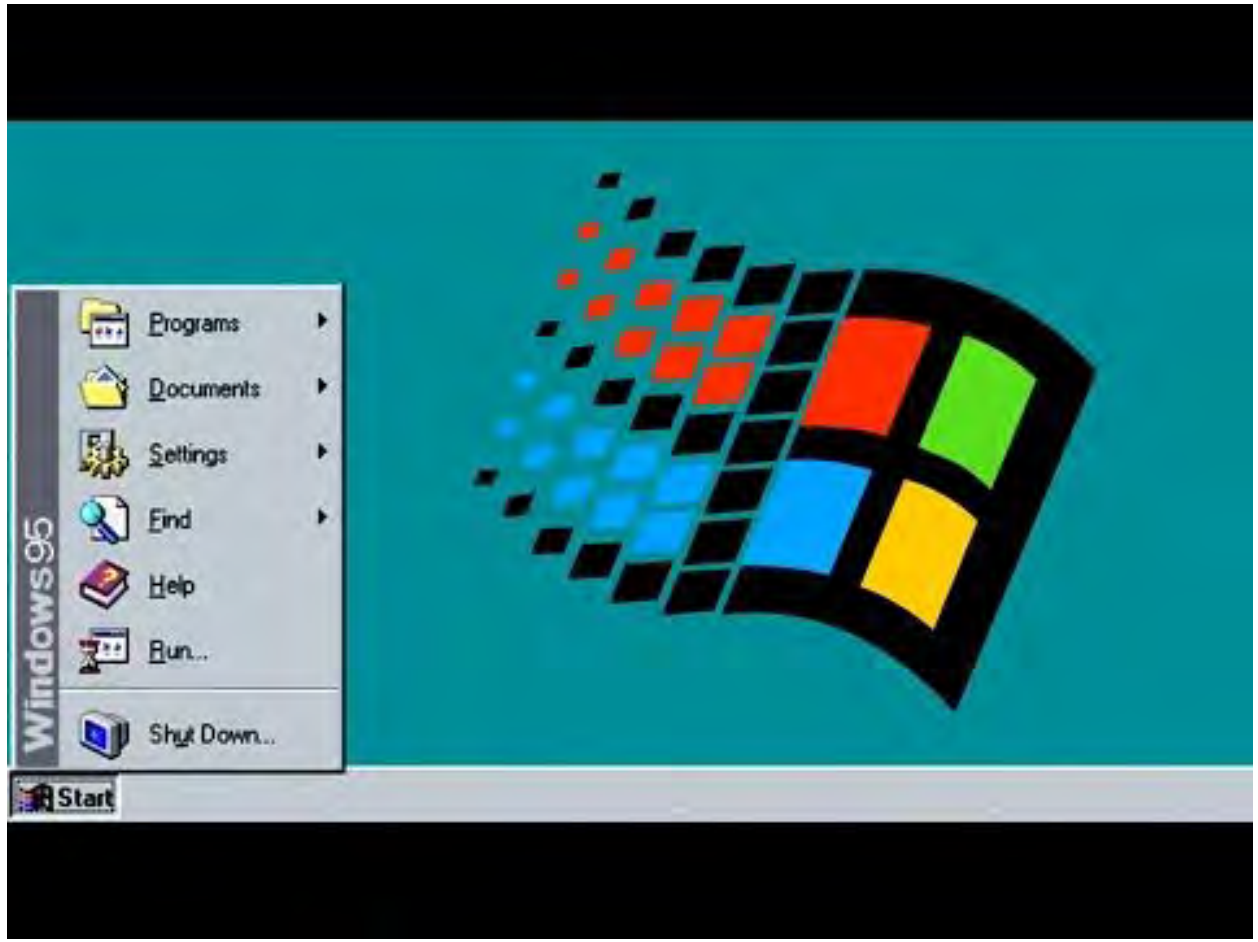
PREDPOL®

Altri casi d'uso:
https://searchcio.techtarget.com/opinion/Ten-big-data-case-studies-in-a-nutshell

Image credit: immagine dal Film "Minority Report" (2002) di Steven Spielberg

Il “primo” problema di Big data

Cerca file in Windows 95



Come funziona un motore di ricerca (1/2)

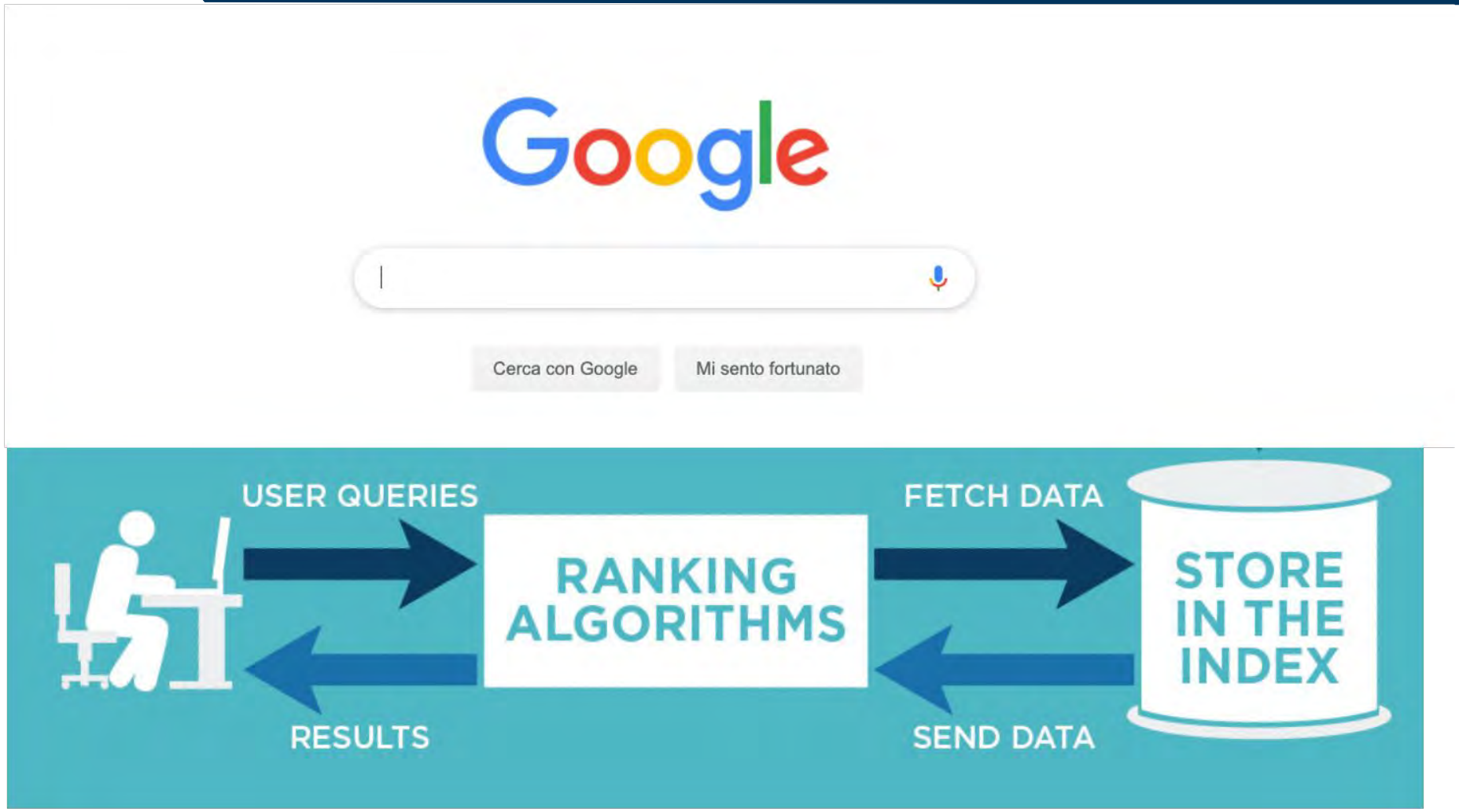


Image credit: https://www.stonetemple.com/wp-content/uploads/2016/12/Crawling-indexing-ranking-hero-image.jpg

Come funziona un motore di ricerca (2/2)

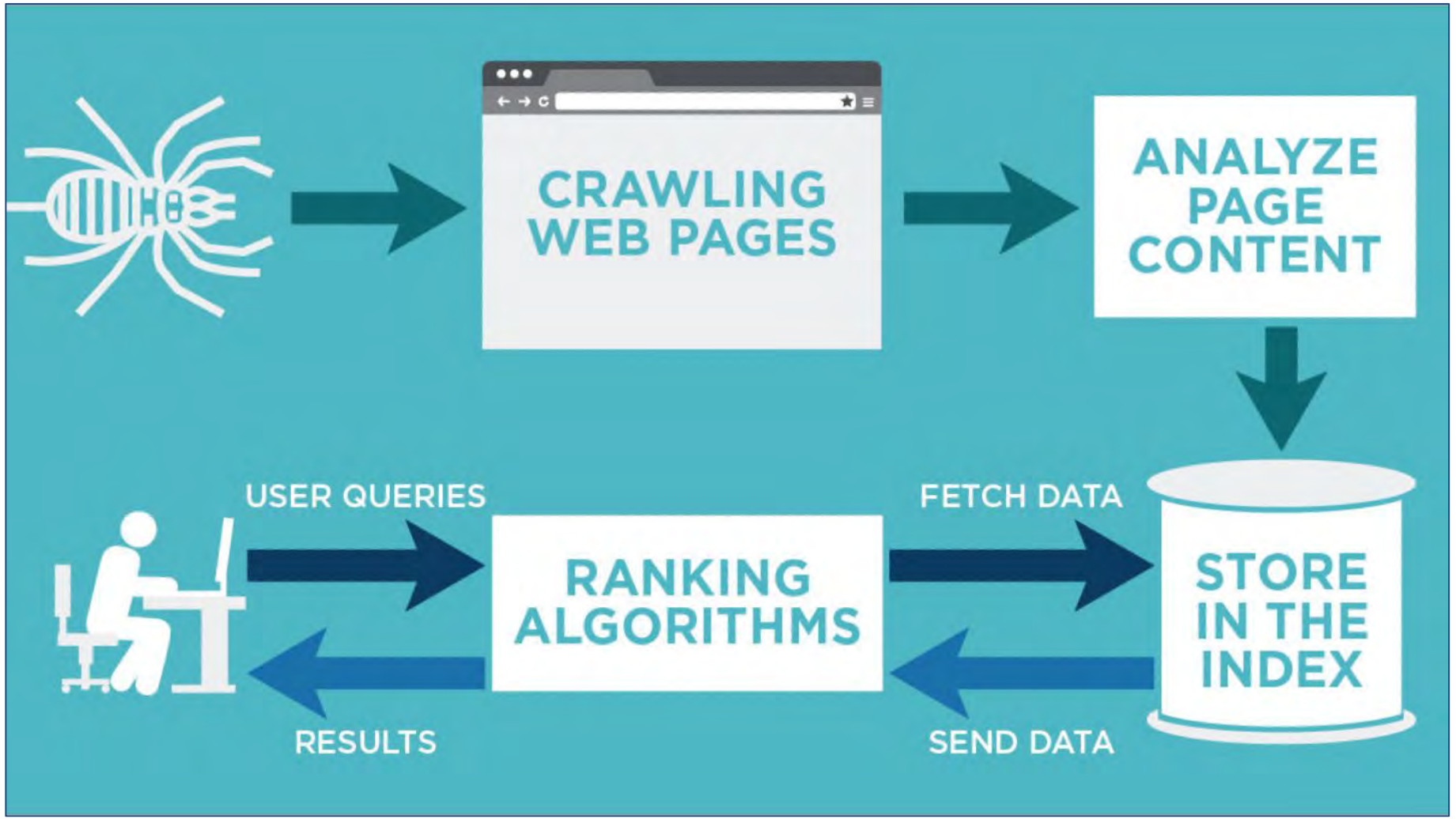


Image credit: https://www.stonetemple.com/wp-content/uploads/2016/12/Crawling-indexing-ranking-hero-image.jpg

Google (~1997)

Image credit: https://betounix.wordpress.com/2011/10/21/los-primeros-servidores-de-google/

Laboratorio di Standford:

- 2 Processori Pentium II 300mhz, 512mb di RAM e 5 dischi da 9 GB.
- 2 Processori Pentium II 300mhz, 512mb di RAM e 4 dischi da 9 GB.
- 4 Processori PPC 604 333mhz, 512mb di RAM e 8 dischi da 9 GB.
- 2 Processori UltraSparc II 200mhz, 256mb di RAM, 3 dischi da 9 GB e 6 da 4 GB.
- 18 dischi da 9 GB come espansione.

Totale

- 1792 MB di memoria RAM.
- 357 GB di disco.
- 2933 Mhz in 10 CPUs.

Scalabilità verticale vs scalabilità orizzontale

- **Scalabilità**: caratteristica di un sistema software o hardware di adattarsi facilmente nel caso di variazioni notevoli della mole o della tipologia dei dati trattati.

Image credit: http://tutorials.jenkov.com/software-architecture/scalable-architectures.html

Vantaggi delle scalabilità orizzontale

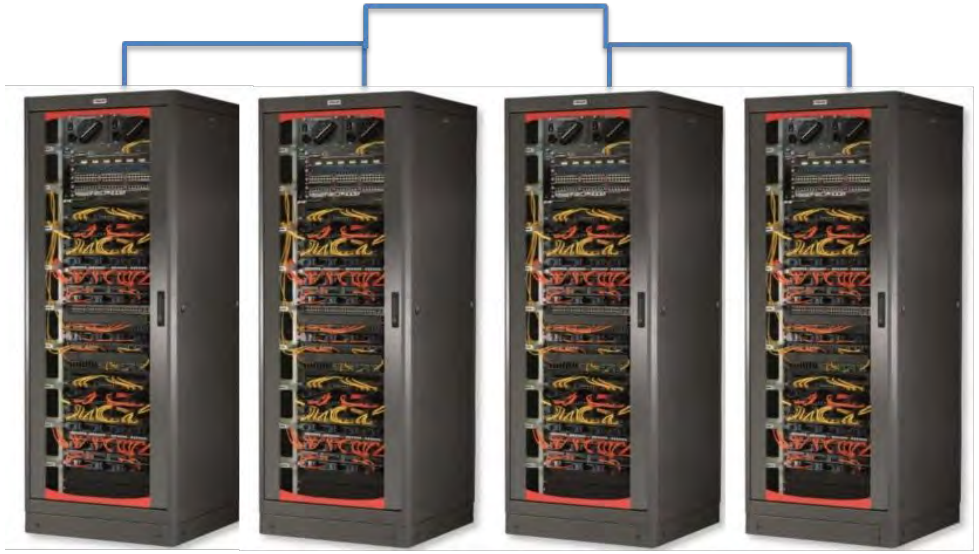
Commodity Hardware

- Nessun lock-in con dispositivi molto costosi
- E' possibile scegliere hardware più economico e disponibile sul mercato
- Commodity non significa di bassa qualità
  (bassa qualità → maggiore tasso di fallimento)

On premise vs. cloud
Principali provider di servizi big data:
AWS, Microsoft Azure e Google

Cluster computing

- Nodi computazionali posizionati in rack
- Più rack connessi tra loro
- Nodi computazionali possono fallire



Google: datacenter di Lenoir, Carolina del Nord

Gartner ha stimato che Google nel 2016 avesse 2.5 milioni di server

Google ha una ventina di datacenter sparsi per il mondo

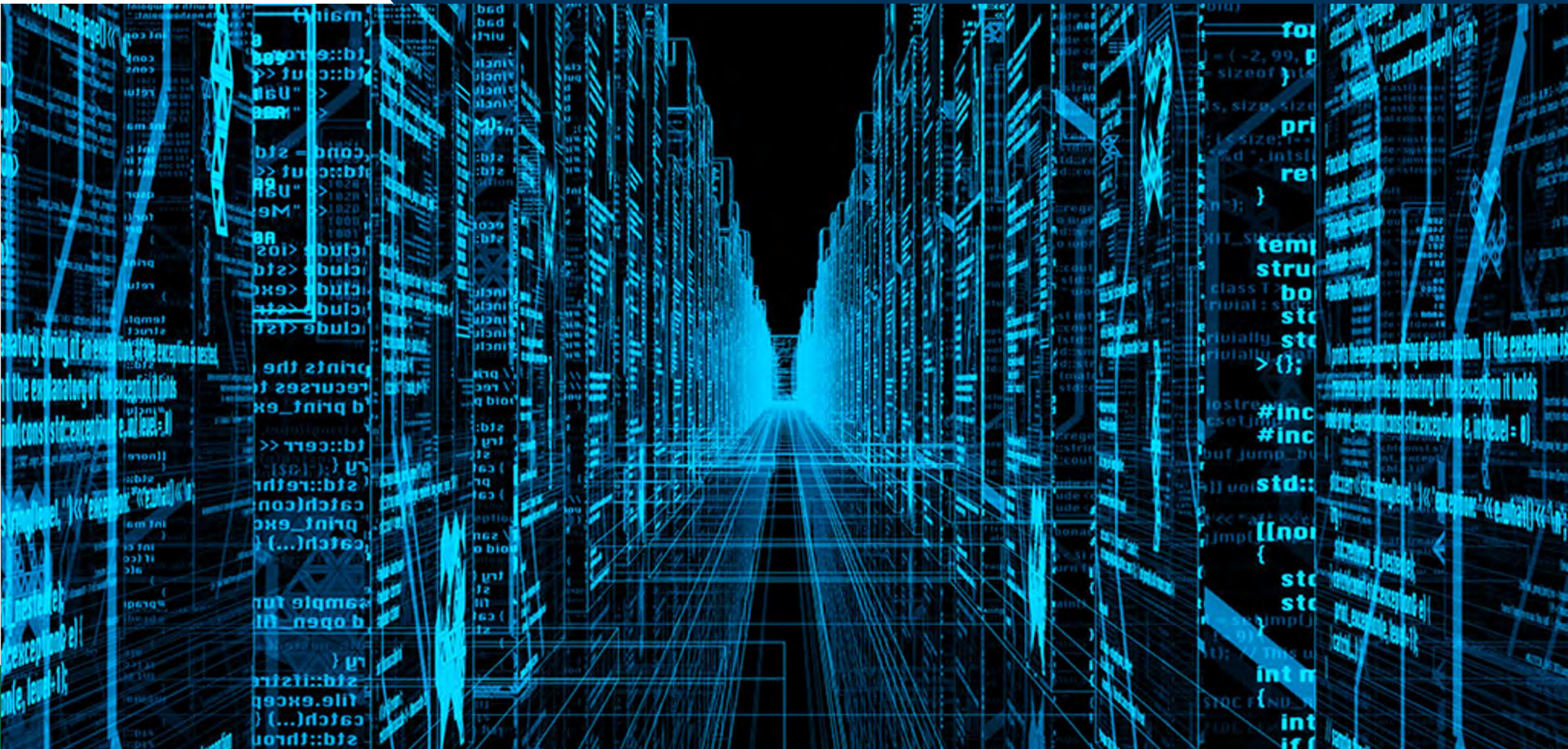
Visita all'interno del datacenter:
https://www.youtube.com/watch?v=avP5d16wEp0

https://www.google.com/intl/it_all/about/datacenters/inside/streetview/


Google
google.com/datacenters

FUB
Fondazione Ugo Bordoni
Ricerca e Innovazione

COME RISOLSE GOOGLE IL PROBLEMA DEL CRAWLING?



MapReduce e Google File System (~ 2003)

Map Reduce: Simplified Processing on Large Clusters

Jeffrey Dean and Sanjay Ghemawat
Google, Inc.
OSDI '04: 6th Symposium on Operating Systems Design and Implementation

The Google File System

Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung

Proceedings of the nineteenth ACM symposium on Operating systems principles, October 19-22, 2003, Bolton Landing, NY, USA

The Google File System

Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung
Google*

MapReduce: Simplified Data Processing on Large Clusters

Jeffrey Dean and Sanjay Ghemawat
jeff@google.com, sanjay@google.com
Google, Inc.

Abstract

MapReduce is a programming model and an associated implementation for processing and generating large data sets. Users specify a *map* function that processes a key/value pair to generate a set of intermediate key/value pairs, and a *reduce* function that merges all intermediate values associated with the same intermediate key. Many real world tasks are expressible in this model, as shown in the paper.

Programs written in this functional style are automatically parallelized and executed on a large cluster of commodity machines. The run-time system takes care of the details of partitioning the input data, scheduling the program's execution across a set of machines, handling machine failures, and managing the required inter-machine communication. This allows programmers without any experience with parallel and distributed systems to easily utilize the resources of a large distributed system.

Our implementation of MapReduce runs on a large cluster of commodity machines and is highly scalable: a typical MapReduce computation processes many terabytes of data on thousands of machines. Programmers find the system easy to use: hundreds of MapReduce programs have been implemented and upwards of one thousand MapReduce jobs are executed on Google's clusters every day.

1 Introduction

Over the past five years, the authors and many others at Google have implemented hundreds of special-purpose computations that process large amounts of raw data, such as crawled documents, web request logs, etc., to compute various kinds of derived data, such as inverted indices, various representations of the graph structure of web documents, summaries of the number of pages crawled per host, the set of most frequent queries in a given day, etc. Most such computations are conceptually straightforward. However, the input data is usually large and the computations have to be distributed across hundreds or thousands of machines in order to finish in a reasonable amount of time. The issues of how to parallelize the computation, distribute the data, and handle failures conspire to obscure the original simple computation with large amounts of complex code to deal with these issues.

As a reaction to this complexity, we designed a new abstraction that allows us to express the simple computations we were trying to perform but hides the messy details of parallelization, fault-tolerance, data distribution and load balancing in a library. Our abstraction is inspired by the *map* and *reduce* primitives present in Lisp and many other functional languages. We realized that most of our computations involved applying a *map* operation to each logical "record" in our input in order to compute a set of intermediate key/value pairs, and then applying a *reduce* operation to all the values that shared the same key, in order to combine the derived data appropriately. Our use of a functional model with user-specified map and reduce operations allows us to parallelize large computations easily and to use re-execution as the primary mechanism for fault tolerance.

The major contributions of this work are a simple and powerful interface that enables automatic parallelization and distribution of large-scale computations, combined with an implementation of this interface that achieves high performance on large clusters of commodity PCs.

Section 2 describes the basic programming model and gives several examples. Section 3 describes an implementation of the MapReduce interface tailored towards our cluster-based computing environment. Section 4 describes several refinements of the programming model that we have found useful. Section 5 has performance measurements of our implementation for a variety of tasks. Section 6 explores the use of MapReduce within Google including our experiences in using it as the basis

To appear in OSDI 2004

1

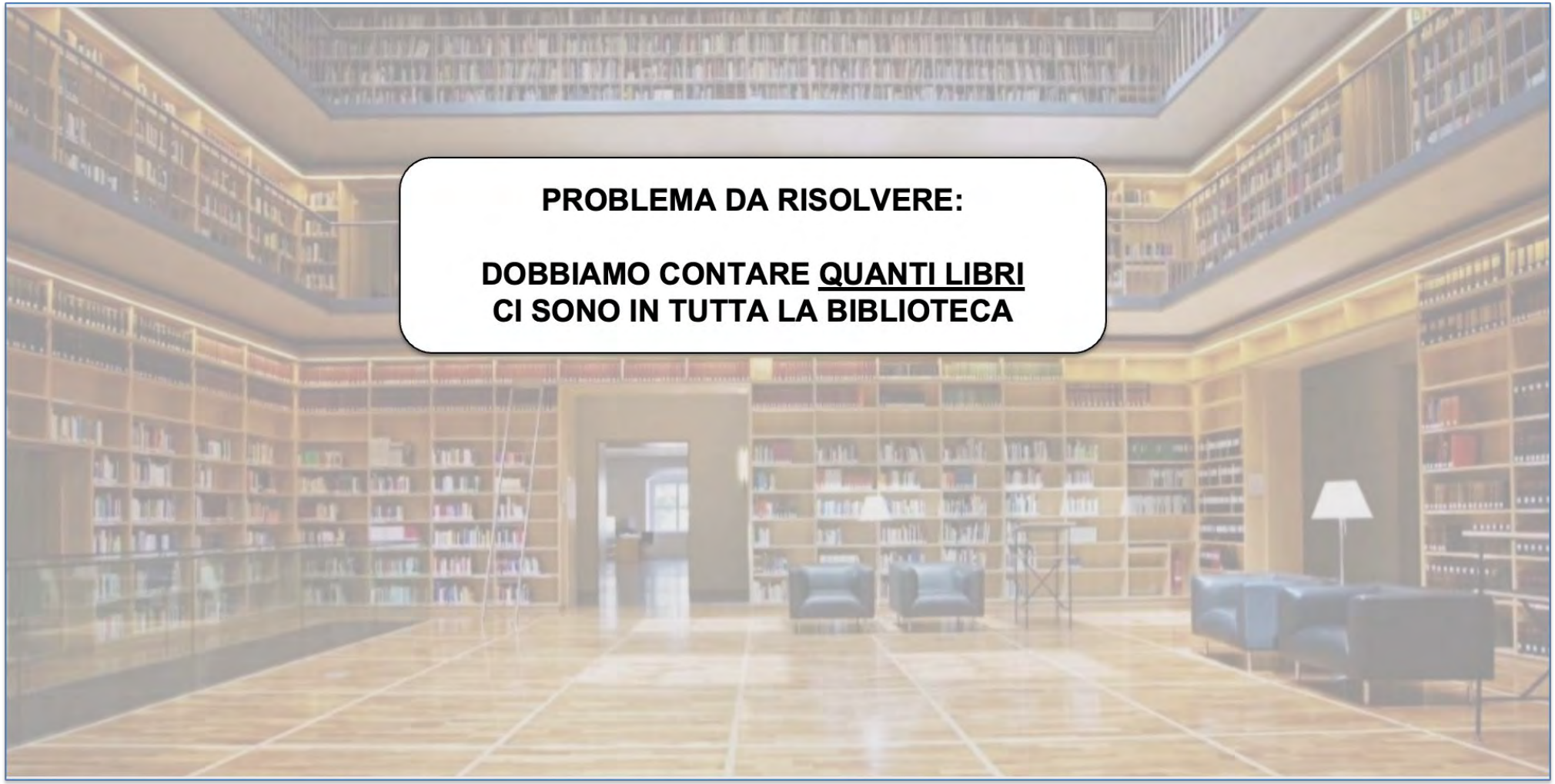
MapReduce

- **Framework di programmazione distribuita**
- **Adotta un paradigma di programmazione basato su due fasi:**
  1. **Fase di MAP: il problema viene suddiviso in tanti piccoli sotto-problemi che sono risolti singolarmente**
  2. **Fase di REDUCE: i risultati dei sotto-problemi vengono raccolti per calcolare il risultato finale del problema iniziale**
- **Principio di località: la computazione viene eseguita più vicino possibile ai dati → si limita lo spostamento dei dati**

Capire MapReduce con un'analogia

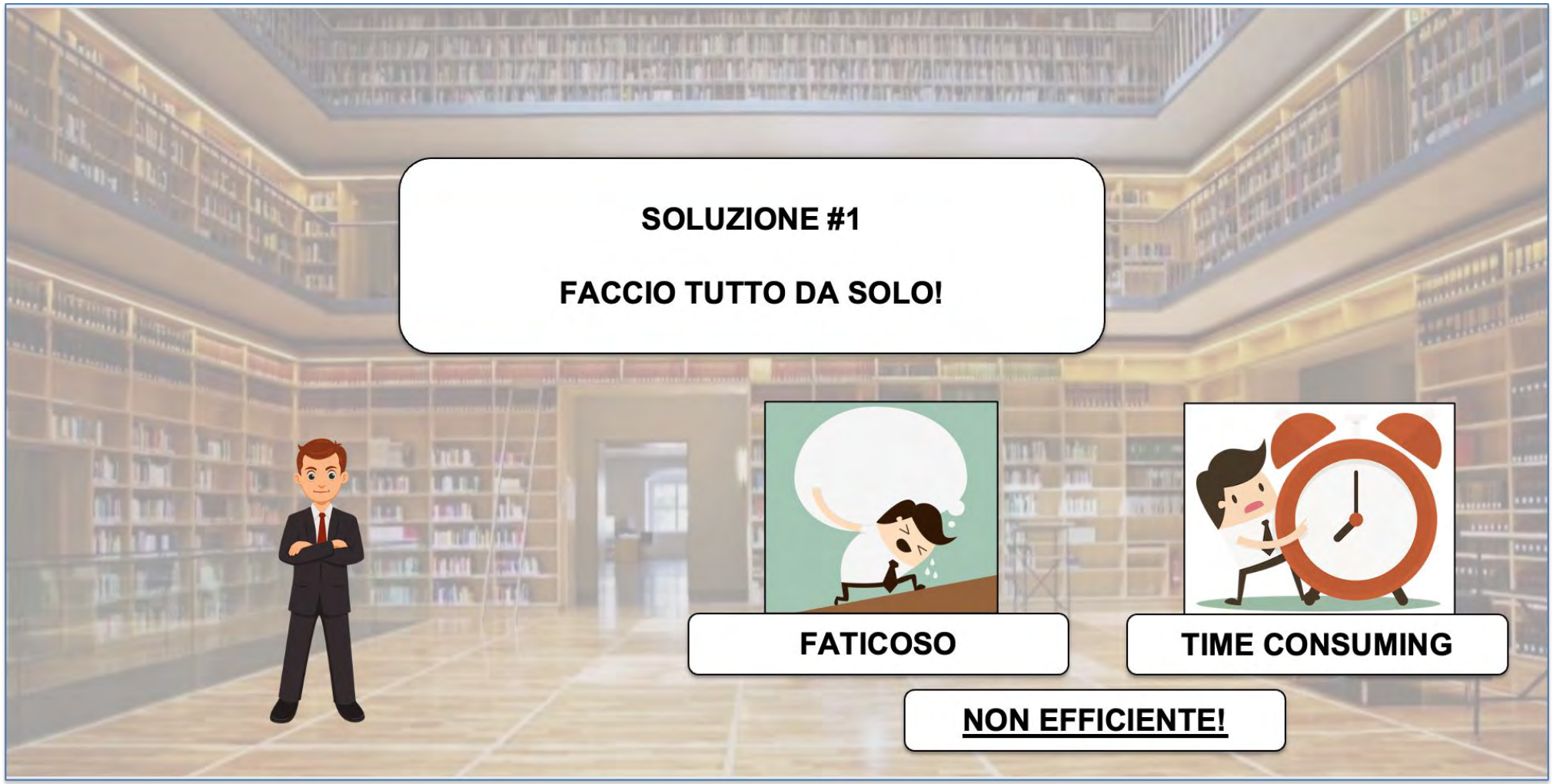
Esempio tratto e adattato da: https://www.slideshare.net/Simplilearn/mapreduce-in-hadoop-mapreduce-explained-mapreduce-architecture-mapreduce-tutorial-simplilearn

Capire MapReduce con un'analogia



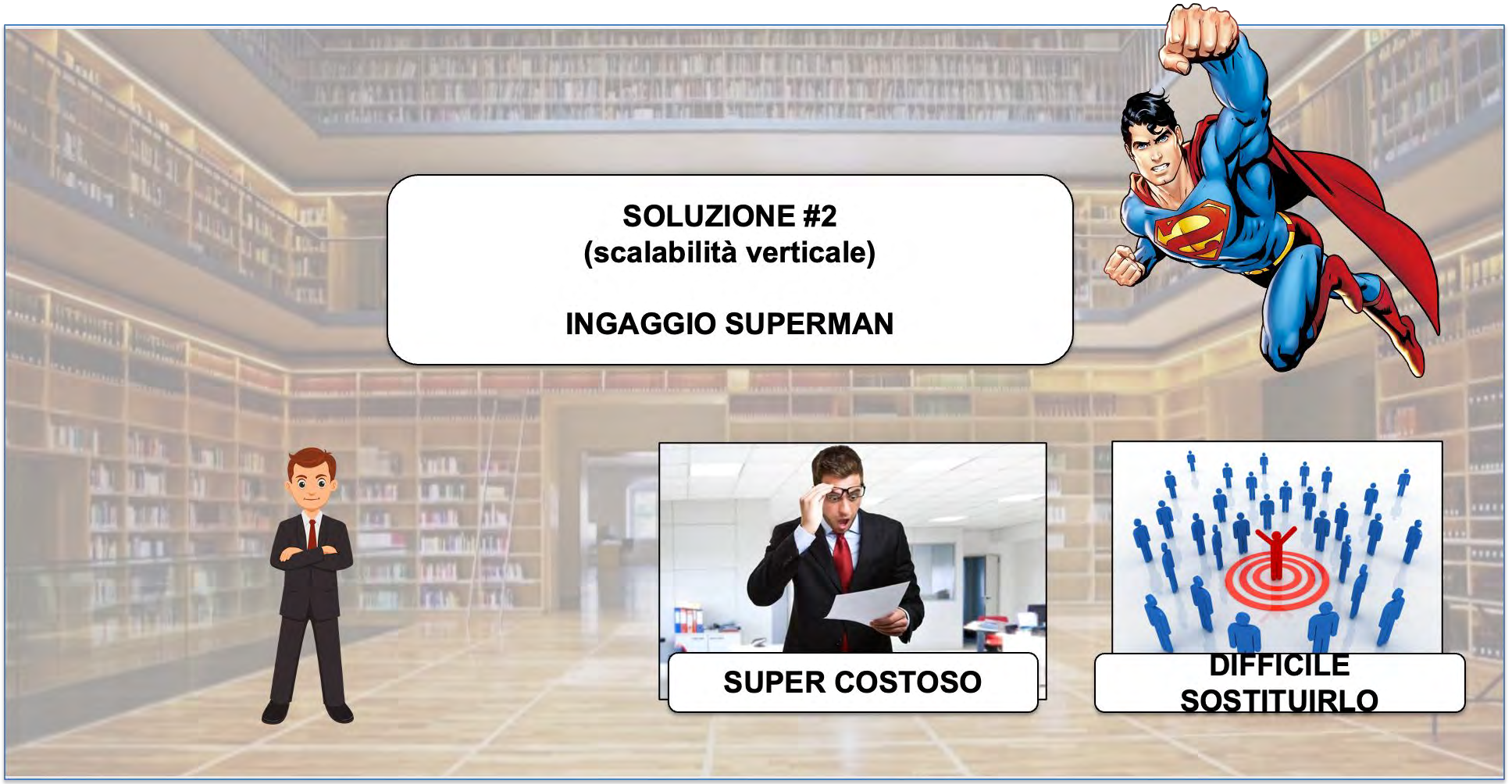
Capire MapReduce con un'analogia

Images credits:
https://www.kissclipart.com/time-consuming-clipart-municipal-account-meter-sol-xnvpal/
https://it.freepik.com/dooder



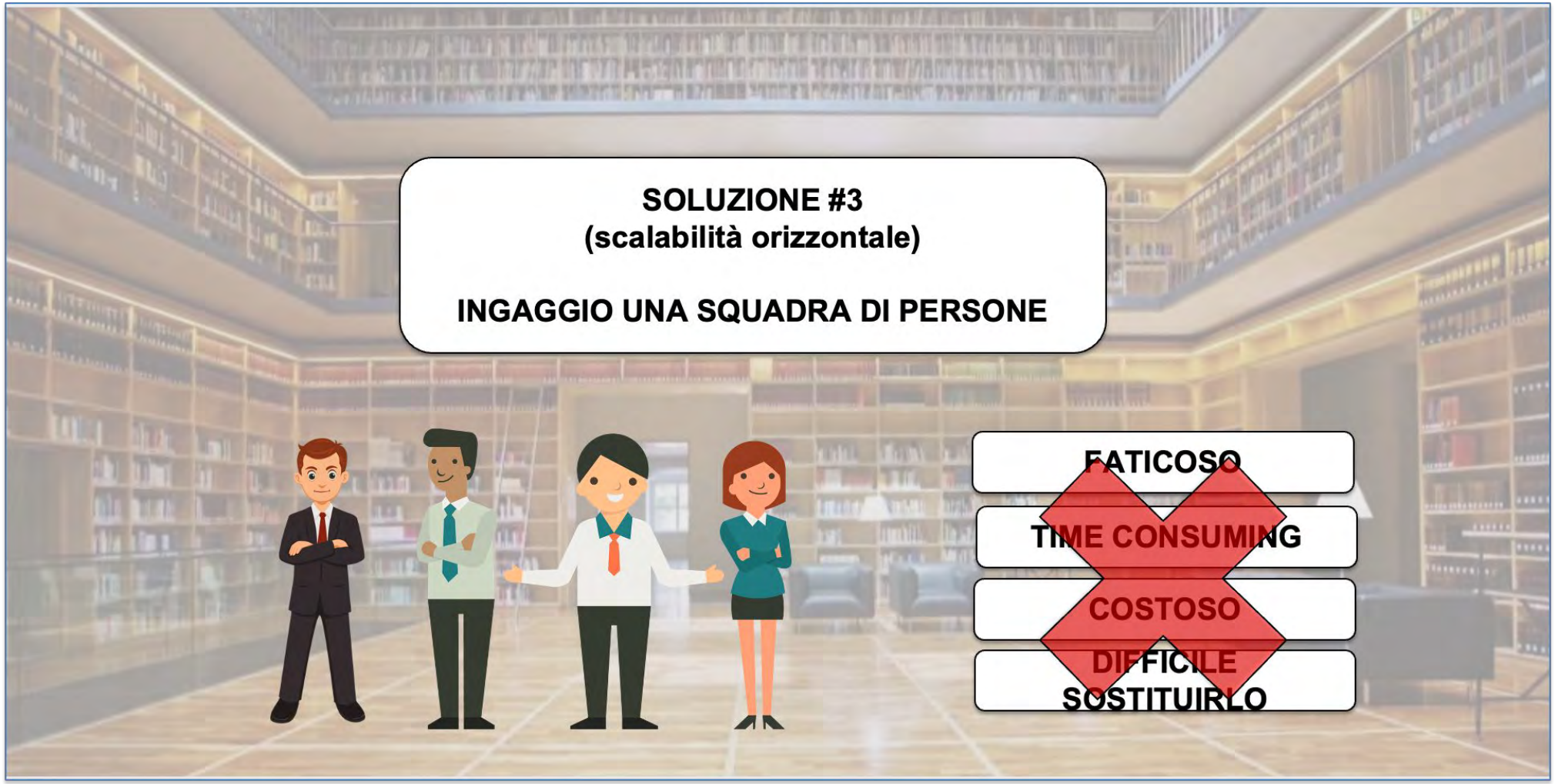
Capire MapReduce con un'analogia

Images credits:
https://clipart.info/superman-png-hd-new-7816
http://greenguardpestrepeller.com/wp-content/uploads/2016/08/Expensive-Bill.jpg



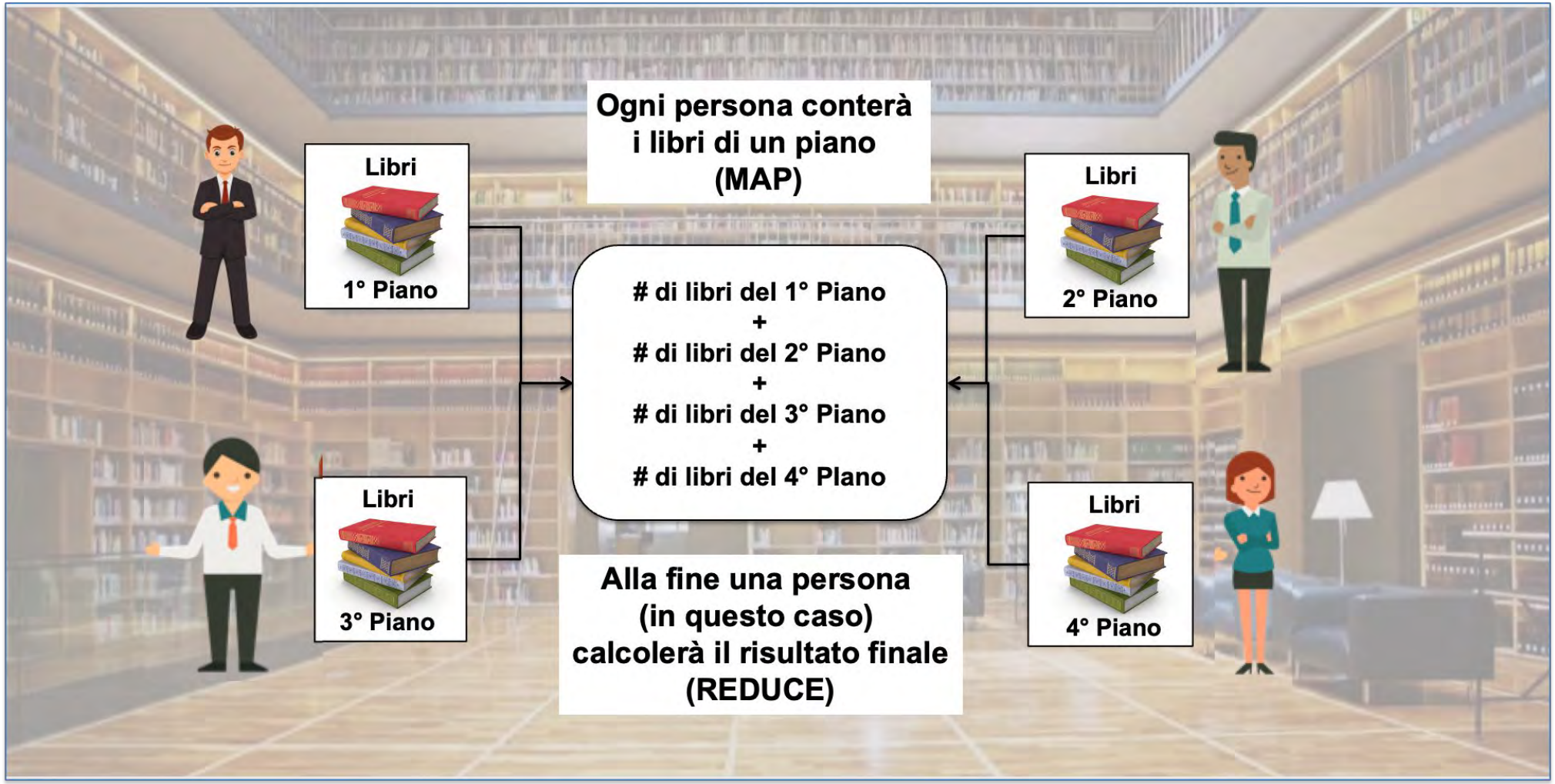
Capire MapReduce con un'analogia

Images credits:
https://www.stickpng.com/img/people/groups/cartoon-people



Capire MapReduce con un'analogia

Images credits:
https://www.stickpng.com/img/people/groups/cartoon-people



Architettura Big data: una definizione

Un'architettura Big Data è un insieme strutturato di componenti, processi e tecnologie progettato per gestire e sfruttare dati di grandi "dimensioni" in modo efficace ed efficiente.

Image credit: https://www.knowledgehut.com/blog/big-data/big-data-architecture

Architetture Big data: aspetti da gestire

- **Raccolta dei dati**: processo di acquisizione di dati provenienti da diverse fonti, come sensori, dispositivi IoT, social media, ecc.
- **Archiviazione dei dati**: i dati raccolti vengono conservati in un sistema di archiviazione scalabile (*data lake*) che può memorizzare grandi quantità di dati in modo efficiente.
- **Elaborazione dei dati**: i dati vengono trasformati e preparati per l'analisi (es. pulizia dei dati, aggregazione, normalizzazione ecc.).
- **Analisi dei dati**: applicazione di algoritmi e strumenti di analisi per scoprire modelli, tendenze e informazioni utili nei dati.
- **Visualizzazione dei dati**: informazioni estratte rappresentate attraverso grafici, dashboard e report comprensibili, in modo che gli utenti possano interpretare facilmente i risultati.
- **Governance e sicurezza dei dati**: garantire la qualità dei dati, il rispetto delle normative sulla privacy e la sicurezza delle informazioni sensibili.
- **Scalabilità e flessibilità**: le architetture Big Data sono progettate per essere altamente scalabili, ossia per gestire facilmente la crescita dei dati nel tempo e per adattarsi alle esigenze in continua evoluzione.

Image credit: https://www.knowledgehut.com/blog/big-data/big-data-architecture

Architetture Big data: schematizzazione

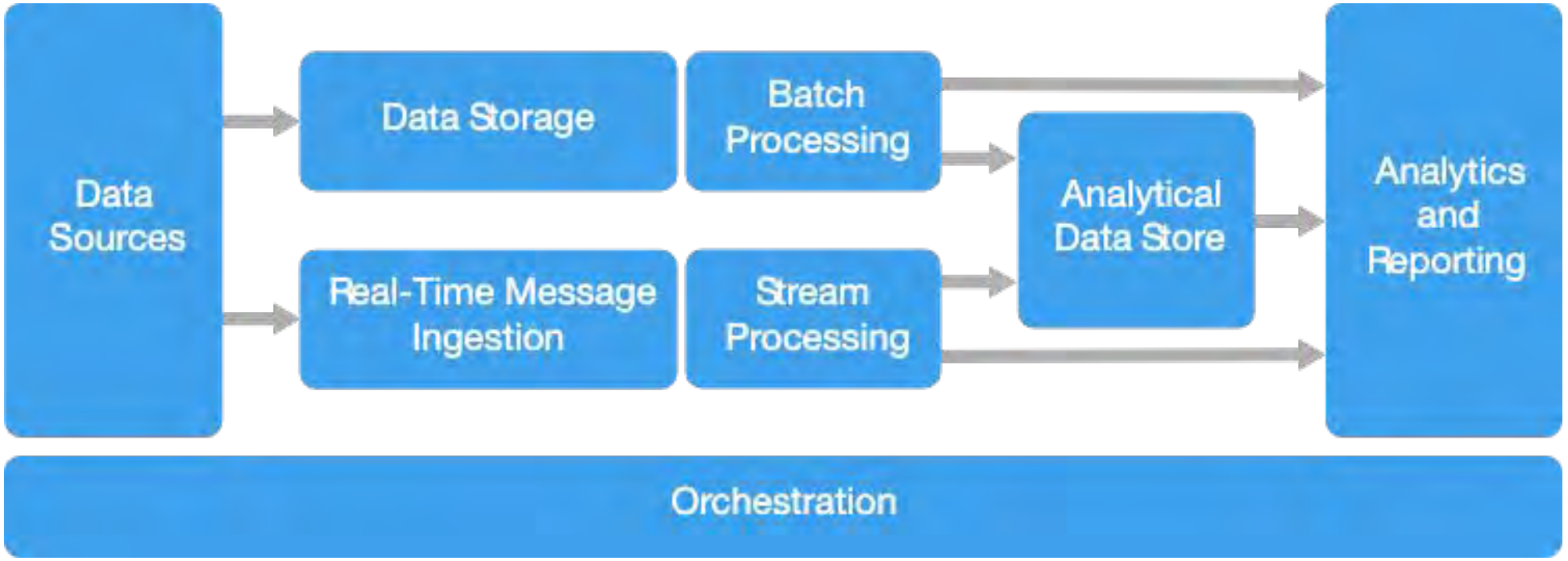
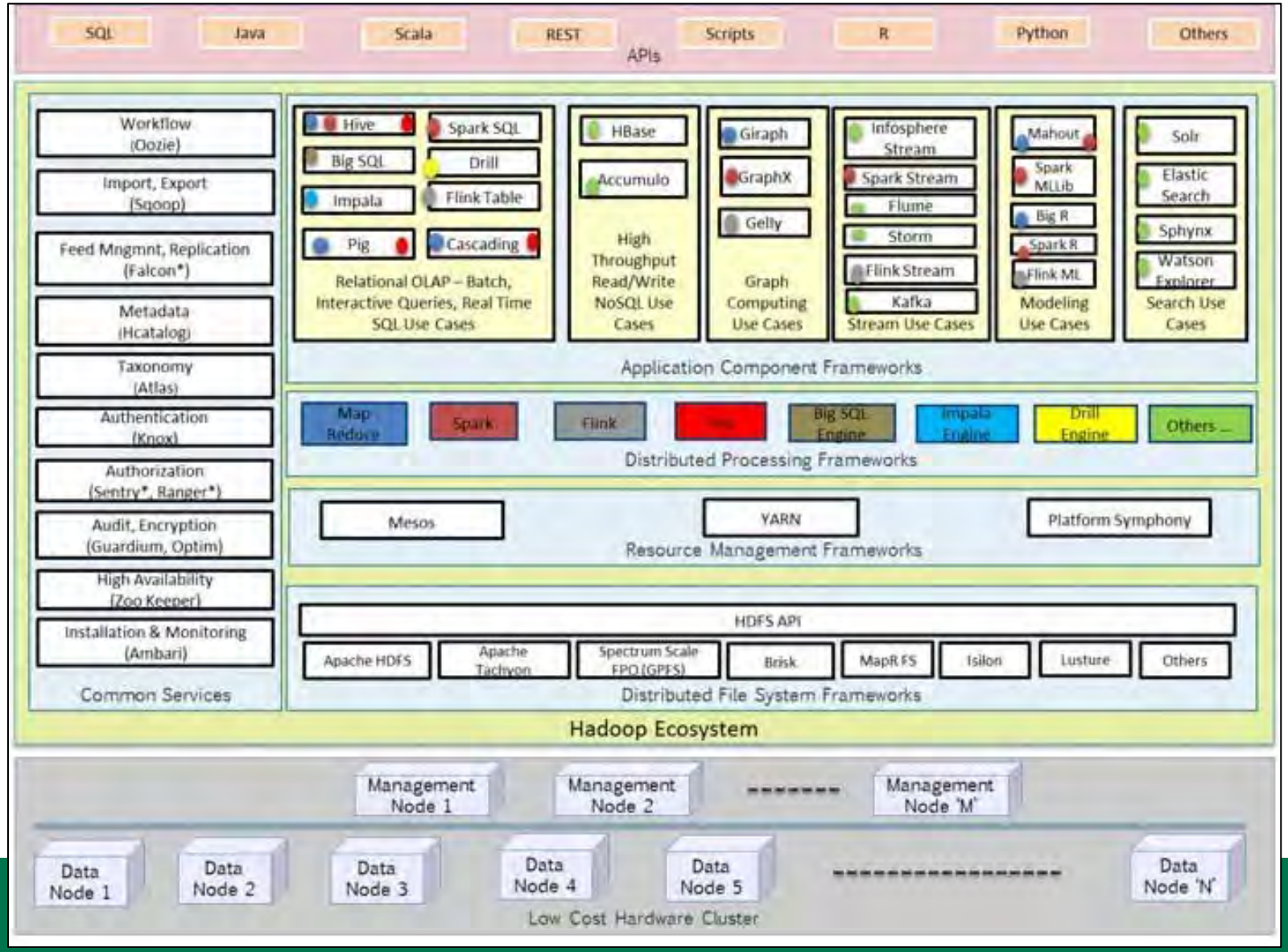


Image credit: https://learn.microsoft.com/en-us/azure/architecture/guide/architecture-styles/big-data

Sistemi operativi big data: struttura a layer



Riferimento:
Hadoop as Big Data Operating System -- The Emerging Approach for Managing Challenges of Enterprise Big Data Platform
Sourav Mazumdar, Subhankar Dhar
Published 30 March 2015
Computer Science
2015 IEEE First International Conference on Big Data Computing Service and Applications

FUB
Fondazione Ugo Bordoni
Ricerca e Innovazione

Ecosistema delle tecnologie Big Data

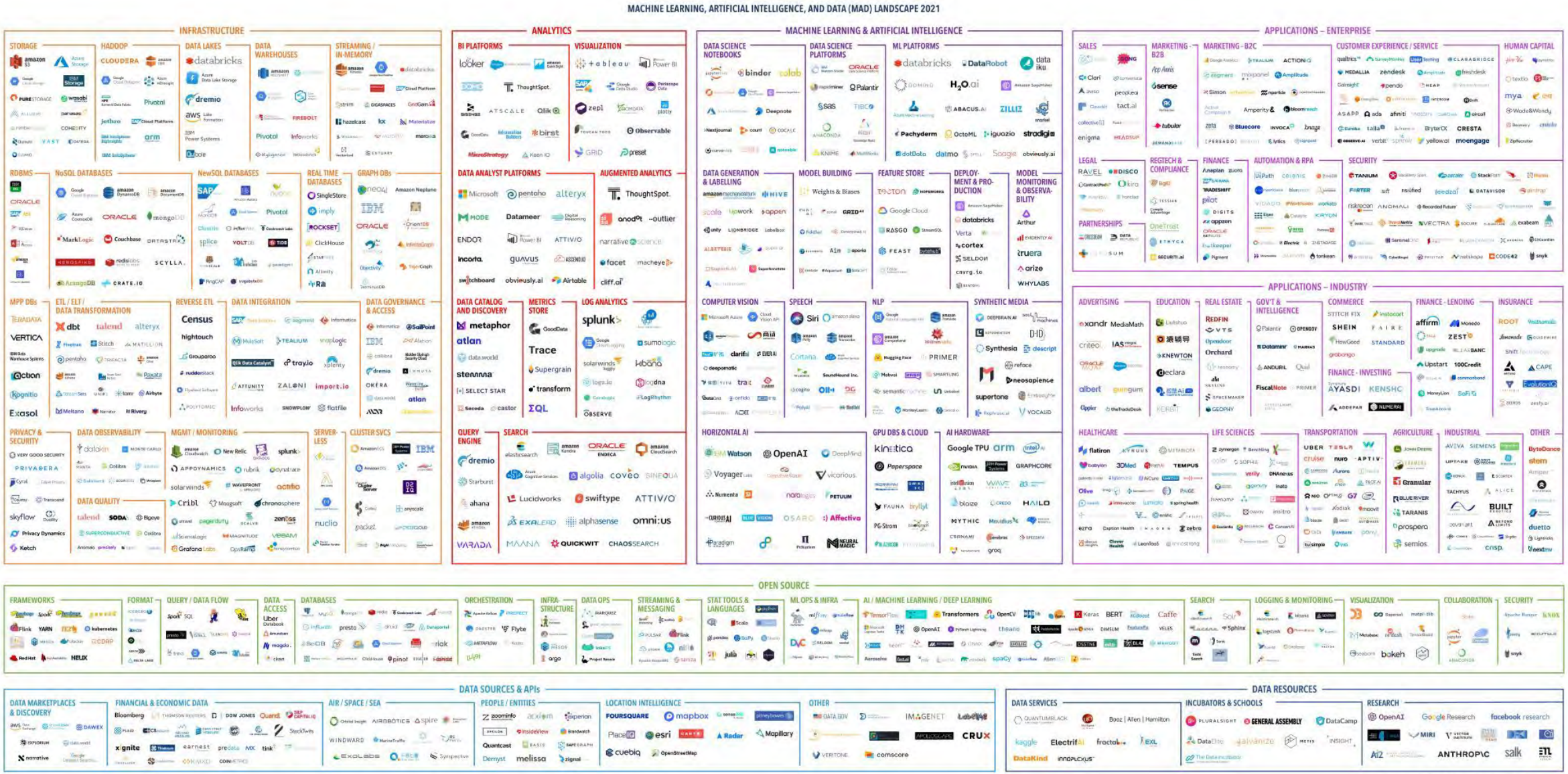


Figure professionali Big data

- **Data Scientist** (Scienziato dei dati): esperti nell'analisi dei dati.
- **Data Engineer** (Ingegnere dei dati): responsabili della progettazione e dell'implementazione delle infrastrutture e delle pipeline di dati.
- **Data Analyst** (Analista dei dati): responsabili dell'elaborazione, dell'analisi e della visualizzazione dei dati per aiutare le aziende a prendere decisioni informate.
- **Big Data Architect** (Architetto dei big data): responsabili della progettazione di soluzioni complesse di archiviazione e gestione dei dati, spesso basate su tecnologie come Hadoop e Spark.
- **Data Governance Manager** (Responsabile della Governance dei dati): si occupano di stabilire politiche, procedure e norme per garantire la qualità, la sicurezza e la conformità dei dati aziendali. Gestiscono anche l'accesso ai dati e la privacy.
- **Chief Data Officer (CDO)**: dirigente di alto livello responsabile della **strategia dei dati in un'organizzazione**.

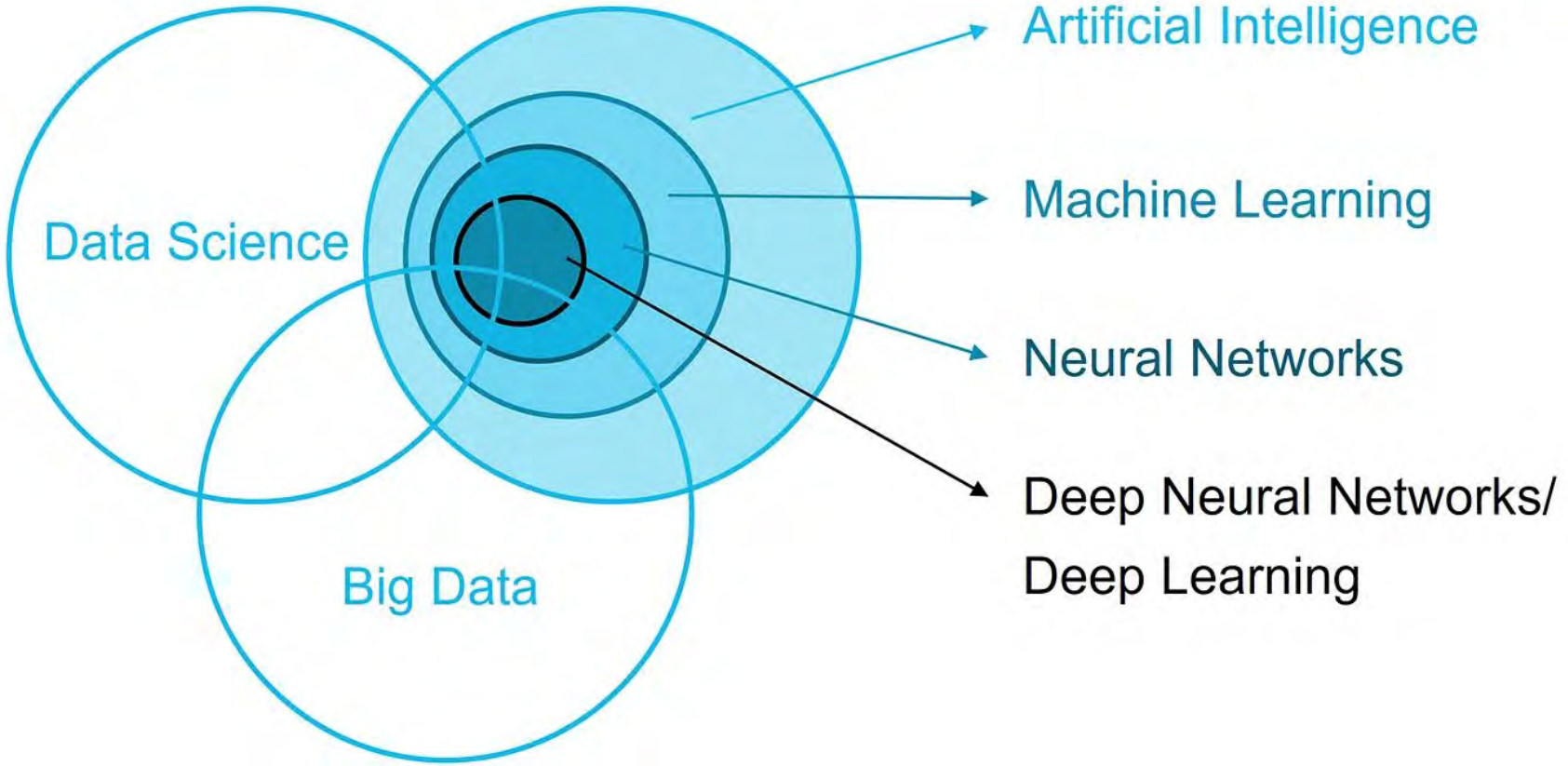
https://data-flair.training/blogs/big-data-job-roles/

FUB
Fondazione Ugo Bordoni
Ricerca e Innovazione

Big data in relazione a IA, IoT, Industrial IoT, edge computing

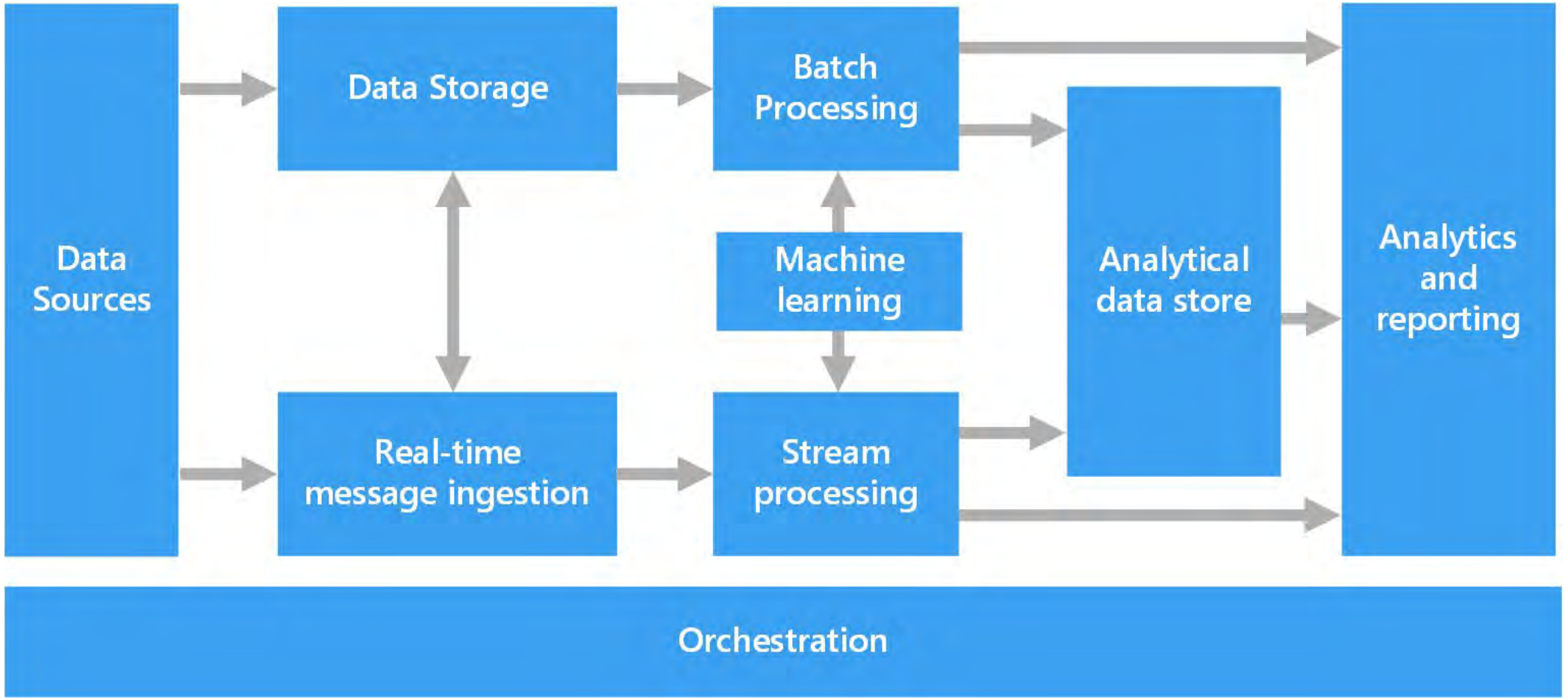
Image credit:
https://scet.berkeley.edu/big-data-solutions-for-small-firms/word-cloud-big-data/

Data science, Intelligenza artificiale e Big data



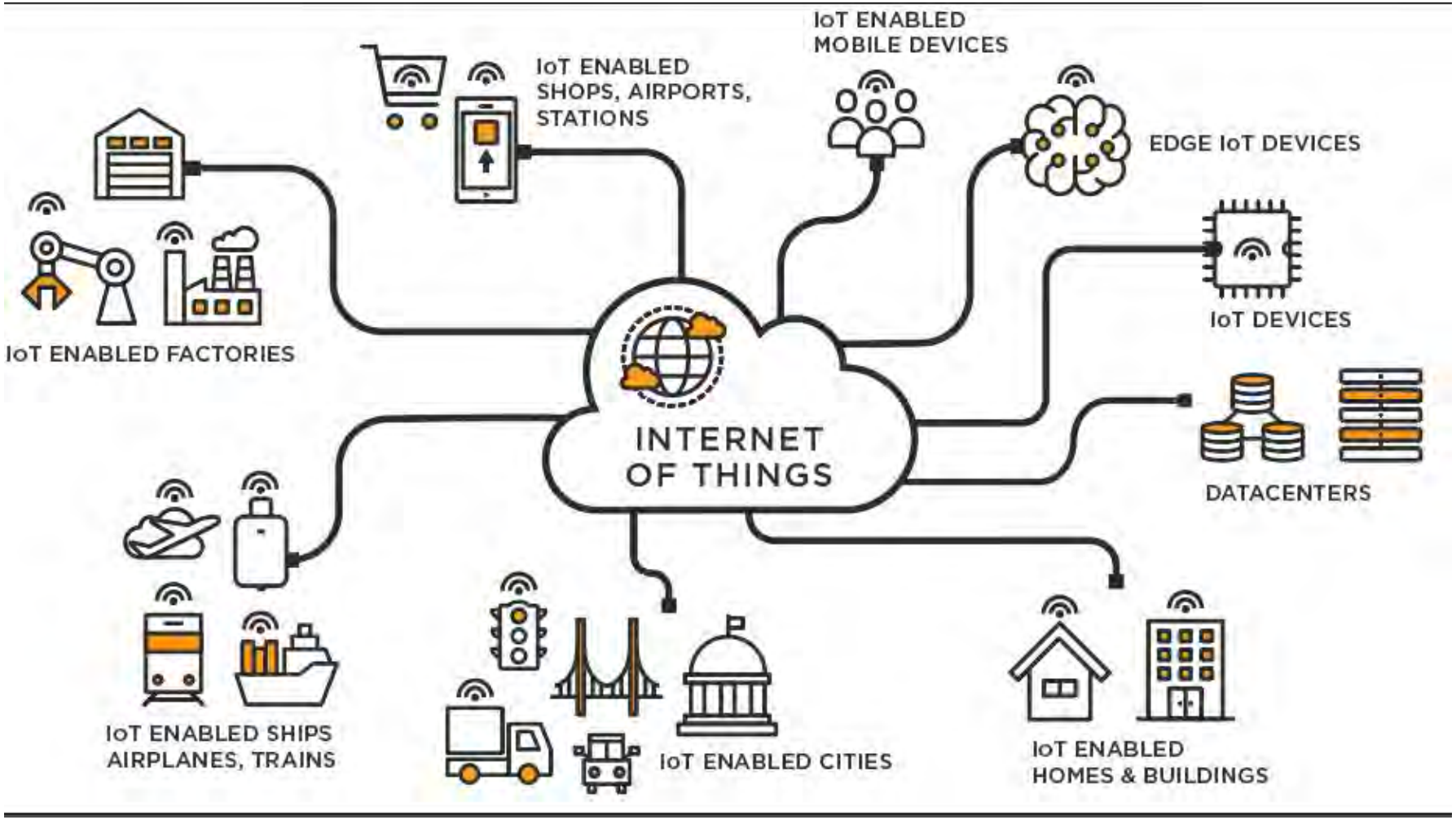
https://towardsdatascience.com/role-of-data-science-in-artificial-intelligence-950efedd2579

Big data e Intelligenza Artificiale



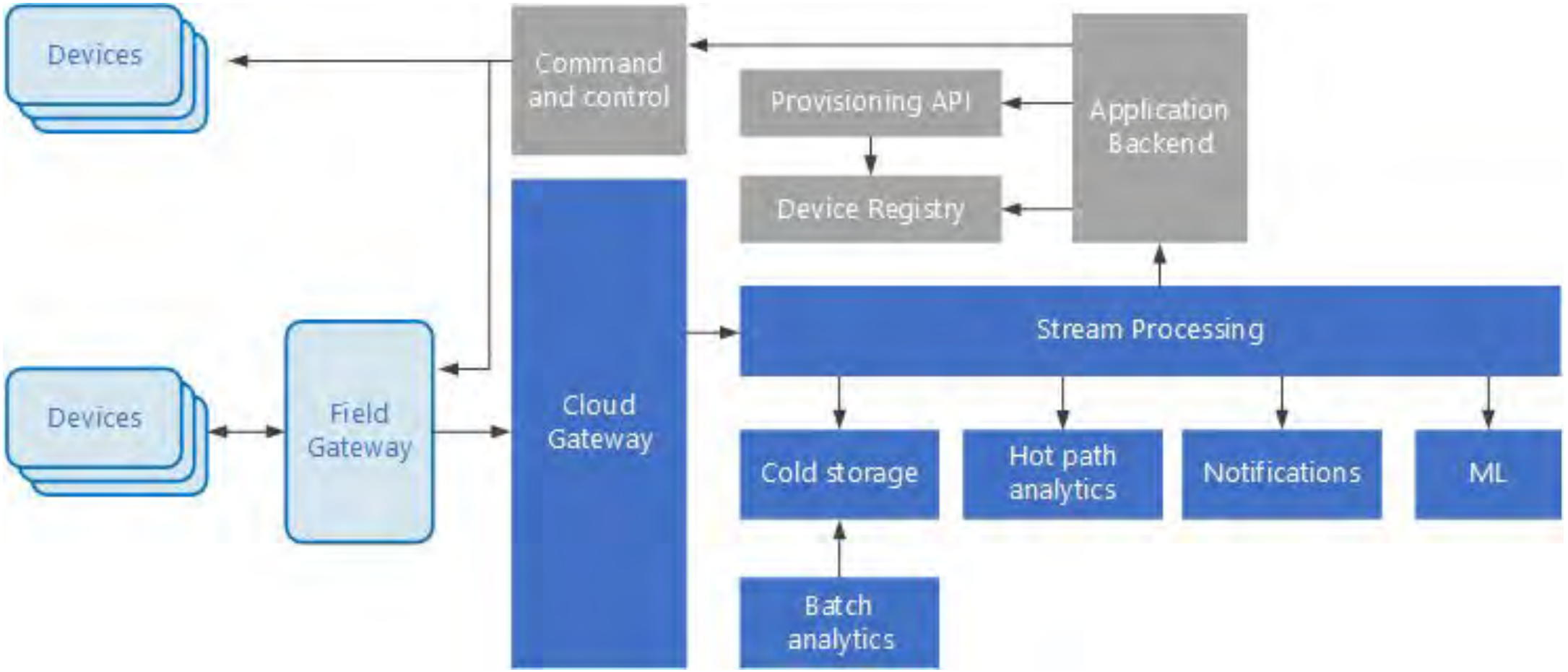
Fonte immagine: docs.microsoft.com

Internet of things



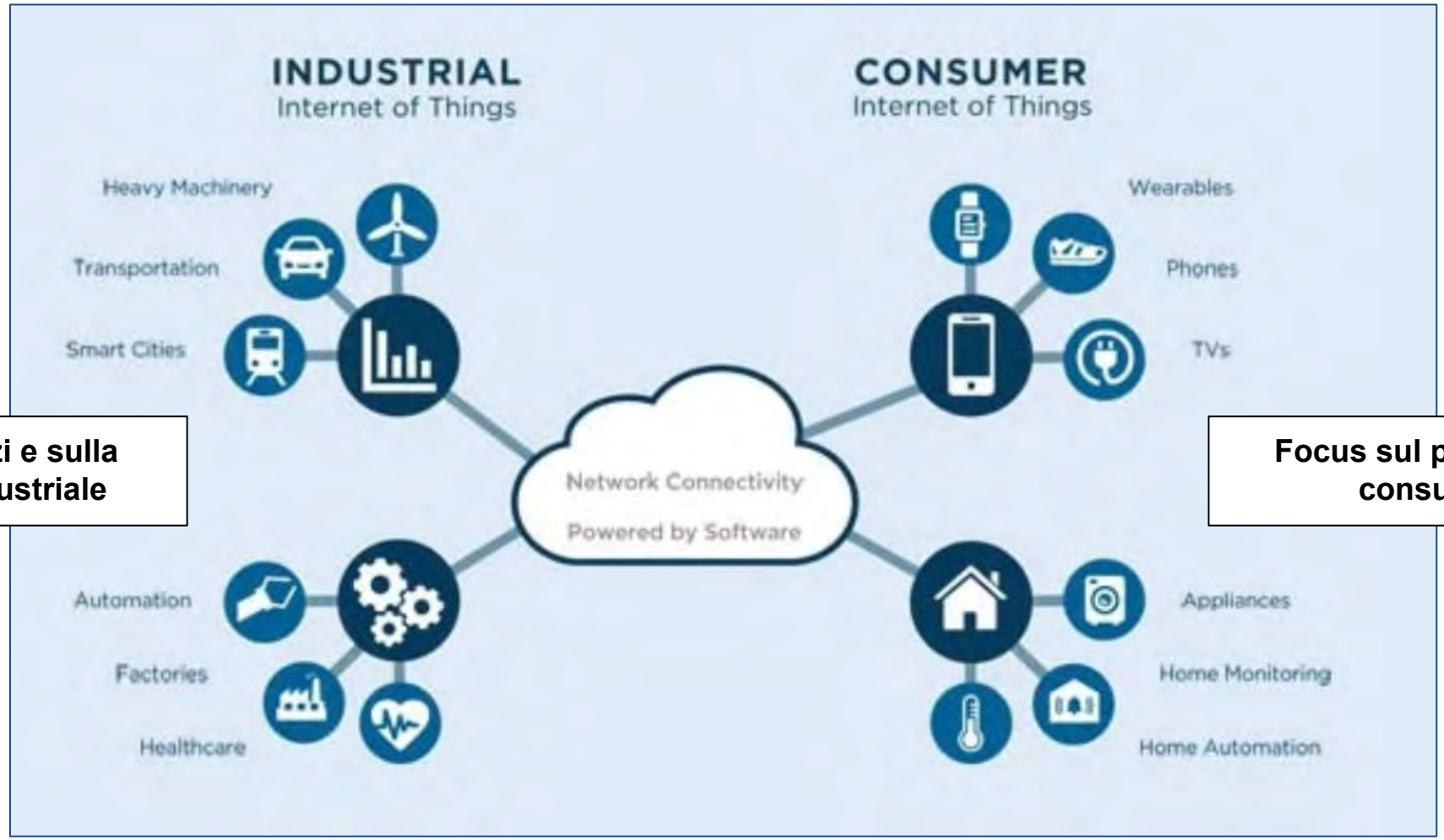
Mukherjee, Suprakash & Gupta, Shashank & Rawlley, Oshin & Jain, Siddhant. (2022). Leveraging Big Data Analytics in 5G-Enabled IoT and Industrial IoT for the Development of Sustainable Smart Cities. Transactions on Emerging Telecommunications Technologies. 33. 10.1002/ett.4618.

Architettura IOT - caso particolare di soluzione Big data



https://learn.microsoft.com/en-us/azure/architecture/guide/architecture-styles/big-data

Industrial IoT e Customer IoT



Focus sui servizi e sulla produzione industriale

Focus sul prodotto e sul consumatore

Image credit: https://www.isipc.it/differenze-tra-internet-of-things-e-industrial-internet-of-things/

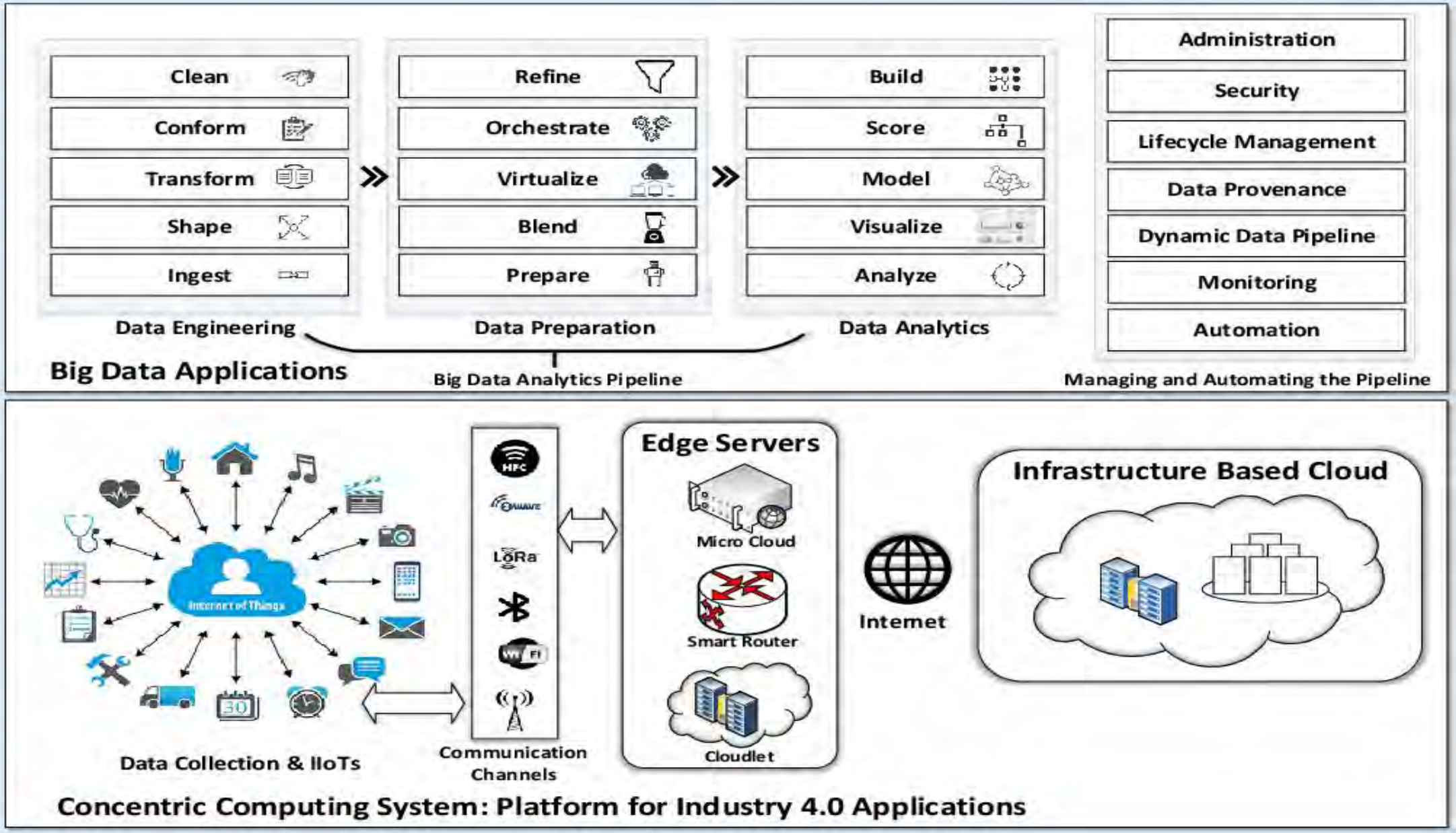
Edge computing

- **Modello di elaborazione** dei dati e delle informazioni che si basa sulla distribuzione di capacità di calcolo e di archiviazione dei dati più vicino al punto in cui questi dati vengono generati o utilizzati, anziché inviarli a un data center o a un cloud remoto per l'elaborazione.

- Ampiamente utilizzato in settori come:
  - sicurezza (sorveglianza video)
  - automazione industriale (manutenzione predittiva)
  - automotive (guida autonoma)
  - salute (monitoraggio pazienti)
  - videogiochi e streaming multimediale

L'edge computing sposta la computazione "ai margini" o "ai bordi" della rete, vicino alle fonti dei dati o agli utenti finali.

- Le caratteristiche principali dell'edge computing includono:
  - **Prossimità ai dispositivi o alle sorgenti dei dati**
  - **Latenza ridotta**
  - **Risparmio di larghezza di banda**
  - **Affidabilità** (in casi di connettività instabile o limitata)
  - **Privacy e sicurezza**



Prakash Jayaraman, Charith Perera, The role of big data analytics in industrial Internet of Things, Future Generation Computer Systems,Volume 99, 2019, Pages 247-259, ISSN 0167-739X, https://doi.org/10.1016/j.future.2019.04.020.

Casi d'uso dei Big data in alcuni comparti industriali

Image credit: https://www.artslife.com/wp-content/uploads/2018/02/yourstory_BigData-590x295.jpg

Big data nel settore agricolo

- Agricoltura di precisione
- Gestione del bestiame
- Previsioni meteorologiche agricole
- Gestione delle risorse idriche
- Gestione delle malattie delle piante
- Mercati e previsioni dei prezzi
- Automazione agricola
- Sviluppo di colture geneticamente migliorate

Image credit:
https://www.businessintelligencegroup.it/agricoltura-4-0-tra-tecnologie-abilitanti-droni-e-big-data/

Big data nel comparto manifatturiero

- Manutenzione predittiva
- Ottimizzazione della produzione
- Gestione delle scorte
- Controllo della qualità
- Progettazione assistita dai dati
- Ottimizzazione della supply chain interna
- Sicurezza delle fabbriche
- Riduzione degli sprechi
- Personalizzazione dei prodotti
- Gestione energetica

Big data nella logistica

- Ottimizzazione delle rotte di trasporto
- Gestione delle flotte
- Tracciamento delle merci in tempo reale
- Gestione degli stock in magazzino
- Previsione della domanda
- Gestione dei resi e dei rifiuti
- Tracciabilità dei prodotti
- Monitoraggio delle condizioni delle merci
- Gestione dei tempi di attesa
- Gestione delle consegne last mile
- Sicurezza delle catene di approvvigionamento

Image credit:
https://blog.locus.sh/towards-unified-indian-logistics-with-data-science-and-ai/

FUB
Fondazione Ugo Bordoni
Ricerca e Innovazione

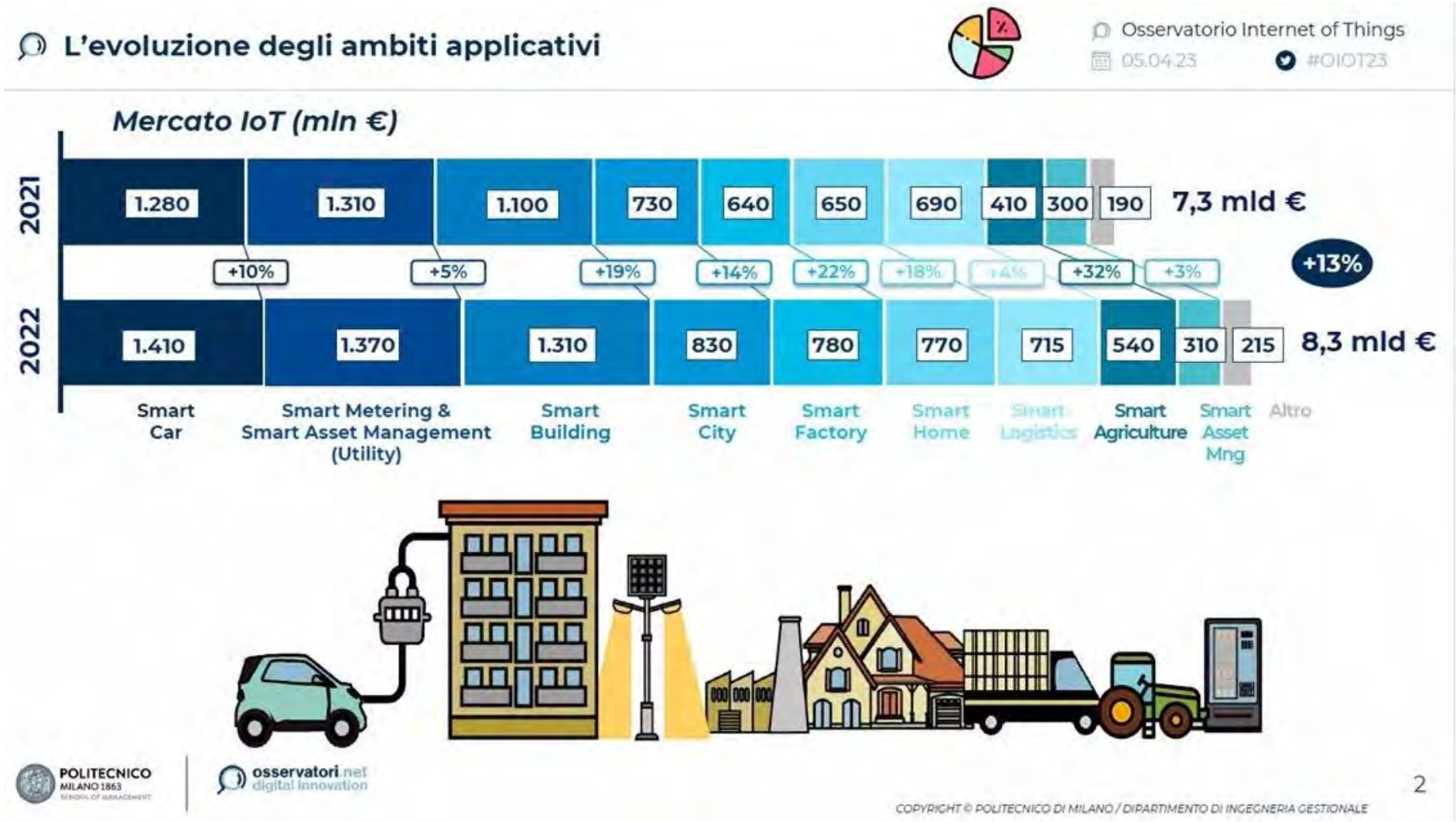
Mercato, data strategy e PMI

Image credit: https://blog.osservatori.net/it_it/mercato-big-data-analytics-in-italia

Il mercato IoT in Italia nel 2022

FUB
Fondazione Ugo Bordoni
Ricerca e Innovazione

Evoluzione del mercato IoT in Italia



https://www.corrierecomunicazioni.it/industria-4-0/iot/internet-of-things-in-italia-ancora-tutta-da-giocare-la-partita-industriale/

Mercato dei Big data in Italia nel 2022

- Il mercato italiano dei Big Data vale 2,41 miliardi di euro, +20% rispetto al 2021
- Nel 2022 forte crescita della spesa in software (+25%).
- Sperimentazione di Advanced Analytics nel 65% delle grandi imprese (54% nel 2021). Nelle PMI si conferma un livello medio inferiore di adozione di tecnologie.
- Permangono importanti differenze tra il livello di maturità delle medie (50-249 addetti) e piccole (10-49 addetti) imprese. Le imprese di medie dimensioni hanno un livello medio di adozione delle tecnologie più alto delle piccole

"Nonostante le difficoltà dello scenario globale, nel 2022 le imprese italiane continuano a mostrare grande interesse per gli Analytics. - afferma **Carlo Vercellis, Responsabile Scientifico dell'Osservatorio Big Data & Business Analytics** - *Cresce la maturità delle organizzazioni verso una cultura data-science-driven e insieme il mercato, che ha lasciato alle spalle il periodo nero. Ma la sfida di chi ha avviato sperimentazioni o progetti di Advanced Analytics ora è quella dell'industrializzazione dei processi per garantire efficienza e governance dei dati in tutti i livelli".*

"La spesa delle aziende italiane è tornata stabilmente a crescere, ancor più velocemente per le realtà più in ritardo, mentre si consolidano i progetti delle aziende più mature. - spiega **Alessandro Piva, Responsabile della Ricerca dell'Osservatorio Big Data & Business Analytics** - *Ma il forte interesse per le soluzioni di analytics non corrisponde sempre a un cambio di rotta complessivo: sono ancora una minoranza le organizzazioni con una Data Strategy di livello corporate. Ora è necessario trasformare le organizzazioni nel profondo, creando ponti tecnologici, organizzativi e culturali tra le opportunità di analisi avanzate, le applicazioni intelligenti e le competenze e attività quotidiane dei lavoratori".*

https://www.osservatori.net/it/ricerche/comunicati-stampa/mercato-big-data-crescita

Data strategy

- Una "data strategy" o "strategia dei dati" è un piano strategico aziendale che definisce come un'organizzazione intende acquisire, gestire, utilizzare e sfruttare i dati per raggiungere i propri obiettivi aziendali.
- È un componente essenziale della gestione aziendale moderna, specialmente in un'epoca in cui i dati sono diventati un asset critico per molte organizzazioni.
- Considera i dati come una nuova “materia prima”

Alcuni elementi chiave di una data strategy:

- obiettivi aziendali
- acquisizione dei dati
- gestione dei dati
- analisi dei dati
- privacy e sicurezza dei dati
- governance dei dati

https://www.digital4.biz/executive/data-strategy-che-cose-e-come-costruirla-in-modo-agile/

Big data e PMI

Il rapporto tra PMI e italiane può essere analizzato in base a queste due variabili:

- percezione del fenomeno
- maturità tecnologica e di gestione dei dati

Risultati (2018):

- tradizionali (10%)
- in preparazione (31%)
- inconsapevoli o bloccate (42%)
- pronte (10%)
- lanciate (7%)

Maggiori ostacoli alla diffusione dei Big Data nelle PMI:

- difficoltà di stimare i benefici degli investimenti;
- mancanza di competenze adeguate, tanto scarse internamente quanto difficili da reperire all'esterno;

https://blog.osservatori.net/it_it/big-data-e-pmi-italiane

PMI: vantaggi dall'uso dei Big Data

- Miglioramento dell'efficienza operativa
- Miglioramento delle decisioni aziendali
- Creazione di nuovi prodotti/servizi

→ Benefici economici e operativi

→ Miglioramento della competitività

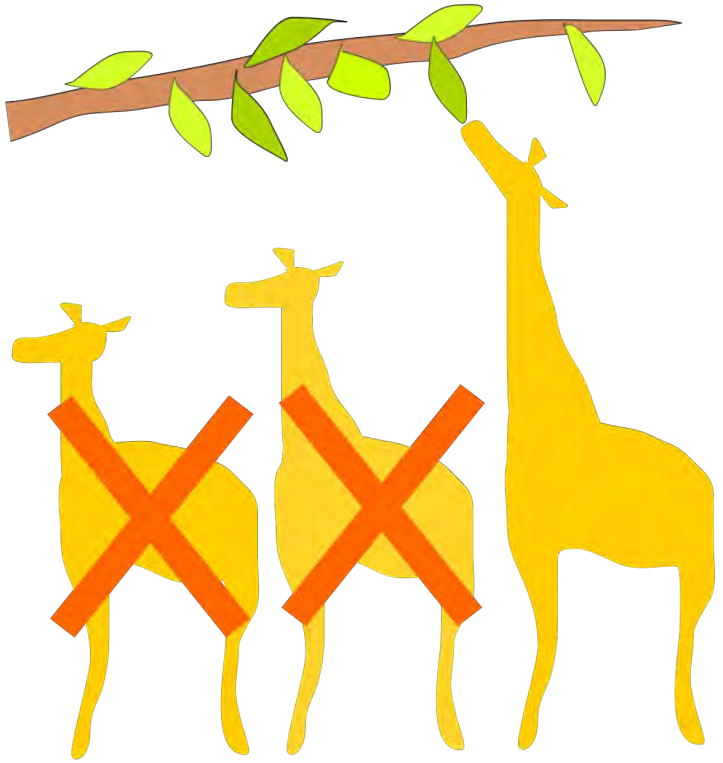


Image credit:
https://it.wikiversity.org/wiki/Le_teorie_evoluzionistiche_%28scuola_media%29#/media/File:Selection.svg

In conclusione: PMI e Big data… primi passi (1/2)

1. **Definire l'obiettivo principale**

- identificare chiaramente l'obiettivo o il problema che vuole risolvere utilizzando l'analisi dei dati.

- **obiettivo allineato agli obiettivi aziendali generali** (es. migliorare l'efficienza operativa, aumentare le vendite o ottimizzare la gestione delle risorse).

2. **Creare una strategia dei dati**

- Chief Data Officer (CDO) o consulenza esterna

- sviluppare una strategia chiara per la gestione dei dati. Ciò include:

  - la raccolta (individuare le fonti di dati - interne ed esterne)
  - l'archiviazione, l'elaborazione, l'analisi
  - tutela della privacy e della sicurezza dei dati

In conclusione: PMI e Big data… primi passi (2/2)

3. **Prendere decisioni data-driven ed implementarle in modo efficace**:

   - possibili modifiche nei processi aziendali o nelle operazioni quotidiane

   - possibile necessità di formazione del personale

   → superare ostacoli organizzativi e culturali

4. **Monitorare, valutare, migliorare**:

   - definire KPI efficaci per misurare i risultati e valutare l'efficacia delle decisioni prese attraverso attività di monitoraggio sistematiche

   - l'analisi dei dati dovrebbe essere un processo continuo, e la PMI dovrebbe essere disposta a regolare le sue strategie in base ai nuovi dati e alle nuove informazioni disponibili.

Conclusione

Argomenti trattati:

- **Introduzione ai problemi "Big Data" con esempi applicativi** ✓
  - **Caratteristiche di un problema big data ed esempi applicativi**
  - **Scalabilità verticale vs. scalabilità orizzontale**
  - **Map-reduce, architetture Big data, "sistemi operativi big data", figure professionali**
- **Big data in relazione a IA, IoT, Industrial IoT, edge computing** ✓
- **Casi d'uso dei Big data in alcuni comparti industriali** ✓
  - **settore agricolo**
  - **manifattura**
  - **logistica**
- **Mercato dei data Big data, *data strategy* e Big data nelle PMI** ✓

FUB
Fondazione Ugo Bordoni
Ricerca e Innovazione

Grazie per l'attenzione!

Domande?

Esempi da esperienze professionali… (3/3)

- **Programma strategico Smarter Italy prevede la definizione ed il lancio di gare d'appalto innovative. Ha lo scopo di accelerare la crescita del Paese e soddisfare le esigenze espresse dalle comunità, città e borghi, iniziando con quattro aree d'intervento:**
  - la **smart mobility** per permettere alle persone delle aree urbane di muoversi in modo più veloce, agile e sostenibile grazie all'uso della tecnologia e dell'innovazione;
  - la **valorizzazione dei beni culturali** (Cultural Heritage) per la valorizzazione economica e turistica delle aree di rilevanza storica e artistica;
  - il **benessere sociale e delle persone** (Wellbeing) per migliorare la salute psico-fisica delle persone;
  - la **salvaguardia dell'ambiente** per il miglioramento della situazione ambientale in tutti i suoi aspetti.
- **Il programma è aperto a tutte le Amministrazioni e soggetti pubblici interessati, che possono proporre fabbisogni di innovazione, co-finanziare il programma e mettere a disposizione campi operativi di sperimentazione.**

AGID Agenzia per l'Italia Digitale

Smarter Italy
Con l'innovazione si cresce

Cos'è Smarter Italy

È il programma del **Ministero delle Imprese e del Made in Italy**, del **Ministero dell'Università e della Ricerca** e del **Dipartimento per la Trasformazione Digitale**, attuato dall'**Agenzia per l'Italia Digitale**, che si basa sullo strumento degli appalti innovativi.

https://smarteritaly.agid.gov.it/