



*Sistemi informativi: averne fiducia e trarne valore*

**Rome Chapter**

## ***Big Data: Il Potere dei Dati nel Mondo Digitale***

Alessia Valentini

Cybersecurity manager, Advisor, Auditor, Giornalista.

Roma 12/12/2025

# Agenda

- Presentazione relatore
- Cosa sono i Big Data?
- Tecnologie e strumenti principali
- Applicazioni pratiche dei Big Data
- Normativa europea Big Data
- Sfide
- Il futuro dei Big Data
- Q&A



# Agenda

- ➔ • Presentazione relatore
- Cosa sono i Big Data?
- Tecnologie e strumenti principali
- Applicazioni pratiche dei Big Data
- Normativa europea Big Data
- Sfide
- Il futuro dei Big Data
- Q&A



# Presentazione relatore



- **Professionista di ICT e sicurezza informatica (Cybersecurity) da 29 anni (Cyber 18)**
- **Certificata come Auditor dei sistemi (CISA/ISACA 2017).**
- ***Membro SheLeadsTech del board Information Systems Audit and Control Association®***
- **Membro dello Strategic Advisory Board dell'Osservatorio Sulla Sicurezza Nazionale (OSSN),** afferente al “Centro di Ricerca in Analisi Dati, Intelligence, Sicurezza, Informazione” del Dipartimento di Management ed Economia della UNIPEGASO istituito a Febbraio 2025
- **Docente al Master Spazio di Tor Vergata sui temi della Cybersecurity per gli asset spaziali (2024-2023-2022)**
- **Giornalista pubblicista presso l'Ordine dei Giornalisti del Lazio (dal 2013) sui temi della Cybersecurity: Cybersecurity 360 sezione THE OUTLOOK e Il SOLE 24ore**
- **Digital Coach per la sensibilizzazione *alla cybersecurity* dedicata alle famiglie dal 2021 a oggi**
- **Cyberlady della community delle Women For Security del Clusit**
- **Consigliere direttivo in Armed Forces Electronic Association (AFCEA) (2014- 2016)**
- ***Membro dell'Osservatorio Sicurezza Nazionale presso CASD fino al 2013.***

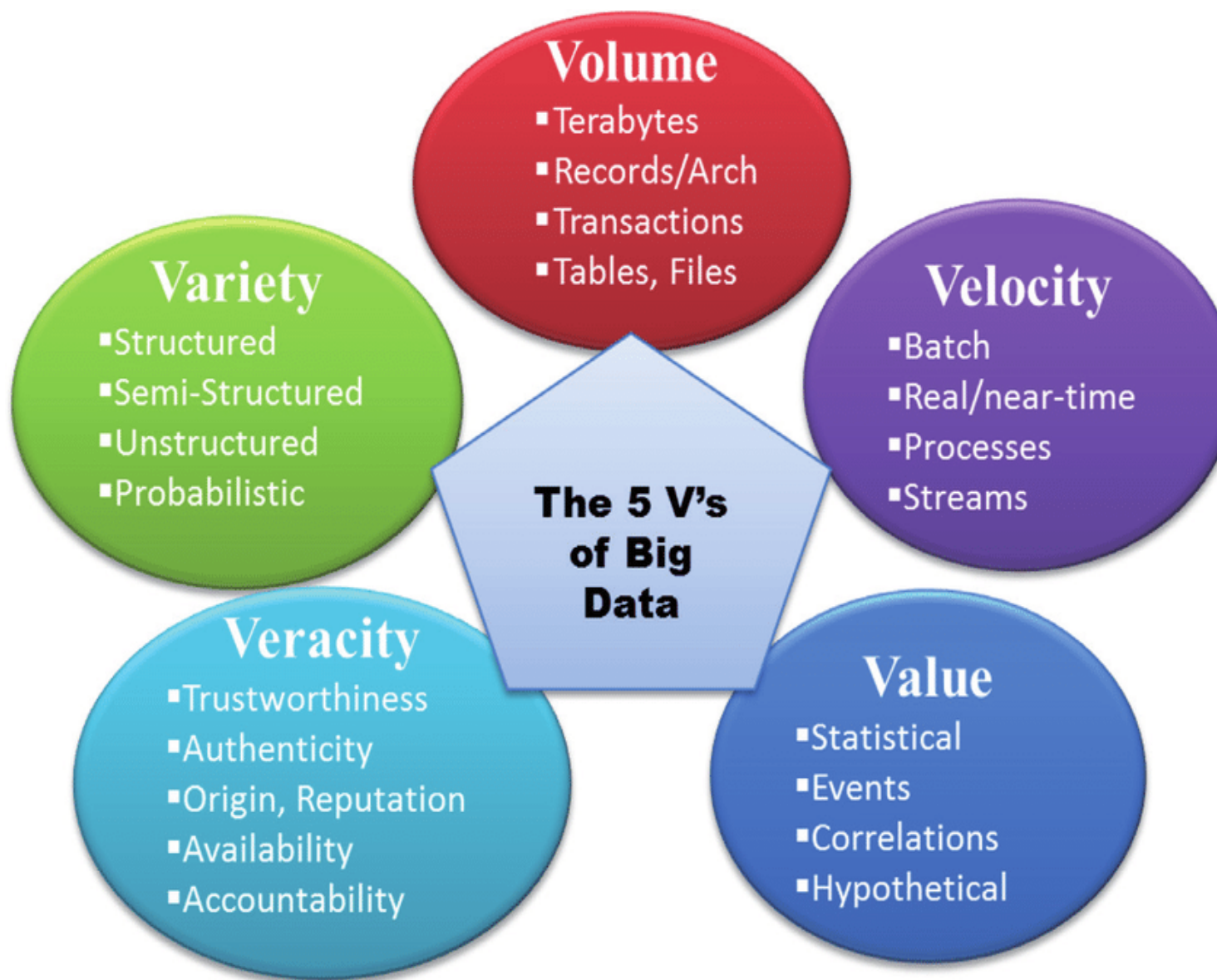
# Agenda

- Presentazione relatore
- ➔ • Cosa sono i Big Data?
- Tecnologie e strumenti principali
- Applicazioni pratiche dei Big Data
- Normativa europea Big Data
- Sfide
- Il futuro dei Big Data
- Q&A



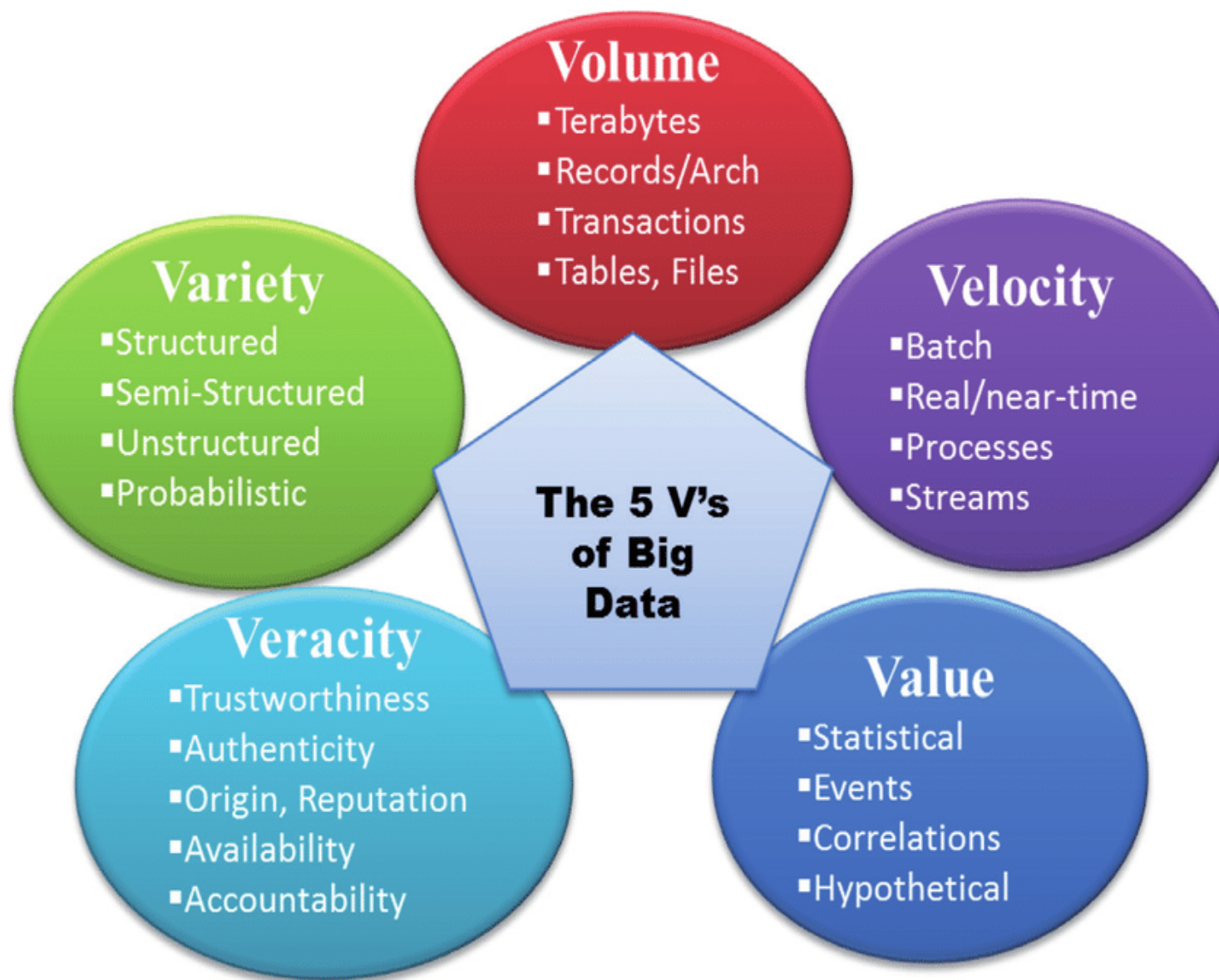
# Cosa sono i Big Data?

I Big Data rappresentano una raccolta di dati così estesa, ampia e diversificata di dati, strutturati e non, da richiedere tecnologie e metodi analitici specifici per l'analisi e l'estrazione di valore. (Fonte [Apogeo](#))



# Cosa sono i Big Data?

Solitamente i grandi volumi di dati si presentano in differenti formati (spesso privi di struttura) e con caratteristiche eterogenee; sono prodotti e diffusi generalmente con una elevata frequenza, e cambiano spesso nel tempo. Ma devono essere sempre affidabili.



# Cosa sono i Big Data?

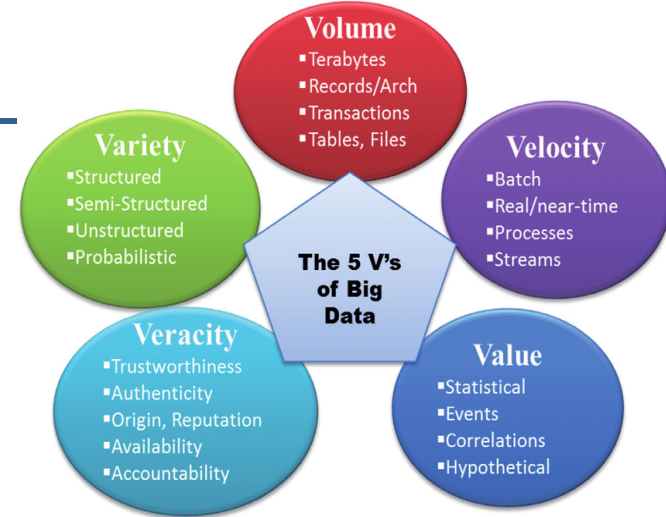
Per questo motivo i BIG DATA sono identificati con le 5V (+3) ovvero Volume, Velocità, Varietà, Veridicità, Valore a cui si può aggiungere: Viscosità, Visualizzazione, Viralità.



# Cosa sono i Big Data?

- **Volume:** caratteristica più immediata, dal momento che si tratta di **dati presenti in grandi quantità** generati in **continua crescita**, superando le capacità di gestione dei sistemi tradizionali.

- Es . Totalità dei dati prodotti da ogni social a cui aggiungere dati prodotti da tutte le altre “attività digitali». Insieme generano corpus di dati e informazioni a loro volta incrociabili, che allettano chi vuole sfruttarli per trarne valore o dedurre fatti, abitudini,



Le unità di misura, vanno oltre i Gigabyte (miliardi di Byte) che oggi è considerato piccolo per i Big Data):

Terabyte (TB): migliaia di GB (circa 1024 GB).

Petabyte (PB): migliaia di TB, l'ordine di grandezza in cui si inizia a parlare di Big Data.

Exabyte (EB): migliaia di PB (miliardi di GB).

Zettabyte (ZB): un miliardo di miliardi di byte (miliardi di GB o migliaia di EB).

Yottabyte (YB) dimensione inimmaginabile...

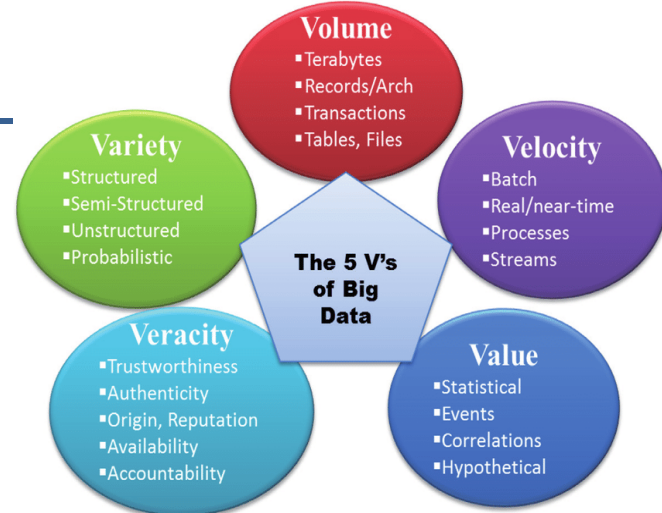
**Ma in generale dipende da cosa si conta...**

# Cosa sono i Big Data?

- **Velocità:** si riferisce all'elevata frequenza con cui i dati vengono generati, a volte anche in tempo reale (che si ripercuote sulla quantità (Volume)).

- Il secondo aspetto della velocità riguarda i tempi di accesso e analisi a questi dati grazie alle capacità di calcolo sempre crescenti.

Maggiore è la velocità di accesso ai dati e di loro analisi per sfruttamento e maggiore sarà la velocità in un processo decisionale e maggiore/migliore competitività sui diversi panorami del mercato.



# Cosa sono i Big Data?

## •Velocità:

La velocità può variare:

- Tipologia dati legacy prodotti nei **batch notturni**



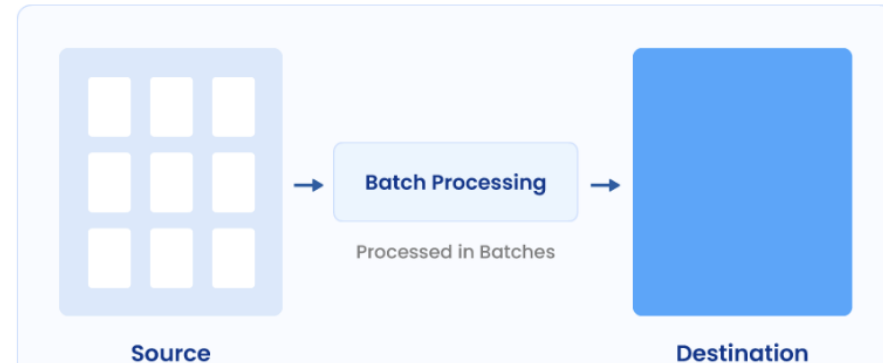
- Dati **prodotti in real-time** o near real-time



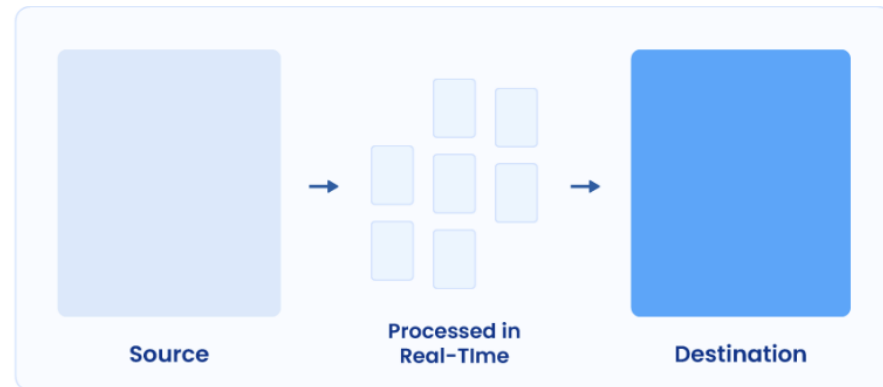
- Fino al flusso continuo in cui i dati vengono generati a un ritmo elevatissimo (es. milioni di ricerche al giorno, post sui social media 24/7).



## Batch Processing



## Stream Processing



# Cosa sono i Big Data?

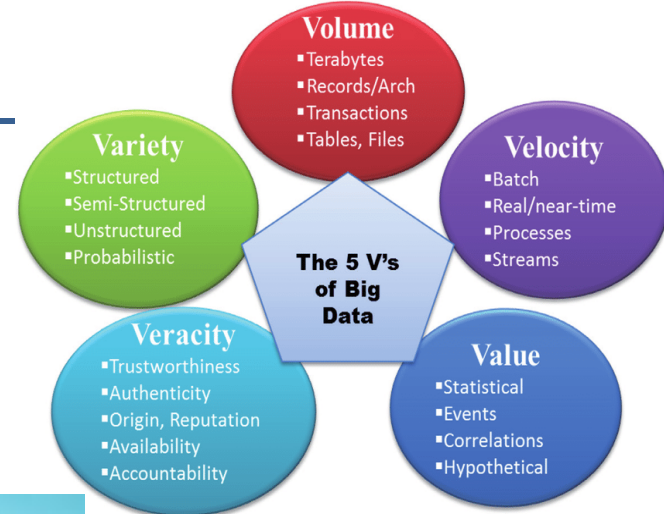
• **Varietà:** è la caratteristica che ha a che fare con la forma in cui i dati si presentano. Nel contesto Big Data le informazioni da trattare sono dati strutturati, non-strutturati (o semi-strutturati) ma anche probabilistici. Gli ultimi 3 non adatti ad essere lavorati con le tecniche tradizionali dei database relazionali.

**Strutturati** (come database, righe/colonne),

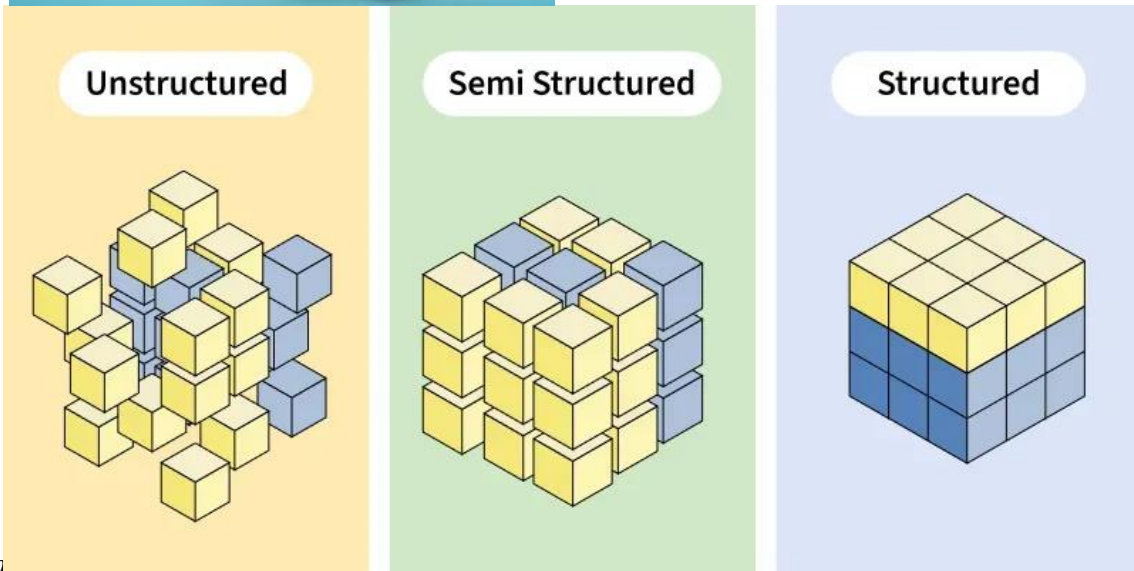
**Non strutturati** (testi liberi, audio, video, immagini),

**Semi-strutturati** (XML, JSON, CSV-comma separated Values)

**Dati probabilistici** (o fuzzy)



Probabilistico



# Cosa sono i Big Data?

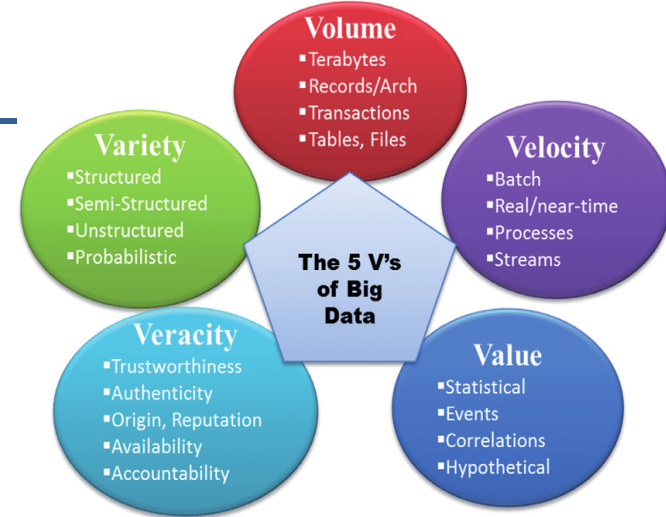
## •Varietà:

## Dati Strutturati

Definizione: Dati organizzati in un formato fisso e prevedibile, come tabelle in database relazionali (SQL).

Esempi: Dati transazionali, anagrafiche clienti, inventari.

Gestione: Facile da gestire con query tradizionali e modelli relazionali.



ANAGRAFICHE CLIENTI

Tabella: Anagrafica

| Anagrafica             | Note                           | Destinazioni Diverse                                                                                                                                                    | Rubrica Organico | Fatturazione PA | Documenti Fiscali | Automezzi del Cliente | Documenti PDF Cliente | Statistiche Fatturato |
|------------------------|--------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|-----------------|-------------------|-----------------------|-----------------------|-----------------------|
| Codice Cliente         | +1                             | 0011                                                                                                                                                                    | Conto Clienti    | Conto Vendita   |                   |                       |                       |                       |
| Rag.Sociale/ Cogn.Nome | Computer System snc            |                                                                                                                                                                         |                  |                 |                   |                       |                       |                       |
| Indirizzo              | Via del Fante 9a/b e c         |                                                                                                                                                                         |                  |                 |                   |                       |                       |                       |
| Cap                    | 37100                          | Città                                                                                                                                                                   | Verona           | Provincia       | VR                | Stato                 | IT                    |                       |
| Partita IVA            | 00966950230                    |                                                                                                                                                                         | Codice Fiscale   | 00966950230     |                   |                       |                       |                       |
| Telefono               | 045123456789                   |                                                                                                                                                                         | Fax              | 045123456777    |                   |                       |                       |                       |
| Cellulare              | 333123123213                   |                                                                                                                                                                         | E-Mail           | info@compsys.it |                   |                       |                       |                       |
| Tipo IVA               | Normale                        | Codice IVA / Esenzione                                                                                                                                                  | 22               | IVA 22 %        |                   |                       |                       |                       |
| Codice Pagamento       | 2                              | RIBA 30-60-90-120 FM DF                                                                                                                                                 |                  |                 |                   |                       |                       |                       |
| Sconto Ricambi         | 10 + 5                         | Sconto Manodopera                                                                                                                                                       | 20               | Listino Numero  | 1                 |                       |                       |                       |
| Banca Appoggio         | PROVA BANCA APPOGGIO           |                                                                                                                                                                         |                  |                 |                   |                       |                       |                       |
| IBAN                   | 123456789012345678901234567812 |                                                                                                                                                                         |                  |                 |                   |                       |                       |                       |
| Tipo Cliente           | Rent                           | Sesso                                                                                                                                                                   |                  | Data di Nascita | 30/06/1985        | Privacy               |                       |                       |
| Responsabile           | Rossi Paolo                    |                                                                                                                                                                         |                  |                 |                   |                       |                       |                       |
| Nazionalità            |                                | <input checked="" type="radio"/> Azienda <input type="radio"/> Privato <input type="radio"/> Intestatario <input type="radio"/> Rent <input type="radio"/> Utilizzatore |                  |                 |                   |                       |                       |                       |

Personalizzazioni

☐ Blocco Amministrativo ☐ Seconda Lingua Articoli

Avvisi e Comunicazioni

☒ Avviso Revisione ☒ SMS ☒ Email ☒ Posta

☒ Avviso Tagliando ☒ SMS ☒ Email

☒ Avviso Compleanno ☒ SMS ☐ Email

☒ Avviso Cambio Gomme ☒ SMS ☐ Email

Fatturazione Elettronica B2B

PEC Destinatario: compsps@pec.it

Codice univoco SdI: 0000000 Esigibilità IVA: ☒ Esigibilità Immediata

Elenco Clienti

| Anagrafica                     | Indirizzo                      | Città                      | P.IVA       | Telefono      | Cellulare    | Codice |
|--------------------------------|--------------------------------|----------------------------|-------------|---------------|--------------|--------|
| MAFFEI ROBERTO                 | Via Verdicchio                 | Milano                     | 0096695     | 045 1234567   | 3389600667   | 0000   |
| Anagrafica Dimostrativa        | Indirizzo Dimostrativo n: 1001 | Città Dimostrativa n: 1001 |             | 045 123456789 |              | 001    |
| Computer System snc            | Via del Fante 9a/b e c         | Verona                     | 00966950230 | 045123456789  | 333123123213 | 0011   |
| Anagrafica Dimostrativa n:1002 | Indirizzo Dimostrativo n: 1002 | Città Dimostrativa n: 1002 |             | 045 123456789 | 31313131313  | 002    |
| Anagrafica Dimostrativa n:1003 | Indirizzo Dimostrativo n: 1003 | Città Dimostrativa n: 1003 |             | 045 123456789 |              | 003    |
| Anagrafica Dimostrativa n:1004 | Indirizzo Dimostrativo n: 1004 | Città Dimostrativa n: 1004 |             | 045 123456789 |              | 004    |
| SSSSSSSSSSSSSS                 | Indirizzo Dimostrativo n: 1005 | Città Dimostrativa n: 1005 |             | 3389600776    | 3389600776   | 005    |
| Anagrafica Dimostrativa n:1008 | Indirizzo Dimostrativo n: 1008 | Città Dimostrativa n: 1008 | 00000000000 | 045 123456789 |              | 008    |

Ricerca per Anagrafica

Ricerca

Clienti Totali Trovati: 410

# Cosa sono i Big Data?

## •Varietà:

### **Dati Non Strutturati**

Definizione: Dati senza uno schema predefinito, che costituiscono la maggioranza dei dati organizzativi.

Esempi: Testi, email, post sui social media, video, audio, immagini, file PDF.

Gestione: Richiedono strumenti specifici (NLP, riconoscimento immagini/audio) e data lake per l'analisi.

## Unstructured Data



# Cosa sono i Big Data?

## •Varietà:

### Dati Semi-Strutturati

Definizione: Un mix tra i due (strutturati e non), con una certa organizzazione ma senza uno schema rigido, spesso usando tag (es. JSON, XML).

Esempi: File XML, JSON, dati provenienti da API.

```
<?xml version="1.0"?>
- <birds>
  - <owl id="1201">
    <species>Bubo bubo</species>
    <name>Eagle Owl</name>
    <region>Eurasia</region>
  </owl>
  - <owl id="1202">
    <species>Strix occidentalis</species>
    <name>Spotted Owl</name>
    <region>North America</region>
  </owl>
</birds>
```

```
-- Listing 1: Sample JSON Document
[
{
  "empid":1,
  "lastname":"Davis",
  "firstname":"Sara",
  "title":"CEO",
  "titleofcourtesy":"Ms.",
  "birthdate":"1968-12-08",
  "hiredate":"2013-05-01",
  "address":"7890 - 20th Ave. E., Apt. 2A",
  "city":"Seattle",
  "region":"WA",
  "postalcode":"10003",
  "country":"USA",
  "phone":"(206) 555-0101"
},
{
  "empid":2,
  "lastname":"Funk",
  "firstname":"Don",
  "title":"Vice President, Sales",
  "titleofcourtesy":"Dr.",
  "birthdate":"1972-02-19",
  "hiredate":"2013-08-14",
  "address":"9012 W. Capital Way",
  "city":"Tacoma",
  "region":"WA",
  "postalcode":"10001",
  "country":"USA",
  "phone":"(206) 555-0100",
}
```

# Cosa sono i Big Data?

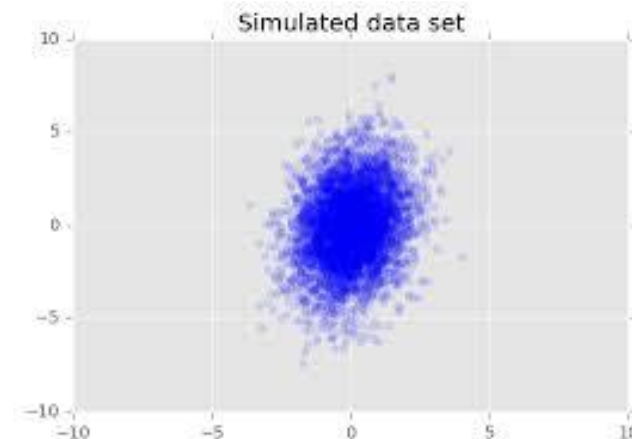
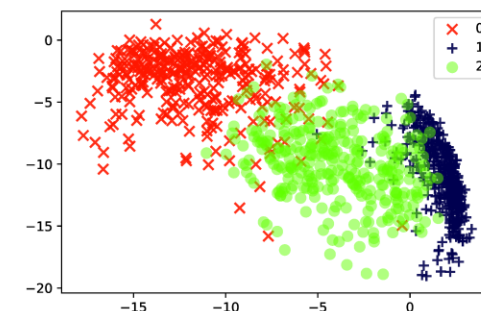
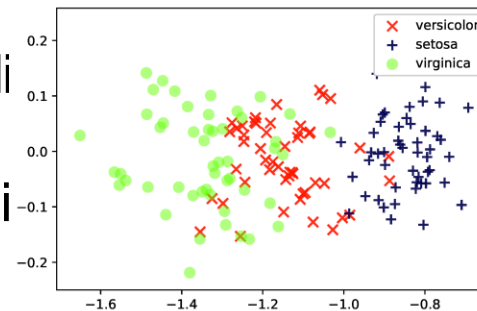
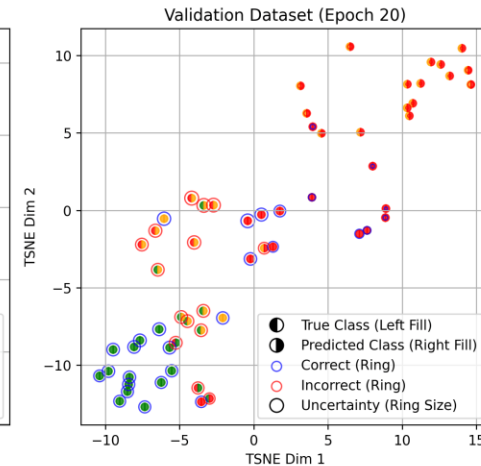
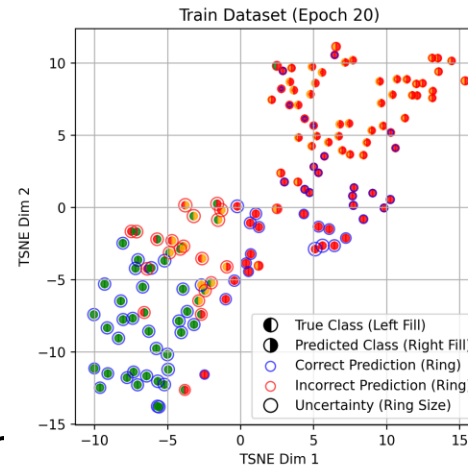
## •Varietà:

### Dati Probabilistici (o Fuzzy)

Definizione: Dati che non hanno un valore esatto ma una distribuzione di probabilità, utilizzati per gestire l'incertezza e l'ambiguità.

Esempi: Dati provenienti da sensori con margine di errore, i punteggi di rischio frode nelle transazioni, le probabilità di click nelle pubblicità, le previsioni del sentiment sui social media, i modelli di traffico per le auto autonome, o le probabilità di successo di una campagna marketing basate sul comportamento utente

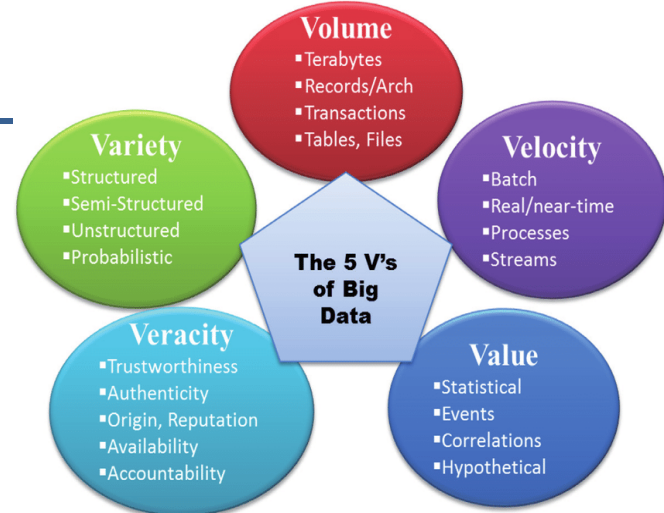
Gestione: Utilizzano algoritmi di machine learning, statistica e fuzzy logic per estrarre valore da informazioni imprecise.



# Cosa sono i Big Data?

- **Veridicità:** Riguarda l'affidabilità, la qualità, l'accuratezza, integrità, concretezza e contestualizzazione dei dati.

- È fondamentale garantire che le informazioni siano veritiere per poter fare analisi affidabili, corrette e utili.

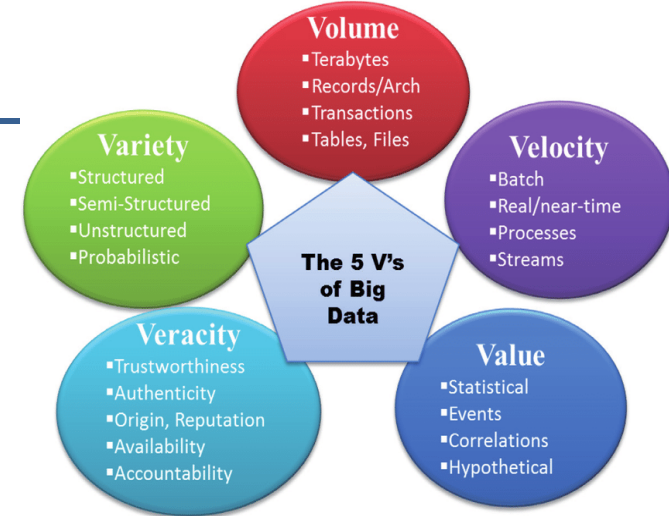


La sfida complessa è data la **grande varietà di fonti (es. web, social media, IoT)** e la **velocità con cui i dati cambiano**, rendendo fondamentale scartare quelli inaffidabili o sintetici o falsi, "rumorosi", incoerenti o rapidamente obsoleti,

I dati devono **essere Validati**.

# Cosa sono i Big Data?

• **Veridicità:** Riguarda l'affidabilità, la qualità, l'accuratezza, integrità, concretezza e contestualizzazione dei dati.



**Come si gestisce**

**Filtraggio:** Scartare i dati non validi a causa della loro origine, formato o contesto di raccolta.

**Controllo qualità:** Monitorare costantemente accuratezza, provenienza e coerenza.

**Contestualizzazione:** Comprendere che il significato dei dati può cambiare rapidamente, influenzando l'affidabilità.



# Cosa sono i Big Data?

• **Valore:** I Big Data nascondono un grande valore da comprendere e gestire per estrarre il potenziale informativo.

• L'obiettivo finale è **estrarre informazioni utili e significative dai dati** per ottenere un vantaggio competitivo, migliorare le decisioni.

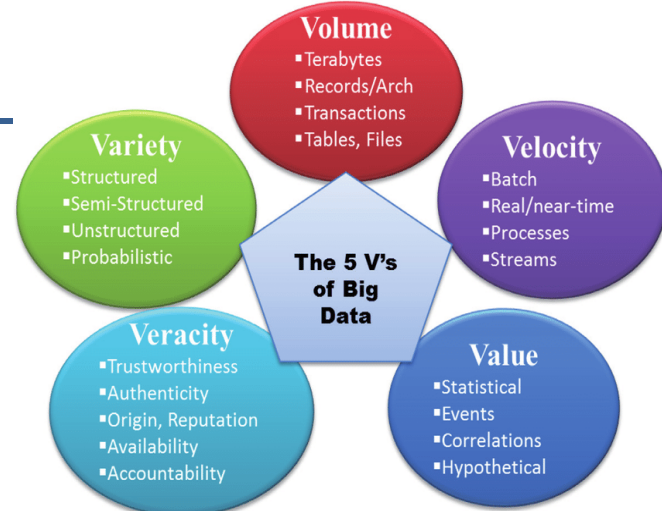
BIG DATA



to



BIG VALUE



Il valore risiede nella **capacità di trasformare enormi volumi di dati grezzi in insight strategici** attraverso l'analisi avanzata (AI, ML), permettendo alle aziende di migliorare decisioni, innovare prodotti/servizi, ottimizzare operazioni, personalizzare l'esperienza cliente e ottenere un vantaggio competitivo, agendo come "il nuovo petrolio" per generare ricavi e ridurre costi. Questo valore si manifesta in maggiore efficienza, migliore gestione del rischio e maggiore agilità in mercati in rapida evoluzione.

# Cosa sono i Big Data?

...altre 3 V

## Viscosità

Definisce la resistenza che si affronta nel **trasformare i dati in insight, ovvero la difficoltà di estrarre valore e muovere i dati**; indica che i dati possono essere "pesanti", difficili da elaborare, trasformare e integrare, proprio come un fluido viscoso che scorre lentamente, richiedendo infrastrutture potenti e processi complessi per ottenere insight e agire rapidamente, ostacolando l'agilità aziendale.



Esempi sono **log di server, dati IoT grezzi, feed di social media**. Richiedono una pulizia, trasformazione e preparazione intensa prima di poter essere analizzati, rallentando i processi e aumentando i costi. Le moderne architetture di Big Data cercano di risolvere con tecnologie di streaming e AI per la qualità dei dati.

# Cosa sono i Big Data?

## Visualizzazione

La rappresentazione grafica di enormi e complessi set di dati attraverso grafici, mappe e diagrammi per renderli comprensibili, identificare pattern, trend e anomalie che sarebbero invisibili in formato grezzo.

L'aspetto visuale è rilevante per prendere decisioni o per stimolarle. **Gli insights devono essere rappresentabili efficacemente** per poter acquisire valore per il business.

## Virilità

Misura la rapidità e ampiezza della propagazione (come una reazione virale) e la capacità di analisi di identificare contenuti che si diffondono rapidamente online, come un hashtag. Quindi misura la **capacità e la velocità dell'insights di diffondersi**.



# Agenda

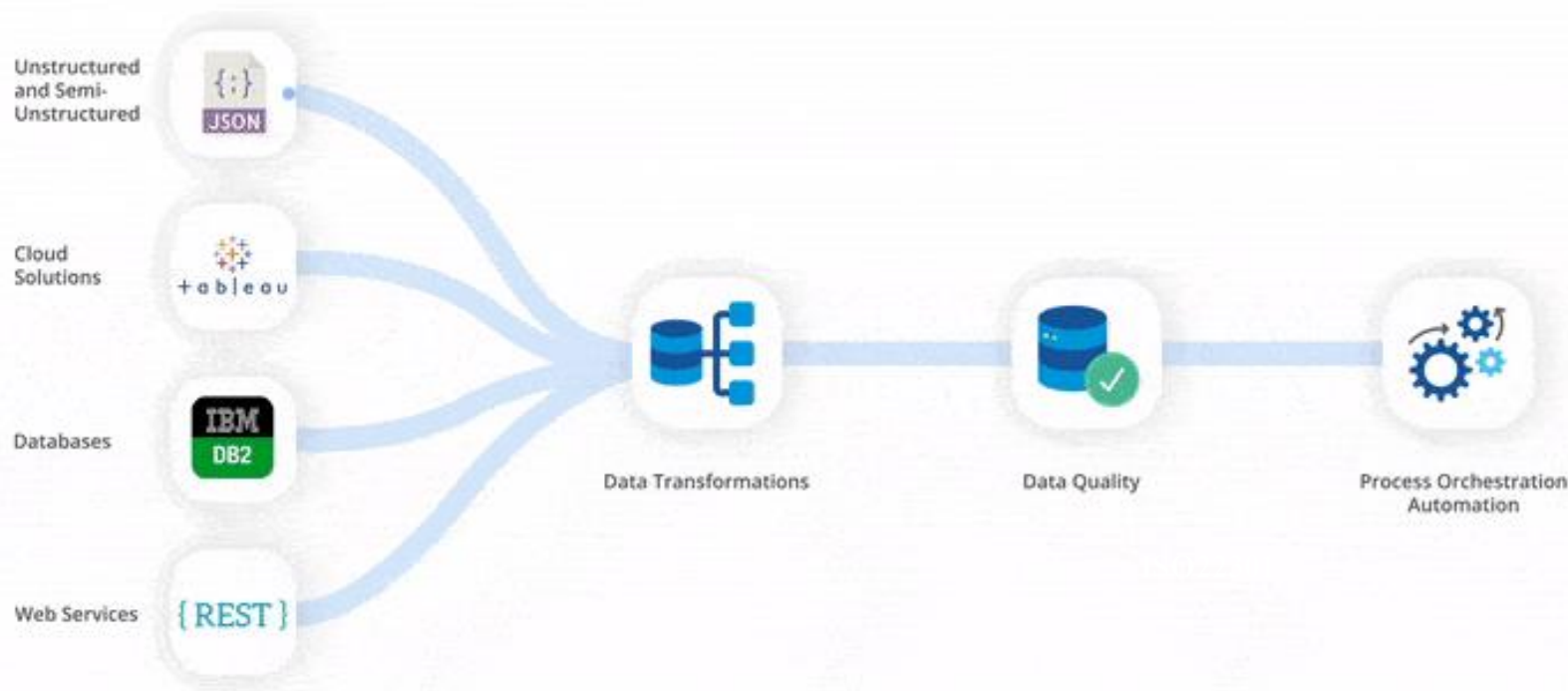
- Presentazione relatore
- Cosa sono i Big Data?
- ➔ • Tecnologie e strumenti principali
- Applicazioni pratiche dei Big Data
- Normativa europea Big Data
- Sfide
- Il futuro dei Big Data
- Q&A



# Tecnologie e strumenti principali- Automazione

Secondo l'International Data Center (IDC), la sfera globale dei dati crescerà fino a **163 zettabyte nel 2025**. Equivale a 163 trilioni di gigabyte, 10 volte di più rispetto della quantità registrata 10 anni fa.

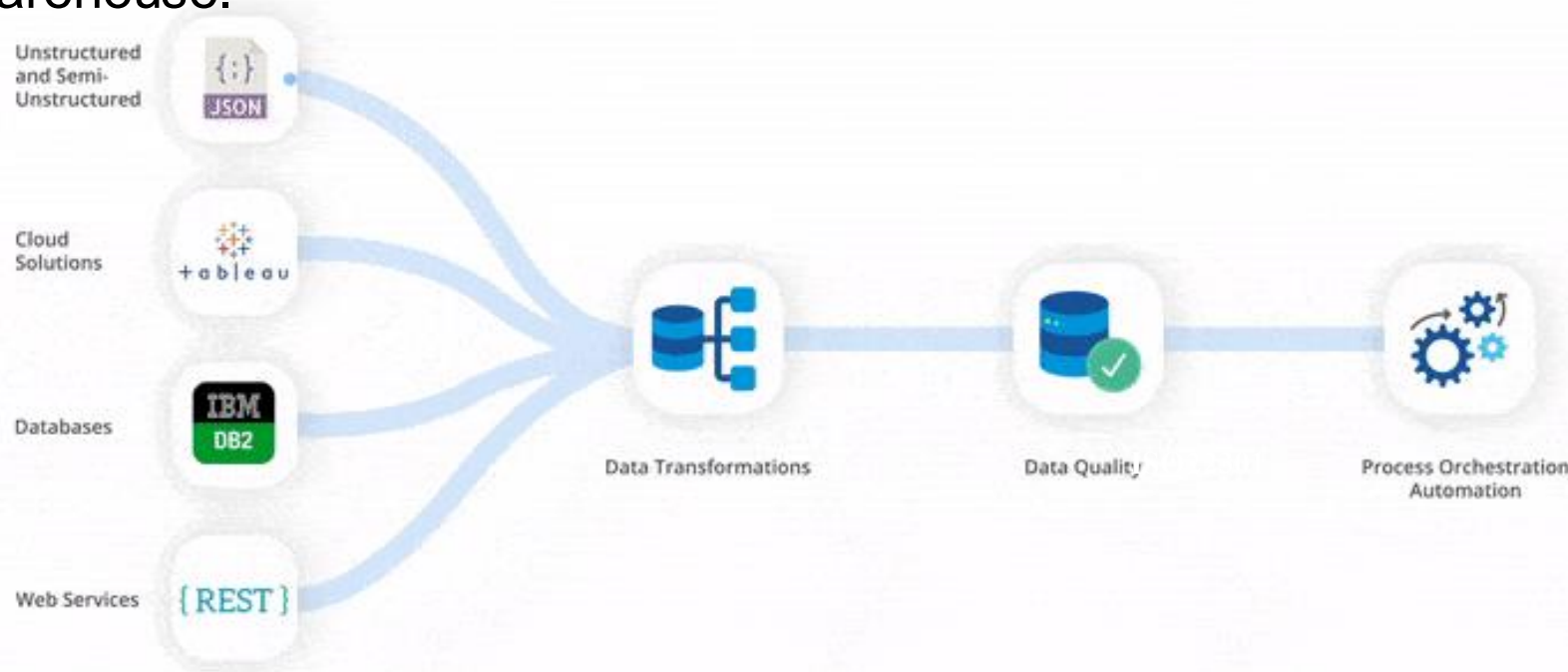
L'automazione dei dati è cruciale per la sostenibilità aziendale, soprattutto considerando la crescita esponenziale del volume di dati.



# Tecnologie e strumenti principali- Automazione

L'automazione è il processo di caricamento, gestione ed elaborazione dei dati tramite strumenti di elaborazione dati automatizzati anziché eseguire manualmente tutte queste attività. Coinvolge tre azioni : Estrai, Trasforma, and Carico (Extract, Transform & Load - ETL):

- **Estrazione** dati provenienti da una o più fonti.
- **Trasformazione** nel formato richiesto dal sistema di destinazione tramite applicazione di trasformazioni, come ordinamento, filtro, ecc.
- **Caricamento** nel sistema di destinazione, come un database o data warehouse.



# Tecnologie e strumenti principali

Per gestire BIG data l'ETL si trasforma.  
Non bastano i Data Warehouse.

## Differenze chiave con l'ETL tradizionale

**Scala:** Gestisce volumi (Volume), velocità (Velocity) e varietà (Variety) dei Big Data, che vanno oltre i Data Warehouse tradizionali.

**Strumenti:** Utilizza piattaforme cloud (AWS, Google Cloud, Azure) e strumenti distribuiti come Hadoop/Spark, spesso con interfacce user-friendly e drag-and-drop.

**ELT:** Diventa comune caricare i dati grezzi direttamente nel Data Lake (ELT) e trasformarli dopo (in-place) grazie alla potenza di calcolo di questi sistemi, offrendo maggiore flessibilità.



# Tecnologie e strumenti principali

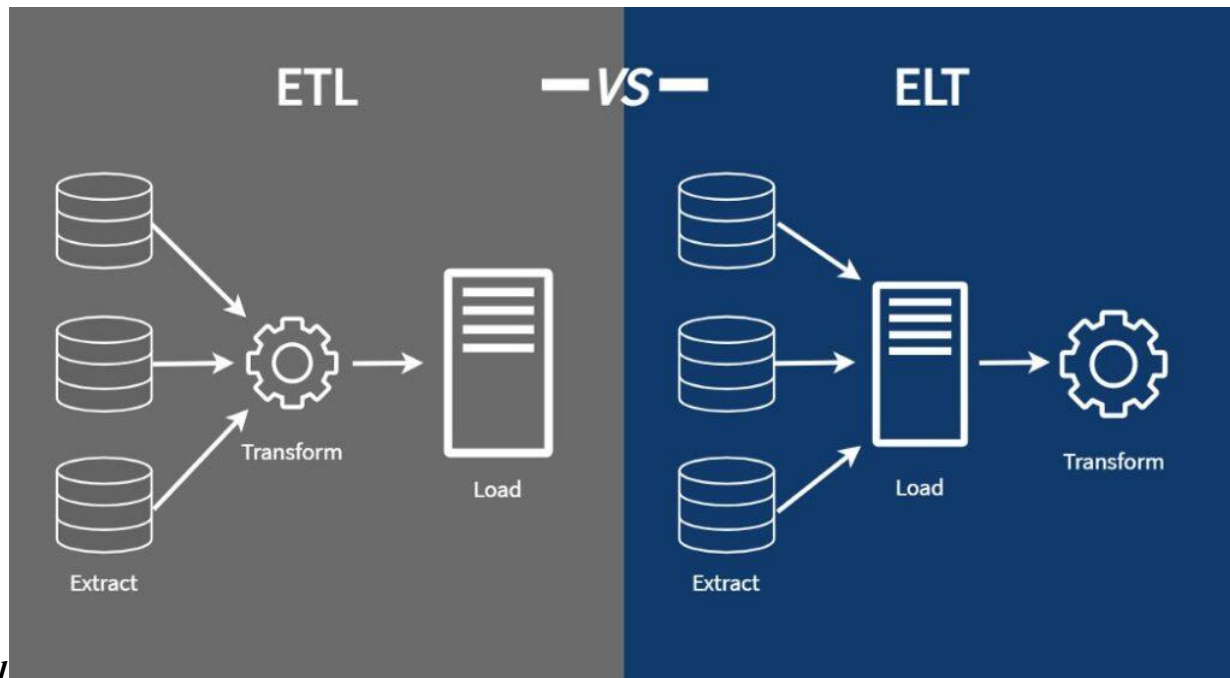
## Nell' Evoluzione per i Big Data:

**Estrazione (Extract):** Si raccolgono dati da molteplici fonti (database, API, file, IoT, social media) sfruttando capacità distribuite per gestire il volume.

**Caricamento (Load):** I dati trasformati vengono caricati in sistemi di destinazione (Data Warehouse, Data Lake, Data Lakehouse) per l'analisi, il reporting e il ML.

**Trasformazione (Transform):** I dati vengono puliti, standardizzati, arricchiti, aggregati e conformati alle regole aziendali, spesso in modo massivamente parallelo.

ETL (Extract, Transform, Load) pre-elabora i dati prima di caricarli, mentre ELT (Extract, Load, Transform) carica i dati grezzi e li trasforma nel sistema di destinazione rendendolo ideale per i Big Data per la sua scalabilità, velocità e flessibilità, sfruttando la potenza di calcolo del cloud per trasformazioni in parallelo su volumi enormi di dati strutturati e non strutturati,



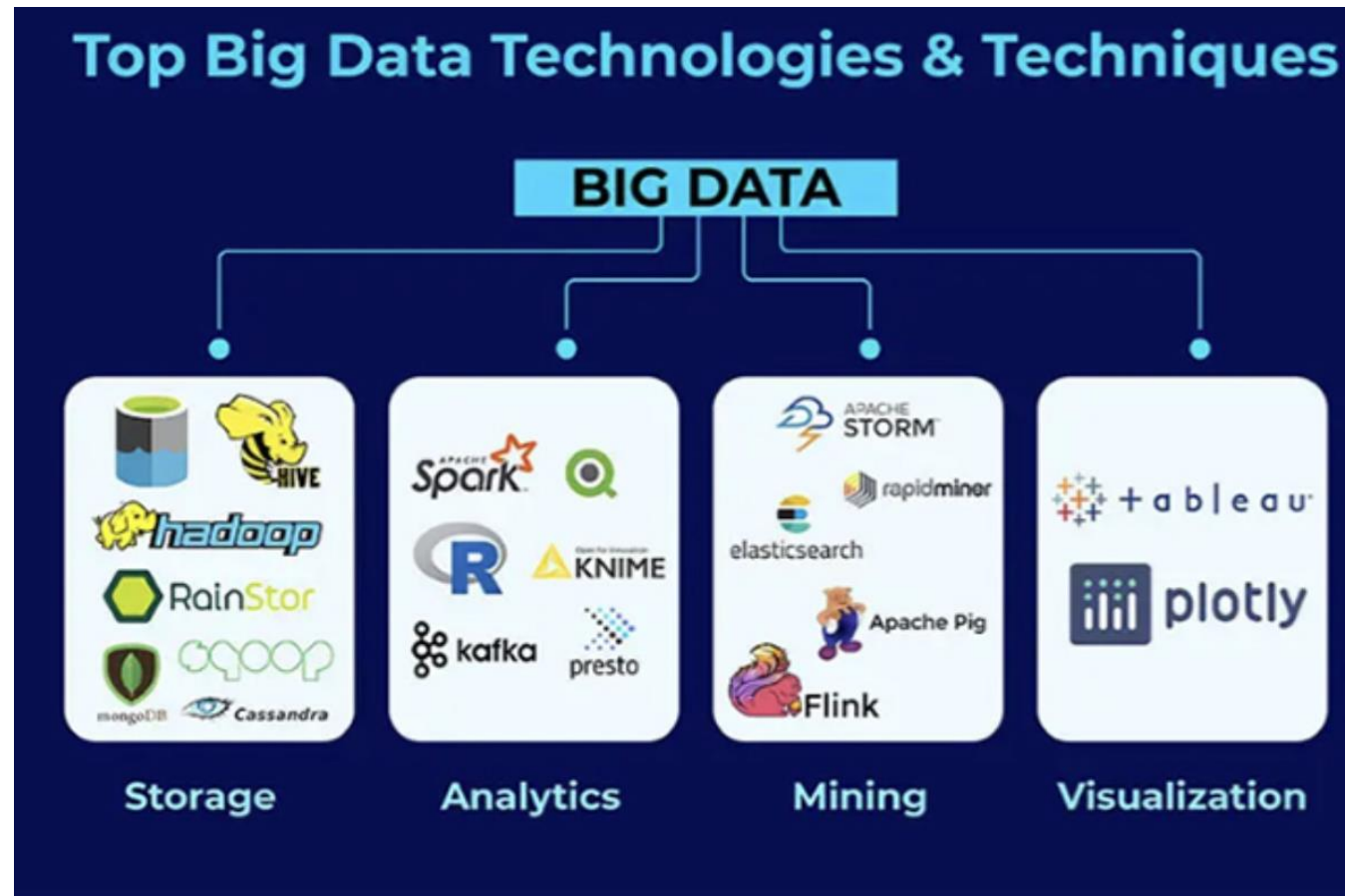
# Tecnologie e strumenti principali

Le principali tecnologie per i big data includono 4 capacità: **memorizzazione, analisi, estrazione di significati e visualizzazione** ed altrettanti rispettivi framework. Es.

**Hadoop Apache Spark DB NoSQL** Ma ne esistono anche altri.

Per creare Data lake su Cloud sono utilizzabili le funzioni sui dati di **AWS, Azure e Google Cloud e Oracle Cloud**.

**Quindi il paradigma per i big data diventa**



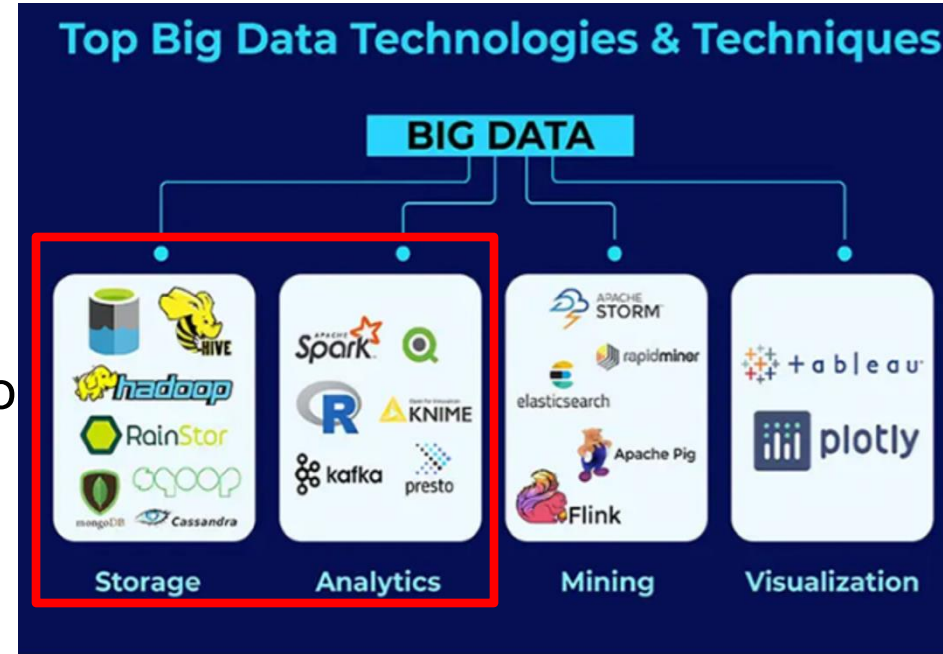
*Big Data technology is divided into four main types – Image: analytixlabs.com*

**Vedremo le tecnologie principali e più note**

# Tecnologie e strumenti

**Hadoop** è un framework open-source progettato per l'**archiviazione e l'elaborazione** distribuita di grandi set di dati su cluster di commodity hardware. In Hadoop tipicamente lo storage è gestito da **HDFS (Hadoop Distributed File System)** ed è abbinato al motore MapReduce per l'elaborazione e di fatto è il modello di programmazione utilizzato da Hadoop (nel deployment classico) per suddividere un problema di elaborazione in due fasi principali: map (mappa) + reduce (riduci) di coppie chiave/valore

**Quando usarlo:** scenari di grandi dati batch, archiviazione massiva, elaborazioni non real-time.

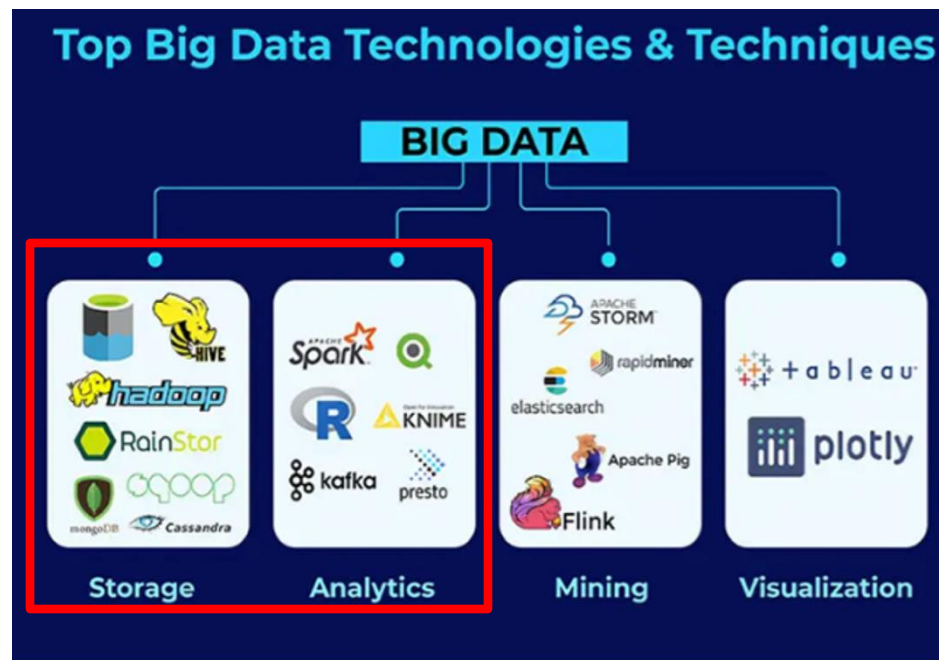


*Big Data technology is divided into four main types – Image: analytixlabs.com*

# Tecnologie e strumenti

**Apache Spark** è un motore per elaborazione distribuita di dati, che supporta carichi batch, streaming, SQL/BI, machine learning, grafi. E' open-source e creato per superare alcuni limiti di Hadoop MapReduce (Elaborazione in memoria che riduce latenza (in-memory), più carichi di lavoro e maggiore velocità ). Rispetto a Hadoop/MapReduce è migliore per scenari interattivi, iterative, real-time.

**Quando usarlo:** analytics in tempo quasi reale, machine learning a scala, elaborazioni complesse su big data.



*Big Data technology is divided into four main types – Image: analytixlabs.com*

**Splunk** è un altro popolare strumento di analisi dei big data per ricavare informazioni da grandi set di dati. È in grado di generare grafici, diagrammi, report e dashboard.

Splunk consente inoltre agli utenti di integrare l'intelligenza artificiale (IA) nei risultati dei dati.

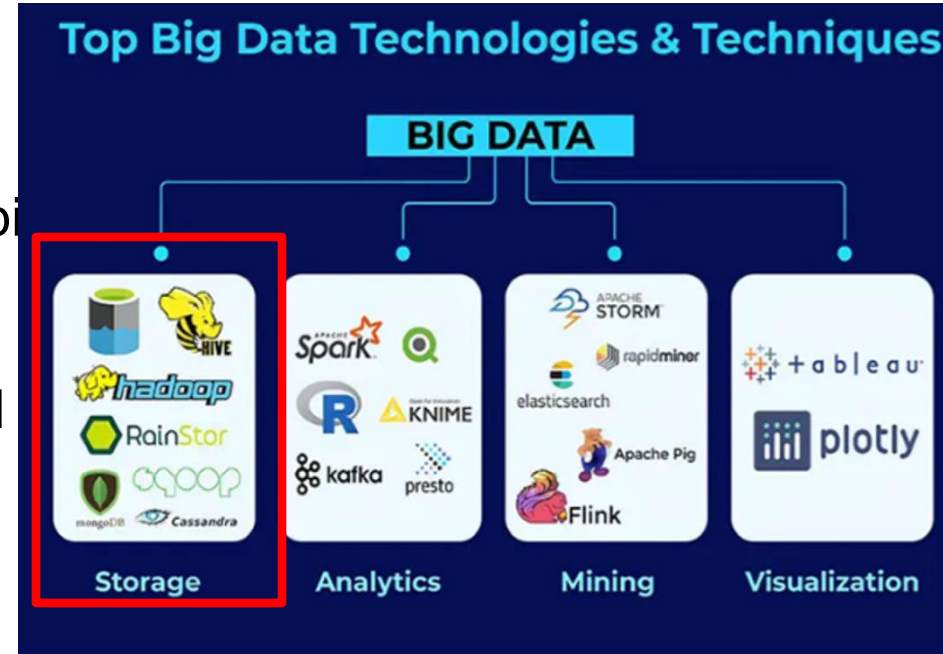
# Tecnologie e strumenti

“NoSQL” si riferisce a **database non relazionali**, che permettono modelli di dati flessibili, scalabilità orizzontale, distribuzione su cluster. Ne sono esempi es. **MongoDB, Cassandra**)

**MongoDB** è un database NoSQL che memorizza dati in formato JSON/BSON

**Punti chiave:** modello di schema flessibile (“schemaless” o con schema evolutivo) in cui i documenti di una collezione possono avere strutture differenti. Supporta query ad hoc, indicizzazione, aggregazioni, replica, Sharding orizzontale (consente di scalare su cluster distribuiti).

**Quando usarlo:** Adatto per applicazioni web, mobile, per memorizzare dati con struttura variabile.



*Big Data technology is divided into four main types – Image: analytixlabs.com*

# Tecnologie e strumenti

**Cassandra** è un database NoSQL open-source progettato per gestire grandi quantità di dati distribuiti su più data center con alta disponibilità

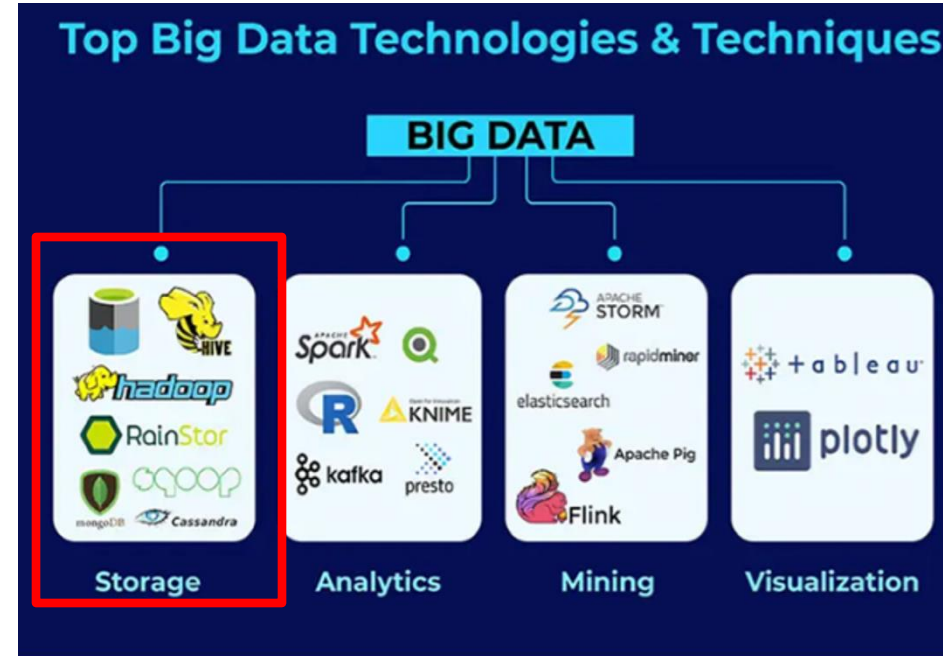
## Punti chiave:

Architettura master-less (ogni nodo è eguale) → nessun singolo punto di fallimento.

Replicazione multi-datacenter e tolleranza ai guasti.

Scalabilità lineare: l'aggiunta di nuovi nodi incrementa le prestazioni sia in lettura che in scrittura.

**Quando usarlo:** scenari con requisiti stringenti di disponibilità, latenza bassa, scritture elevate, distribuzione geografica.



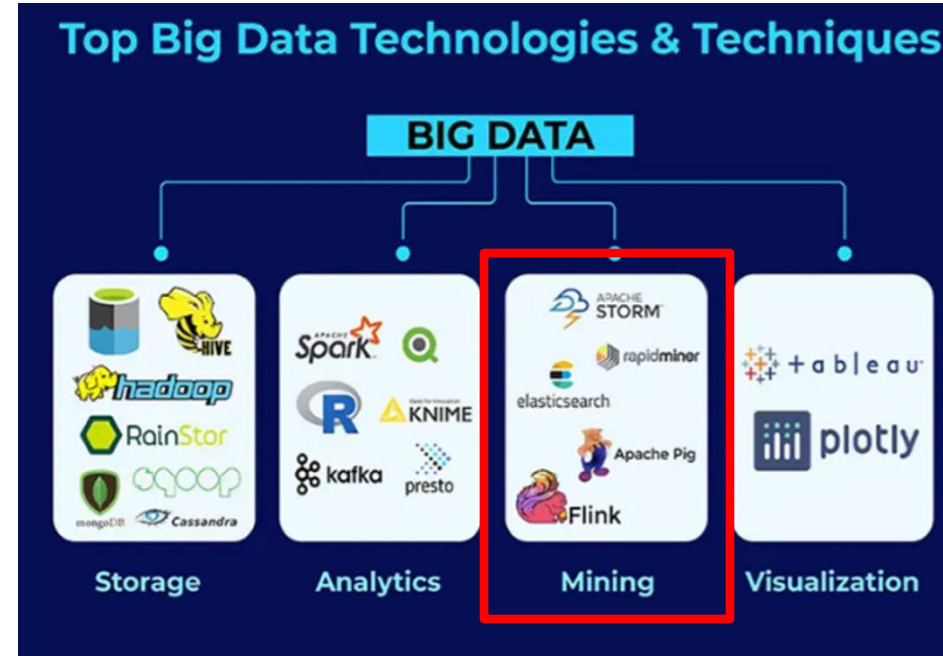
*Big Data technology is divided into four main types – Image: analytixlabs.com*

# Tecnologie e strumenti

Il data mining estrae modelli e tendenze utili dai dati grezzi così da trasformare dati strutturati e non strutturati in informazioni utilizzabili.

**RapidMiner** è uno strumento di data mining che consente di creare modelli predittivi. Si basa su elaborazione e preparazione dei dati e creazione di modelli di apprendimento automatico e deep learning. Il modello end-to-end consente di utilizzare analisi avanzate e automazione dell'intelligenza artificiale.

**Presto:** Presto è un motore di query open source originariamente sviluppato da Facebook per eseguire query analitiche sui suoi ampi set di dati. Ora è ampiamente disponibile. Una query su Presto può combinare dati provenienti da più fonti all'interno di un'organizzazione ed eseguire analisi su di essi in pochi minuti.



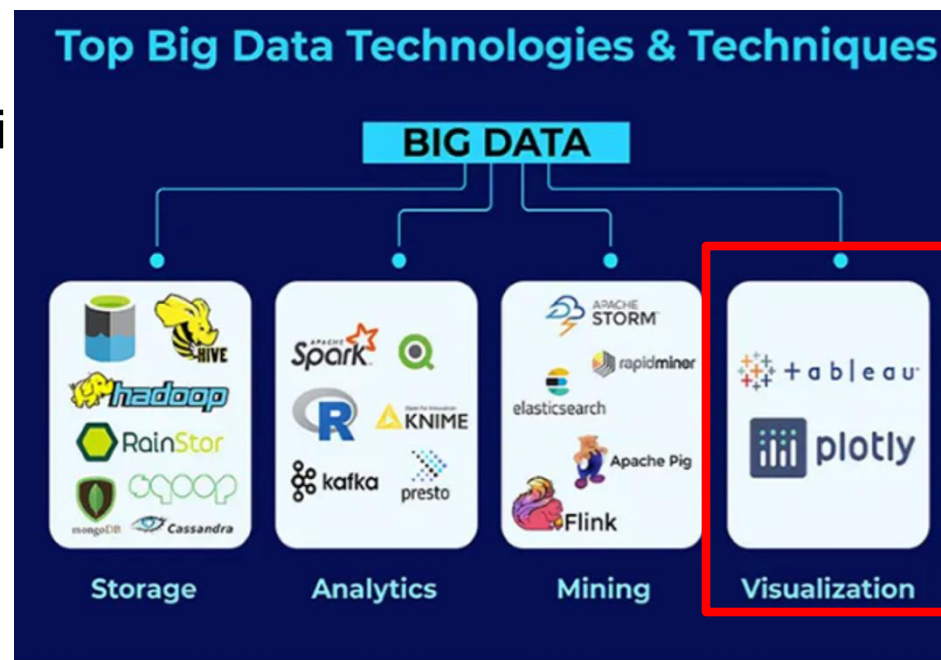
*Big Data technology is divided into four main types – Image: analytixlabs.com*

# Tecnologie e strumenti

La **visualizzazione dei dati** è una **competenza utile per presentazioni agli stakeholder** (redditività, operazioni aziendali), raccontando una storia d'impatto sui dati con un grafico.

**Tableau:** molto popolare nella visualizzazione dei dati: la sua interfaccia drag-and-drop semplifica la creazione di grafici a torta, grafici a barre, box plot, diagrammi di Gantt. È una piattaforma sicura anche per condividere visualizzazioni e dashboard in tempo reale.

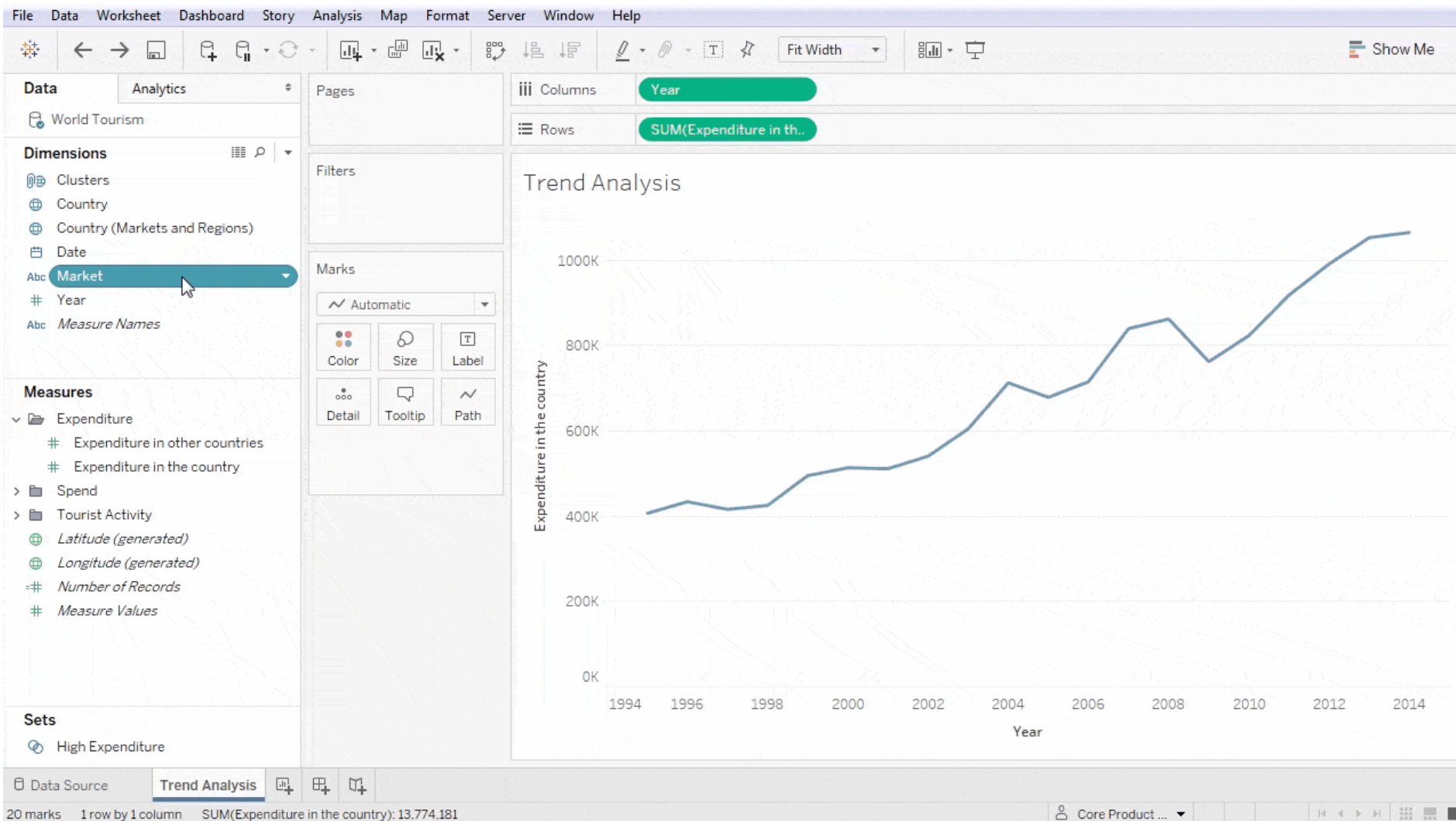
**Looker** è uno strumento di business intelligence (BI) utilizzato per comprendere l'analisi dei Big Data e condividere informazioni con altri team. Con una query, è possibile configurare grafici, diagrammi, dashboard e altre visualizzazioni di dati, come il monitoraggio settimanale del brand engagement tramite l'analisi dei social media.



*Big Data technology is divided into four main types – Image: analytixlabs.com*

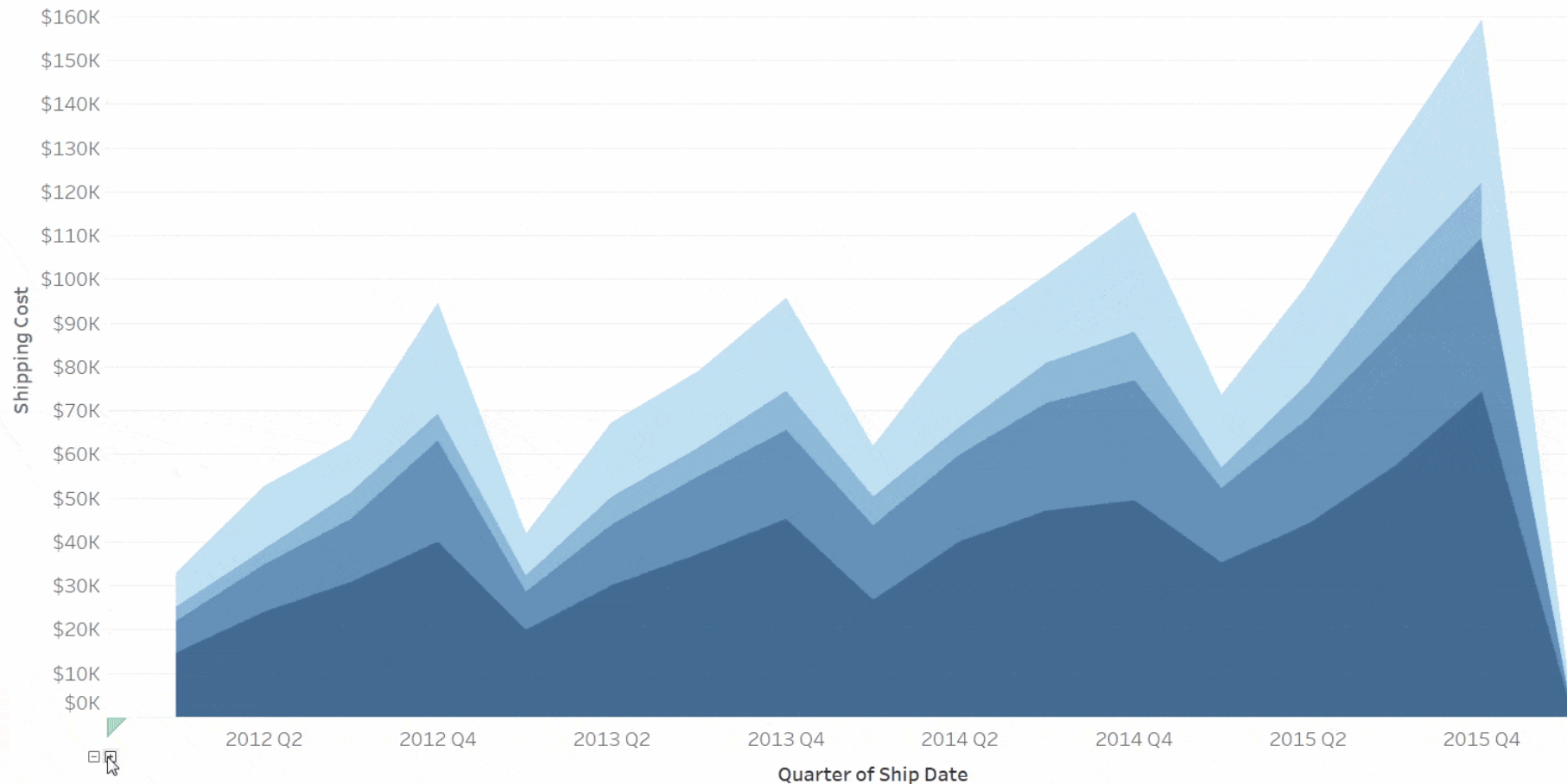
# Tecnologie e strumenti

Esempi con **Tableau**. La galleria pubblica di Tableau contiene moltissime visualizzazioni realizzate con lo strumento gratuito Tableau Public, come anche alcune dashboard aziendali da usare come modelli;



# Tecnologie e strumenti

**Tableau esempio** La galleria pubblica di Tableau contiene moltissime di visualizzazioni realizzate con lo strumento gratuito Tableau Public, come alcune dashboard aziendali da usare come modelli;



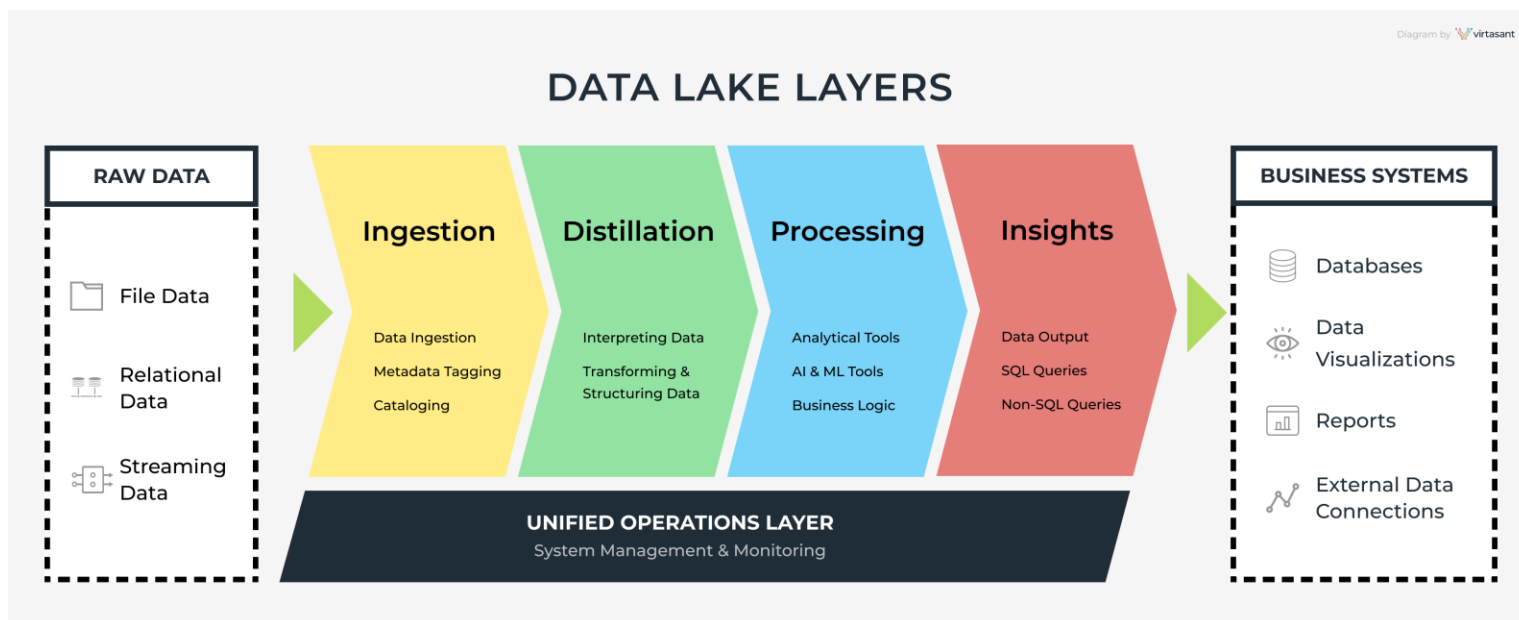
# Tecnologie e strumenti principali

Un **data lake** è un repository centralizzato che consente di immagazzinare dati grezzi (raw), rigidi (tabellari), semi-strutturati (CSV, JSON, log), non strutturati (documenti, immagini, audio/video) -formato file o oggetti -senza la necessità di uno schema fisso definito in anticipo.

Può essere on-premises oppure nel cloud ed è utilizzato per analytics, BI, machine learning, data exploration — poiché conserva dati “raw” che possono essere trasformati secondo necessità.

**Vantaggi:** flessibilità, costi potenzialmente ridotti (storage di oggetti), supporto a molteplici casi d'uso.

**Sfide:** governance, qualità dei dati, sicurezza, catalogazione/metadati, gestione accessi.



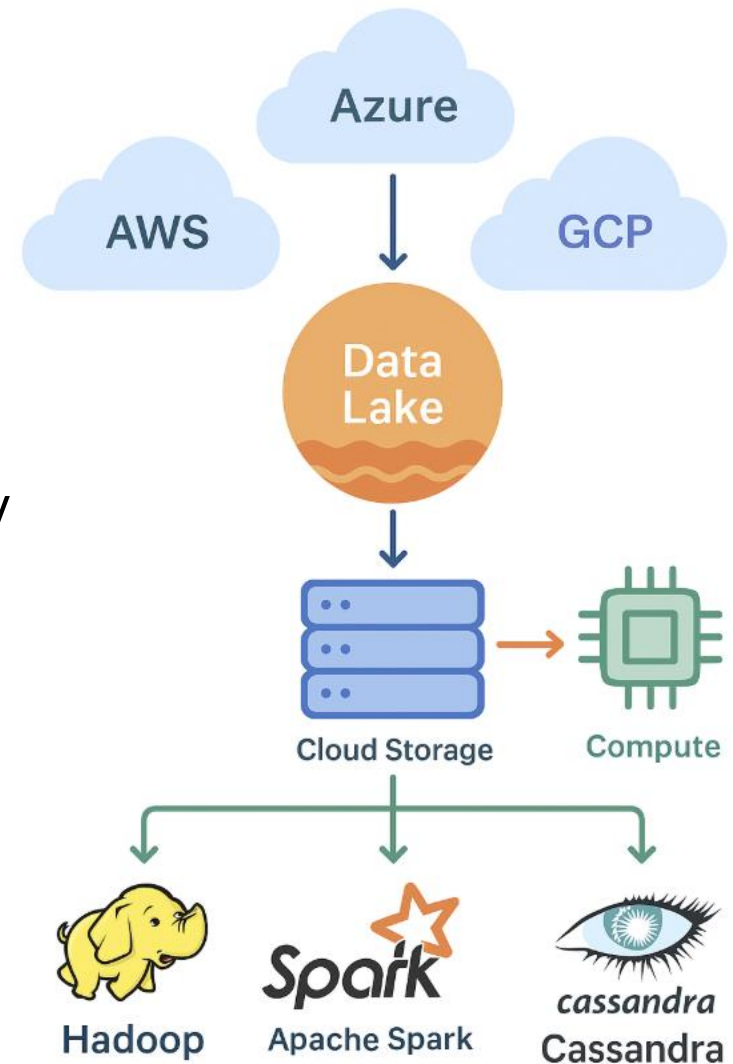
# Tecnologie e strumenti principali

Cloud provider per Data Lake:

## AWS (Amazon Web Services)

Amazon S3 (storage oggetti per data lake), Amazon Elastic MapReduce-EMR (cluster Hadoop/Spark), Amazon Redshift (data warehouse), AWS Glue (Extract, Transform, Load -ETL serverless), Amazon Athena (query SQL su S3)

**Vantaggi:** ampia maturità, ecosistema ricco, integrazione con analytics/ML, opzioni serverless.

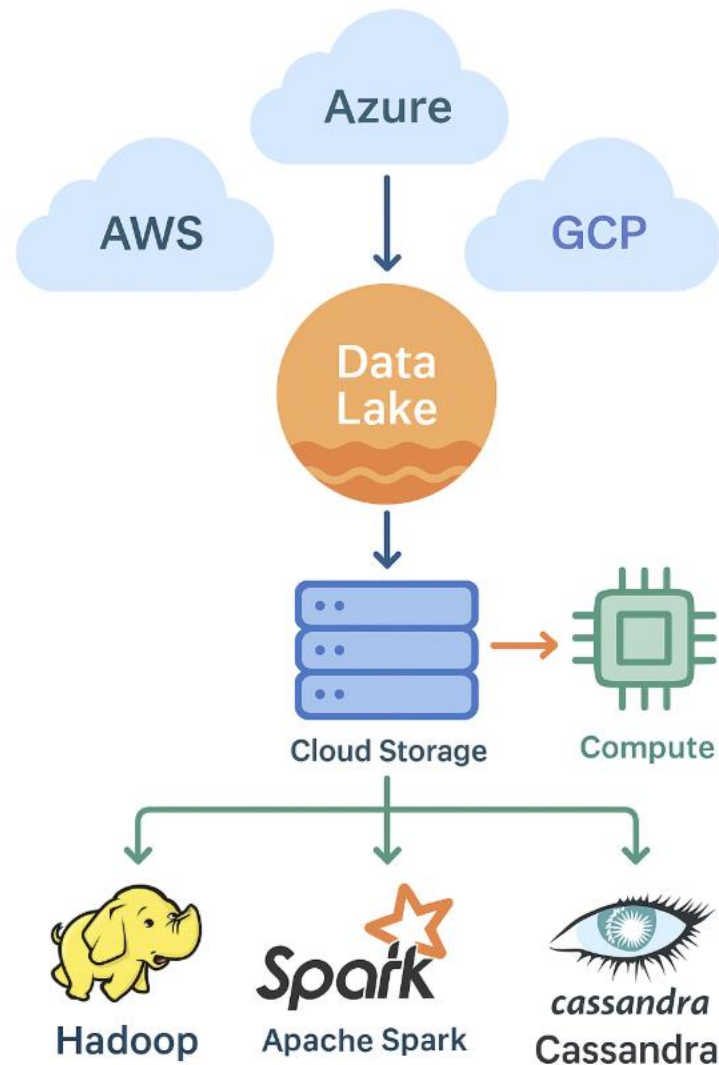


# Tecnologie e strumenti principali

## Microsoft Azure

Azure Data Lake Storage (storage scalabile per data lake) Wikipedia, Azure Synapse Analytics, Azure Databricks, Azure HDInsight, Azure Data Factory

**Vantaggi:** forte integrazione con Microsoft stack, BI (Power BI), capacità ibrida on-premises/cloud.

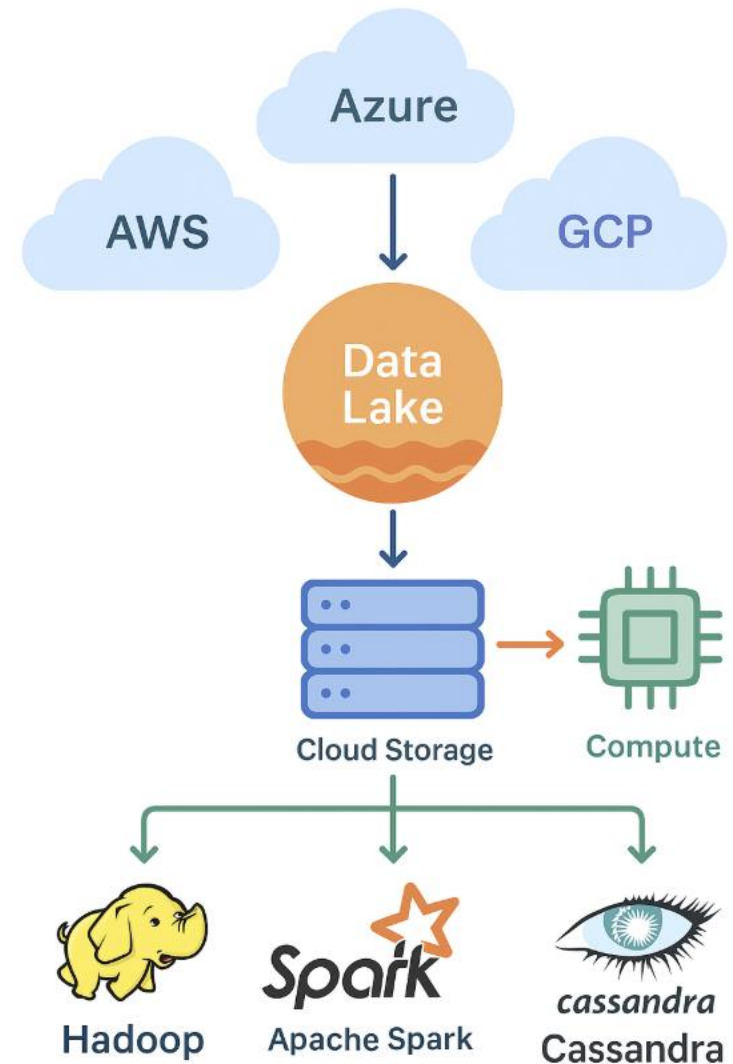


# Tecnologie e strumenti principali

## Google Cloud Platform (GCP)

Google Cloud Storage (object storage),  
BigQuery (data warehouse serverless),  
Dataproc (Hadoop/Spark managed), Dataflow  
(stream/batch), Pub/Sub (messaggistica)

**Vantaggi:** forte orientamento analytics/ML,  
serverless data warehouse, integrazione con  
Google AI/ML.



# Tecnologie e strumenti principali

## Quando usare cosa

| Esigenza principale                                 | Tecnologia consigliata                                                |
|-----------------------------------------------------|-----------------------------------------------------------------------|
| <b>Elaborazione batch su petabyte di dati</b>       | Hadoop / Amazon Elastic MapReduce-EMR                                 |
| <b>Analisi e ML in tempo reale</b>                  | Spark / Databricks / Dataflow                                         |
| <b>Database per dati semi-strutturati</b>           | MongoDB                                                               |
| <b>Alta disponibilità e scritture distribuite</b>   | Cassandra                                                             |
| <b>Archiviazione di dati grezzi e multi-formato</b> | Data Lake (S3, Azure Data Lake Storage-DLS, Google Cloud Storage-GCS) |
| <b>Infrastruttura flessibile e scalabile</b>        | Cloud (AWS, Azure, GCP)                                               |

# Tabella comparativa – Tecnologie Big Data & Cloud

| Categoria / Tecnologia                   | Vantaggi principali                                                                                                                                                                                                   | Limiti principali                                                                                                                                                              | Casi d'uso tipici                                                                              | Maturità / Complessità                                 |
|------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|--------------------------------------------------------|
| <b>Hadoop (HDFS + MapReduce)</b>         | <ul style="list-style-type: none"> <li>- Scalabilità orizzontale su hardware economico</li> <li>- Affidabilità e tolleranza ai guasti (replica dati)</li> <li>- Ecosistema esteso (Hive, Pig, HBase, etc.)</li> </ul> | <ul style="list-style-type: none"> <li>- Lento per analisi iterative o interattive</li> <li>- Configurazione e gestione complessa</li> <li>- Non adatto a real-time</li> </ul> | Batch processing, archiviazione massiva, analisi log e clickstream, Extract-Transform-Load-ETL | <p>Alta maturità</p> <p>Alta complessità operativa</p> |
| <b>Apache Spark</b>                      | <ul style="list-style-type: none"> <li>- Elaborazione in-memory molto veloce</li> <li>- Supporto per batch, streaming, ML e grafi</li> <li>- API intuitive (Python, Scala, Java, R)</li> </ul>                        | <ul style="list-style-type: none"> <li>- Uso intensivo di memoria</li> <li>- Richiede tuning accurato per cluster grandi</li> </ul>                                            | Analisi in tempo reale, machine learning, data pipeline, ETL avanzato                          | <p>Alta maturità</p> <p>Media complessità</p>          |
| <b>MongoDB (NoSQL documentale)</b>       | <ul style="list-style-type: none"> <li>- Schema flessibile (JSON-like)</li> <li>- Scalabilità orizzontale tramite sharding*</li> <li>- Query e aggregazioni potenti</li> </ul>                                        | <ul style="list-style-type: none"> <li>- Non ideale per transazioni complesse</li> <li>- Consistenza eventuale</li> </ul>                                                      | Applicazioni web, mobile, cataloghi prodotti, IoT                                              | <p>Alta maturità</p> <p>Bassa-media complessità</p>    |
| <b>Apache Cassandra (NoSQL columnar)</b> | <ul style="list-style-type: none"> <li>- Alta disponibilità (no single point of failure)</li> <li>- Ottima per scritture ad alto volume</li> <li>- Scalabilità lineare su più datacenter</li> </ul>                   | <ul style="list-style-type: none"> <li>- Modello dati rigido (schema chiave-valore per colonne)</li> <li>- Query limitate (no JOINS complessi)</li> </ul>                      | Logging, IoT, messaggistica, dati temporali distribuiti                                        | <p>Alta maturità</p> <p>Media complessità</p>          |

\*sharding è una tecnica di divisione di un database o di una rete blockchain in parti più piccole e gestibili, chiamate "shard", per migliorarne la scalabilità e le prestazioni.

# Tabella comparativa – Tecnologie Big Data & Cloud

| Categoria / Tecnologia             | Vantaggi principali                                                                                                                                                                               | Limiti principali                                                                                                                        | Casi d'uso tipici                                       | Maturità / Complessità               |
|------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------|--------------------------------------|
| <b>AWS Big Data Stack</b>          | <ul style="list-style-type: none"> <li>- Servizi maturi (S3, EMR, Glue, Redshift, Athena)</li> <li>- Integrazione profonda con AI/ML (SageMaker)</li> <li>- Ecosistema molto ampio</li> </ul>     | <ul style="list-style-type: none"> <li>- Costi variabili elevati se non ottimizzati</li> <li>- Vendor lock-in</li> </ul>                 | Data lake, ETL, analytics serverless, ML a scala        | Molto alto / media complessità       |
| <b>Azure Big Data Stack</b>        | <ul style="list-style-type: none"> <li>- Integrazione con Microsoft tools (Power BI, Synapse)</li> <li>- Ottimo per ambienti ibridi</li> <li>- Servizi gestiti (Databricks, HDInsight)</li> </ul> | <ul style="list-style-type: none"> <li>- Prezzi e configurazioni meno trasparenti</li> <li>- Dipendenza dallo stack Microsoft</li> </ul> | BI aziendale, data lake ibridi, pipeline dati corporate | Alto / media complessità             |
| <b>Google Cloud Big Data Stack</b> | <ul style="list-style-type: none"> <li>- Servizi serverless nativi (BigQuery, Dataflow)</li> <li>- Performance elevate su query analitiche</li> <li>- Fortemente integrato con AI/ML</li> </ul>   | <ul style="list-style-type: none"> <li>- Limitata compatibilità con tool on-prem</li> <li>- Lock-in nei servizi GCP</li> </ul>           | Analytics real-time, ML, data warehouse moderno         | Molto alto / bassa-media complessità |

La scelta del provider e degli strumenti dipende da fattori come: **compatibilità con strumenti esistenti, integrazione con altri servizi, latenza, costi di servizio e governance dei dati, sicurezza, regioni geografiche (localizzazione/sovranità).**

# Agenda

- Presentazione relatore
- Cosa sono i Big Data?
- Tecnologie e strumenti principali
- ➔ • Applicazioni pratiche dei Big Data
- Normativa europea Big Data
- Sfide
- Il futuro dei Big Data
- Q&A



# Applicazioni pratiche dei Big Data

## Come possono migliorare la nostra vita i big data?

### Ambiente

Soluzioni innovative per mitigare il cambiamento climatico

### Salute

Diagnosi migliori e cure più efficaci

### Industria

Prodotti innovativi, maggiore produttività, crescita economica

### Agricoltura

Maggiore sicurezza alimentare e utilizzo di risorse naturali

### Settore pubblico

Maggiore efficienza e trasparenza

### Trasporti

Regolazione del flusso del traffico e prevenzione degli ingorghi stradali

**Monitoraggio, profilazione, modelli previsionali per ottimizzazione**

# Applicazioni pratiche dei Big Data



## Industria

I *big data* consentono **analisi migliore delle esigenze e delle attese delle persone** per creare prodotti completamente nuovi (**ottimizzazione produttività**); Un utilizzo migliore dei dati industriali **aumenta il controllo della qualità** e aiuta a **tagliare i costi**, es. prevedendo le vendite e la conservazione dei prodotti negli stabilimenti intelligenti.

## Retail Marketing

La disponibilità in tempo reale dei dati ( interazione del cliente con il negozio, canali di comunicazione del brand, analisi e profilazione sulla base di uno storico comportamentale) puo' generare **incentivi all'acquisto automatici (coupon promozionali personalizzati)**. Il **real time marketing** fondamentale per il retail (che ha costi di locali e personale)

## Digital Marketing e Vendite

marketing e vendite e customer care, beneficiano dall'analisi dei dati per costruire **strategie su misura per ogni cliente**, (+acquisizioni, + profittevoli e fidelizzati nel lungo periodo). Dati di relazione con il cliente gestite con Client Relationship Management (CRM) sono alla base di miglioramento prodotti e servizi offerti.

# Applicazioni pratiche dei Big Data



## Assicurazioni

L'analisi di Big Data dati, riguarda la valutazione del rischio, per generare le condizioni ed i premi delle polizze. **Profilazione cliente** e offerte personalizzate.

## Media e nell'intrattenimento

Raccolta big data in tempo reale su consumi (**profilazione**) per ripensare raccomandazioni per streaming e contenuti. In generale analisi fruizione utente per migliorare palinsesti, tipi di trasmissioni/film/serie.

## Logistica e distribuzione

Analisi big data da sensori GPS e RFID forniscono dati per migliorare il **monitoraggio** delle spedizioni. L'analisi di big data consente **l'efficientamento delle catene di approvvigionamento** dell'industria e della grande distribuzione: es. ottimizzare le forniture dei supermercati, riducendo al minimo le esigenze di magazzino per lo stoccaggio dei prodotti e i rischi di sovrastima e sottostima degli approvvigionamenti. I **modelli previsionali** sono accurati, si basano sullo storico dei dati e dei consumi in relazione a tantissime variabili, capaci di considerare anche le condizioni eccezionali, come i consumi a ridosso dei festivi o nel caso di particolari condizioni meteorologiche.

# Applicazioni pratiche dei Big Data



## Istruzione-education

si raccolgono e analizzano Big Data dettagliati sul **comportamento degli studenti**: frequenza ai contenuti, tempi di risposta, interazioni online e progressi nei voti. Obiettivo: individuare pattern di apprendimento, prevedere il rischio di abbandono scolastico e personalizzare l'offerta formativa (Tutoring intelligenti con materiali personalizzati in base alle difficoltà o ai punti di forza dello studente). Percorsi di formazione su misura.

## Settore pubblico

Big data e analisi approfondite possono aumentare l'efficienza e l'efficacia dei servizi pubblici, con **servizi su misura dei cittadini** e maggiore trasparenza (es. modello smart city per il sistema di servizi interconnessi e monitorati con strumenti IoT: parcheggi, illuminazione, trasporti, densità popolazione... ).

## Turismo

Big Data (**profilazione utenti**) per anticipare le tendenze, personalizzare le offerte e migliorare l'esperienza dei viaggiatori.

## Sicurezza informatica

Da big data su telemetrie log e dati tecnici si arriva a analisi anomalie (Anomaly detection) e **identificazione pattern ricorrenti** per prevenire minacce.

# Applicazioni pratiche dei Big Data



## Ambiente

I dati raccolti dai satelliti possono migliorare la ricerca e aiutare l'UE a ridurre le emissioni di gas a effetto serra, a prevenire e a rispondere ai disastri naturali come gli incendi. **Monitoraggio dati di inquinamento** e individuazione dei maggiori produttori di inquinanti.

## Energia

Raccolta big data in tempo reale su consumi, produzione, condizioni climatiche e funzionamento delle infrastrutture. L'analisi permette di bilanciare in modo dinamico la domanda e l'offerta di energia, evitando sprechi e blackout. Nel campo delle energie rinnovabili, i Big Data aiutano a **prevedere** la produzione solare o eolica in base alle condizioni meteo e a ottimizzare la distribuzione dell'energia prodotta.

Anche la **manutenzione predittiva** delle centrali e delle reti elettriche è resa possibile da algoritmi che rilevano in anticipo eventuali guasti.

## Agricoltura

Gli agricoltori possono utilizzare i dati dei satelliti e dei sensori per utilizzare meglio le risorse come acqua e luce, oltre ad adattare la coltivazione al mutare delle circostanze.

# Applicazioni pratiche dei Big Data



## Trasporti

I big data raccolti dal GPS e dai social media possono mitigare la congestione del traffico, che a sua volta contribuisce a risparmiare tempo e carburante oltre a diminuire le emissioni di CO2.

## Salute

I Big Data sono fondamentali nei processi diagnostici accurati, cure e **previsionali** del rischio malattia, e per sviluppo di medicinali a costi ridotti, mediante analisi di cartelle cliniche rese anonime e dati immessi nelle app mediante strumenti wearable di misura/monitoraggio parametri vitali.

## Ricerca Scientifica

La decodifica del DNA e altre ricerche sono accelerate grazie alla potenza computazionale dell'analisi dei Big Data.

## Fintech

Usa i Big Data Analytics, per le **analisi predittive** sui titoli al fine di ottimizzare la gestione di un portfolio di investimenti grazie a previsione di rendimento titoli in tempo reale (mercato di riferimento, condizioni azienda, propensione al rischio). Questo implica soluzioni e prodotti finanziari quasi personalizzati per ogni utente

# Agenda

- Presentazione relatore
- Cosa sono i Big Data?
- Tecnologie e strumenti principali
- Applicazioni pratiche dei Big Data
- Normativa europea Big Data
- Sfide
- ➔ • Il futuro dei Big Data
- Q&A



# Il futuro dei Big Data.....

## .....abilitato dal Cloud



L'insieme dei dati aziendali potrà essere gestito come un set di Big data. MA Richiede **progettazione di un'architettura digitale** e una **strategia di gestione** per trarre vantaggio dalla loro analisi.

**Il cloud è l'infrastruttura che meglio si presta ad ospitare Big Data** (implementazione nel cloud di cluster di big data gestiti).

3 elementi necessari: **gestione dei dati attenta e accurata, visibilità in tempo reale e sicurezza delle informazioni raccolte.**

**I tipi di dati aziendali d'interesse** possono essere: Registri di reti virtuali, Schemi ed eventi di sicurezza, Schemi di traffico di reti globali, Rilevamento e risoluzione di anomalie, Informazioni di conformità, Monitoraggio del comportamento e delle preferenze dei clienti, Dati di geolocalizzazione, Dati provenienti da canali social per il monitoraggio del sentiment del brand, Monitoraggio dei livelli delle scorte e delle spedizioni, Altri dati specifici che **possono avere un impatto importante sul business.**

# Il futuro dei Big Data....

## ....per l'AI

**Intelligenza Artificiale:** Big Data essenziali per addestrare i modelli di IA per riconoscimento immagini e l'elaborazione del linguaggio naturale. Ottimizzano tutte le attuali applicazioni dei Big Data mediante automazione e velocità di elaborazione.

## Manifattura 4.0

L'impiego di **modelli predittivi basati sul machine learning** può gestire i Big Data provenienti dai sensori di un impianto per mettere a punto una strategia di **manutenzione predittiva**, capace di prevedere i guasti e i dannosi fermo macchina che ne deriverebbero con notevoli risparmi rispetto alla manutenzione preventiva, in quanto permette di effettuare soltanto gli interventi effettivamente utili a mantenere l'impianto nelle migliori condizioni funzionali.

Capacità di ottimizzare una supply chain, piuttosto che efficientare il ciclo di vita di un prodotto.

# Il futuro dei Big Data

## Sicurezza informatica

Anomaly detection potenziata da AI anche per la difesa Cyber, analisi dei pattern di attacco per mitigazione con MTTR minimale.

## Sicurezza transazioni bancarie

Grazie a sistemi di AI basati su training da big data, i sistemi per la gestione dei pagamenti sono in grado di **rivelare in tempo reale delle possibili anomalie rispetto alle procedure di routine**, che potrebbero corrispondere ad un tentativo di frode da parte di un malintenzionato. Tali sistemi si basano sull'analisi di uno streaming di dati rilevato dai POS e, in caso di anomalie, sono in grado di attivare automaticamente le procedure anti-frode, che, in una condizione praticamente immediata, sono in grado di far procedere l'operazione richiesta o arrestarla preventivamente.

# Il futuro dei Big Data

In generale **BIG DATA + CLOUD + AI** = abilitazione **all'analisi in tempo reale** (in diverse aree di applicazione). L'analisi può portare a:

1. Forme di **analisi predittiva** per varie finalità favorendo un **approccio preventivo e proattivo** (es verso clienti nel marketing, in caso di attacchi nella difesa)
2. **supporto alle decisioni A/B**
3. **Decisioni automatizzate in tempo reale**



# Agenda

- Presentazione relatore
- Cosa sono i Big Data?
- Tecnologie e strumenti principali
- Applicazioni pratiche dei Big Data
- ➔ • Normativa europea Big Data
- Sfide
- Il futuro dei Big Data
- Q&A



# Normativa europea Big Data

Lista di principali normative europee rilevanti per i big data:

**Regolamento (UE) 2016/679 – GDPR (General Data Protection Regulation)**

applicabile dal **25 maggio 2018**. Disciplina la protezione dei dati personali di cittadini UE, impone obblighi a chi tratta dati (controllori e processor), garantisce diritti (accesso, cancellazione, portabilità...) e regola le trasferimenti di dati fuori dall'UE.

**Regolamento (UE) 2022/868 – Data Governance Act (DGA)** adottato il **30 maggio 2022**, GU UE il **3 giugno 2022**, ed è applicabile dal **24 settembre 2023**. Stabilisce un quadro per la condivisione sicura dei dati, promuove l'intermediazione dati (data-intermediaries), facilita l'altruismo dei dati e rafforza la fiducia nel riutilizzo di dati sia personali che non personali.

**Regolamento (UE) 2023/2854 – Data Act** adottato il **13 dicembre 2023**, pubblicato GU UE il **22 dicembre 2023** direttamente applicabile a partire dal **12 settembre 2025**. Regola l'accesso equo e l'uso dei dati generati da dispositivi connessi ("connected products") e servizi digitali; punta a distribuire il valore dei dati in modo più giusto tra chi genera dati, chi li possiede e chi li usa

# Normativa europea Big Data

**Data: Regolamento (UE) 2022/1925 - Digital Markets Act (DMA)** adottato il **14 settembre 2022**, entrato in vigore l'**1 novembre 2022** (applicabile dal **2 maggio 2023**). Impone regole precise alle grandi aziende tecnologiche (i "gatekeeper" come Google, Apple, Meta) per garantire concorrenza leale, trasparenza e protezione dei dati, vietando pratiche anticoncorrenziali, facilitando la portabilità dei dati e consentendo agli utenti di disinstallare app preinstallate, con l'obiettivo di un mercato digitale più equo e aperto. Pur non essendo una norma "big data" in senso stretto, può influenzare l'uso dei dati da parte delle grandi piattaforme.

**Regolamento (UE) 2025/327 – European Health Data Space (EHDS)** adottato l'**11 febbraio 2025**, pubblicato il **5 marzo 2025**, in vigore il **26 marzo 2025**. Mira a creare uno spazio europeo per i dati sanitari, favorendo la condivisione di dati di salute (anche per scopi secondari, come ricerca), mantenendo al contempo la protezione dei dati personali sanitari.

**EU Cloud Code of Conduct per i fornitori di servizi cloud.** E' un codice di condotta previsto dall'art. 40 del GDPR. Definisce requisiti per i cloud provider (IaaS, PaaS, SaaS) affinché dimostrino conformità al GDPR quando agiscono come "processor" di dati personali.

# Normativa europea Big Data

**Directive 2002/58/CE – Direttiva ePrivacy – da aggiornare rispetto alla prima versione del 2002.** Riguarda la protezione dei dati nella comunicazione elettronica (es. confidenzialità delle comunicazioni, traffico dati, cookie, spam), complemento del GDPR per il settore delle telco. (bozza [agg UE 2017](#))

**DIGITAL OMNIBUS:** proposta iniziale Omnibus digitale comprende una serie di modifiche tecniche a un ampio corpus di legislazione digitale. ([pubblicata il 19 novembre](#)): razionalizza norme dell' «acquis digitale»: dal Data Act al GDPR, dall'ePrivacy Directive alla direttiva NIS2 (semplificate e consolidate) e revisione del Regolamento UE sull'AI (2024/1689). E' una sorta di «Codice Unico del digitale europeo». Punta alla disciplina dei dati, razionalizzazione delle segnalazioni di incidenti informatici e chiarimenti per implementazione regole sull'AI.

L'iter è all'inizio. Prox passi: Consultazione pubblica e confronto con gli stakeholder. Eventuali revisioni della commissione, discussione e voto in Parlamento Europeo, nel Consiglio e dopo loro allineamento, la pubblicazione e il periodo transitorio di adeguamento.

# Normativa europea Big Data

---

Il 19 novembre, è stata presentata anche la Data Union Strategy: La strategia per l'Unione dei dati prevede azioni in tre aree prioritarie:

- Aumentare l'accesso ai dati per l'intelligenza artificiale per garantire alle aziende UE l'accesso a dati di alta qualità necessari per l'innovazione
- Semplificare le norme sui dati per garantire certezza giuridica alle aziende e ridurre i costi di conformità
- Salvaguardare la sovranità dell'UE sui dati per rafforzare la nostra posizione globale sui flussi internazionali di dati (fonte UE)

# Agenda

- Presentazione relatore
- Cosa sono i Big Data?
- Tecnologie e strumenti principali
- Applicazioni pratiche dei Big Data
- Normativa europea Big data
- ➔ • Sfide
- Il futuro dei Big Data
- Q&A



# Sfide

## Sfide diverse: privacy, etiche, tecnologiche, analitiche, costi

### Protezione dei diritti di privacy

A causa dell'enorme quantità di dati generati:

- La maggior parte delle organizzazioni non riesce a garantire la privacy per lacune nei controlli regolari.
- I fornitori di servizi digitali hanno più informazioni sui propri utenti rispetto a quante ne abbiano gli utenti sui fornitori → la profilazione eccessiva porta rischi di Privacy. → Sbilanciamento non regolamentato causa sfruttamento di preferenze, e debolezze utenti per fini commerciali o politici. → Rischi di manipolazione delle persone o valutazioni arbitrarie es. su salute o altri dati sensibili con potenziale esclusione da opportunità professionali o copertura sanitaria.

È necessario che le aziende creino una mappa dei dati ed eseguire audit regolari per informare sulle modifiche in materia di sicurezza e privacy e garantire trasparenza e protezione.

# Sfide

**Sfide diverse: privacy, etiche, tecnologiche, analitiche, costi**

## **Sfide etiche**

Raccogliere così tante informazioni significa aumentare la probabilità che contengano dati personali identificabili.

I dati possono contenere pregiudizi. E usati per training di AI possono portare a un'intelligenza artificiale distorta che amplifica ulteriormente i pregiudizi umani.

È necessario che le aziende anonimizzino i dati. Inoltre prima di usarli come dataset per AI dovrebbero essere ripuliti da elementi di genere, razza, e orientamento sessuale o qualsiasi altro stigma che possa creare pregiudizio. Rispetto alla quantità si dovrebbe usare solo il quantitativo strettamente necessario.

# Sfide

## Sfide diverse: privacy, tecnologiche, analitiche, costi

### **Tecnologiche: Condivisione e accesso ai dati**

È la sfida più comune che tutti si trovano ad affrontare nella gestione dei big data. Nonostante i numerosi sforzi, i dati esterni provenienti da determinate fonti sono inaccessibili.. La disponibilità dei dati è un fattore cruciale per accedervi in modo completo, accurato e tempestivo, poiché è necessaria per prendere decisioni informate.

### **Tecnologiche: Scalabilità**

Le sfide in termini di scalabilità ed elasticità includono sia la scalabilità verticale (aggiunta di più risorse a una singola macchina) sia quella orizzontale (aggiunta di più macchine a un cluster).

# Sfide

**Sfide diverse: privacy, tecnologiche, analitiche, costi**

## **Analitiche**

sfide chiave legate alle complessità analitiche nei Big Data includono dimensionalità, raccolta dati, pre-elaborazione dati, selezione algoritmica, risorse computazionali, archiviazione dati e analisi in tempo reale.

**Costi:** mantenere un'infrastruttura Big Data può essere costoso. Gestire attentamente i costi associati ad hardware, software e personale



# Contatti

---



A.valentini@isacaroma.it

*Grazie...*