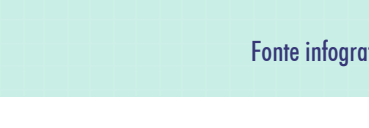
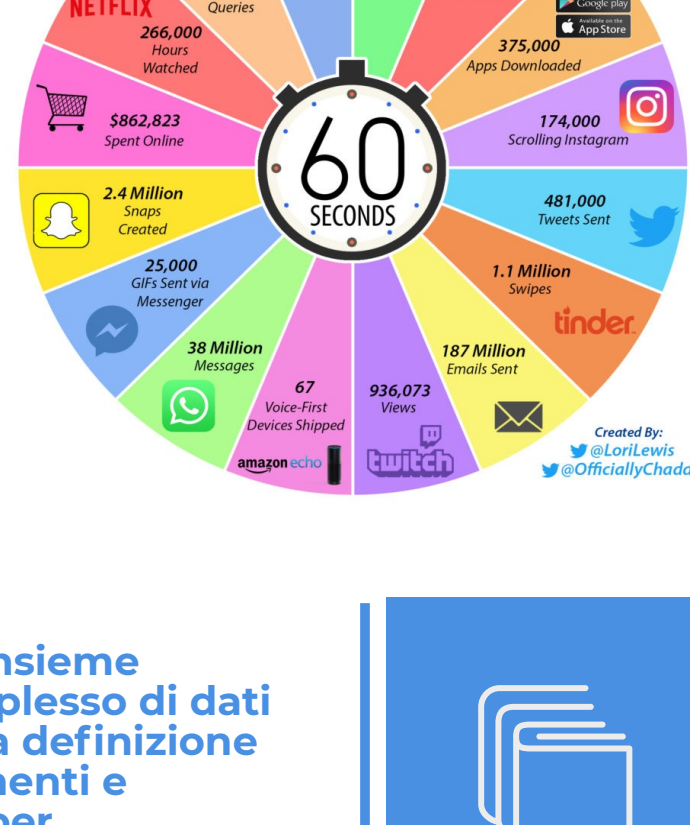


VALORIZZARE I BIG DATA

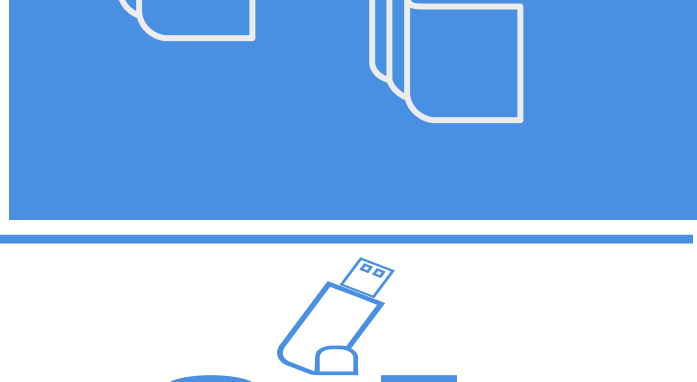


Fonte infografica: Big Data e Analytics - Ing. Roberto Piermarini

Internetcosa succede in un minuto



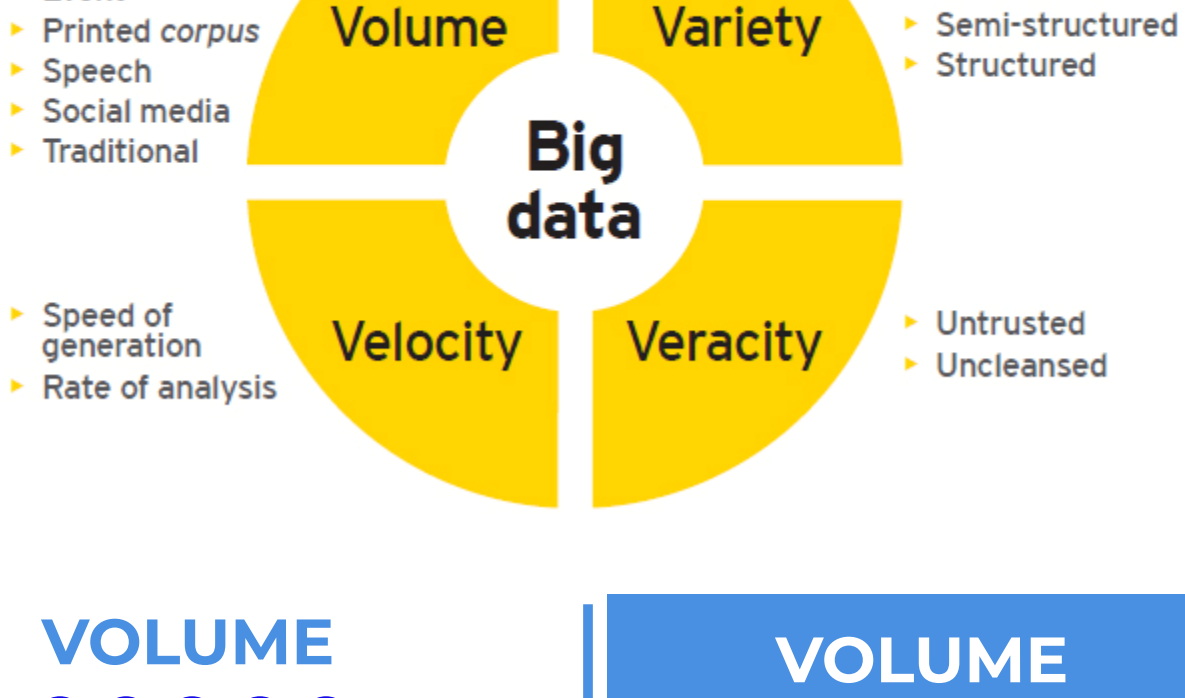
Big Data, un insieme grande e complesso di dati che richiede la definizione di nuovi strumenti e metodologie per estrarre, gestire e processare la mole di informazioni..



DEFINIZIONE: "big data" come bene informativo caratterizzato da alto volume, elevata velocità, ampia varietà, che richiede forme economiche ed innovative per l'elaborazione di informazioni, al fine di ottenere una maggiore comprensione e prendere decisioni entro un tempo ragionevole.



LE 4 V DEL BIG DATA



VOLUME



Dati a riposo
- Da Terabyte a Exabyte di dati da processare

VOLUME

Quantità di dati generati ogni secondo, da sorgenti eterogenee. Questi volumi mettono in crisi le tecniche tradizionali di stotrage ed analisi, fatti con gli odierni Sistemi di gestione di basi di dati (DBMS).

VELOCITA'

Si intende quella di produzione dei dati stessi, che aumenta anche in funzione del numero in crescita di sorgenti che li producono (PC, smartphone, sensori interconnessi, wearable, etc.)

VELOCITA'

Dati in movimento
- Data streaming
- Risposte a richieste che vanno da millisecondi a secondi

VARIETA'



Dati in molte forme
- Strutturati
- Non strutturati
- Testo
- Multimedia

VARIETA'

La varietà dei Big Data è dovuta alla loro mancata strutturazione: tra essi sono infatti inclusi anche documenti di vario genere e sono vari anche nelle fonti

VERIDICITA'



Connotazione qualitativa dei dati, punto di partenza delle attività di analisi e di interpretazione.

VERIDICITA'

Dati in dubbio
- Inconsistenza dei dati
- Ambiguità
- Latenza, frodi
- Approssimazioni del modello

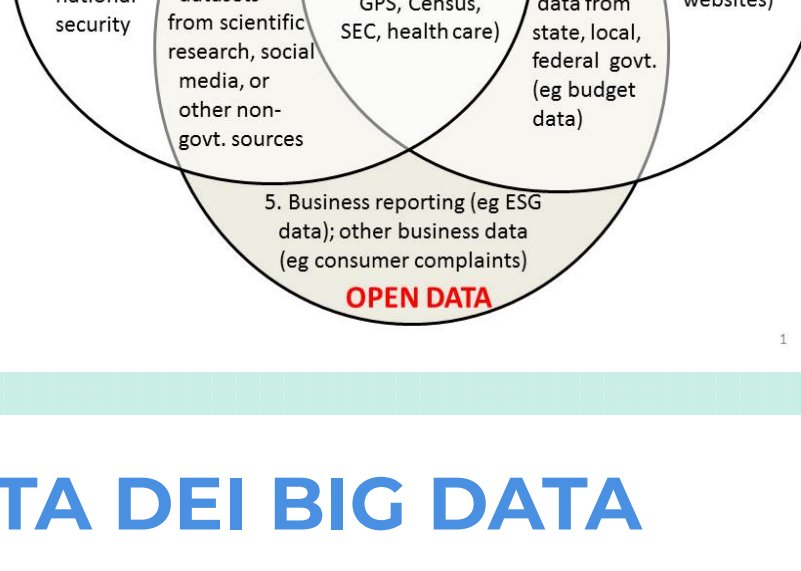
Big Data vs Open Data

OPEN DATA= DATI APERTI (accessibili e riutilizzabili)

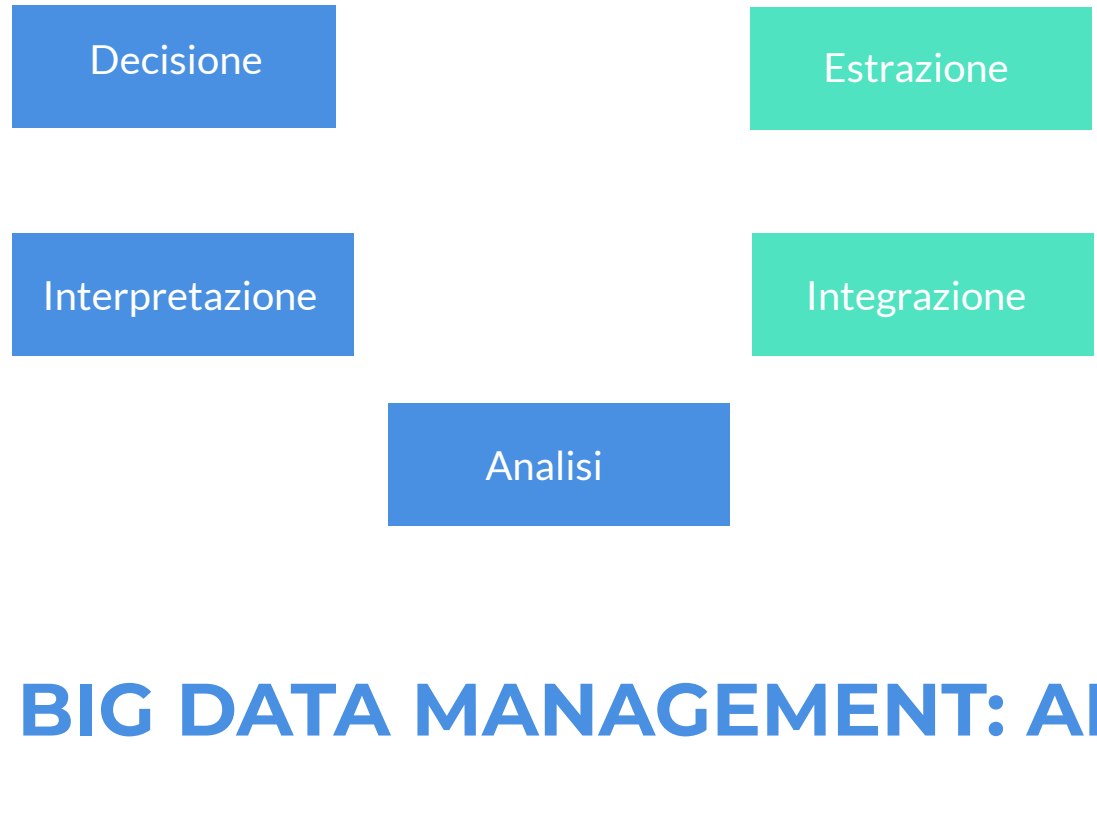
- Trasparenza
- Partecipazione
- Collaborazione

datiopen.it iniziativa alla quale partecipa Unioncamere ed alcune Camere di commercio.

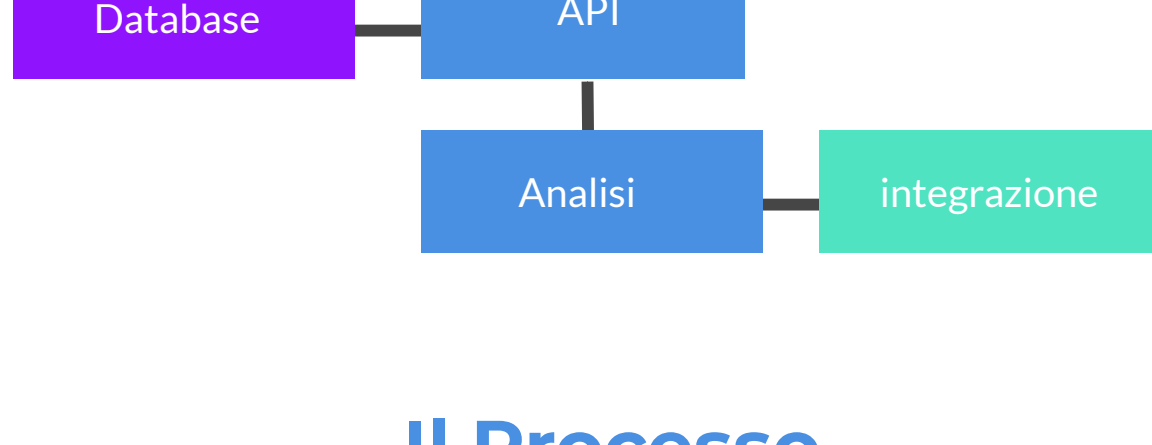
camcom.gov.it raccoglie una sezione dataset.



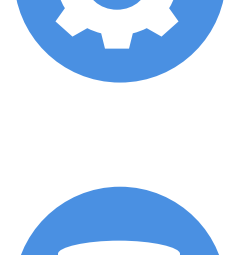
CICLO DI VITA DEI BIG DATA



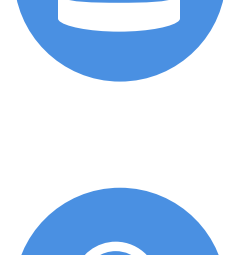
BIG DATA MANAGEMENT: API



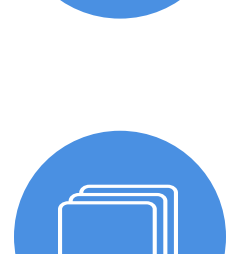
Il Processo



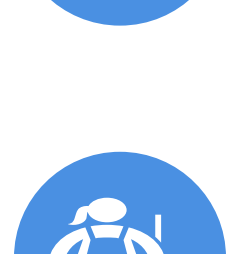
WEB SCRAPING
Con il termine di Web Scraping si indicano le metodologie che consentono di estrarre e collezionare dati e informazioni da Internet



ACQUISIZIONE DATA WHAREHOUSE
Importando i dati da database relazionali, non relazionali o da altre sorgenti in strumenti di ETL, già usati ampiamente per la movimentazione di dati in sistemi di Data Warehousing e Data Mart



ACQUISIZIONE IN STREAMING
Leggendo flussi continui di dati mediante sistemi capaci di catturare eventi, elaborarli e salvarli su un database in modo efficiente



ARCHIVIAZIONE DATI
In seguito all'acquisizione dalle sorgenti, i dataset sono trasferiti a sistemi di archiviazione locali o remoti attraverso mezzi di trasmissione ad alta velocità



ESTRAZIONE E PULIZIA DELLE INFORMAZIONI
Bisogna creare un processo di estrazione che prelevi le informazioni richieste dalla grande fonte dei dati e le rappresenti in una forma standard e strutturata pronta per l'analisi. La progettazione del processo di estrazione dipende principalmente dall'area di applicazione

Modelli di Analytics

Descriptive Analytics, insieme di strumenti orientati a descrivere la situazione attuale e passata dei processi aziendali e/o aree funzionali. Tali strumenti permettono di accedere ai dati secondo viste logiche flessibili e di visualizzare in modo sintetico e grafico i principali indicatori di prestazione

Predictive Analytics, strumenti avanzati che effettuano l'analisi dei dati per rispondere a domande relative a cosa potrebbe accadere nel futuro; sono caratterizzati da tecniche matematiche quali regressione, forecasting, modelli predittivi, ecc.

Prescriptive Analytics, tool avanzati che, insieme all'analisi dei dati, sono capaci di proporre al decision maker/strategie sulla base delle analisi svolte

Automated Analytics, capaci di implementare autonomamente l'azione proposta secondo il risultato delle analisi svolte.

Data Scientist

Riescono a dare un senso ai Big Data, grazie alla loro capacità di unire la conoscenza delle tecnologie allo stato dell'arte con la conoscenza, la competenza e la capacità di analisi del contesto.

Figura professionale dotata di abilità "integrate" che spaziano dalla matematica, all'apprendimento automatico, alla statistica, al data mining, ai database e all'ottimizzati one

MATH & STATISTICS

- Machine learning
- Statistical modeling
- Experiment design
- Bayesian inference
- Statistical learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- Computer science fundamentals
- Scripting language e.g. Python
- Statistical computing packages e.g. R
- Databases: SQL, NoSQL, Hadoop
- Relational algebra
- Parallel databases and parallel query processing
- MapReduce concepts
- Flinking and Storm
- Custom resolvers
- Experience with tools like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- Passionate about the business
- Curious about data
- Inference without authority
- Industry context
- Problem solver
- Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- Ability to engage with senior management
- Clear thinking skills
- Transferable data driven insights into business and actions
- Visual design
- Fluency in any of visualization tools e.g. Power BI, Tableau