



**UNIVERSITÀ
DI TORINO**

Introduzione ai big data, che cosa sono e le diverse tipologie

Ph.D Luca Giraldi

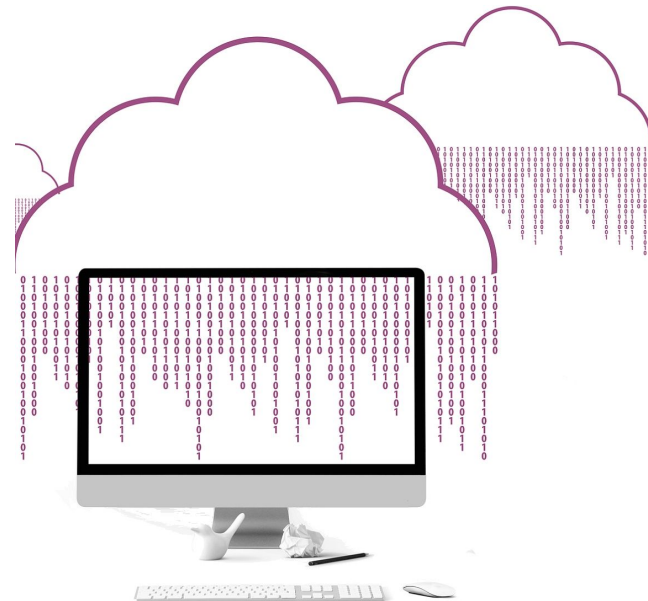
Big data è un termine che descrive grandi **volumi** di dati ad alta **velocità**, **complessi e variabili** che richiedono tecniche e tecnologie avanzate per consentire l'acquisizione, l'archiviazione, la distribuzione, la gestione e l'analisi delle informazioni per **migliorare la comprensione e il processo decisionale**.



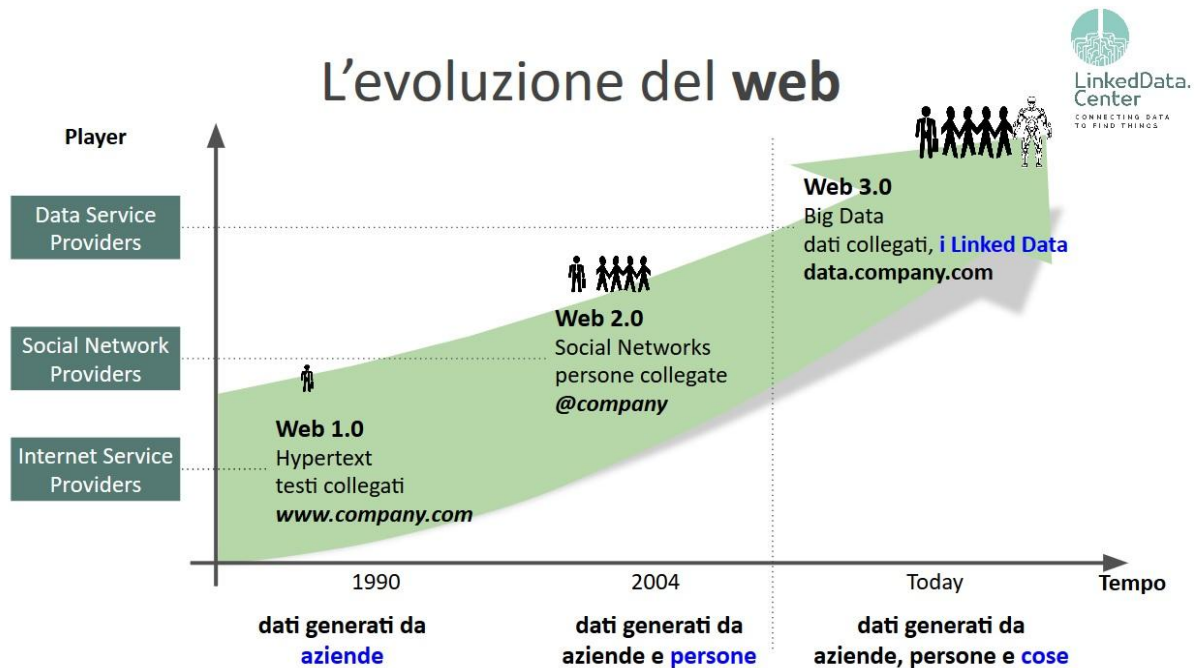
- 1.il termine «grandi dati» ;
- 2.Aumento della potenza di elaborazione dei computer;
- 3.Vari formati di dati



Nei primi anni 2000 le tecnologie di storage sono state sopraffatte da una enorme quantità di dati



Evoluzione dei BIG DATA



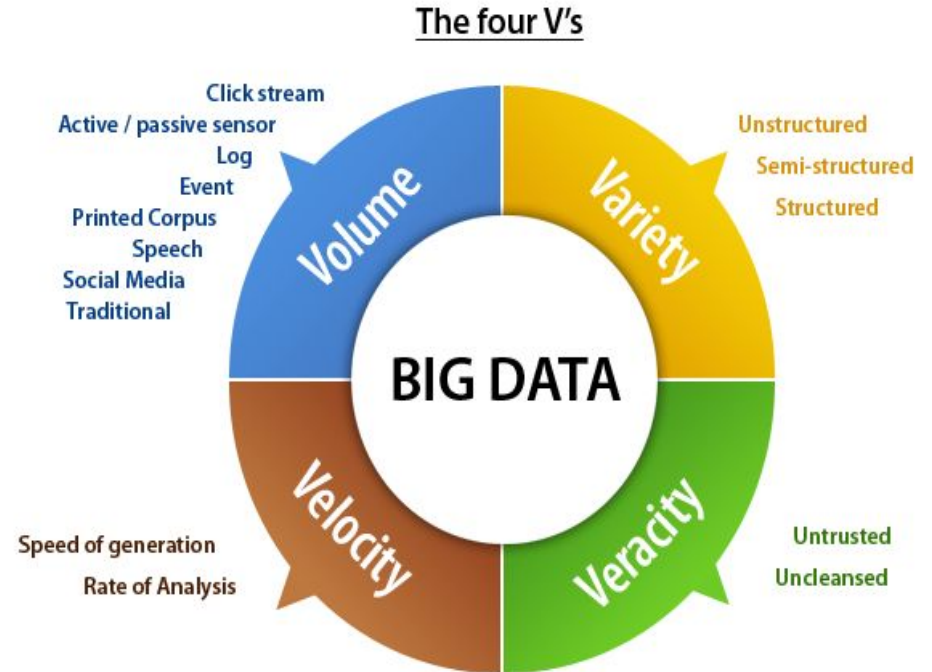
● Attributi dei Big Data:

□ Volume

□ Varietà

□ Velocità

□ Veridicità.



Big Data → il Volume

Volume

Everyday, we
create about
**2.5
Quintillion
Bytes**
of data

=

 X 10 million



proprietà Primaria dei Big Data

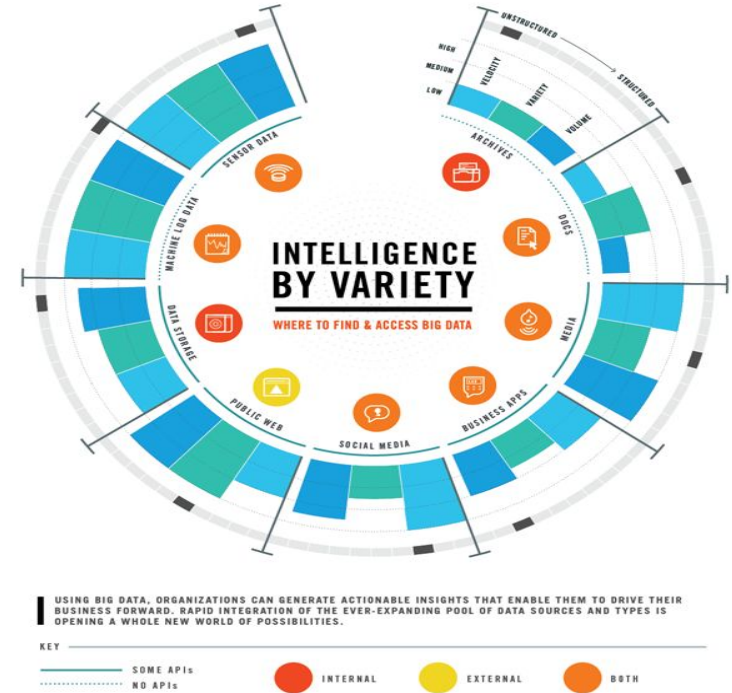
Big Data → il Volume

- Questa caratteristica è in assoluto la più semplice da notare e assimilare: “Big Data” significa avere a che fare con un'enorme quantità di dati. Vediamo qualche esempio che lo dimostra:
- ❖ Nel 2016 nel mondo 2.08 miliardi di persone hanno un telefono cellulare. La popolazione totale è di circa 7 miliardi di persone.
- ❖ Si stima che ogni giorno nel mondo vengano creati 2,5 quintilioni di byte (pari a 2,3 trilioni di gigabyte) che sono sufficienti per riempire 10 milioni di dischi Blu-Ray.
- ❖ Il 90 per cento dei dati disponibili ad oggi è stato prodotto nel corso degli ultimi due anni.
- ❖ Facebook conta 1,71 miliardi di utenti attivi. Ogni giorno nella piattaforma vengono visualizzati 8 miliardi di video.
- ❖ Twitter ha 320 milioni di utenti attivi. Si contano in media 500 milioni di tweet al giorno.

Questo enorme **volume** di dati comporta anche una grande **varietà** degli stessi, proprietà che approfondiremo di seguito.

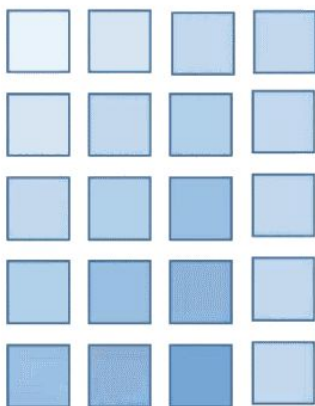
Big Data → Varietà

- Grande varietà di dati:
- da fonti diverse;
- di origine diversa;
- di tipo diverso.

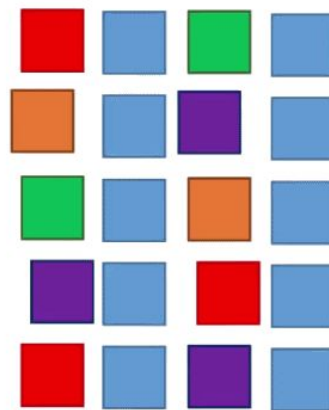


Big Data → Varietà

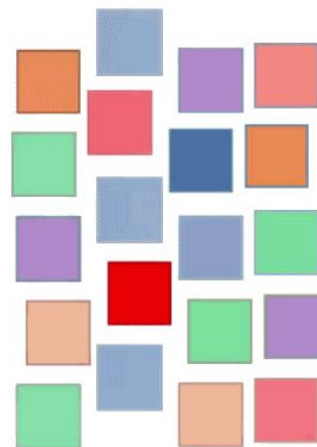
Structured Data



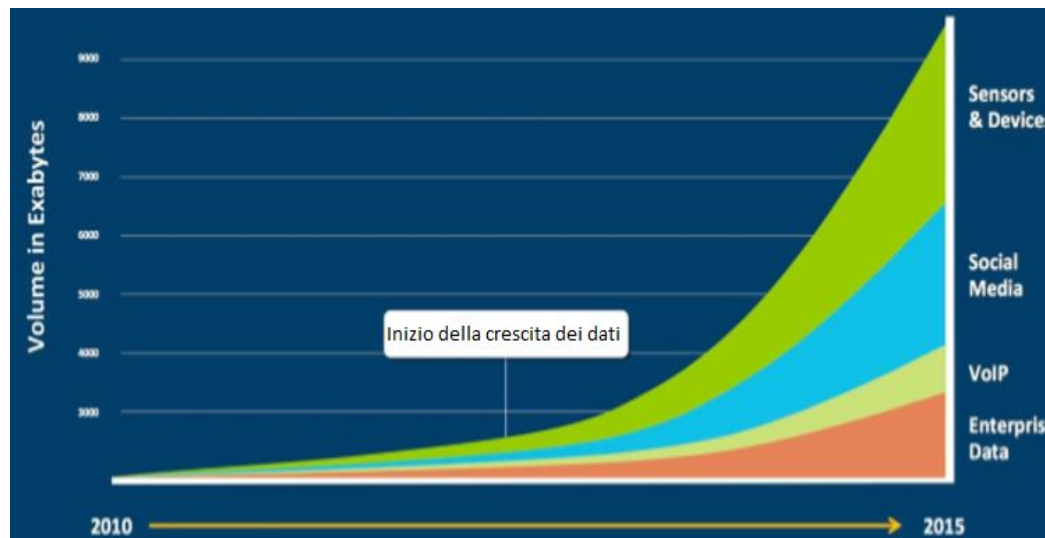
Semi-Structured Data



Unstructured Data



Big Data → Varietà



***Enterprise data:** Sono i dati aziendali, essi sono dati condivisi generalmente tra i dipartimenti dell'azienda.

***Voice over IP** (in italiano "Voce tramite protocollo Internet", in acronimo **VoIP**), in telecomunicazioni e informatica, indica una tecnologia che rende possibile effettuare una conversazione telefonica sfruttando una connessione Internet o una qualsiasi altra rete dedicata.

Big Data → Velocità

I dati devono essere:

- ☐ raccolti
- ☐ Elaborati
- ☐ archiviati
- ☐ analizzati



In breve tempo



VELOCITY

Big Data → Velocità

- Per capire cosa si intende per velocità, basti pensare che:
 - ❖ Si stima che nel 2018 il traffico internet globale raggiunga i 50 mila GB al secondo.
 - ❖ Ogni 60 secondi, si spediscono 204 milioni di email e si caricano 72 ore di video su YouTube e 216 mila post su Instagram.

Big Data → Veridicità

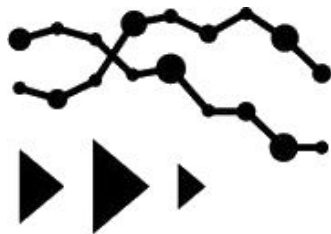
- La veridicità. IBM ha coniato Veracity come la quarta V, che rappresenta l'inaffidabilità insita in alcune fonti di dati.
- Nel caso della “**veridicità**” dei dati, ci si riferisce alla loro attendibilità, ossia dovrebbero essere affidabili e di qualità.
- Tuttavia è necessario dire che anche i dati apparentemente non affidabili possono, in determinate circostanze, certo, essere presi come riferimento assolutamente utile per l'analisi. Sono i cosiddetti **dati “sporchi”**.

Big Data → Veridicità

I dati sporchi
possono essere
fondamentali!



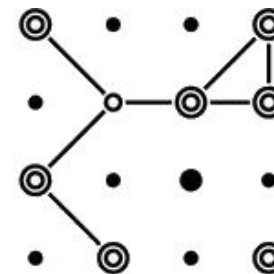
Altre proprietà dei Big Data



variability



Un stesso dato può avere
significati differenti



complexity

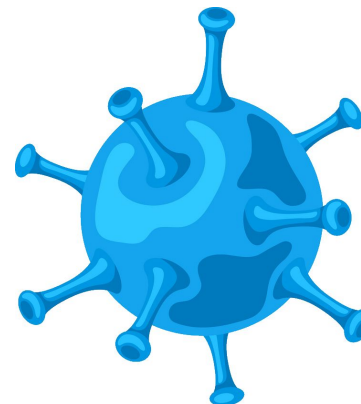


Maggiore è la dimensione
del dataset, maggiore è la
complessità dei dati

Altre proprietà dei Big Data

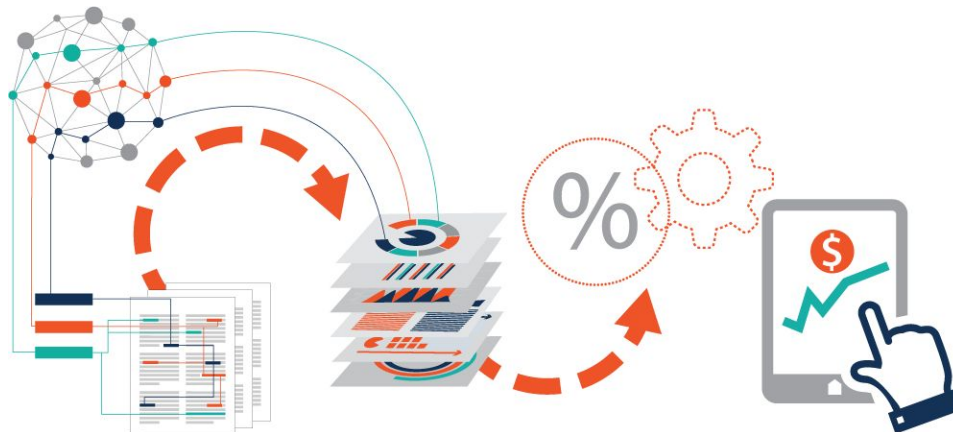


l'importanza dei Big Data



Velocità di trasmissione

DIFFERENZA TRA ANALYTICS E ANALYSIS



i dati raccolti non hanno di per sé valore;
l'analisi che potrebbe essere fatta su di essi genera un enorme valore su di essi!

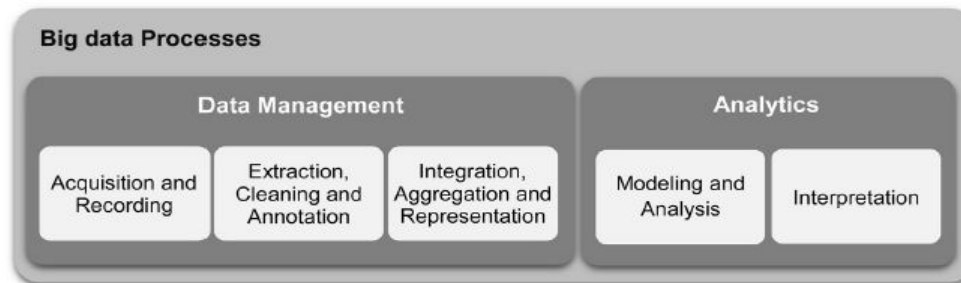
DIFFERENZA TRA ANALYTICS E ANALYSIS



Nel corso del tempo le tecnologie per l'analisi dei dati si sono evolute:

- 1970s nascita dei primi sistemi di supporto alle decisioni;
- 1990s nascita della Business Intelligence;
- 2010 nascita dell'analytics.

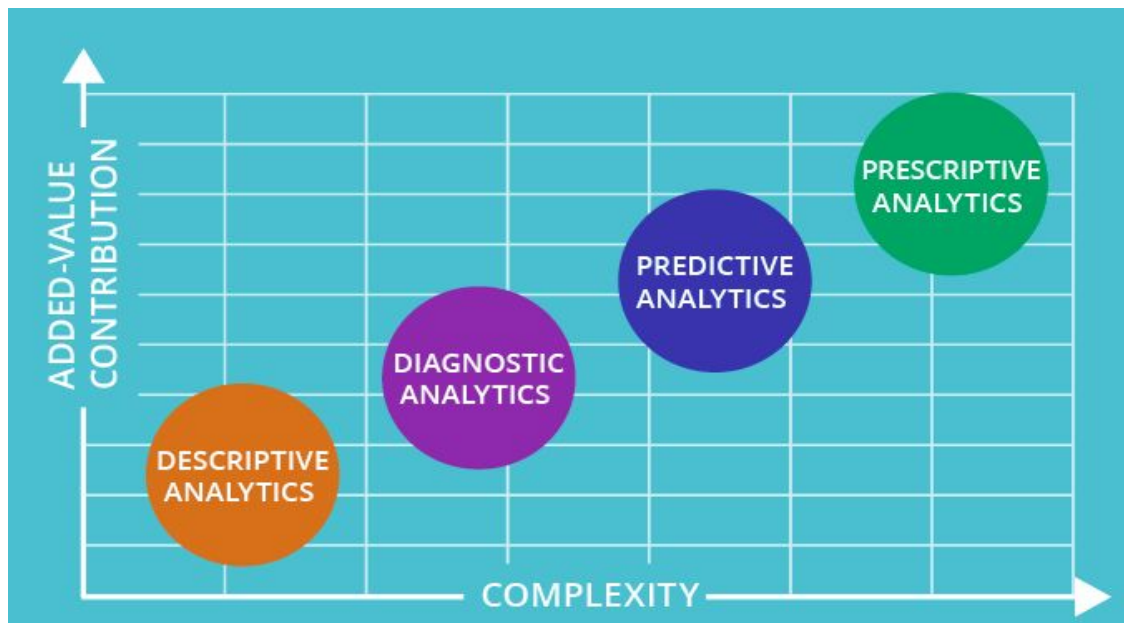
DIFFERENZA TRA ANALYTICS E ANALYSIS



- Sono 5 le fasi del processo dei Big Data:

1. **Acquisizione;**
2. **estrazione, pulizia, annotazioni;**
3. **integrazioni e rappresentazioni;**
4. **analisi;**
5. **Interpretazione.**

DIFFERENTI TIPI DI ANALYTICS



Quale domanda ci poniamo?



che cosa è successo?

Quali tecnologie possiamo utilizzare?

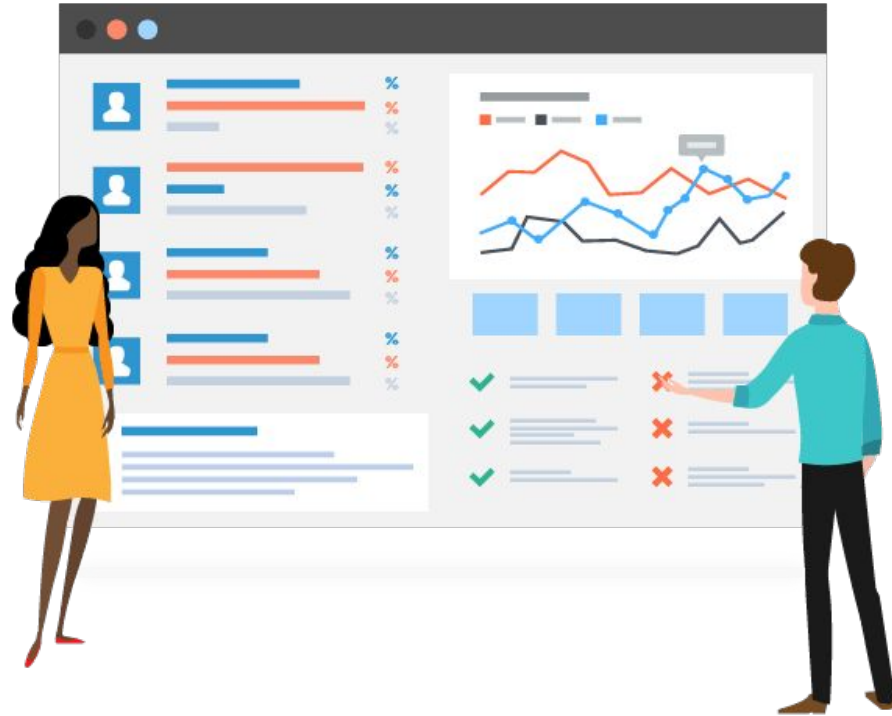


Dashboard, Scorecard, Report, Data
Warehouse

Quali risultati otteniamo?



Definizione del problema



Quale domanda ci poniamo?



Perchè è successo?

Quali tecnologie possiamo utilizzare?



Business reporting; Dashboard; Data
warehouse.

Quali risultati otteniamo?



capacità di approfondire fino alla causa
principale

Essa risponde alle domande sul perché. Perché le cose stanno accadendo? Cosa spinge le cose ad andare su, o giù? Le **funzioni** della **diagnostic analytics** rientrano in linea di massima in **tre** categorie:

- **Identificare le anomalie:** Sulla base dei risultati dell'analisi descrittiva, gli analisti devono identificare le aree che richiedono ulteriori studi perché sollevano domande alle quali non si può rispondere semplicemente guardando i dati. Queste potrebbero includere domande come perché le vendite sono aumentate in una regione dove non c'è stato alcun cambiamento nel marketing, o perché c'è stato un improvviso cambiamento nel traffico verso un sito web senza una causa ovvia.
- **Esercitare l'analisi (scoperta):** Gli analisti devono identificare le **fonti** di dati che li aiuteranno a spiegare queste anomalie e per identificare le correlazioni e determinare se alcune di esse sono di natura causale.
- **Determinare le relazioni causali:** Le relazioni nascoste sono scoperte osservando gli eventi che potrebbero aver portato alle anomalie identificate. La teoria della probabilità, l'analisi della regressione, il filtraggio e l'analisi dei dati delle serie temporali possono essere utili per scoprire storie nascoste nei dati.

Quale domanda ci poniamo?



che cosa è probabile che accada?

Quali tecnologie possiamo utilizzare?



Data mining, web mining

Quali risultati otteniamo?



Proiezioni, stime di eventi

Esempio 1



Analisi di cosa può accadere

Esempio 2



Analisi delle future attività

Quale domanda ci poniamo?



Cosa posso fare?

Quali tecnologie possiamo utilizzare?



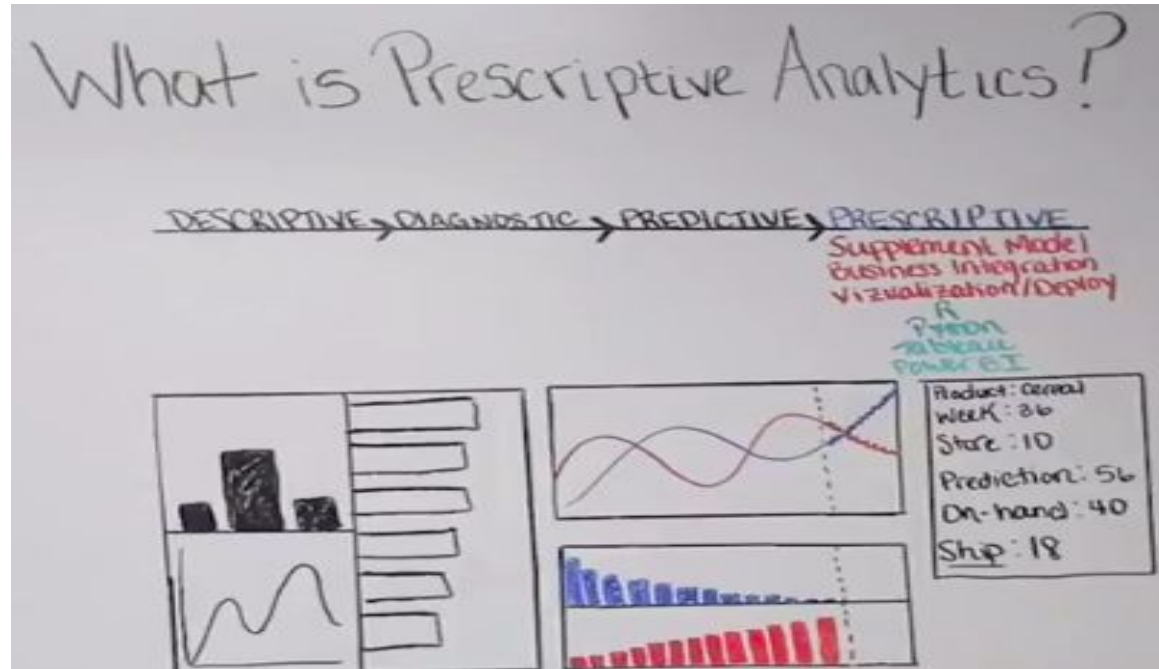
Simulation, optimization

Quali risultati otteniamo?



Prospettiva di business possibile

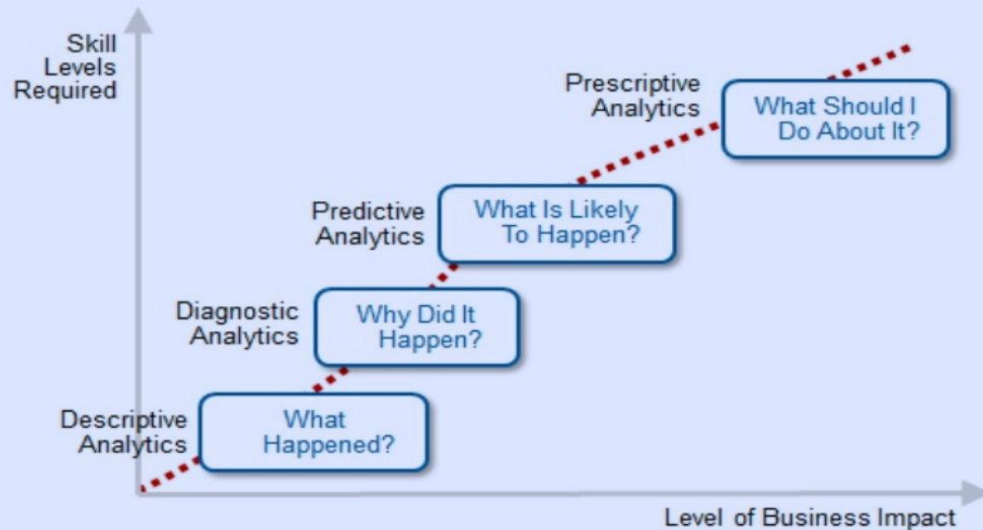
Esempio



<https://www.youtube.com/watch?v=Z9-0ei8HnYU>

The Four Types of Analytics

A cosa serve generare e raccogliere tanti dati?



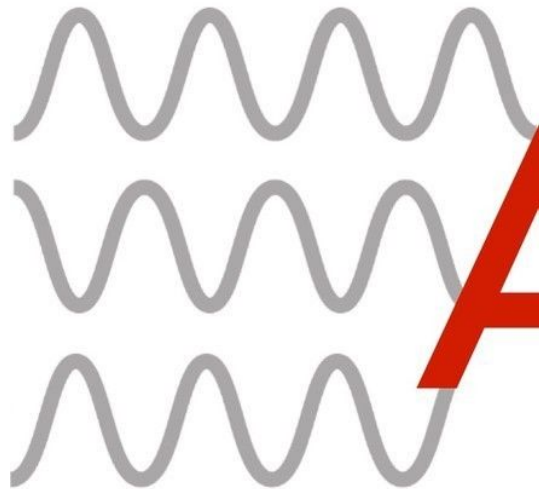
Source: Gartner (January 2013)



UNIVERSITÀ
DI TORINO

AUTOMATED
ANALYTICS

L'ULTIMA FRONTIERA DELL'ANALYTICS



AUTOMATED ANALYTICS

0001010010101011
0010010101001101
000101000101
010001001001
010110110010100
010010101011101
010101001101011

L'adozione degli Analytics nelle grandi imprese

OSSERVATORI.NET
digital innovation

DESCRIPTIVE ANALYTICS

100%



Strumenti orientati a descrivere la situazione attuale e passata dei processi aziendali e/o aree funzionali

PREDICTIVE ANALYTICS

73%



Strumenti avanzati che effettuano l'analisi dei dati per rispondere a domande relative a cosa potrebbe accadere nel futuro

PRESCRIPTIVE ANALYTICS

33%



Tool avanzati che, insieme all'analisi dei dati, sono capaci di proporre al decision maker soluzioni operative/strategiche sulla base delle analisi svolte

AUTOMATED ANALYTICS

11%



Tool capaci di implementare autonomamente l'azione proposta secondo il risultato delle analisi svolte

Campione: 159 organizzazioni italiane (>249 addetti)

Big Data is now; tomorrow is too late

22.11.17



#OBDA17



Network Digital360 - Events

- **Assicurazione auto** - pricing, analisi del rischio clienti, individuazione delle frodi, elaborazione più rapida dei sinistri.
- **Telecomunicazioni** - analisi dei modelli di servizi sui social network, redditività dei social network dei clienti, riduzione al minimo dei costi di produzione.
- **Distribuzione e vendita al dettaglio** - monitoraggio della disponibilità degli scaffali, valutazione dell'impatto dei display promozionali, valutazione dell'efficacia delle campagne promozionali, gestione dell'inventario, pricing, analisi avanzata del clickstream.
- **Trasporti e logistica** - gestione della flotta in tempo reale, RFID per il monitoraggio delle risorse.
- **Utilities** - analisi dei dati smart grid per determinare modelli di prezzo variabili, contatori intelligenti per prevedere la domanda di energia, piani tariffari personalizzati per i clienti.
- **Gaming** - analisi del gioco per fornire un feedback ai produttori di giochi, opportunità, offerte.
- **Forze dell'ordine** - identificazione di persone legate a gruppi di problemi noti, determinando la posizione di individui e gruppi individuazione di soggetti pericolosi.

VANTAGGI PRINCIPALI DELL'ANALISI DEI BIG DATA:

- **Vantaggio competitivo;**
- **Marketing social - influencer;**
- **Riconoscimento delle vendite;**
- **Numero maggiore e più accurato di intuizioni;**
- **Comprensione dei cambiamenti aziendali;**
- **Migliore pianificazione e previsione;**
- **identificazione delle cause dei costi;**
- **Rilevazione delle frodi;**
- **Quantificazione dei rischi**



PRINCIPALI BARRIERE:

- Inadeguatezza del personale;
- Difficoltà di progettare una grande sistema di analisi dei dati;
- Problema nel rendere i Big Data utilizzabili dagli utenti finali;
- Mancanza di sponsorizzazione;
- Costi complessivi elevati;
- Problemi di scalabilità dei Big Data;



PUNTI DI FORZA

- LE 4 V DEI BIG DATA
- ESPANSIONE DELLA PORTATA E DELLA PROFONDITA' DEI DATI
- MAGGIORE FLESSIBILITA' TECNICA

PUNTI DI DEBOLEZZA

- ASSENZA DI PROFESSIONISTI QUALIFICATI
- INADEGUATEZZA DELLE STRUTTURE
- ASSENZA DI UNA PROSPETTIVA CHIARA

OPPORTUNITA'

- PRODOTTI E SERVIZI PIU' «SMART»
- MARKETING PIU' EFFICACE
- RIDUZIONE DELLE FRODI
- PERSONALIZZAZIONE

MINACCE

- RIDUZIONE DELLA PRIVACY
- CORRELAZIONI ERRONEE DI DATI

grazie