

PIANO DEL SISTEMA CONFINDUSTRIA EMILIA-ROMAGNA “VERSO INDUSTRIA 4.0”

BIG DATA MANUFACTURING NELLA TRASFORMAZIONE DIGITALE DEI PROCESSI AZIENDALI

Bologna, 24 ottobre 2017



Operazioni Rif. PA. N.2016-5455/RER, Rif. PA. N.2016-5456/RER, Rif. PA. N.2016-5457/RER
coordinate con le Operazioni Rif. PA. N.2016-5452/RER, Rif. PA. N.2016-5453/RER, Rif. PA. N.2016-5454/RER
approvate dalla Regione Emilia-Romagna con DGR n. 1450/2016 del 12/09/2016
e finanziate con fondi POR FSE 2014/2020 - Obiettivo tematico 8

Big Data: tra mito e realtà

L'ANALISI DATI AL TEMPO DEI BIG DATA

Il Business Intelligence Group

Il Business Intelligence Group dal 1997 svolge ricerche legate alle metodologie, tecniche e Tecnologie nell'ambito dell'analisi dati.

- Attualmente include 5 ricercatori

Più nel dettaglio:

- Business Intelligence
- Data Warehouse
- Simulazione
- Pervasive BI
- Collaborative BI

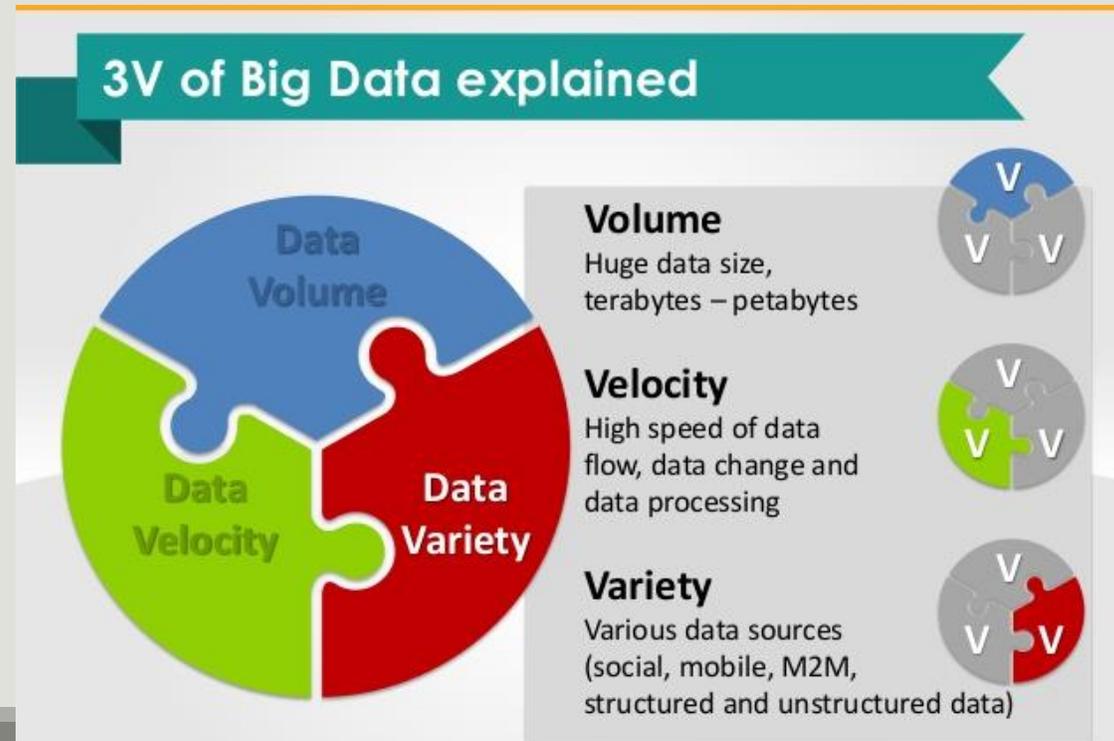
I nostri attuali temi di ricerca:

- Social BI
- Big Data & NOSql DBMS
- Semantic Data Warehousing
- Data mining



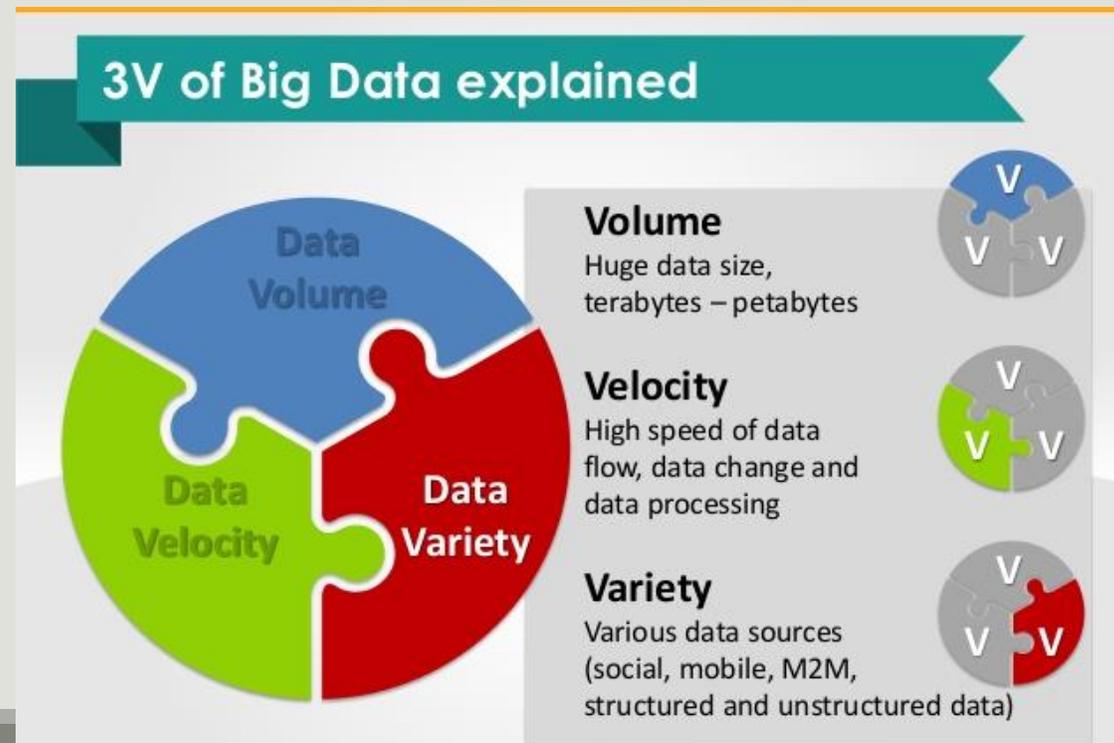
Big Data: una definizione

Big Data sono data set che hanno alcune delle seguenti caratteristiche, *ma non necessariamente tutte!*



Big Data: una definizione

Big Data sono data set che hanno alcune delle seguenti caratteristiche, *ma non necessariamente tutte!*



Mito: si parla di Big data solo quando si generano, trasferiscono, e analizzano enormi moli di dati

Big Data: volume

Volume: Terabyte o Petabyte così da eccedere la capacità di elaborazione dei sistemi tradizionali.

- Alcuni esempi:
 - Walmart: 1 milione di transazioni per ora (2010)
 - eBay: il data throughput raggiunge 100 petabyte al giorno (2013)
 - Facebook: 40 miliardi di foto (2010); 250PB data warehouse con 600TB aggiunti ogni giorno (2013)
 - 500 milioni di tweet al giorno (in 2013)

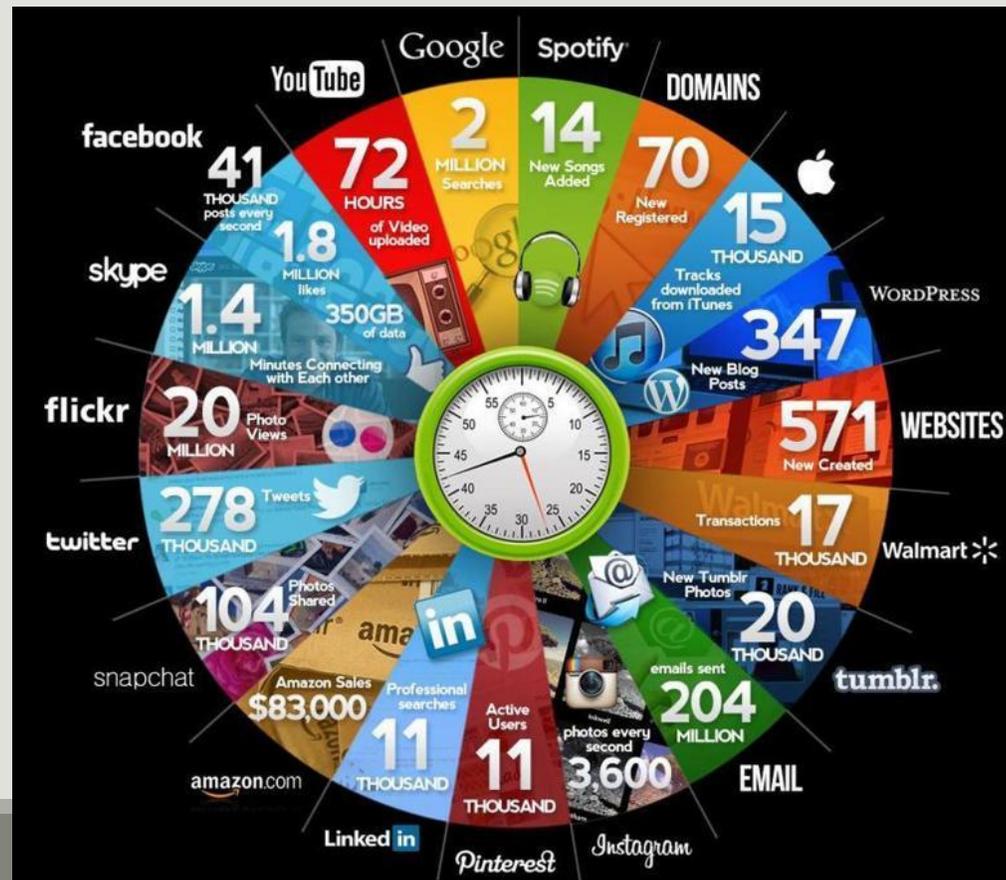
**L'hard disk di un desktop PC
memorizza in media 1 TB**

terabyte	TB	10^{12}	1 HD
petabyte	PB	10^{15}	1000 HD
exabyte	EB	10^{18}	1 milion of HD
zettabyte	ZB	10^{21}	1 bilion of HD

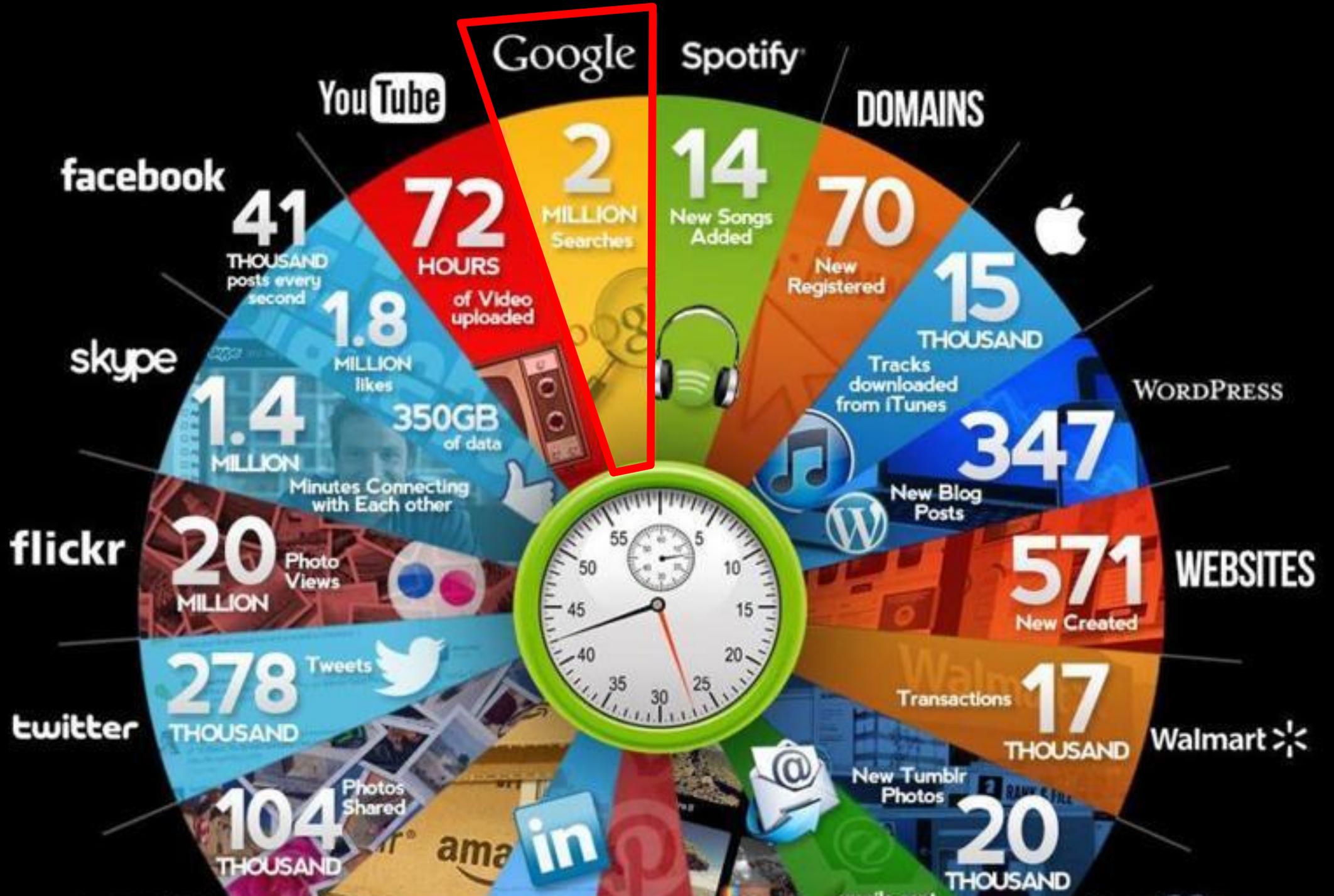
Big Data: velocity

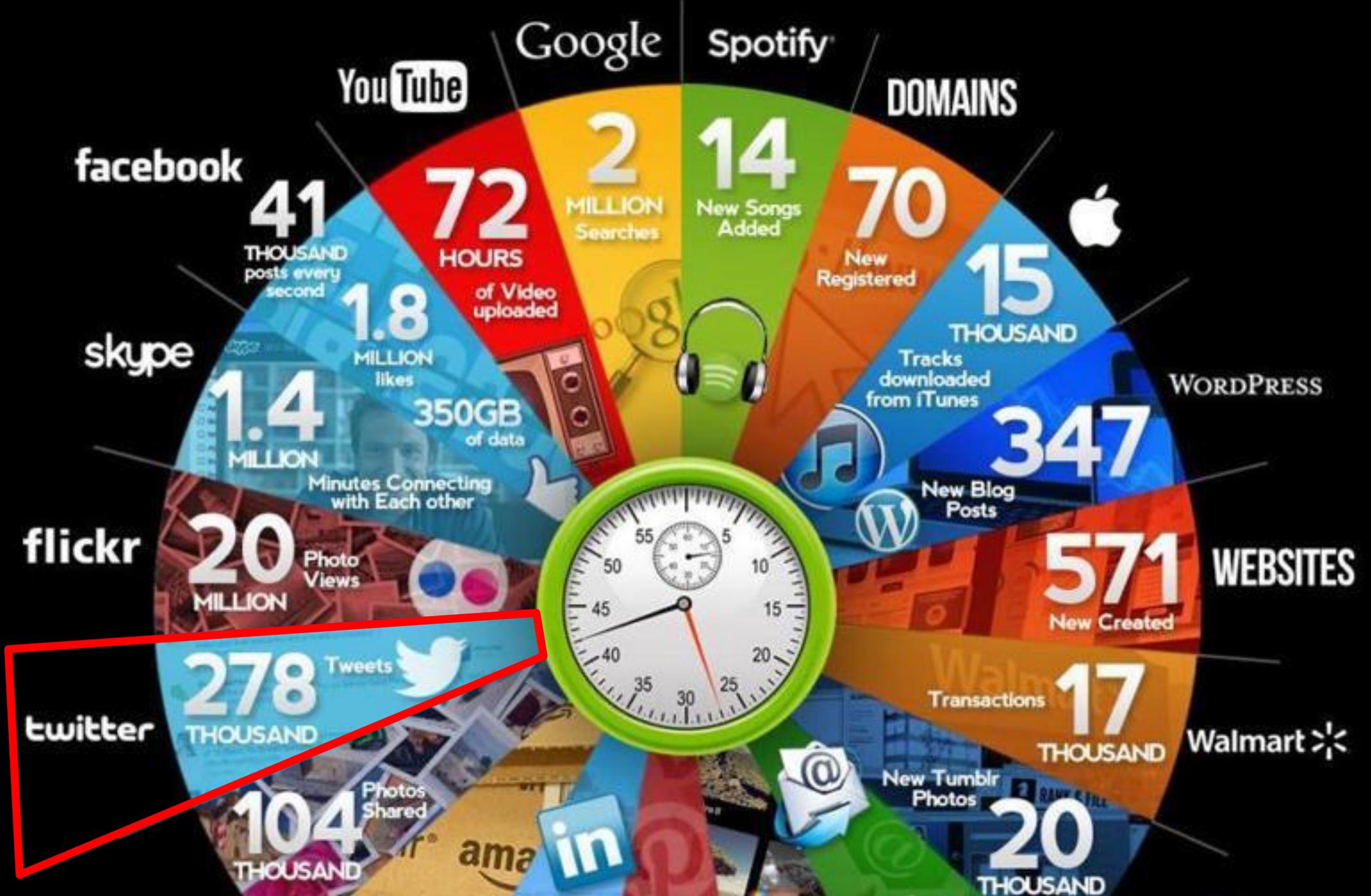
Velocità: dispositivi mobile e personali e transazioni IoT producono dati a una velocità maggiore dei sistemi informativi tradizionali

Cosa succede in internet in 60 secondi?





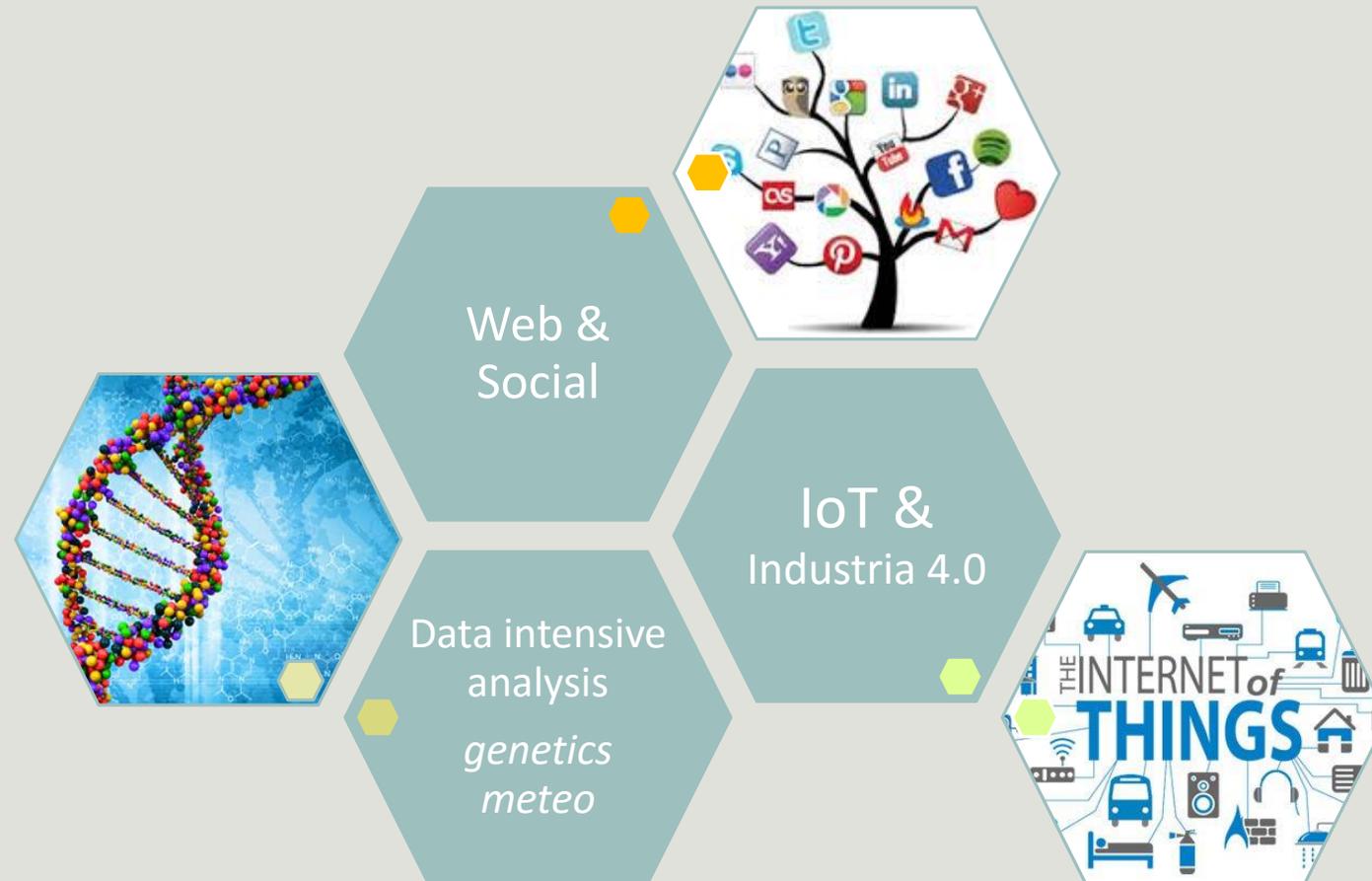




Big Data: Variety

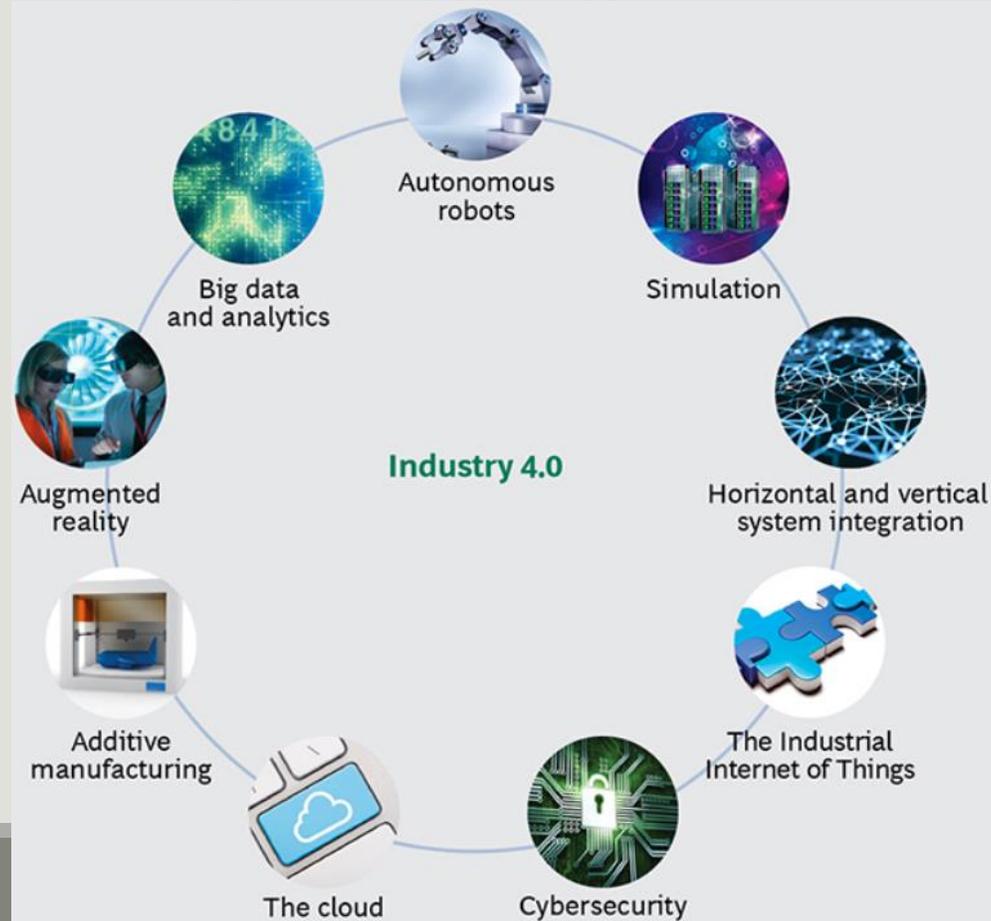
- **Varietà:** I dati sono estremamente eterogenei, nel format in cui sono rappresentati, ma anche nel modo in cui rappresentano le informazioni
 - E.g., testi dai social networks, dati di sensori, log di web application, database, XML documents, RDF data, etc.
 - Il format dei dati può essere **structured** (es. relational database) **semistructured** (es. XML document), **unstructured** (es. text document)
- **Veracità:** il livello qualitative dei dati è eterogeneo e in molti casi non è previsto uno schema ben definito
 - Il concetto di **schema less** evidenzia questa eterogeneità

Chi genera i big data?



Big Data e Industria 4.0

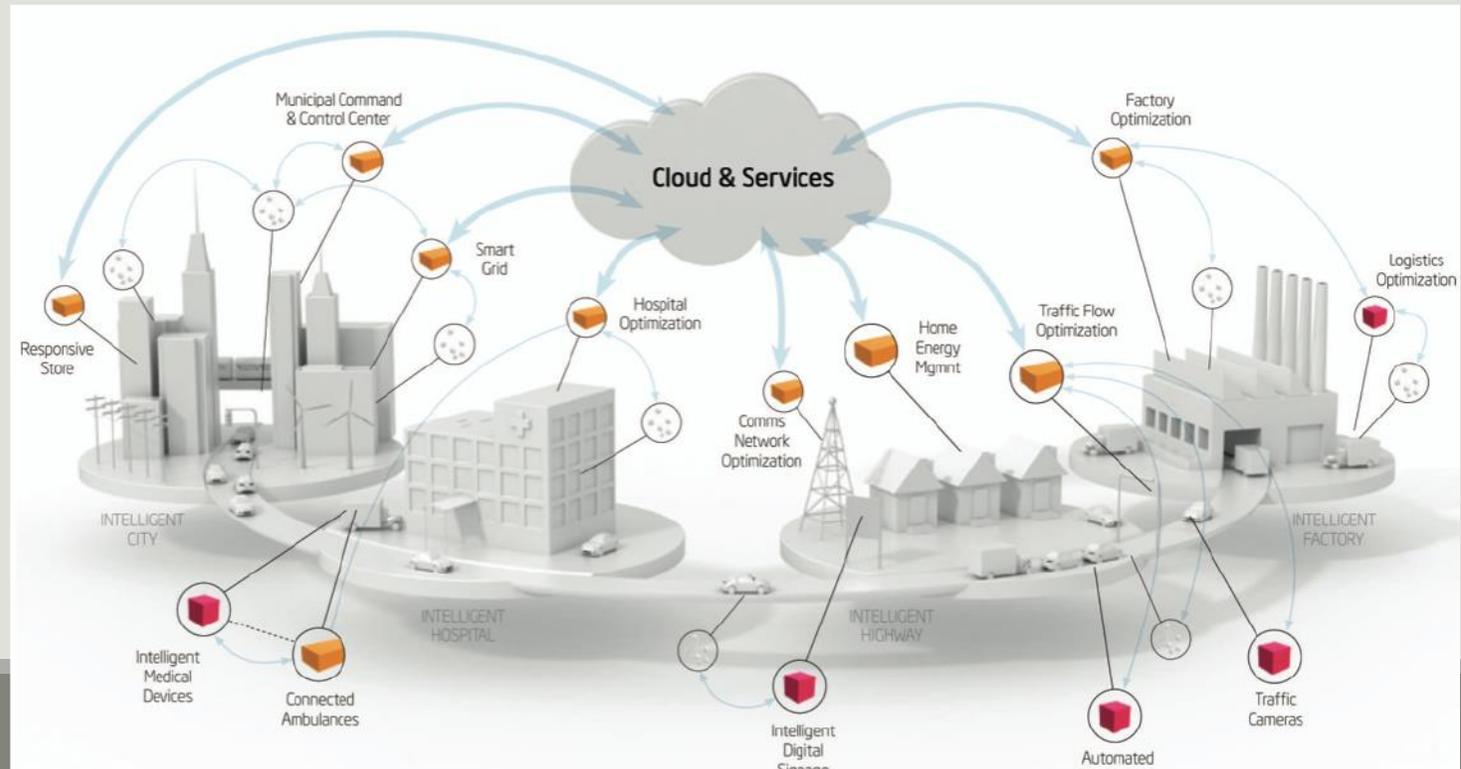
Sono considerati una delle tecnologie abilitanti, ma di fatto i dati Big (e Small) rappresentano la materia prima, il collante e il risultato di un processo o un'applicazione di Industria 4.0



Big Data e Industria 4.0

Sono considerati una delle tecnologie abilitanti, ma di fatto i dati Big (e Small) rappresentano la materia prima, il collante e il risultato di un processo o un'applicazione di Industria 4.0

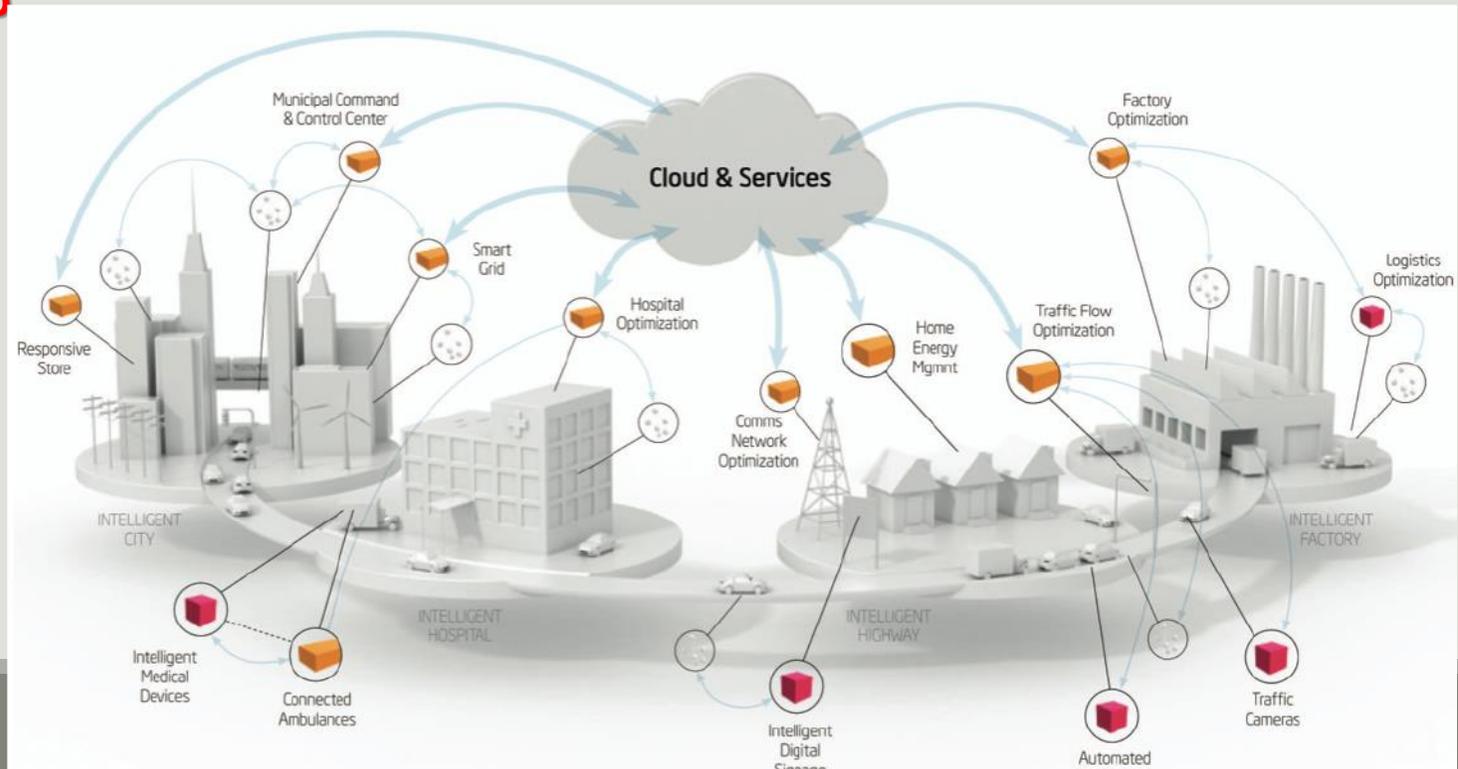
- I sensori producono dati
- L'interconnessione è basata sullo scambio dei dati
- Il monitoraggio dei sistemi è basato sui dati



Big Data e Industria 4.0

Sono considerati una delle tecnologie abilitanti, ma di fatto i dati Big (e Small) rappresentano la materia prima, il collante e il risultato di un processo o un'applicazione di Industria 4.0

- I sensori producono dati
- L'interconnessione è basata sullo scambio dei dati
- **Il monitoraggio dei sistemi è basato sui dati**



Una definizione per Data Science e Data Scientist

Data Science Indica la disciplina che mira ad estrarre conoscenza da grandi moli di dati utilizzando tecniche e tecnologie proprie della statistica e dall'information technology

Il **data scientist** è un esperto (o un gruppo di esperti) con approfondite conoscenze in suddetti ambiti e con una grande attitudine a comprendere il business e i dati. E' in grado di trasformare le informazioni nascoste nei dati in un vantaggio competitivo. Il suo obiettivo finale è quello di nuovi modelli di business nella **data-driven economy**

La Data Science è un'evoluzione della **Business Intelligence** della quale supera i seguenti limiti:

Business Intelligence	Data Science
Works on Structured data only	Works on Structured and unstructured data
Requires a complex/rigid ETL process	Is based on a more flexible/agile/extemporary extraction process
Implemented by Computer Engineers Exploited by Business Users and data analysts	Implemented and exploited by Data Scientists and Data Enthusiasts
Implemented on Relational DBMS	Implemented on Big Data and NoSQL technologies

Perché data science?



Esigenze di
business



Disponibilità
di grandi
quantità di
dati

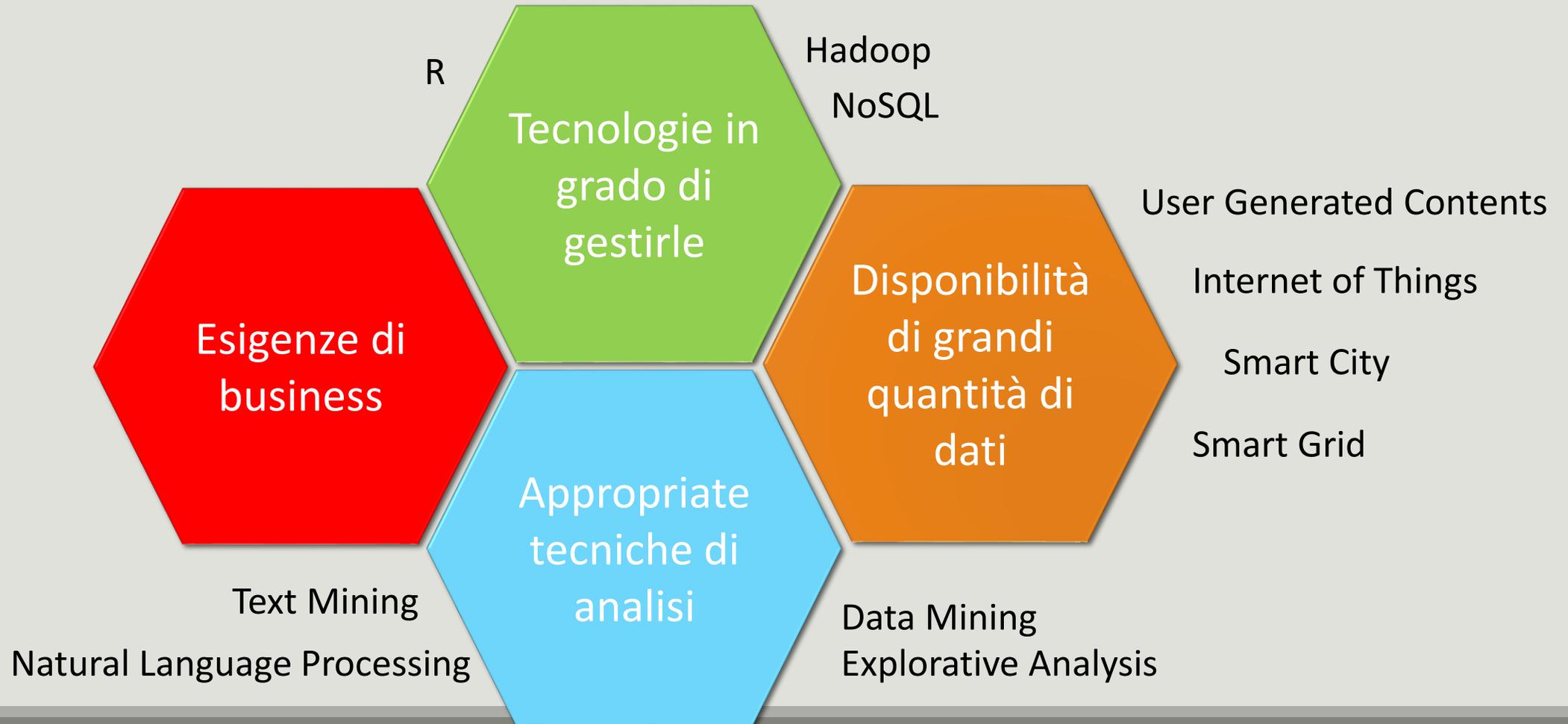
User Generated Contents

Internet of Things

Smart City

Smart Grid

Perché data science?



Big Data: il rinascimento della Business Intelligence

Il principale effetto del fenomeno dei Big Data è un rinnovato interesse per la Data Analysis. Mentre il gap tecnologico può essere colmato velocemente, colmare il gap culturale per la Data Analysis può richiedere anni (o addirittura una nuova generazione di manager).

- La maggior parte delle **tecnologie** legate ai Big Data sono **NUOVE**
- La maggior parte delle **tecniche** legate ai Big Data sono **note** e sono state adattate alle nuove tecnologie

Big Data: il rinascimento della Business Intelligence

Il principale effetto del fenomeno dei Big Data è un rinnovato interesse per la Data Analysis. Mentre il gap tecnologico può essere colmato velocemente, colmare il gap culturale per la Data Analysis può richiedere anni (o addirittura una nuova generazione di manager).

- La maggior parte delle **tecnologie** legate ai Big Data sono **NUOVE**
- La maggior parte delle **tecniche** legate ai Big Data sono **note** e sono state adattate alle nuove tecnologie

Mito: le tecniche legate ai Big data sono una *scoperta recente*

Analytics

[Wikipedia] **Analytics** indica i sistemi software finalizzati alla scoperta, interpretazione e comunicazione di pattern interessanti nei dati. Gli analytics si basano sull'utilizzo simultaneo di applicazioni di statistica, programmazione e ricerca operativa. Gli analytics utilizzano spesso tecniche avanzate di visualizzazione dei dati.

Gli **analytics** stanno alla BI 2.0 come il **data mining** sta alla BI 1.0

I sistemi di Data Mining hanno avuto una diffusione molto inferiore a quelli di Data Warehousing a causa della:

- complessità e costi di realizzazione elevati
- necessità di un esperto per l'interpretazione dei risultati
- mancanza di certezza del risultato

Analytics

Gli analytics stanno avendo una diffusione decisamente maggiore grazie a:

- Maggiore disponibilità dei dati
- Maggiore potenza di calcolo
- Maturazione della cultura aziendale e aumento della pressione competitiva
- **Forte attenzione alla user-experience**
- **Soluzioni maggiormente industrializzate e pacchettizzate**
 - Riduzione dei costi di realizzazione
 - Aumento della semplicità di fruizione
 - Maggiore focus sul problema

Adottare una soluzione di Data Analytics

Immagina

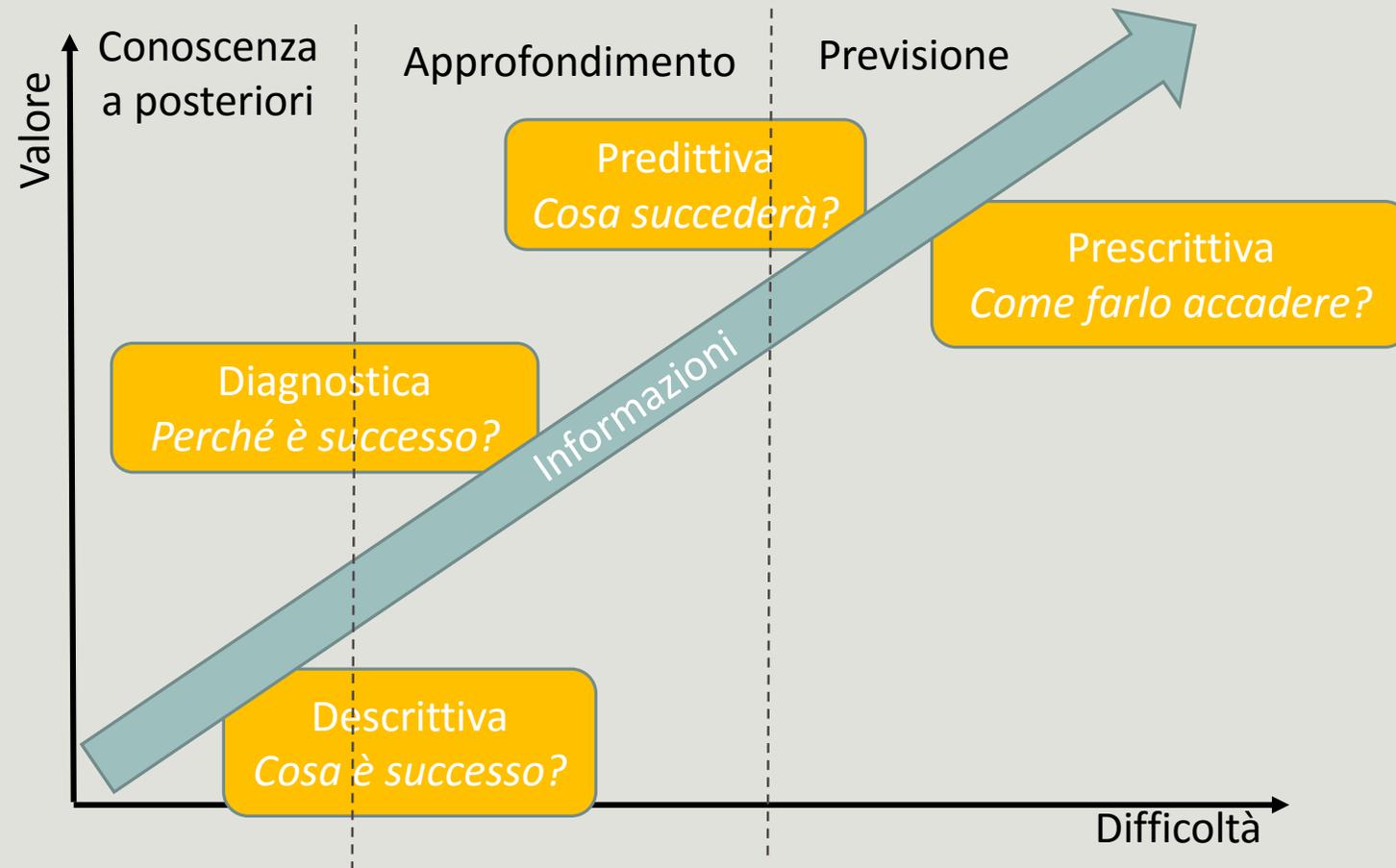
Trasforma

Implementa

Quali applicazioni Big Data sono applicabili al mio Business?

Quali impatti avranno sul mio business?

Quale risposta mi serve?



Manutenzione di attrezzature

Manutenzione Failure-driven (run-to-failure maintenance)

Può essere molto costosa se eseguita in emergenza oppure quando il down-time implica alti costi oppure se implica alti costi per il mantenimento di un magazzino di materiali di ricambio

Manutenzione Time-based (preventive maintenance)

- Riduce il rischio di effettuare manutenzione in emergenza
- Può essere difficile stimare il time-to-failure in condizioni di funzionamento non stazionarie

Manutenzione Condition-based reduce l'incertezza sulla manutenzione basandosi sulle condizioni di utilizzo delle attrezzature

Un caso di studio sui Big Data Analytics: manutenzione predittiva di attrezzature

Stimare il tempo mancante al prossimo malfunzionamento funzionale sulla base sintomi rappresentati dei dati raccolti in tempo reale dai radar (stato e malfunzionamenti minori)



Stato1 Fault2 Stato3 Stato4 Stato5 Stato6 Fault2 Stato7

Remaining Useful Life Estimator

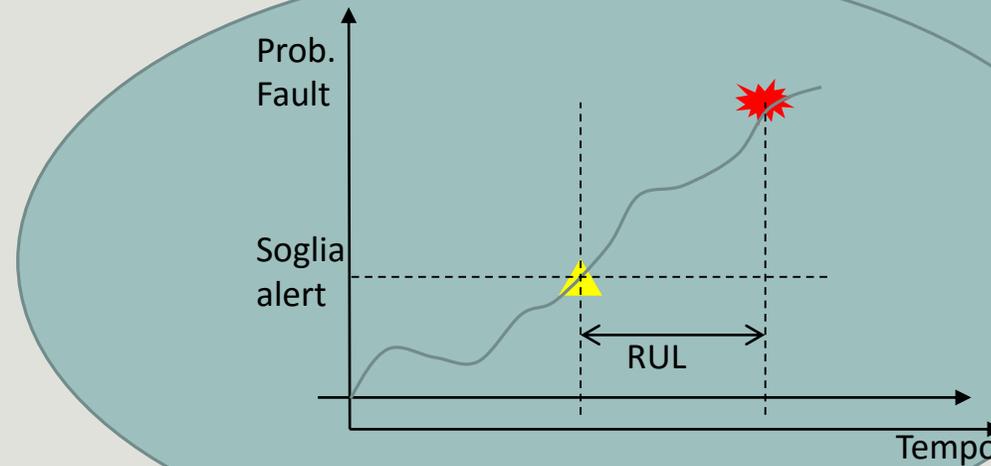


Manutenzione predittiva di attrezzature

Stimare il tempo mancante al prossimo malfunzionamento funzionale sulla base *sintomi* rappresentati dei dati raccolti in tempo reale dai radar (stato e malfunzionamenti minori)



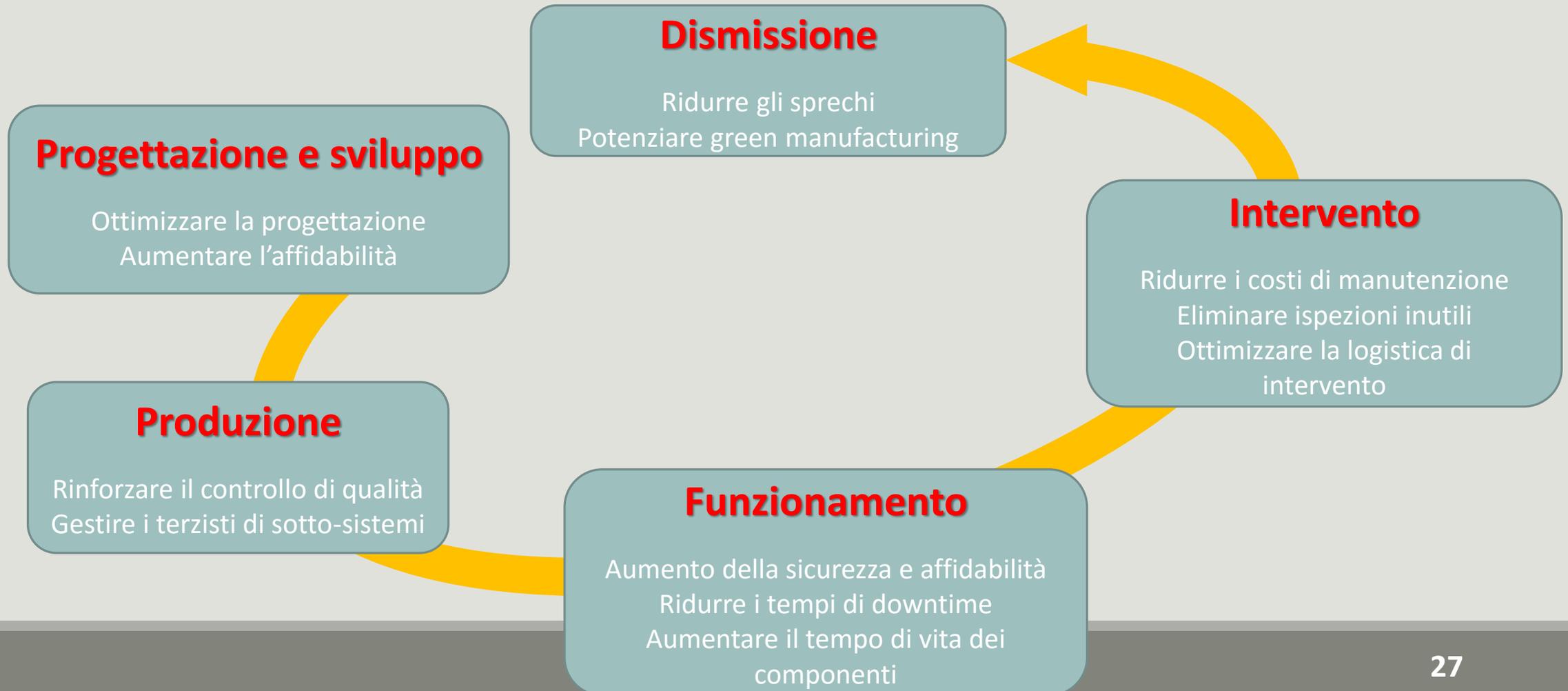
Stato1 Fault2 Stato3 Stato4 Stato5 Stato6 Fault2 Stato7



Remaining Useful Life Estimator



Manutenzione predittiva di attrezzature



Fasi della manutenzione predittiva

1) **Acquisizione e processing dei dati**

- Quali variabili monitorare?
- Quali sensori usare?
- Quali trasformazioni effettuare sui dati per renderli utilizzabili?

2) **Diagnosticare un mal funzionamento**

- Finalizzata a comprendere quali valori delle variabili monitorate determinano un mal funzionamento
- Distinguere tra diversi tipi di mal funzionamento e relativa *severity*

3) **Predire un mal funzionamento**

- E' molto più complesso predire un mal funzionamento rispetto a riconoscerlo

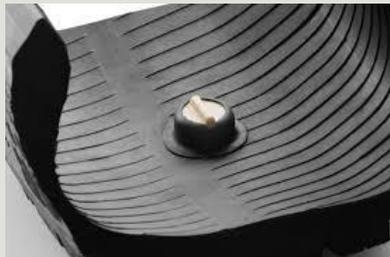
Predizione di mal funzionamenti

1. **Expert-based**: in assenza di una case history la conoscenza degli esperti di dominio può essere utilizzata per modellare le condizioni di previsione (DSS);
2. **Artificial intelligence**: ci si basa sulla storia dei mal funzionamenti per costruire un modello di previsione. Esistono molte tecniche in AI per ottenere questo risultato
 - Si assume che ci sia una relazione tra mal funzionamenti tecnici minori (che sono individuati) e mal funzionamenti funzionali (che devono essere prevenuti)
3. **Model-based**: inferiscono un comportamento anormale in quanto si differenzia da un comportamento normale che è completamente modellato in base alle caratteristiche fisiche del sistema
 - Costruire un modello fisico accurate può essere difficile e costoso

Le domande da porsi

Per chi produce il sistema

- Le machine sono già dotate di sensori?
- Esiste una case history dei guasti?
- Quali servizi possono essere venduti ai clienti?
 - Manutenzione predittiva
 - Valutazione della modalità di utilizzo



- Cambio del modello di business da prodotto a servizio

Per chi utilizza il sistema

- Qual è il ROI?
- Quali altre trasformazioni/investimenti si rendono necessari

Adottare una soluzione di Data Analytics

Immagina

Trasforma

Implementa

Quale percorso è necessario per riuscire a realizzare l'applicazione?

Quali skill servono per realizzare il cambiamento?

Il progetto Big Data in Technogym

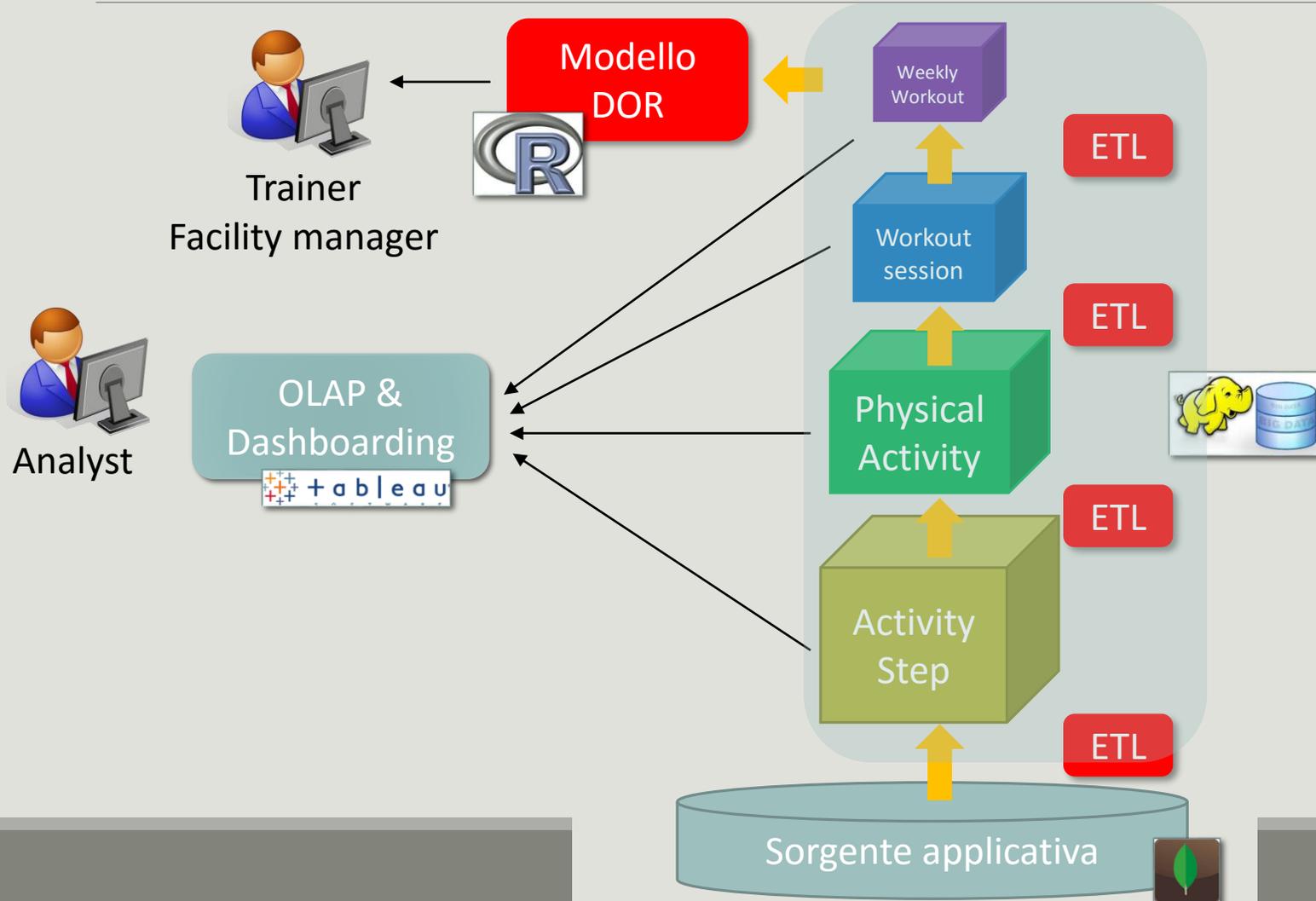
Technogym raccoglie oggi i dati di ogni singola attività svolta sui propri macchinari. Tale possibilità nasce in risposta a un processo di digitalizzazione dei suoi sistemi

1. Aggiunge software e sensori alle proprie attrezzature che sono stand-alone o connesse a livello di palestra
2. Costruisce un ambiente social (social network, Apps) fornendo ulteriori servizi ai propri clienti
3. **Sposta il proprio software nel cloud**

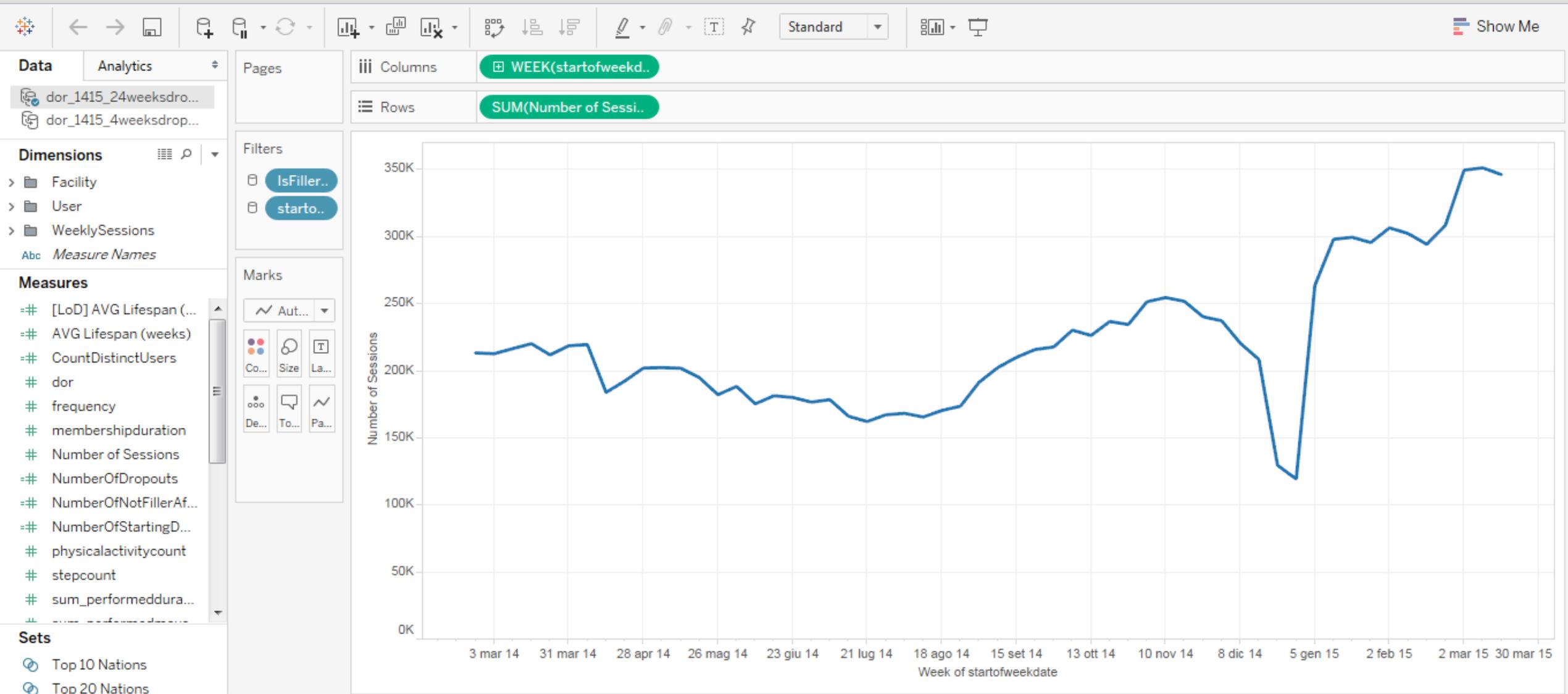


Miliardi di dati con il dettaglio della singola Physical Activity Step diventa così disponibile per:
Profilare i clienti, Analisi scientifiche, Prevenzione di abbandoni, etc.

Architettura della soluzione



- Quali sono gli esercizi di tipo StrengthLoad più di utilizzati tra le donne nelle diverse nazioni?
- Quale applicazione è maggiormente utilizzata da un certo tipo di utente/In una certa nazione/in un certo tipo di Facility?



Data | Analytics

dor_1415_24weeksdro...
 dor_1415_4weeksdro...

Dimensions

- Facility
- User
- WeeklySessions
- Measure Names

Measures

- [LoD] AVG Lifespan (...)
- AVG Lifespan (weeks)
- CountDistinctUsers
- dor
- frequency
- membershipduration
- Number of Sessions
- NumberOfDropouts
- NumberOfNotFillerAf...
- NumberOfStartingD...
- physicalactivitycount
- stepcount
- sum_performeddura...

Sets

- Top 10 Nations

Pages

Columns | WEEK(startofweekd..)

Rows | Nation (Facility) | SUM(Number of Sessi..)

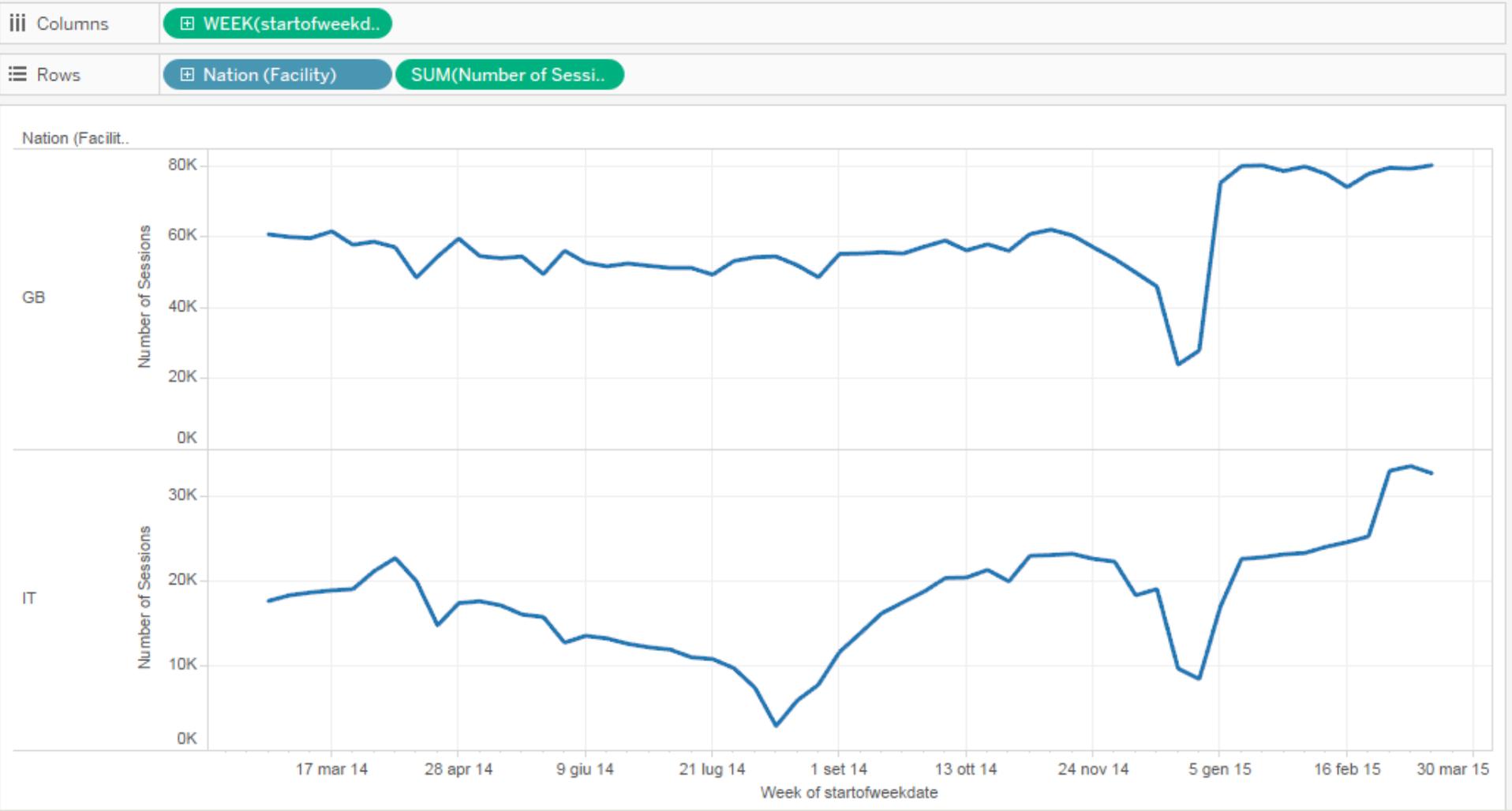
Filters

- IsFillerAfte..
- Nation (Fac..)
- startofwee..

Marks

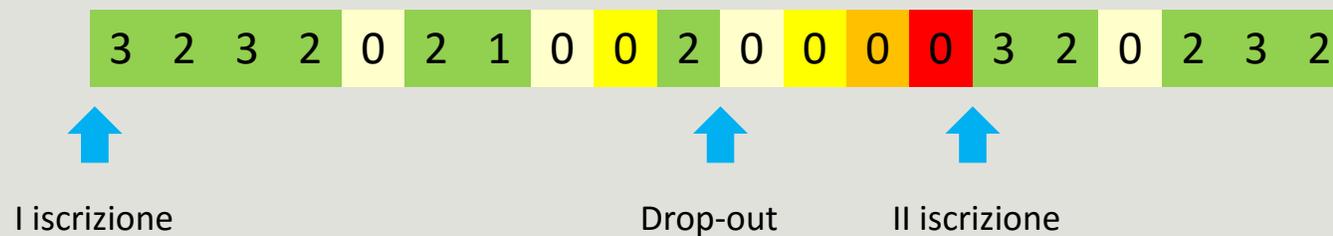
Automatic (dropdown)

Color, Size, Label, Detail, Tool..., Path



Definizione del concetto di Drop Out

- In mancanza di comunicazione da parte dell'utente si ha un drop-out se l'utente non si allena per X settimane in palestra
 - $X = 4$ il cliente potrebbe allenarsi in modo diverso (estate/vacanza) o essere ammalato
 - $X = 24$ il cliente non ritorna più



- Oltre alla frequenza di allenamento per ogni utente e per ogni settimana sono registrate una serie di feature

Adottare una soluzione di Data Analytics

Immagina

Trasforma

Implementa

Qual è lo stack tecnologico più adatto a realizzare l'applicazione?

*Esistono partner tecnologici affidabili?
Sono mature le tecnologie necessarie?*

Le tecnologie dei Big Data

Da una recente statistica è emerso che:

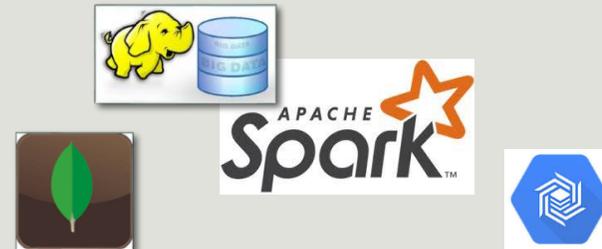
- Il 72% dei CEO hanno affermato che i prossimi 5 anni porteranno a cambiamenti tecnologici più impattanti di quelli degli ultimi 50 anni
- Solo il 5% delle aziende si sente pronta ad affrontare i cambiamenti tecnologici

Molte tecnologie non sono ancora mature ma la loro velocità di crescita è maggiore di quella delle aziende ad acquisirle e ogni tecnologia ne abilita altre

- Nel 2018 primi test del 5G ai Mondiali in Russia + Olimpiadi invernali in Corea → Realtà Virtuale

Altri sono maturi ma in continua evoluzione

- Machine learning
- Tecnologie per i dati e il calcolo
 - Hadoop
 - Spark
 - MongoDB



Prof. Matteo Golfarelli (Phd)
Computer Science & Engineering
University of Bologna
Tel: +39 0547 338 862
e-mail: matteo.golfarelli@unibo.it
skype: matteo.golfarelli

www: <http://bias.csr.unibo.it/golfarelli/>
BIG: <http://big.csr.unibo.it/>

