

Università degli Studi di Venezia – Corso di Laurea in Informatica  
Codice Insegnamento: CT0323



**Social media web e smart apps (A.A. 2014/2015)**  
**Gianluigi Cogo**

Lezione 12: Big Data

# di cosa stiamo parlando



# Big Data è un tormentone!

Nell'approccio ‘folkloristico’  
sentiamo dire spesso:  
**‘creeranno milioni di posti di  
lavoro!’**

# Ma l'Italia non è l'America!

Quando diciamo: '**tutto ciò succederà tra poco!**'  
dovremmo contestualizzare: **dove?**



# I fondamentali

Dato e Informazione non sono sinonimi!

Il dato è un elemento basico (o informazione grezza) ed è spesso costituito da simboli non ancora elaborati.



# I fondamentali

L'informazione è un elemento più ricco, che deriva dall'elaborazione di più dati e che restituisce un valore, solitamente **consapevolezza, comprensione dei fatti e verità.**



L'informazione è il risultato di un'elaborazione dati.

# I fondamentali

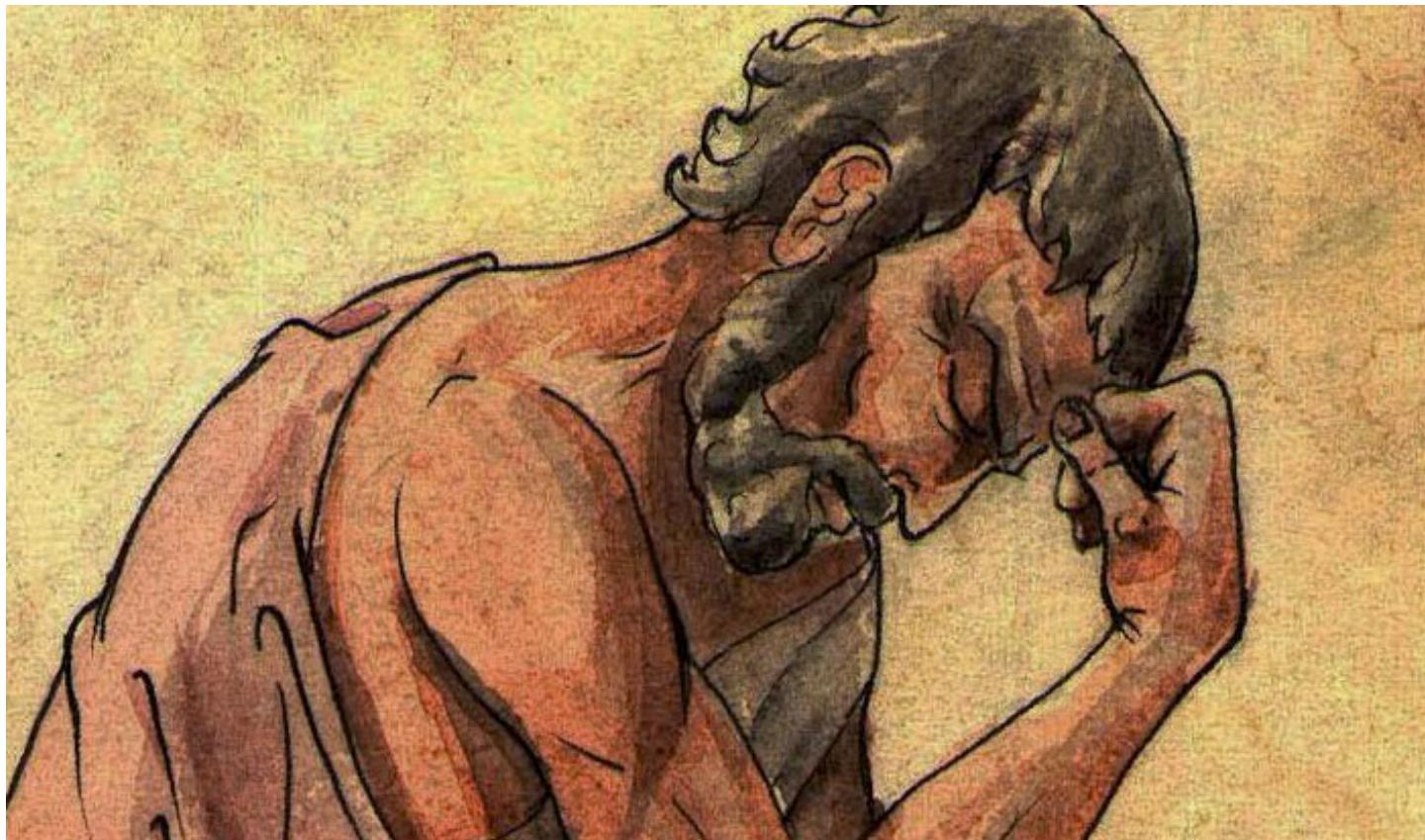
La **CONOSCENZA** è l'autocoscienza del possesso di informazioni connesse tra di loro, le quali, prese singolarmente, hanno un valore e un'utilità inferiori.



# I fondamentali

Il tutto è maggiore della somma delle sue parti!

(Aristotele – l'inventore dell'approccio sistematico)



# I fondamentali

Il concetto di Big Data è proprio del campo dei **database**: il termine indica grandi aggregazioni di dati, la cui mole richiede strumenti differenti da quelli tradizionali, in tutte le fasi del processo (dalla gestione, alla **curation**, passando per condivisione, analisi e visualizzazione).



(Wikipedia)

# I fondamentali

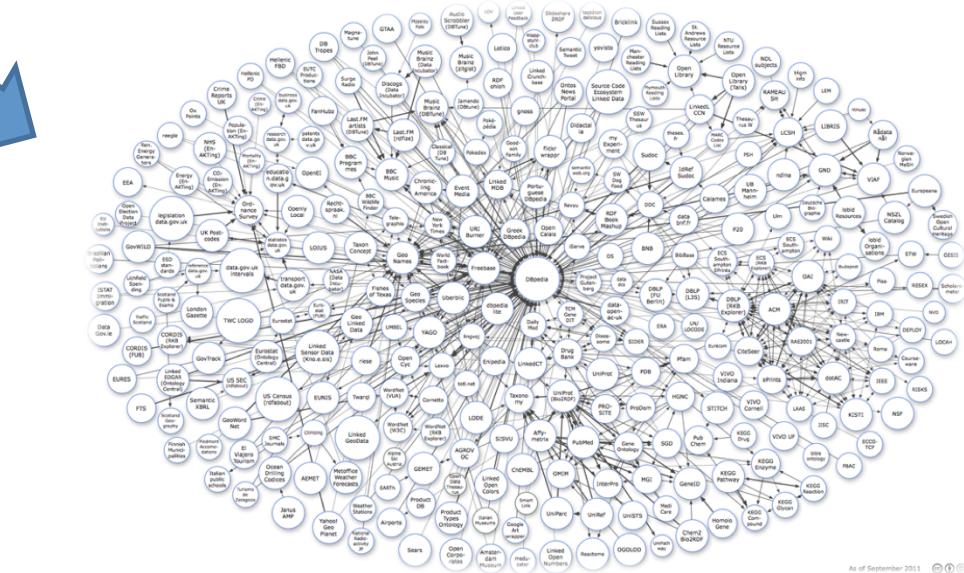
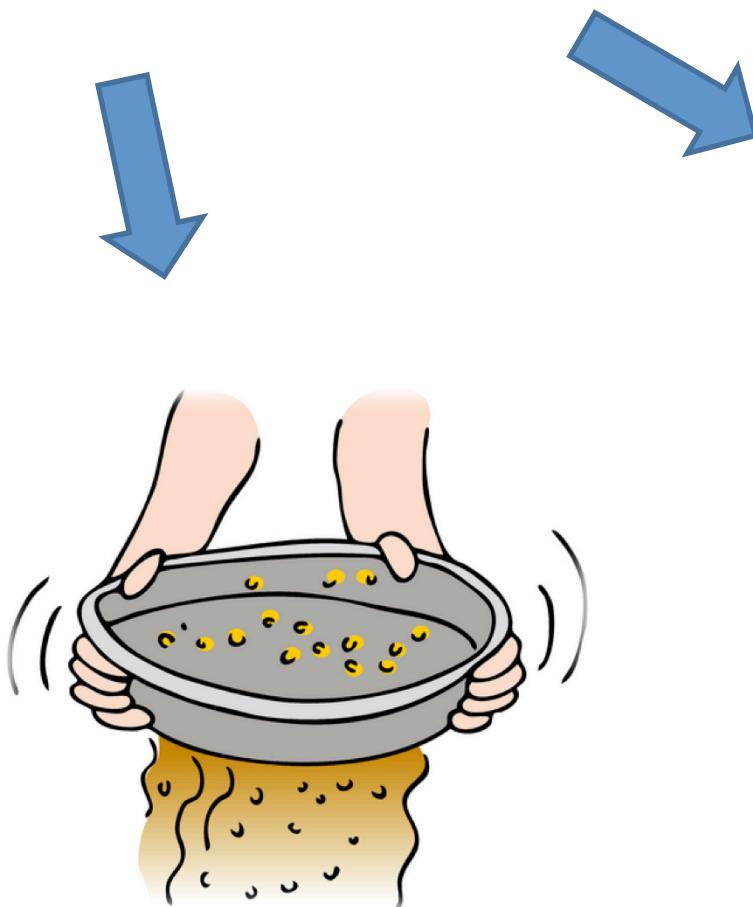
Il progressivo aumento della dimensione dei **dataset** è legato alla necessità di analisi su un unico insieme di dati correlati rispetto a quelle che si potrebbero ottenere analizzando piccole serie con la stessa quantità totale di dati ottenendo informazioni che non si sarebbero potute ottenere sulle piccole serie.



(Wikipedia)

## I fondamentali

Due termini relativamente nuovi:  
**curation** e **dataset**!



As of September 2011

# DEFINIZIONI



<http://ed.ted.com/lessons/exploration-on-the-big-data-frontier-tim-smith/>

# DEFINIZIONI

I Big Data sono l'elemento fondamentale per la creazione di **nuovi livelli di valore per il business**. Grazie a storage integrato, analisi e applicazioni, i Big Data contribuiscono a migliorare efficienza, qualità, prodotti e servizi personalizzati, producendo livelli più elevati di soddisfazione ed esperienza del cliente. (EMC<sup>2</sup>)



# DEFINIZIONI

I Big Data sono la grande, enorme massa di dati di cui dispongono oggi le aziende, che costituiscono un **problema se non utilizzati** o usati poco o male, ma che possono trasformarsi in una formidabile opportunità quando vengono sfruttati nel modo corretto. (SAS)



# DEFINIZIONI

Le aziende sono sommerse di dati più che mai. **L'informazione che può fare la differenza per il vostro business è nascosta** in questa mole di dati.

L'analisi dei Big Data, vi aiuterà a trasformare i vostri dati, apparentemente senza significato e sconnessi tra di loro, in informazioni utili creando il vantaggio competitivo. (R. Jacobs).



# DEFINIZIONI



Explaining Big Data



[http://youtu.be/7D1CQ\\_LOizA/](http://youtu.be/7D1CQ_LOizA/)

# DEFINIZIONI

Hot on the heels of Web 2.0 and cloud computing, Big Data may well be the **Next Big Thing in the IT world**. Whereas Web 2.0 links people and things online, and cloud computing is about the transition to an online computing infrastructure, Big Data generates value .....  
**(CONTINUA)**



# RIFLESSIONI

Oggi le aziende devono essere in grado di utilizzare pienamente **tutte** le risorse di dati.

Purtroppo i dati **non strutturati** vengono integrati in quelli strutturati o, molto spesso, **nemmeno soggetti a raccolta** e tantomeno a conservazione.



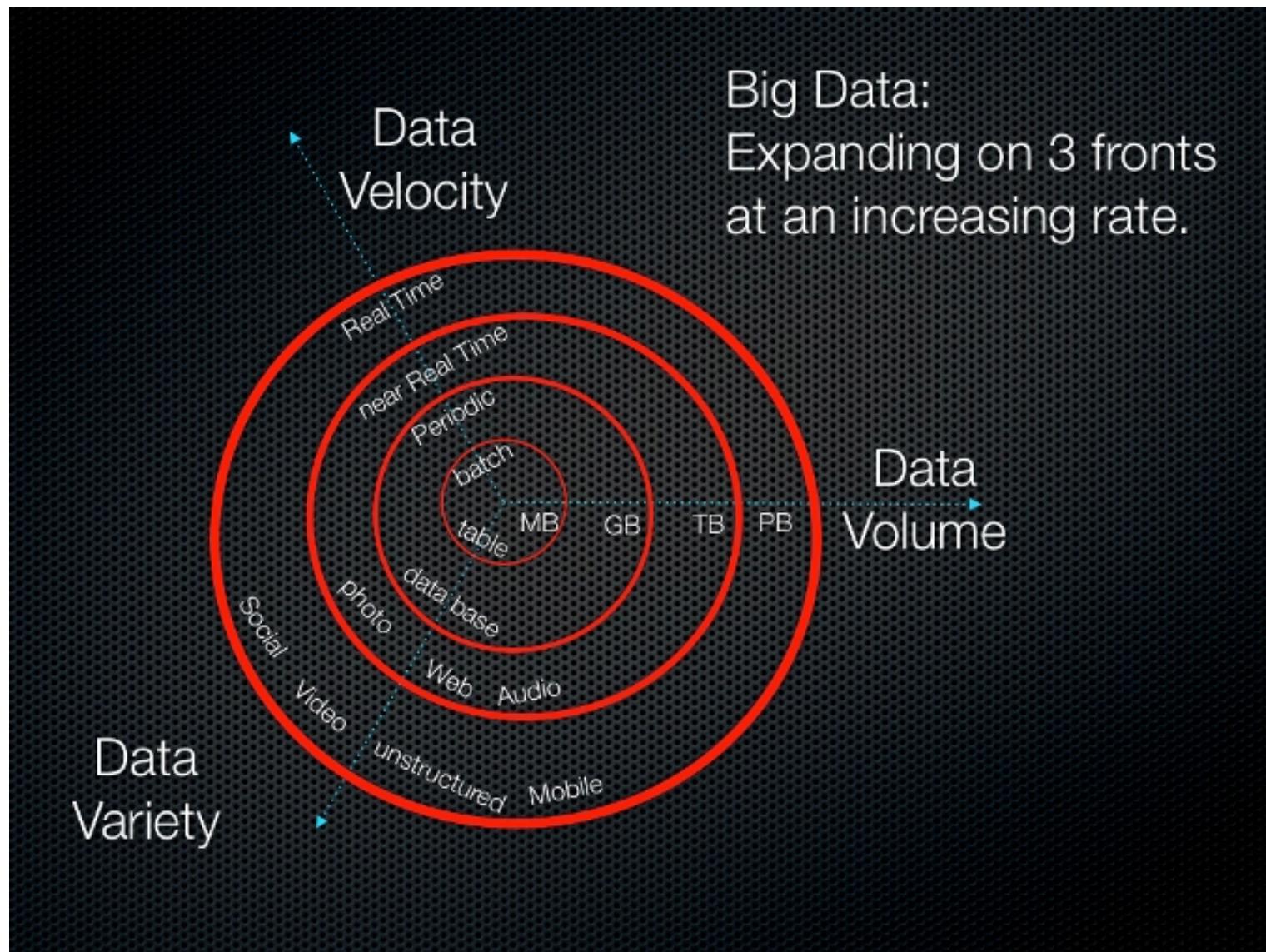
# RIFLESSIONI

L'aumento del **volume, velocità e varietà** dei dati spesso supera la reale capacità delle aziende di gestirli ed elaborarli con efficacia nei tempi utili. Una complessità che rende difficile far fronte alle sempre più urgenti e crescenti esigenze del business.

Il paradigma delle **3V** riassume l'impatto dei big data sulle aziende (SAS)



# 3V = le proprietà dei Big Data



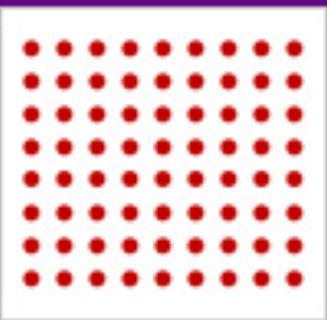
**Volume:** la mole di dati, spesso destrutturati, aumenta in maniera esponenziale. Diventa sempre più difficile individuare per tempo quelli a maggior valore per il business (**Brontobyte?**).

**Varietà:** la tipologia di dati non è più uniforme e legata solo ai sistemi legacy. Ci troviamo di fronte a dati in formato testuale, audio, video, streaming, provenienti da blog, web e social network (**social-unstructured-data > enterprise-structured-data**).

**Velocità:** i dati vengono prodotti con una velocità e frequenza sempre maggiore. Il "time to decision" richiesto all'IT si sta riducendo sempre di più. La sfida è quella di riuscire a gestire ed elaborare informazioni in tempi sempre più rapidi.

**+ Valore o Veridicità:** i modelli analitici sono sempre più complessi e impongono capacità elaborative fino a poco tempo fa impensabili. Diventa determinante sapere individuare i dati a valore rispetto agli altri .

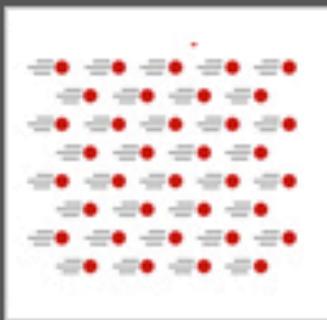
## VOLUME



Data at Rest

Terabytes to exabytes of existing data to process

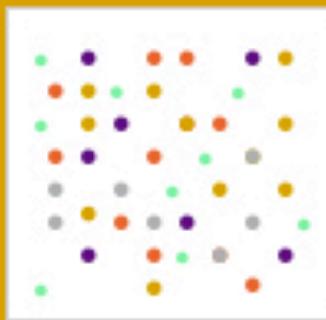
## VELOCITY



Data in Motion

Streaming data, milliseconds to seconds to respond

## VARIETY



Data in Many Forms

structured,  
unstructured,  
text, multimedia

## VERACITY



Data in Doubt

Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception and model approximations



Executive  
Dashboards

Enterprise  
Search

Customer  
Interaction

Predictive  
Analytics

Web  
Engagement

## NEXT GEN INFORMATION PLATFORM



Variety

Velocity

Volume

Terabyte

10,000,000,000,000

Petabyte

10,000,000,000,000,000

Exabyte

10,000,000,000,000,000,000

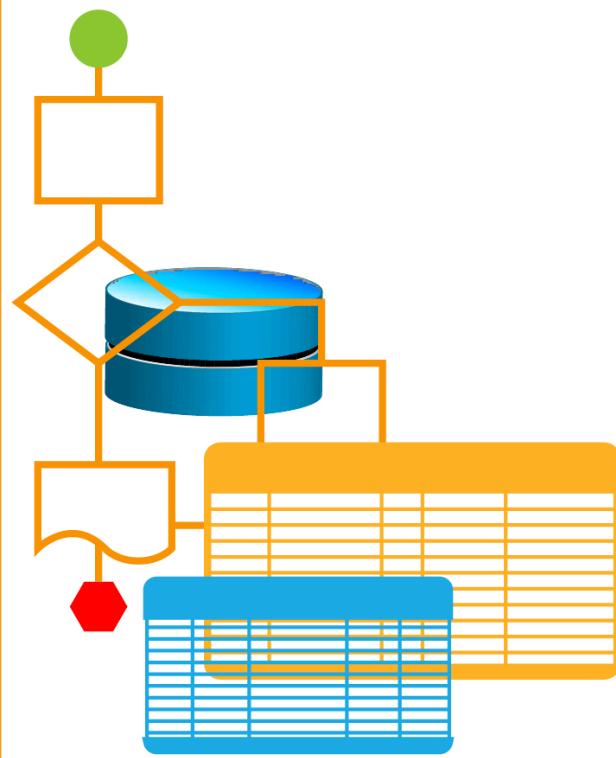
Zettabyte

10,000,000,000,000,000,000,000

Yottabytes

10,000,000,000,000,000,000,000,000

## ENTERPRISE STRUCTURED DATA



New Applications that run on Cloud2.0 have to assimilate, integrate and provide most relevant information to Business

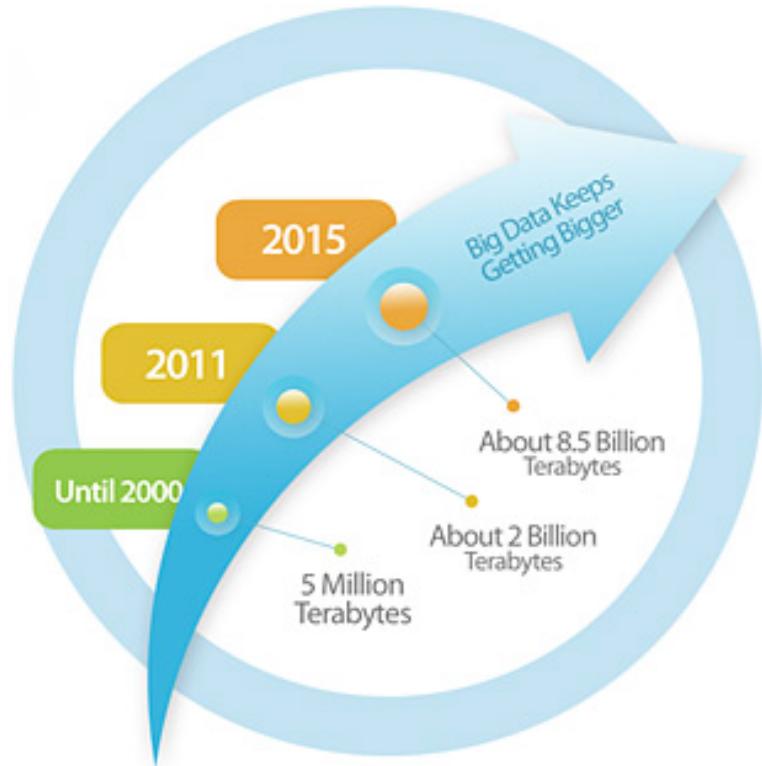
## SOCIAL UNSTRUCTURED DATA



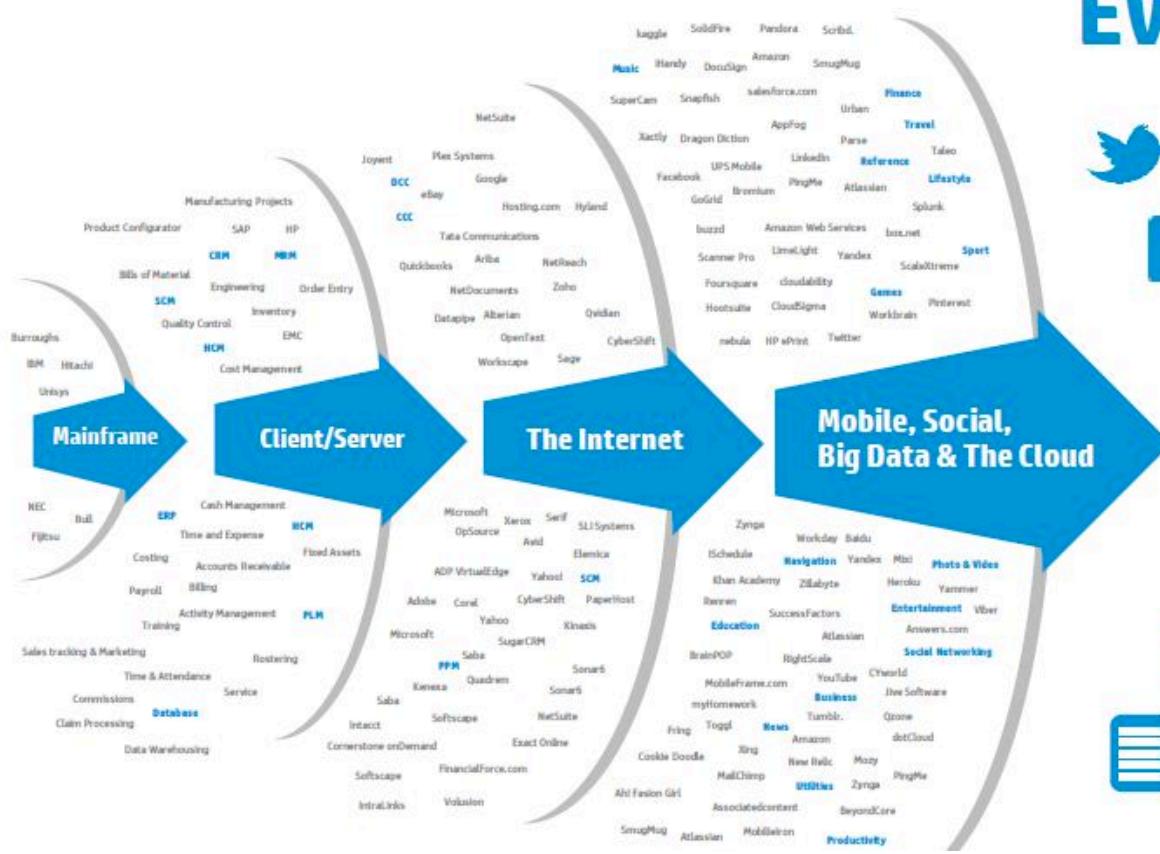
Nel 2000 il 75% delle informazioni era raccolto sulla carta, sulla plastica magnetica, su altri supporti analogici e solo il 25% era in digitale.

Nel 2013, l'analogico è ridotto al 2% mentre il 98% delle informazioni è registrato in digitale.

I dati digitali raddoppiano ogni tre anni.



# A new style of IT emerging



# Every 60 seconds

 98,000+ tweets

 **695,000** status updates

 **11million** instant messages

 698,445 Google searches

 **168 million+** emails sent

**1,820TB** of data created

 217 new mobile web users

# What Happens in an Internet Minute?



And Future Growth is Staggering



Le decisioni che un'organizzazione deve prendere, oggi possono (o meglio **devono**) essere basate anche sui feedback dei clienti, sui report di stato, sulle valutazioni delle prestazioni e non solo sui dati demografici e operativi.



MOLTI VANTAGGI PER LE AZIENDE SONO DUNQUE CONSEGUENZA  
DIRETTA DELLA LORO CAPACITA' DI PREDIZIONE



L'analisi dei dati è un processo di **ispezione**, **pulizia**, **trasformazione** e **modellazione** con il fine di evidenziare informazioni che suggeriscano conclusioni e **supportino** le decisioni strategiche aziendali.



[http://it.wikipedia.org/wiki/Analisi\\_dei\\_dati](http://it.wikipedia.org/wiki/Analisi_dei_dati)

L'analisi predittiva permette alle aziende (o meglio alle organizzazioni) di capire cosa succederà nel futuro e reagire di conseguenza.



ANSWERS  
QUESTIONS

Prevedendo cosa accadrà nel futuro si potranno pianificare e portare avanti strategie che supportino e migliorino il processo decisionale.



## (Data Mining con l'ausilio delle tecnologie)

è il processo di estrazione  
di conoscenza da banche dati  
di grandi dimensioni tramite  
l'applicazione di algoritmi che  
individuano le associazioni  
“nascoste” tra le informazioni  
e le rendono visibili.

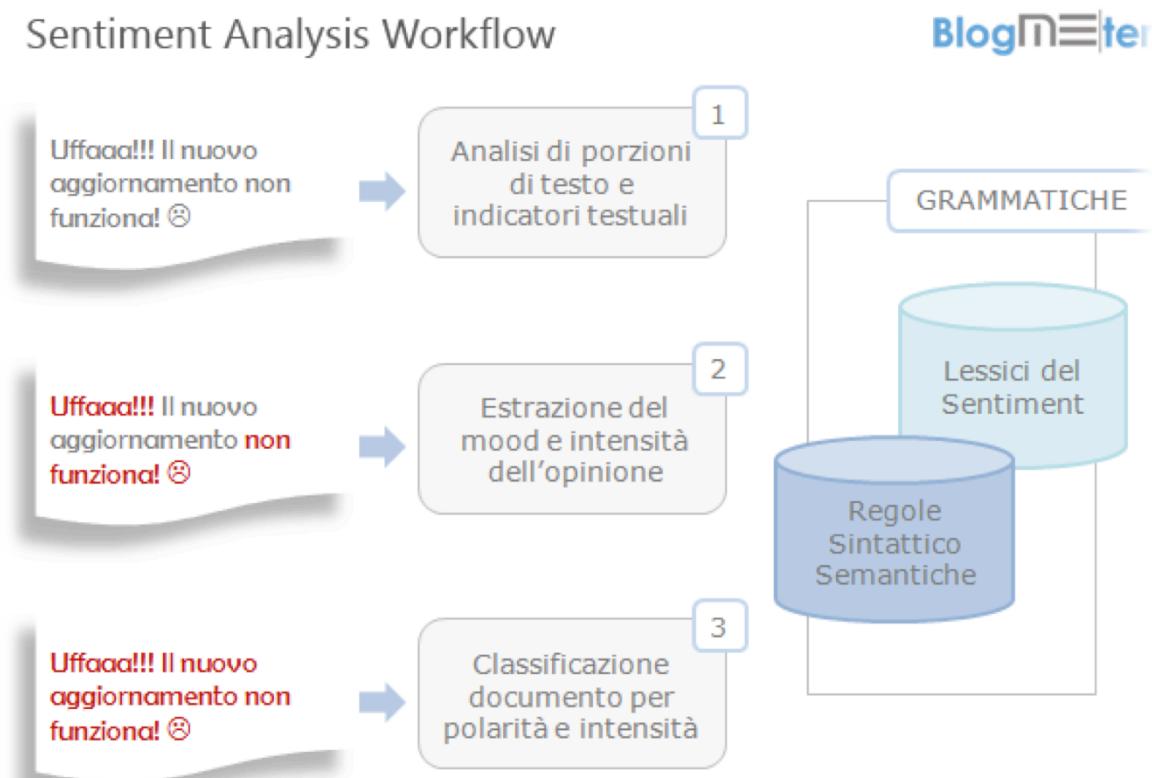


Viene anche detto: 'Knowledge Discovery in Databases' (KDD)

L'analisi predittiva con la '**Sentiment Analysis**' prevede l'analisi qualitativa delle conversazioni in rete e mira a comprendere lo stato d'animo degli utenti rispetto un particolare brand, prodotto, tema, servizio. Viene anche detta: '**Opinion Mining**'



Prendendo in esame le **conversazioni degli utenti** nei diversi spazi della rete (blog, forum, social network) si può determinare come è percepito e considerato un determinato brand o prodotto e orientare le strategie di comunicazione future di conseguenza a queste analisi.



Dunque i dati sono solo **numeri**?

- quanti utenti per quel servizio
- quanto prodotti venduti in quel luogo
- quante chiamate al call-center per assistenza
- ecc.



Il perimetro dei dati, per effetto della consumerization, diventa molto più **ampio!**



E con il mobile  
diventa  
**infinito!**

Verizon 3:31 PM 76%

Top Picks  
Bushwick, New York

THIS PLACE IS ON A LOT OF TO-DO LISTS



**Tortilleria Mexicana Tres Her...**  
TACOS  
9.0 130 ft

Try the chorizo taco at this Bushwick tortilla factory. It's one of NYC's 26 best tacos!

Time Out New York

Friends who have been here:



+11

More popular places nearby >

Friends Explore Talisa

# Social media e consumerization cambiano tutto!

## Social Media Explained with Beer

**facebook**

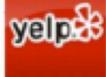
- I like beer

**twitter**

- I am having a #beer

**foursquare**

- this is where I drink beer



- you will like the beer here

**YouTube**

- here I am having a beer

**LinkedIn**

- my skills include beer

**Instagram**

- here is a photo of my beer



- listening to a song about beer

**Pinterest**

- here are beers that I like

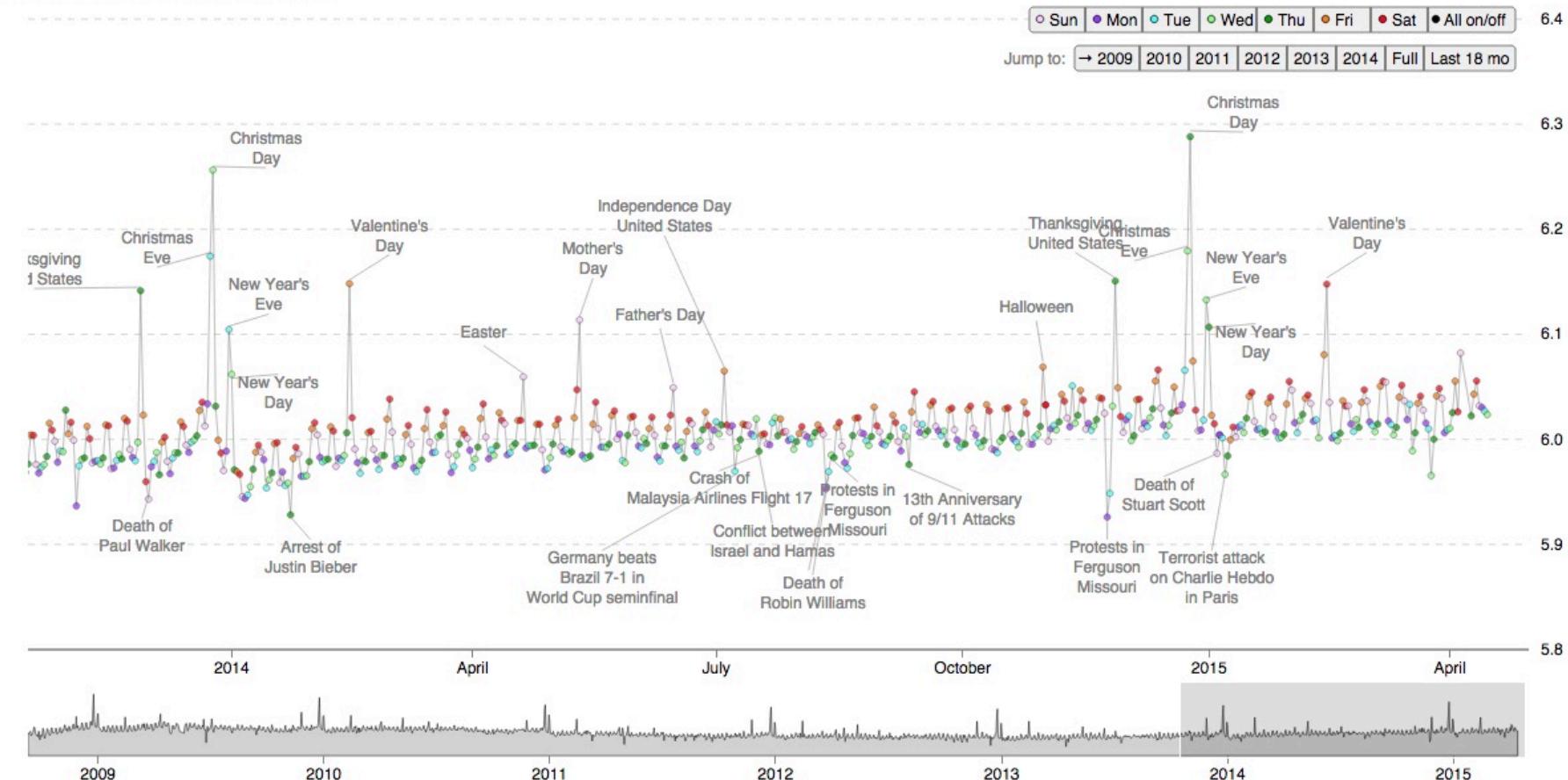


- find half priced beer here

# I social media cambiano tutto!

projects ▾ About Instructions Words Blog Press Papers Talks API Fund us

## Average Happiness for Twitter



<http://hedonometer.org/index.html>

# I social media cambiano tutto!

## La classifica degli ultimi 7 giorni

La città mediamente più felice, quella a metà classifica e quella più arrabbiata degli ultimi 7 giorni. Per l'intera classifica degli ultimi 7 giorni seguire [questo link](#)

FORLI'



86.2%

TORINO



65.6%

VICENZA



26.2%

## Mappe giornaliere della felicità

Le mappe seguenti rappresentano la felicità aggregando i dati provinciali giornalieri per regione e si riferiscono alle sole 24 ore precedenti.

Venerdì



Sabato



Domenica



Lunedì



Martedì



Mercoledì



Giovedì



## Mappa della felicità del 16 aprile

<http://www.blogsvoices.unimi.it/index.html>

# I social media cambiano tutto!



# Esempio: Stasoft (Dell)



<https://www.youtube.com/watch?v=Ga2jMY5nzzY/>

# Esempio: LiveHelpNow



<https://www.youtube.com/watch?v=N5mE-LzIKCI>

# Esempio: digital advertising

Se fino ad oggi il digital advertising era una pratica approssimativa e spesso si procedeva per tentativi, nell'ultimo decennio si è evoluto in un metodo scientifico che permette di realizzare campagne personalizzate per un pubblico targettizzato.



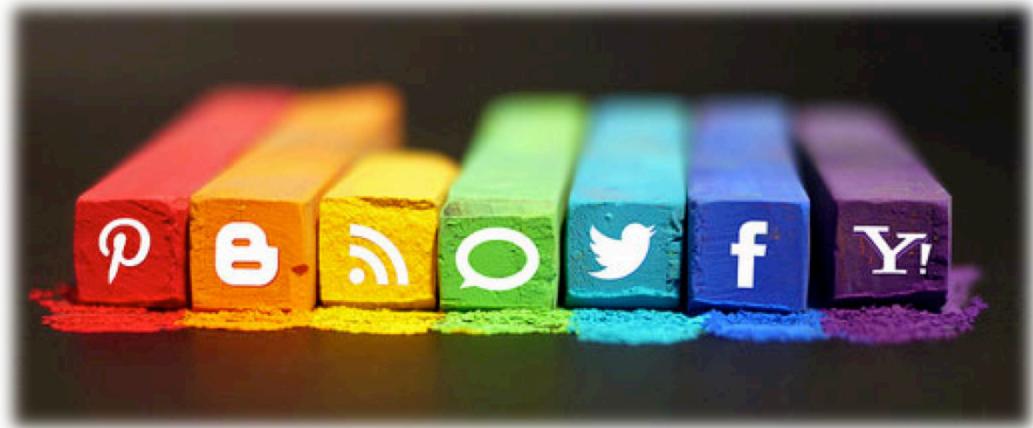
Ogni azienda dispone ormai di enormi quantità di dati relativi ai propri clienti: da quelli delle transazioni raccolti quando un cliente acquista un prodotto online o in store, a quelli acquisiti nel momento in cui un cliente contatta il call center. Per non parlare poi dei dati raccolti dalle newsletter, dalle email, da mobile, ecc.



Questi dati rappresentano una fonte preziosissima di informazioni e costituiscono il punto di partenza per conoscere a fondo e meglio il proprio pubblico/target. Dunque basta capire come poter gestire le informazioni che si hanno a disposizione al fine di ottimizzarle e allinearle agli obiettivi di business per ottenere il miglior ritorno possibile.



L'esempio di Turn, piattaforma leader nel cloud marketing, permette di riassumere l'intero processo fornendo cinque linee guida utili a capire chi sono gli utenti con cui si sta interagendo e come rapportarsi con loro non perdendo d'occhio l'obiettivo finale di intraprendere campagne di successo e replicabili.



- 1. La consapevolezza:** prendere atto dei dati di cui si dispone
- 2. La struttura:** creare un ritratto completo del pubblico
- 3. La strategia:** coinvolgere il pubblico con una comunicazione mirata
- 4. La replicabilità:** ampliare il mercato attraverso un modello Look Alike
- 5. La conclusione:** analizzare, ottimizzare, ripetere





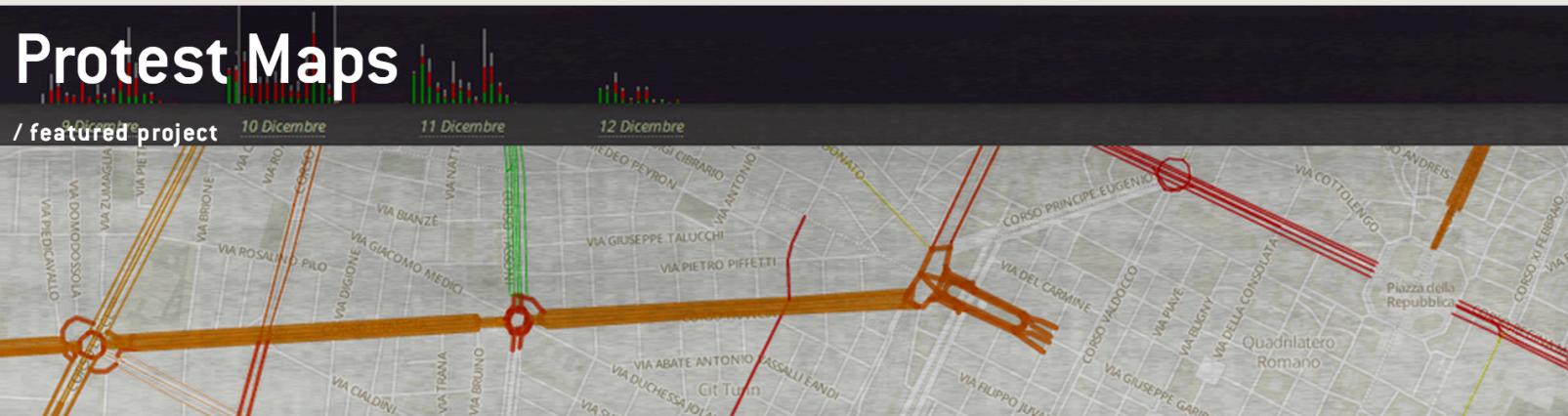
# The Marketer's Challenge



# Esempi: Traffico e ordine pubblico

## DATAINTERFACES

NEWS PROJECTS ABOUT TEAM



[About »](#) [Team »](#)

[Projects »](#)

[News »](#)

### About

Data Interfaces is a research laboratory that merges the competences of communication design, complex systems science, and computer science in the creation of interfaces between data and people.

### Twitter Topic Explorer

[/ project](#)

### Datalninterfaces at Art Verona

[/ news](#)

<http://www.datainterfaces.org/2014/03/protest-maps/>

# Esempi: Polizia investigativa



<http://www.bbc.com/news/technology-22008497>

# Lezione per le imprese

I Big Data aiutano

- ad accelerare il marketing
- favorire la fidelizzazione con l'utente



# Cosa serve?

- Capacità di raccogliere dati
- Capacità di organizzare dati
- Capacità di analizzare dati
- Capacità di reagire



E tutto ciò si fa **ANCHE** con le tecnologie, non **SOLO!**

# Campi applicativi

## Applicazioni tipiche del Big Data Analytics



Smarter Healthcare



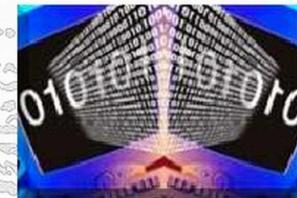
Multi-channel Sales



Finance



Log Analysis



Homeland Security



Traffic Control



Telecom



Search Quality



Manufacturing



Trading Analytics



Fraud and Risk



Retail: Churn, NBO



# Chi vincerà?

Chi saprà integrare **dati strutturati** e **dati non strutturati** e coordinarli/organizzarli per favorire un miglioramento/cambiamento



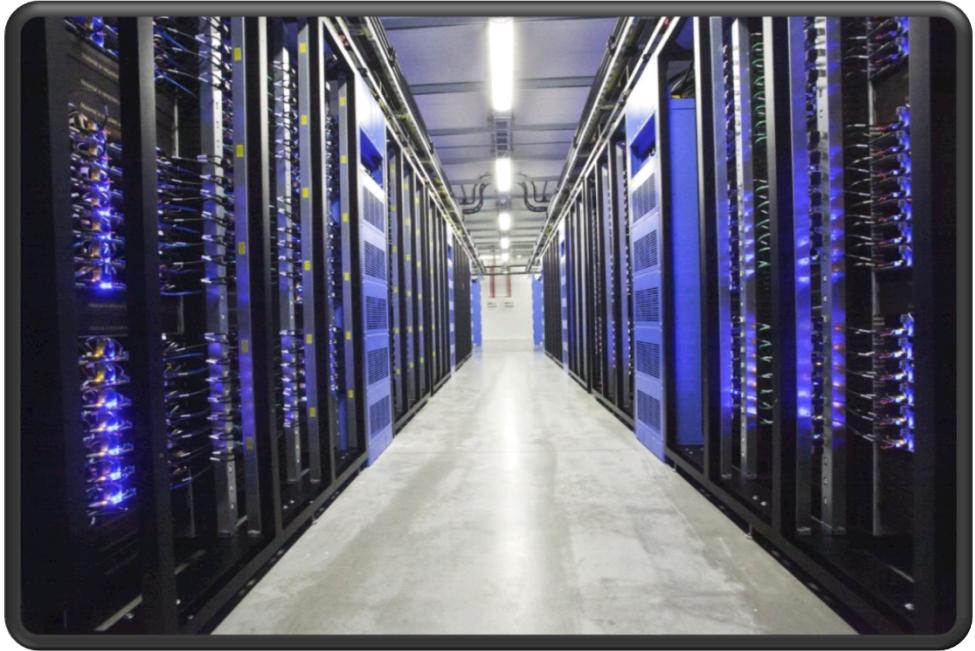
# Tecnologie

Dove stanno i dati?

Chi è obbligato a raccoglieri?

Perchè tenere anche quelli non obbligatori?

UNICO DATO CERTO E' CHE I **DATI CRESCERANNO** E DA  
QUALCHE PARTE VANNO RACCOLTI



# Problemi della tecnologia

- Costi (storage costa meno ma ne serve di più)
- Spazi per i data center .... CLOUD ECONOMY!
- Nuove tecnologie come I dischi SSD ....
- Consumi di energia?



# Big Data Landscape

## Vertical Apps



MYRRIX

## Log Data Apps

splunk &gt; loggly + sumologic

## Ad/Media Apps



TURN



## Business Intelligence

ORACLE | Hyperion



Business Objects



Business Intelligence



COGNOS



Autonomy



MicroStrategy



## Analytics and Visualization



METAMARKETS

TERADATA

ASTER

SAS

TIBCO

panopticon

Real-Time Visual Data Analysis

Datameer

platfora

ClearStory

CIRRO

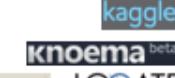
pentaho

alteryx

visual.ly

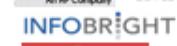
AYATA

## Data As A Service



Everything Location

## Analytics Infrastructure



Exasol - Advanced Data Analytics



## Operational Infrastructure



## Infrastructure As A Service



Google BigQuery

## Structured Databases



SYBASE

hadoop

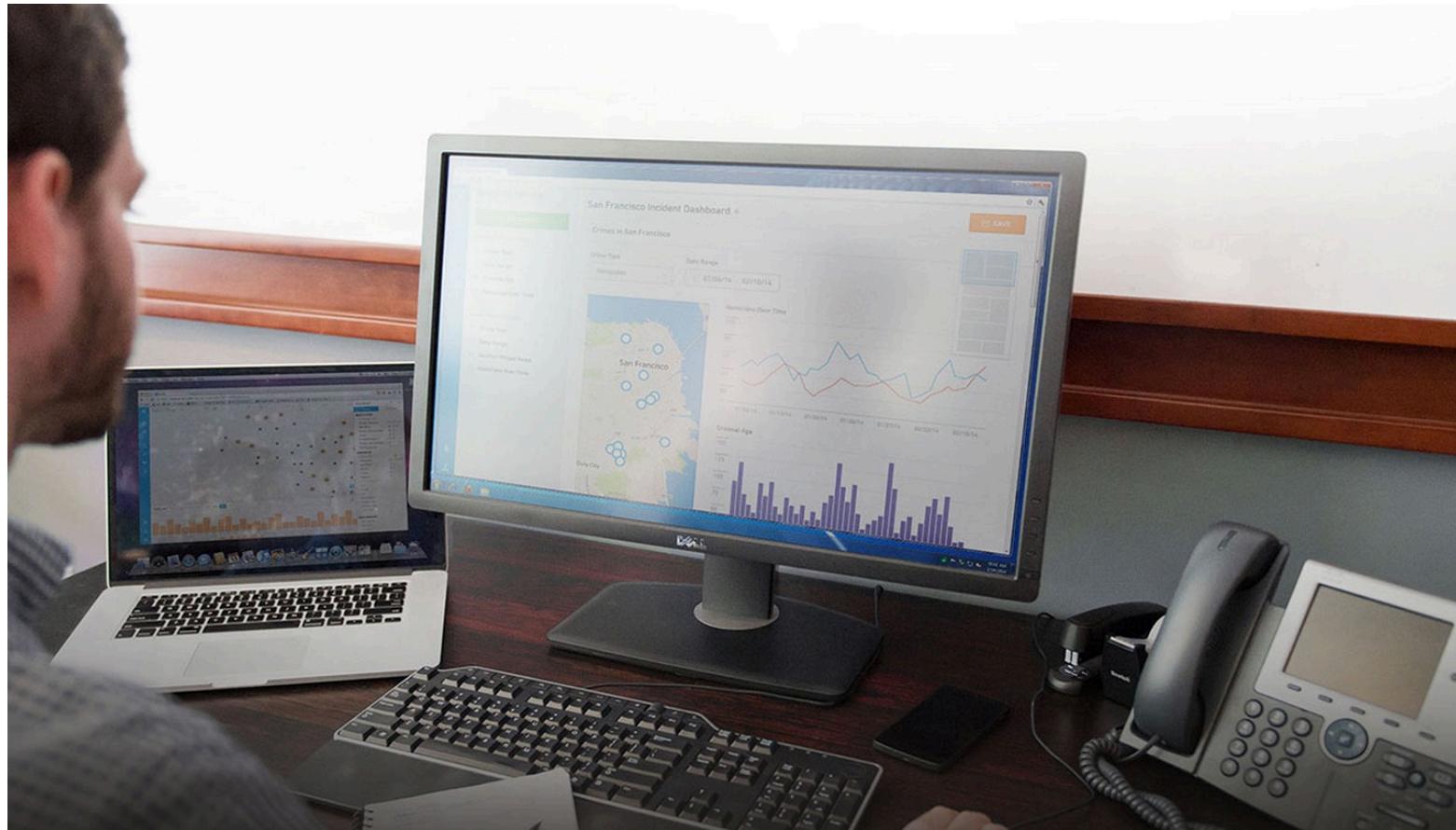
hadoop mapReduce

mahout

APACHE HBASE

Cassandra

**Palantir** analizza dati per risolvere i problemi di sicurezza dalla truffa, al terrorismo. I loro sistemi sono stati sviluppati con finanziamenti della CIA e sono ampiamente utilizzati dal governo degli Stati Uniti e dalle agenzie di sicurezza federali.



Il fatturato dello scorso anno è stato di circa 418 milioni dollari mentre il valore dell'azienda, in odore di IPO, è stato recentemente valutato a \$ 15 miliardi.

<https://www.palantir.com/products/>

Anche l'attività di **Uber** si basa sull'analisi dei dati.

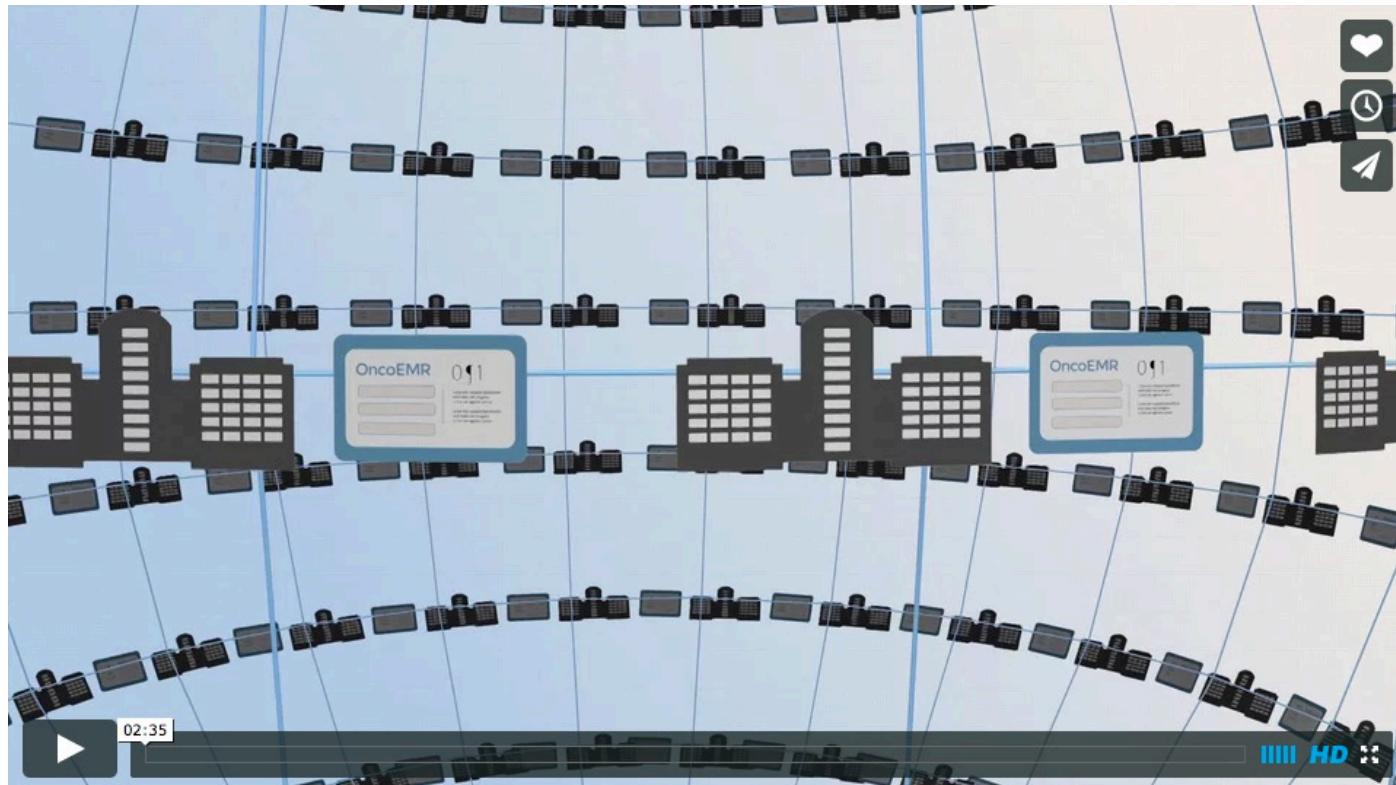


Con i dati rilevati su conducenti e passeggeri, Uber alimenta un algoritmo per trovare corrispondenze opportune e convenienti, nonché le tariffe più adatte.

Uber sta tessendo alleanze con catene di Hotel, negozi fashion e provider musicali.

<http://blog.uber.com/tag/uberdata/>

## Flaitron usa i Big Data contro il cancro.



Analizzando automaticamente nella sua OncologyCloud terabyte di dati raccolti durante diagnosi e cure dei pazienti affetti da cancro, Flaitron spera di fornire dati utili per curare quel 96% di pazienti sui quali non vengono raccolti e trattati dati.

L'anno scorso ha ricevuto 130 milioni di dollari di finanziamento da Google

<http://www.flatiron.com/>

**Affectiva** ha creato una "tecnologia di misura emozionale" che si basa sul riconoscimento facciale e consente, analizzando foto e video, di determinare lo stato d'animo e il sentimento delle persone presenti in un determinato luogo.



La tecnologia può essere utilizzata per valutare la reazione del pubblico alla pubblicità, misurare lo stato d'animo delle persone che interagiscono con un servizio, o giudicare l'umore del pubblico durante un dibattito politico. Coca Cola ha utilizzato Affdex per effettuare analisi di marketing.

<http://www.affectiva.com/>

# Lezione per tutti

Mi serve un dato:

- Ho archiviato tutto per bene e ho un sacco di storage, ma quel dato che cerco non lo trovo;
- Non basta avere tecnologie, spazi e strumenti, è necessario agevolare la ricerca e per poterlo fare bisogna ripensare i modelli stessi di ricerca in funzione di quello che fa il **consumatore** e non a quello che fanno le **aziende**;
- Google è nato per questo. O no?

# Conclusioni

Si tratta di un cambiamento nella pragmatica della conoscenza. I dati sono più numerosi, facili da trovare e meno costosi da archiviare. La fine della scarsità dei dati non riduce il rispetto del loro significato e non annulla la necessità di una profonda consapevolezza epistemologica. Ma certamente favorisce una pratica della sperimentazione matematica alla ricerca di pattern emergenti e correlazioni, piuttosto che un ricorso all'approccio basato sui campioni statistici, le ipotesi causali a priori, le teorie in attesa di verifica. «Meno why e più what».

Si tratta di un cambiamento economico, perché lo sfruttamento dei giacimenti di dati è un grande valore per le mega compagnie che li raccolgono ma anche per le startup che ne individuano nuovi utilizzi.



@webeconoscenza

<http://www.gigicogo.it>