

Si può produrre ricchezza
con i dati?

From big data to
big gains

Oggi quali sono le aziende USA più ricche?

*** Diamo un'occhiata alla classifica di fortune 500**

1. Wal-Mart Stores
2. Exxon Mobil
3. Chevron
4. Berkshire Hathaway

5. Apple
6. Phillips 66
7. General Motors
8. Ford Motor
9. General Electric
10. Valero Energy

Nel settore dell'information technology ieri

2012	
Posizione	Azienda
17	Apple
69	Microsoft
91	Amazon.com
167	Google
Oltre il 500	Facebook Inc. fb

Nel settore dell'information technology oggi

2014	
Posizione	Azienda
5	Apple
34	Microsoft
35	Amazon.com
46	Google
341	Facebook Inc. fb
Fuori dall 500 però: +35% Year	Twitter Inc. twtr

Che cosa hanno in comune?



I dati

Usare i dati

- * Diventa importante saper analizzare grandi quantità di dati utilizzando metodi «non convenzionali»
- * Non basta un data base relazionale con delle query
- * Si parla allora dei **big data**
- * Analizzare i dati in modo ottimale comporta vantaggi competitivi

Cosa sono i big data?

Esistono diverse definizioni

- ❑ “Big data exceeds the reach of commonly used hardware environments and software tools to capture, manage, and process it with in a tolerable elapsed time for its use population.” -Teradata Magazine article, 2011
- ❑ “Big data refers to data sets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze.” -The McKinsey Global Institute, 2012
- ❑ “Big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools.” - Wikipedia, 2012

Quando i dati diventano big?



* IOPS: Input/Output Operations Per Second

L'impetuosa crescita delle informazioni

- * Dal 1453 al 1503 furono stampati in Europa circa 8 milioni di libri
 - * Un numero superiore a quelli prodotti dagli amanuensi, dalla fondazione di Costantinopoli (nei 1200 anni precedenti)
- Il patrimonio informativo (rappresentato dai libri) impiegava circa 50 anni per raddoppiare.
Oggi sono sufficienti solo 3 anni per raddoppiare la quantità di dati disponibili su Internet

Big data

- * Lo sfruttamento dei BIG Data prevede l'applicazione della matematica ad enormi quantità di dati per valutare delle probabilità:
- * nella correzione/traduzione di un testo
- * sulla diffusione di un epidemia influenzale
- * → Evoluzione verso sistemi di valutazione automatica

L'enorme quantità di dati disponibili riduce il rischio di errori

Diamo i numeri

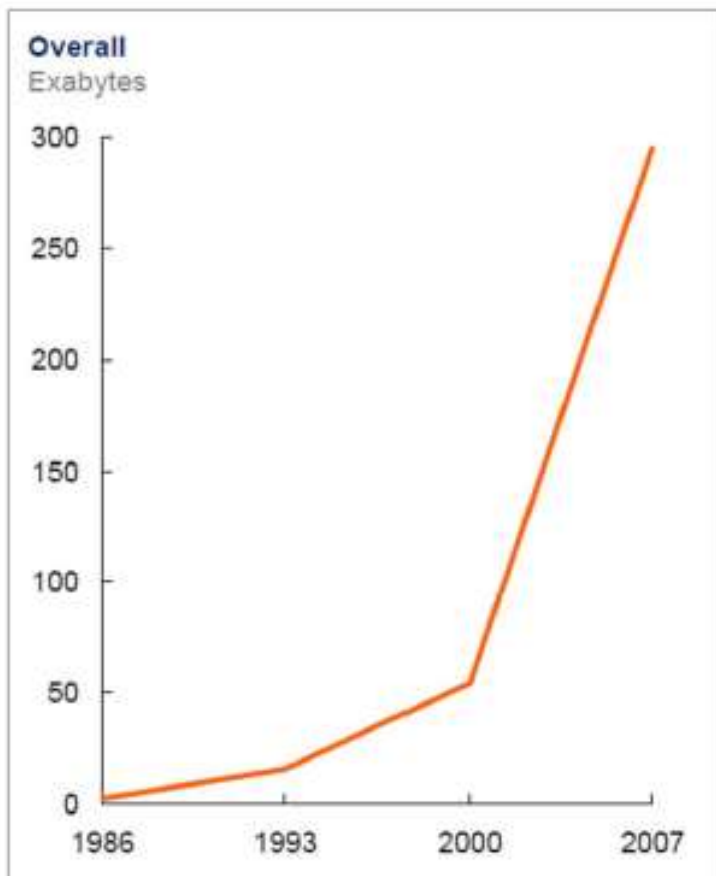
Quanti dati ci sono nel mondo?

1. 800 Terabytes, 2000
2. 160 Exabytes, 2006
3. 500 Exabytes, 2009
4. 2.7 Zettabytes, 2012
5. 35 Zettabytes by 2020

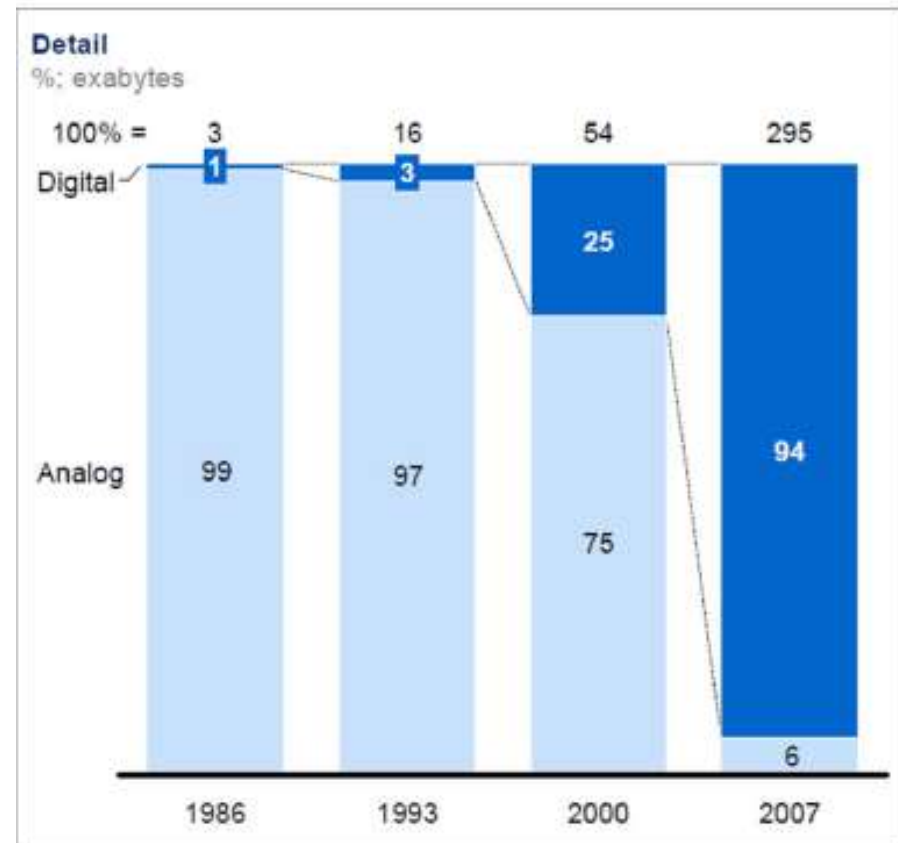
Quanti dati sono generati in un giorno?

1. 7 TB, Twitter
2. 10 TB, Facebook

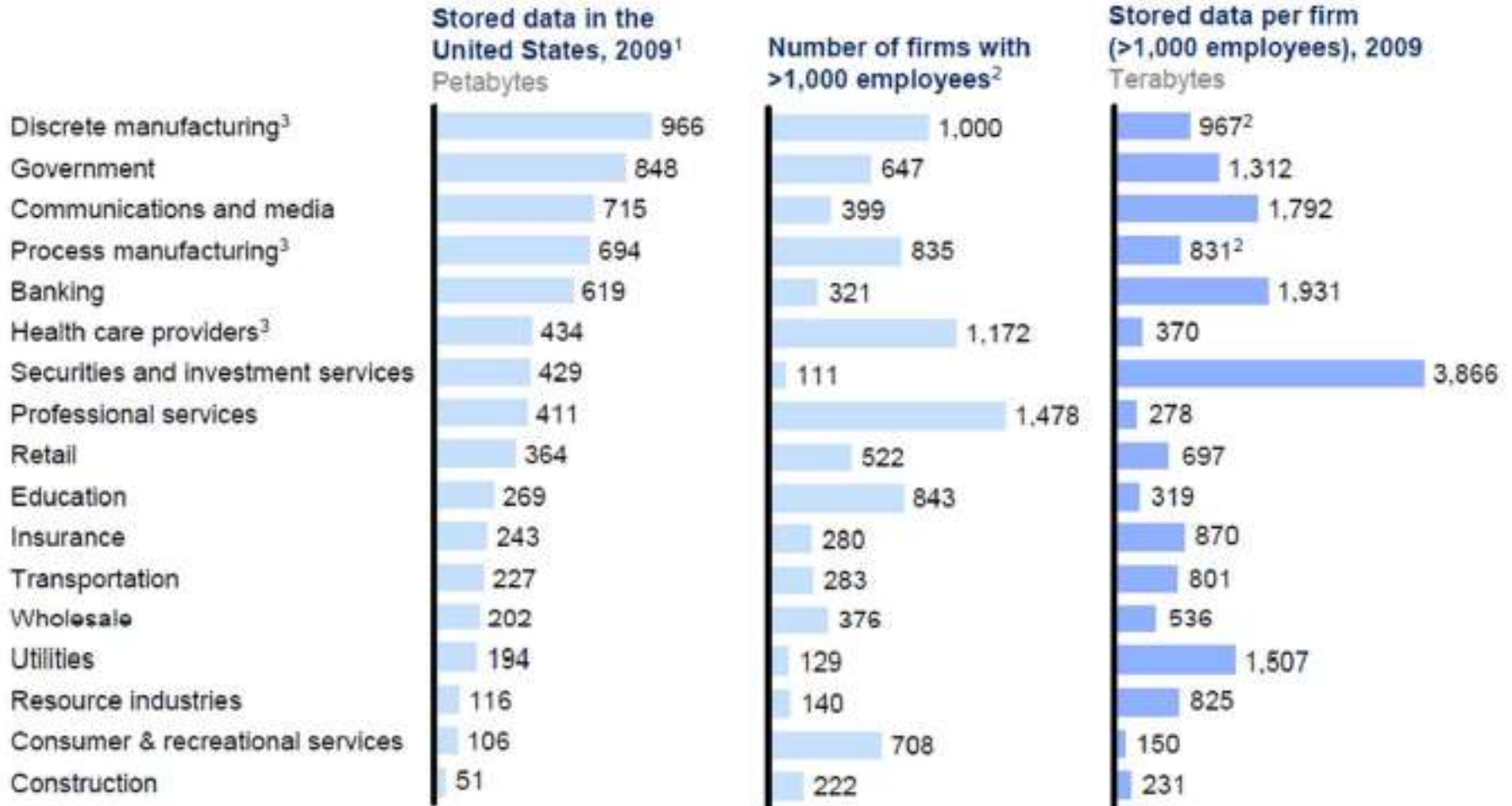
Sono tanti e crescono velocemente



From big data to big gains a.s 2014-15



Nei vari settori



1 Storage data by sector derived from IDC.

2 Firm data split into sectors, when needed, using employment

3 The particularly large number of firms in manufacturing and health care provider sectors make the available storage per company much smaller.

Tipologia di dati

	Video	Image	Audio	Text/ numbers	
Banking	Medium	Medium	Medium	High	
Insurance	Low	Low	Low	High	
Securities and investment services	Low	Low	Low	High	
Discrete manufacturing	Medium	Medium	Low	High	
Process manufacturing	Medium	Medium	Low	High	
Retail	Medium	Low	Low	High	
Wholesale	Low	Low	Low	High	
Professional services	Medium	Medium	Medium	High	
Consumer and recreational services	Medium	Low	Medium	Medium	
Health care	Low	High	Low	High	
Transportation	Medium	Medium	Low	High	
Communications and media ²	High	Medium	High	High	
Utilities	Medium	Medium	Low	High	
Construction	Low	High	Low	Medium	
Resource industries	Medium	Medium	Low	High	
Government	High	Medium	High	High	
Education	High	Medium	High	Medium	

Penetration

- High
- Medium
- Low

Big Data, social network e valore economico

Esempio

WhatsApp all'atto dell'installazione acquisisce la rubrica dell'utente (indicato nei *Terms of service*) ed è libera di trasformare (si parla di *datizzazione*) in metadati economicamente valorizzabili i rapporti inter-personali/inter-aziendali, i legami di amicizia ecc.

- * Il valore della società è legato esclusivamente al valore intrinseco dei Big Data riferiti ad oltre 350 milioni di utenti attivi ogni mese.

BIG DATA e CLOUD Computing

- * La gestione dei Big Data ha generato l'esigenza di conservare ed elaborare enormi quantità di dati...
- * Soluzione tecnologica d elaborazione distribuita
- * **CLOUD Computing**



Le tre V dei big data

- * Non è solo questione di velocità



Le V sono quattro

- **Volume** : Gigabyte(10^9), Terabyte(10^{12}), Petabyte(10^{15}), Exabyte(10^{18}), Zettabyte (10^{21})
- **Varietà**: Structured, semi-structured, unstructured; Text, image, audio, video, record
- **Velocità**: Periodic, Near Real Time (piccolo ritardo dovuto a motivi tecnici), Real Time
- **Valore**: Può generare grossi vantaggi competitivi!

Mettiamo ordine nei numeri: Dati McKinsey

- 600\$ per un disco nel quale memorizzare tutta la musica del mondo
- 5 miliardi di cellulari nel 2010
- 30 miliardi di contenuti al mese su Facebook
- 40% di crescita di dati all'anno
- 5% di crescita di investimenti nell'IT
- 235 TB di dati raccolti dalla US Library of Congress

I Dati nel Cyberspace

- * Piattaforme social media
- * Reti di sensori (che rilevano, creano e comunicano dati)
- * Servizi Cloud

→ creazione di «miniere di dati» nel Cyberspace

Dati pubblici e privati su tutti gli aspetti della vita sociale e privata delle persone

Valore commerciale dei dati = ENORME

Social Media: esempio di elaborazione demografica

Fonte: <http://www.flowtown.com/>

KEY

■ = 1%

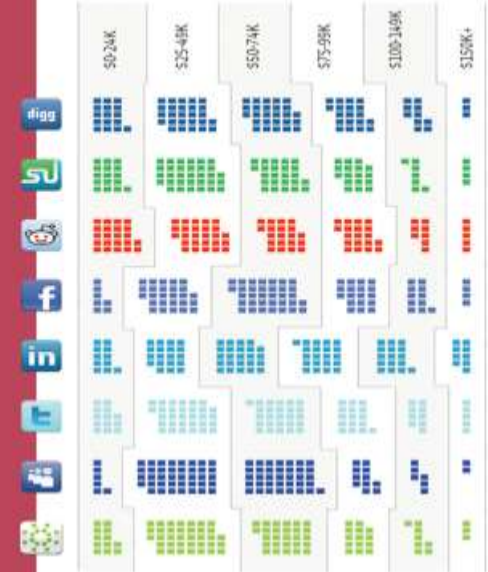
*"Be civil to all;
sociable to many;
familiar with few;
friend to one;
enemy to none."*

Benjamin Franklin

By Gender



By Income Level



By Education Level



Source: Google Ad Planner Data



Flowtown.

By Age



- DIGG** the latest news headlines, videos and images
- STUMBLEUPON** personalized recommendations to help you discover the best of the web
- REDDIT** what's new online!
- FACEBOOK** connect and share with the people in your life
- LINKEDIN** relationships matter
- TWITTER** share and discover what's happening right now, anywhere in the world
- MYSPACE** a place for friends
- NING** create and discover Ning social networks for your interests and passions

Valore potenziale in alcuni settori



US health care

- \$300 billion value per year
- ~0.7 percent annual productivity growth



Europe public sector administration

- €250 billion value per year
- ~0.5 percent annual productivity growth



Global personal location data

- \$100 billion+ revenue for service providers
- Up to \$700 billion value to end users



US retail

- 60+% increase in net margin possible
- 0.5–1.0 percent annual productivity growth

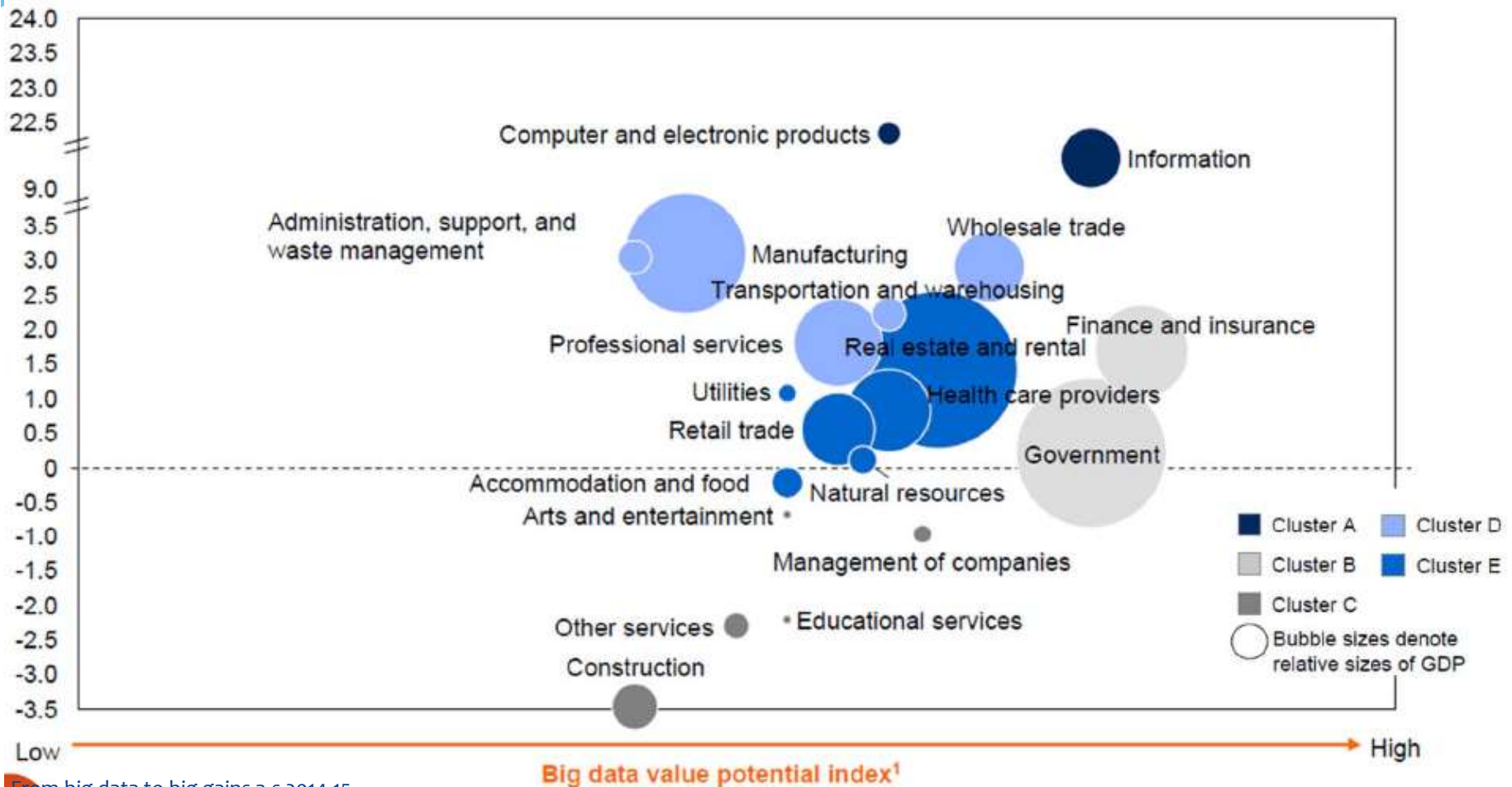


Manufacturing

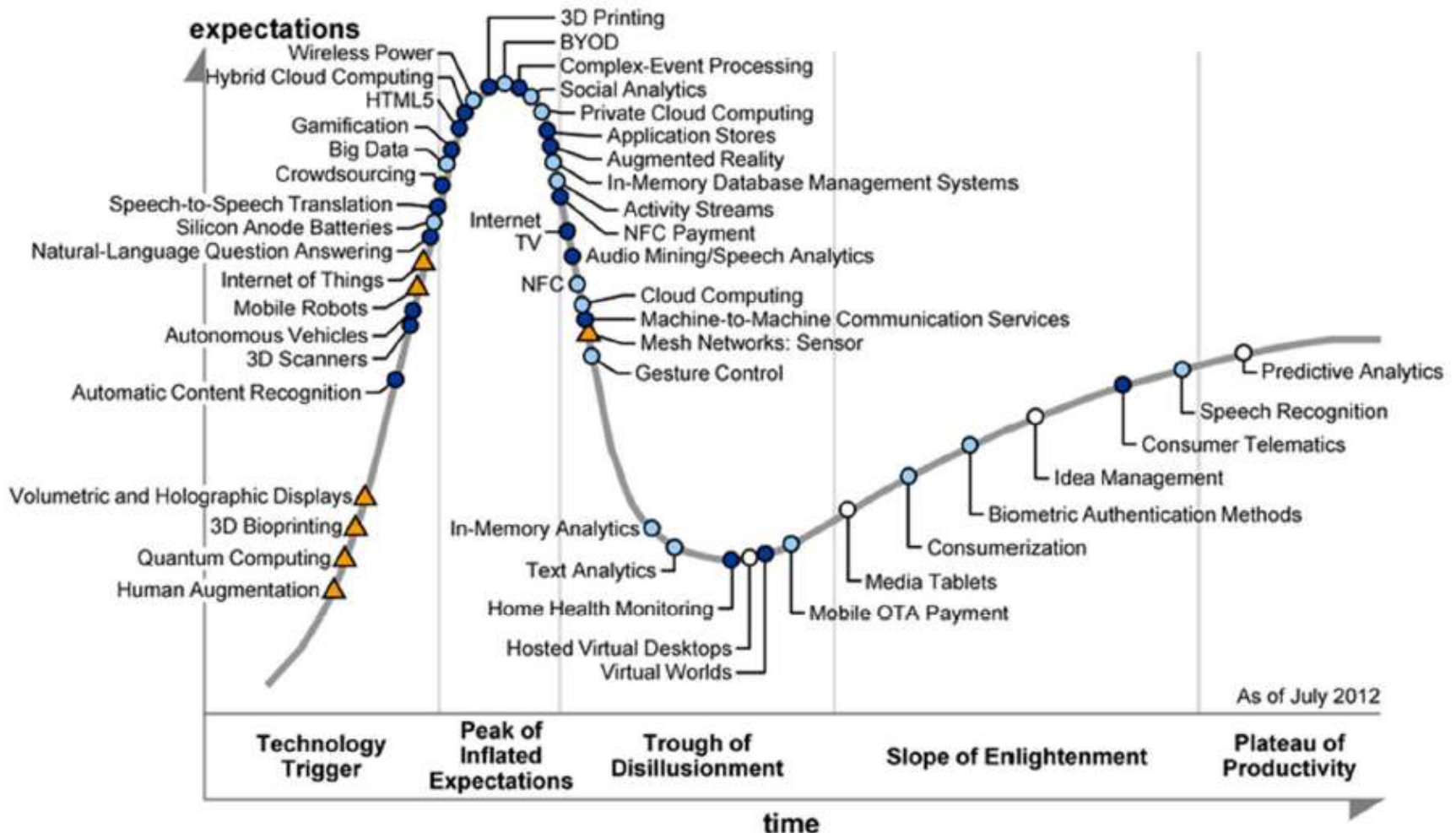
- Up to 50 percent decrease in product development, assembly costs
- Up to 7 percent reduction in working capital

Prospettive di guadagni

Historical productivity growth in the United States, 2000–08



Il ciclo delle aspettative (Gartner 2012)



E' più importante il “big” o il “data”?

1. Il “Big”
2. Il “Data”
3. Entrambi
4. Nessuno dei due
 - Risposta: nessuno dei due

La domanda è : Cosa fanno le organizzazioni con i big data?

«I dati non sono informazioni, le informazioni non sono il sapere, la conoscenza non è la comprensione, la comprensione non è saggezza» *Cliff Stoll (professore a Berkley e commentatore)*

Casi di uso (1)

	Data Source	High-frequency operations	Lower-frequency operations
Financial trade monitoring	Capital markets	Write/index all trades, store tick data	Show consolidated risk across traders
Telco call data record management	Call initiation request	Real-time authorization	Fraud detection/analysis
Website analytics, fraud detection	Inbound HTTP requests	Visitor logging, analysis, alerting	Traffic pattern analytics
Online gaming micro transactions	Online game	Rank scores: <ul style="list-style-type: none"> • Defined intervals • Player "bests" 	Leaderboard lookups
Digital ad exchange services	Real-time ad trading systems	Match form factor, placement criteria, bid/ask	Report ad performance from exhaust stream
Wireless location-based services	Mobile device location sensor	Location updates, QoS, transactions	Analytics on transactions

Casi di uso (2)

Today's Challenge

New Data

What's Possible

Healthcare

Expensive office visits

Remote patient monitoring

Preventive care, reduced hospitalization

Manufacturing

In-person support

Product sensors

Automated diagnosis, support

Location-Based Services

Based on position

Real time location data

Geo-advertising, traffic, local search

Public Sector

Standardized services

Citizen surveys

Tailored services, cost reductions

Retail

One size fits all marketing

Social media

Sentiment analysis segmentation

Un approccio alternativo a quello commerciale?

Open data:

- Iniziativa volta a rendere liberamente accessibili i dati privi di brevetti e non coperti da copyright
- Open government
- Dato come bene comune
- Formato aperto
- Adatti alla elaborazione

Open data in Italia

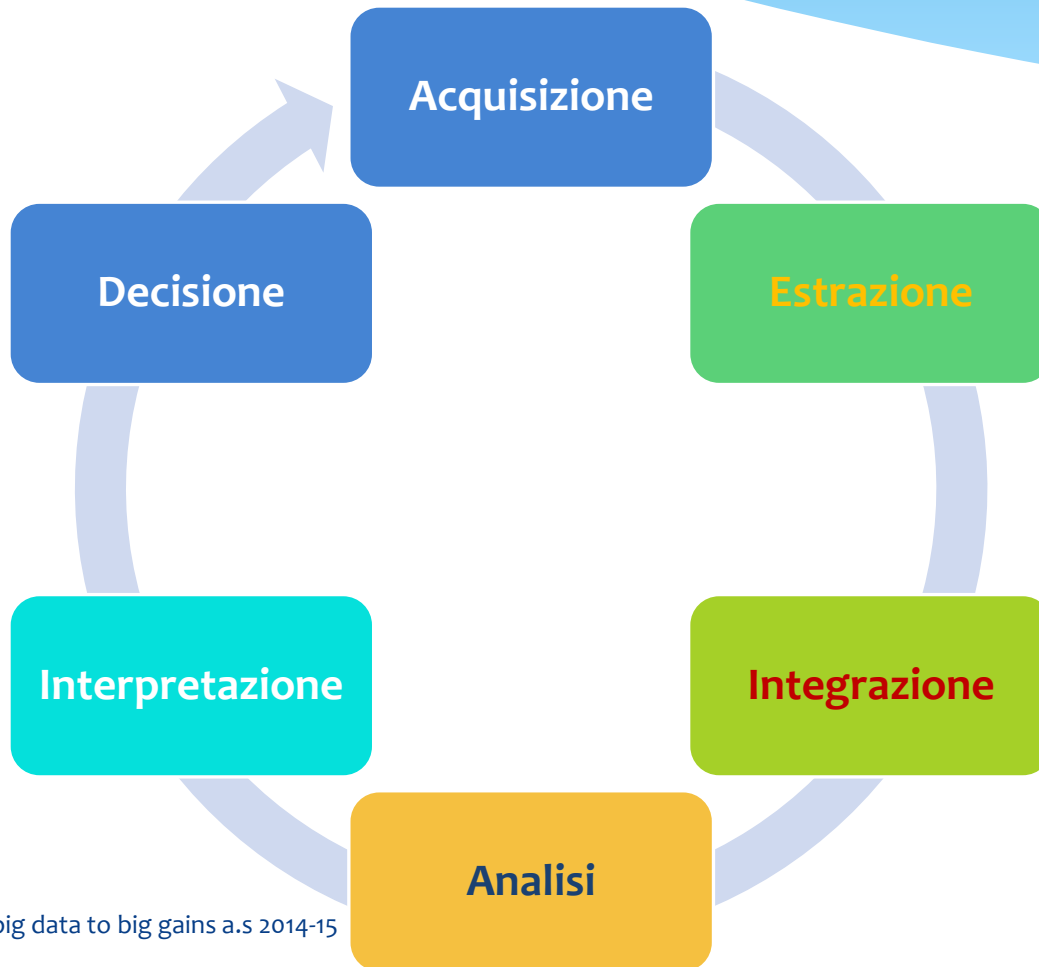
- Siamo in ritardo (dati.gov.it)
- Prevalentemente regionale
- I primi 10 enti coprono il 75% dei dati disponibili

I rischi dei Big Data

- Ingestibilità dei dati a disposizione
- Non scalabilità
- Inefficacia delle analisi
- Costi
- Privacy



Allora come uso i big data, cosa ci posso fare?



Obiettivo:

saper prendere decisioni strategiche efficaci sfruttando la grande disponibilità di dati

Nel dettaglio



Richiede:

- Selezione
- Filtraggio
- Generazione di metadati
- Gestione provenienza

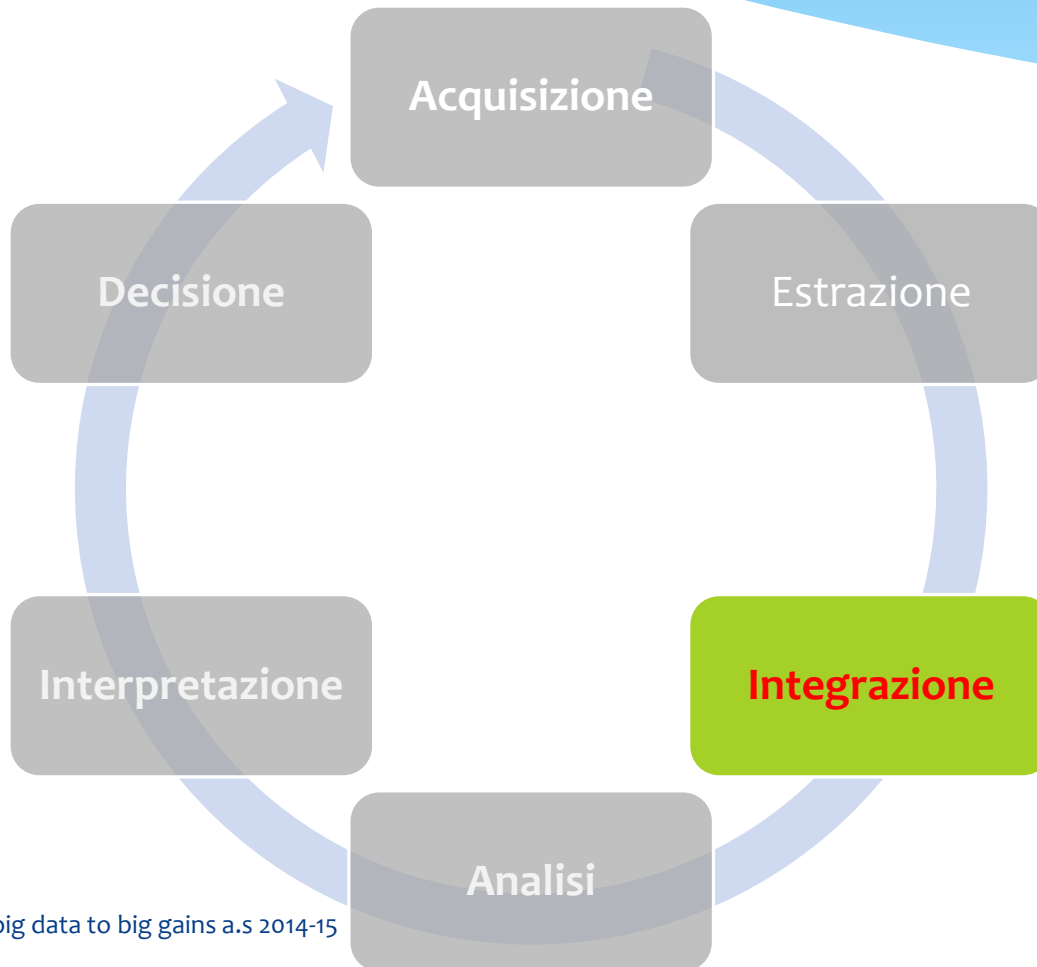
Nel dettaglio



Richiede:

- Trasformazione
- Normalizzazione
- Cleaning (aggiunta di valori mancanti, aggiustamento *dei dati* evidentemente errati.....)
- Aggregazione
- Gestione errori

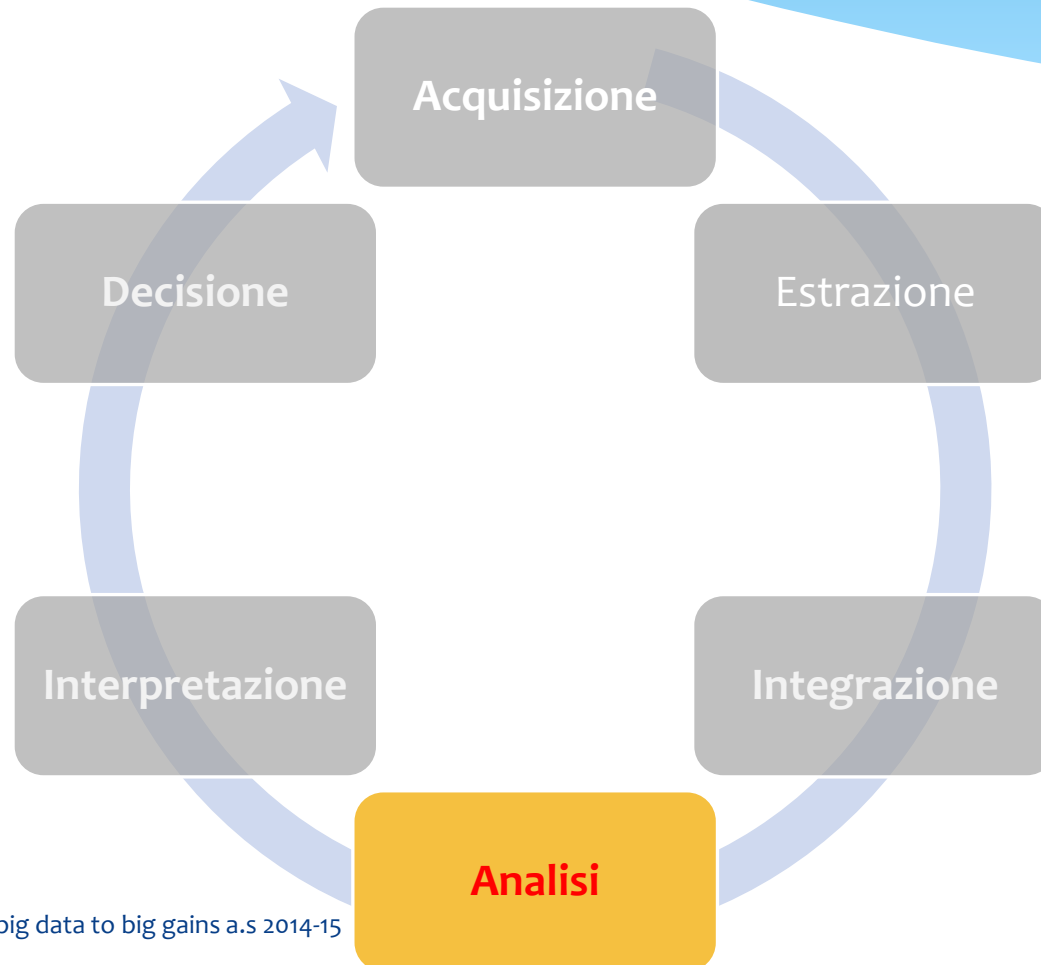
Nel dettaglio



Richiede:

- Standardizzazione
- Gestione conflitti
- Riconciliazione
(analisi delle differenze e tentativo di conciliarle)
- Definizione di mapping (attività di creazione, prelievo, aggiornamento ed eliminazione *dei dati*)

Nel dettaglio



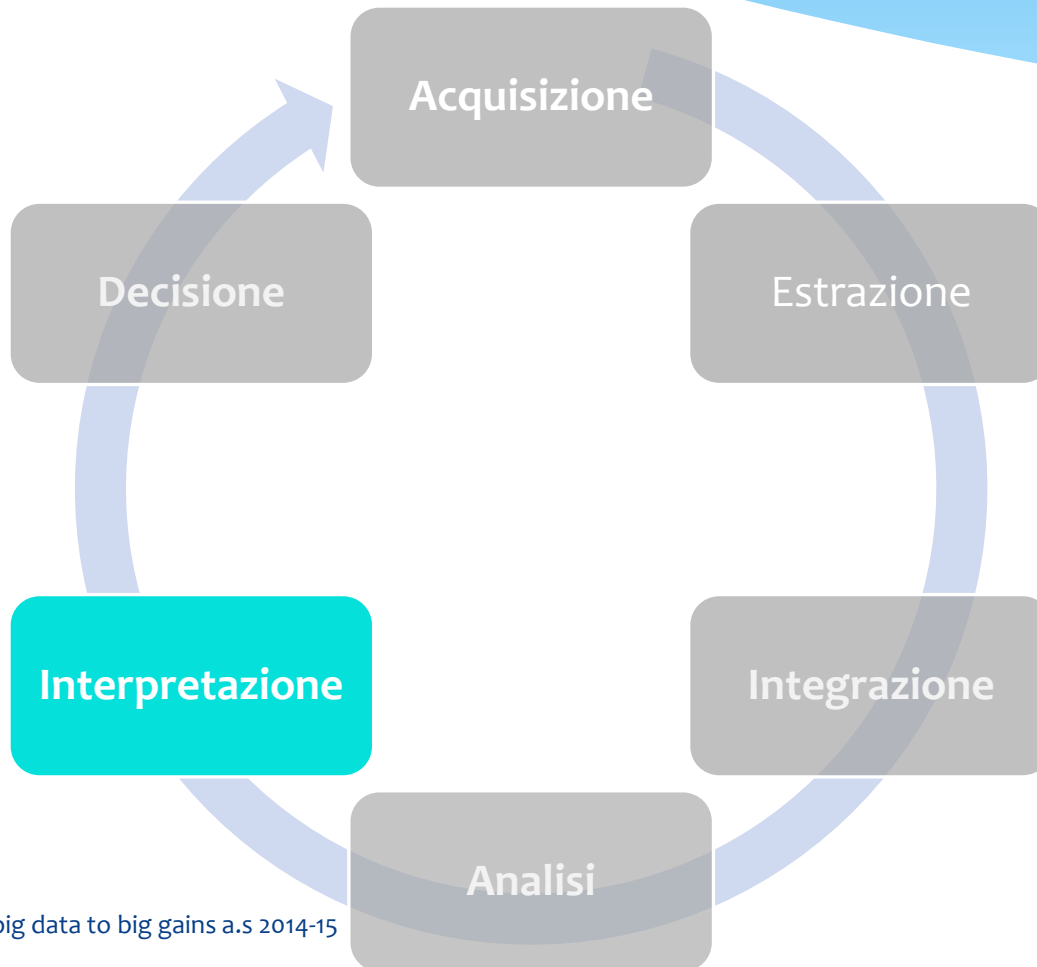
Richiede:

- Esplorazione
- Mining (è l'insieme di tecniche e metodologie che hanno per oggetto l'estrazione di un sapere o di una conoscenza a partire da grandi quantità di dati)
- Apprendimento automatico (tecnica comune per l'analisi statistica *dei dati*)
- Visualizzazione

Nel dettaglio

Richiede:

- Conoscenza del dominio
- Conoscenza della provenienza
- Identificazione pattern di interesse (i vari ambiti di interesse)
- Flessibilità del processo



Nel dettaglio



Richiede:

- Capacità Manageriali
- Miglioramento continuo del processo

Le sfide

- 1) Performance, performance, performance!
- 2) Scalabilità
- 3) Eterogeneità e incompletezza
- 4) Flessibilità
- 5) Usabilità
- 6) Efficacia
- 7) Privacy
- 8) Proprietà
- 9) Collaborazione umana

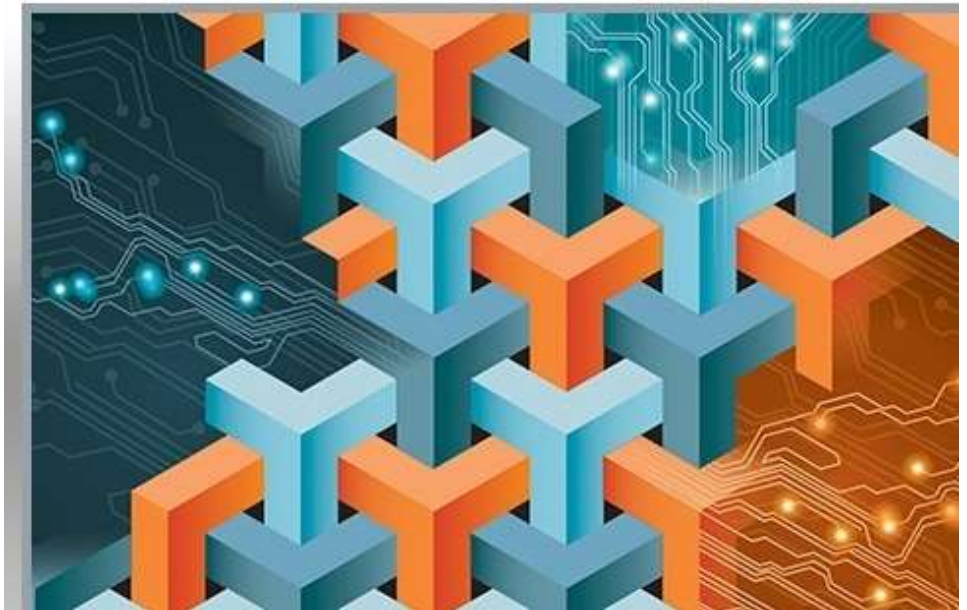
Quali sono le tecniche di analisi?

1. Data mining
 - A. Association rule learning
 - B. Classification
 - C. Cluster analysis
 - D. Regression
2. Crowdsourcing
3. Data fusion and data integration
4. Genetic algorithms
5. Machine learning
 - A. Supervised learning
 - B. Unsupervised learning.

Tecniche per l'analisi dei big data

1. Natural language processing
2. Neural networks
3. Network analysis
4. Pattern recognition
5. Predictive modeling
6. Sentiment analysis
7. Signal processing
8. Spatial analysis
9. Simulation
10. Time series analysis
11. Visualization.

revelytix



Tecnologie per i big data

- 1) Business intelligence (BI)
- 2) Cloud computing
- 3) Distributed system
 - a. Hadoop
- 4) Relational database
- 5) NoSQL systems
 - a) Cassandra
 - b) Dynamo
 - c) Hbase
 - d) Big Table
- 6) Extract, transform, and load (ETL)
- 7) Mashup
- 8) Semi-structured/Unstructured data
- 9) Stream processing
- 10) Visualization



Ma allora?



- * Tranquilli vedremo solo alcune tecniche

NLP

- L'**Elaborazione del linguaggio naturale**, detta anche NLP (dall'inglese *Natural Language Processing*, elaborazione lingua naturale), è il processo di trattamento automatico mediante un calcolatore elettronico delle informazioni scritte o parlate in una lingua naturale.

NLP

- Questo processo è reso particolarmente difficile e complesso a causa delle caratteristiche intrinseche di ambiguità del linguaggio umano. Per questo motivo il processo di elaborazione viene suddiviso in fasi diverse, tuttavia simili a quelle che si possono incontrare nel processo di elaborazione di un linguaggio di programmazione:
 - I. analisi lessicale: scomposizione di un'espressione linguistica in *token* (in questo caso le parole)
 - II. analisi grammaticale: associazione delle parti del discorso a ciascuna parola nel testo
 - III. analisi sintattica: arrangiamento dei *token* in una struttura sintattica (ad albero: *parse tree*)
 - IV. analisi semantica: assegnazione di un significato (semantica) alla struttura sintattica e, di conseguenza, all'espressione linguistica

Neural networks

- * Tradizionalmente il termine **rete neurale** (o **rete neuronale**) viene utilizzato come riferimento ad una rete o ad un circuito di neuroni biologici, tuttavia ne è affermato l'uso anche in matematica applicata con riferimento alle **reti neurali artificiali**, modelli matematici composti di "neuroni" artificiali. L'espressione può acquisire pertanto due significati distinti:
 - I. **Le reti neurali biologiche:** sono costituite dai neuroni biologici, cellule viventi tipiche degli animali connesse tra loro o connesse nel sistema nervoso periferico o nel sistema nervoso centrale.
 - Nel campo delle neuroscienze, sono spesso identificati come gruppi di neuroni che svolgono una determinata funzione fisiologica nelle analisi di laboratorio.

Neural networks

- II. **Le reti neurali artificiali:** sono modelli matematici che rappresentano l'interconnessione tra elementi definiti neuroni artificiali, ossia costrutti matematici che in qualche misura imitano le proprietà dei neuroni viventi.
 - Questi modelli matematici possono essere utilizzati sia per ottenere una comprensione delle reti neurali biologiche, ma ancor di più per risolvere problemi ingegneristici di intelligenza artificiale come quelli che si pongono in diversi ambiti tecnologici (in elettronica, informatica, simulazione, e altre discipline).
 - Una rete neurale artificiale può essere realizzata sia da programmi software che da hardware dedicato (DSP, *Digital Signal Processing*).

Predictive modeling

- * La **modellazione predittiva** è un processo utilizzato in analisi predittiva per creare un modello statistico del comportamento futuro.
- * Analisi predittiva è l'area del **data mining** che si occupa di probabilità e previsione delle tendenze.
- * Un modello predittivo è costituito da un numero di **predittori**, che sono fattori variabili che possono influenzare il comportamento o risultati futuri. Nel marketing, ad esempio, il sesso, l'età, e la cronologia degli acquisti di un cliente potrebbe prevedere la probabilità di una futura vendita.

Tecniche: Extract, transform, load

- **Extract, Transform, Load (ETL)** è un'espressione in [lingua inglese](#) che si riferisce al processo di estrazione, trasformazione e caricamento dei dati in un sistema di sintesi ([data warehouse](#), [data mart...](#)).
- I dati vengono estratti da sistemi sorgenti quali database transazionali ([OLTP](#)), comuni file di testo o da altri sistemi informatici (ad esempio, sistemi [ERP](#) o [CRM](#)).
- Subiscono quindi un processo di trasformazione, che consiste ad esempio nel:
 - a) Selezionare solo quelli che sono di interesse per il sistema
 - b) Normalizzare i dati (per esempio eliminando i duplicati)
 - c) Tradurre dati codificati
 - d) Derivare nuovi dati calcolati
 - e) Eseguire accoppiamenti (join) tra dati recuperati da differenti tabelle
 - f) Raggruppare i dati

Extract, transform, load

- Tale trasformazione ha lo scopo di consolidare i dati (cioè rendere omogenei dati provenienti da sorgenti diverse) e di fare in modo che siano più aderenti alla logica di business del sistema di analisi per cui viene sviluppato.
- Vengono infine memorizzati nelle tabelle del sistema di sintesi (load).
- Occorre prestare particolare attenzione alla granularità delle informazioni da memorizzare nella struttura a valle. Queste infatti, non solo devono essere aggregate in modo da non avere un dettaglio eccessivo (cosa che potrebbe portare ad un decadimento delle prestazioni delle interrogazioni effettuate sul sistema), ma devono anche mantenere una granularità che consenta di effettuare le necessarie analisi sui dati.

Tecniche: Stream processing

- ❑ Lo **stream processing** è un paradigma di **programmazione parallela** di tipo **SIMD** (*Single Instruction, Multiple Data* è un'architettura in cui più unità elaborano dati diversi in parallelo) che permette ad alcune applicazioni di sfruttare semplicemente una forma limitata di elaborazione parallela.
- ❑ Le applicazioni utilizzano delle unità funzionali come **FPU** o **GPU** senza doversi preoccupare di gestirne la sincronizzazione o la comunicazione tra le unità funzionali, sgravando il programmatore da uno dei principali problemi della programmazione parallela.
- ❑ Il paradigma dello stream processing semplifica lo sviluppo dell'hardware e del software limitando le operazioni che possono essere eseguite.
- ❑ Questo paradigma presume di avere i dati da elaborare organizzati in gruppi (stream) e che questi possano essere elaborati applicando ad essi una serie di operazioni. Configurazioni nelle quali una serie di **operazioni** devono essere applicate a tutti i dati sono comuni.

Mash-up

- In [informatica](#) un **mash-up** è un [sito](#) o un'[applicazione web](#) di tipo ibrido, cioè tale da includere dinamicamente informazioni o contenuti provenienti da più fonti.
 - Un esempio potrebbe essere un programma che, acquisendo da un sito web una lista di appartamenti, ne mostra l'ubicazione utilizzando il servizio [Google Maps](#) per evidenziare il luogo in cui gli stessi appartamenti sono localizzati.
- **Mash-up** (letteralmente: "poltiglia"), in termini informatici, indica un'applicazione che usa contenuto da più sorgenti per creare un servizio completamente nuovo.
- Il contenuto dei mash-up è normalmente preso da terzi via [API](#) (in [informatica](#), si indica ogni insieme di [procedure](#) disponibili al [programmatore](#), di solito raggruppate a formare un set di strumenti specifici per l'espletamento di un determinato compito all'interno di un certo [programma](#),) tramite feed (es. [RSS](#) e [Atom](#)) o [Javascript](#). I mash-up stanno rivoluzionando lo sviluppo del web permettendo a chiunque di combinare dati da siti come [Amazon.com](#), [eBay](#), [Google](#), [Windows Live](#) e [Yahoo!](#) in modi innovativi
- Sono semplici da progettare: richiedono minime conoscenze tecniche e quindi sono solitamente creati da contributori inusuali. I mashup sono uno degli **elementi del cosiddetto [web 2.0](#)**.

OLAP - Datawarehouse

- **OLAP** è l'acronimo di *On Line Analytical Processing*. Consiste un insieme di tecniche software per analizzare velocemente grandi quantità di dati, anche in modo complesso.
- Questa è la componente tecnologica di base del data warehouse e serve ad esempio alle aziende per analizzare i risultati delle vendite, l'andamento dei costi di acquisto merci, al marketing per misurare il successo di una campagna pubblicitaria, ed altri casi simili. In definitiva un sistema Olap permette di:
 - a. studiare una grande quantità di dati
 - b. vedere i dati da prospettive diverse
 - c. supportare i processi decisionali.

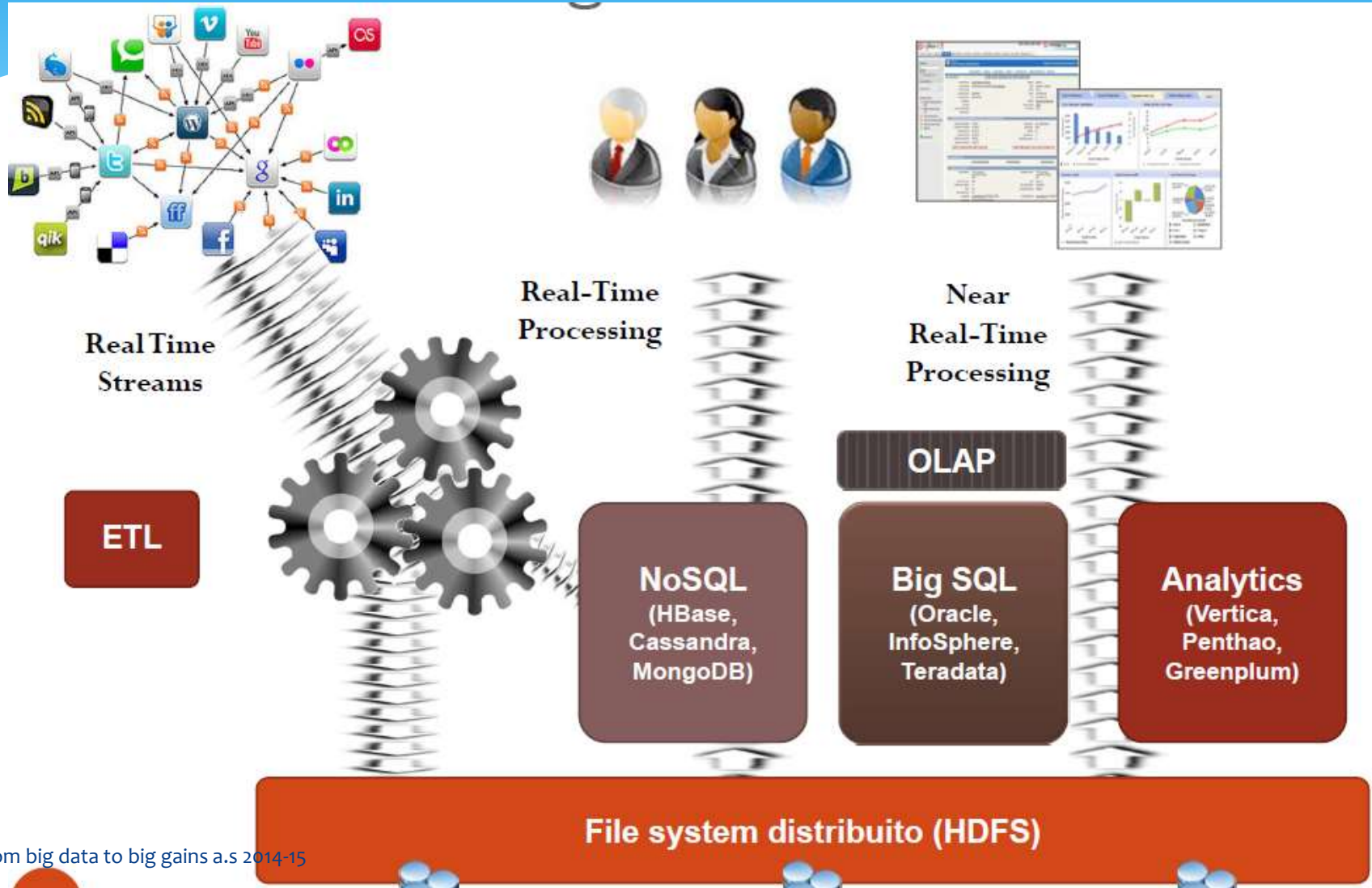
OLAP - Datawarehouse

- ❑ La creazione di un database OLAP consiste nel fare una fotografia delle informazioni (ad esempio quelle di un database relazionale) in un determinato momento e trasformare queste singole informazioni in dati multidimensionali.
- ❑ Eseguendo successivamente delle interrogazioni sui dati così strutturati è possibile ottenere risposte in tempi ridotti rispetto alle stesse operazioni effettuate su **db** relazionali.
- ❑ Una struttura OLAP creata per questo scopo è chiamata "cubo" multidimensionale.
- ❑ Ci sono diversi modi per creare un cubo ma il più conosciuto è quello che utilizza uno schema "a stella"; al centro c'è la tabella dei "fatti" che elenca i principali elementi su cui sarà costruita l'interrogazione, e collegate a questa tabella ci sono varie tabelle delle "dimensioni" che specificano come i dati saranno aggregati.

OLAP - Datawarehouse

- ❑ Con la struttura OLAP, un archivio di clienti può essere raggruppato per città, provincia, regione; questi clienti possono essere relazionati con i prodotti ed ogni prodotto può essere raggruppato per categoria.
- ❑ Il calcolo delle possibili combinazioni di queste aggregazioni rappresenta tale struttura che, potenzialmente, potrebbe contenere tutte le risposte per ogni singola combinazione.
- ❑ Nella realtà solo un numero predeterminato di combinazioni viene memorizzato, mentre le rimanenti vengono ricalcolate solo al momento in cui quella richiesta viene materialmente effettuata.

Il flusso dei Big Data



I requisiti per sfruttare i big data

- Politiche
- Tecnologie e tecniche
- Riorganizzazioni
- Ricerca del talento
- Accesso ai dati
- Struttura aziendale

Un caso di uso: agenzia delle entrate

- Elaborazione dati sui redditi
- Analisi di mercato
- Monitoraggio utenti
- Confronto dati
- Fraud detection



Opportunità

- * Creare trasparenza
- * Scoprire i bisogni e individuare variabilità
- * Personalizzare le azioni
- * Supporto alle decisioni
- * Innovare prodotti e servizi



Come agire

- * Spesso i dati sono già disponibili!
- * Adottare opportune tecnologie
- * Reclutare e addestrare personale di talento
- * Gestire il cambiamento nell'organizzazione
- * Condividere i dati tra agenzie diverse
- * Introdurre la cultura del rinnovamento e miglioramento
- * Serve un supporto politico

Big Data, il nostro futuro sono i loro guadagni (Corsera 7/2014)

- * Prevedere le esigenze di chi naviga studiando le abitudini digitali. Così si vende di più e la pubblicità è personalizzata
- * Google, Apple e Amazon puntano a offrirci quello che ci serve prima di averne bisogno

di Greta Sclaunich

BTO 2014 Live – Big Data, Measurement & Analytics

- * *“A dispetto di quello che si possa pensare, **nel turismo non c’è niente di più umano dei big data**“*, che non sono solo numeri e dati, ma sono le tracce che gli esseri umani lasciano ogni momento online, attraverso ogni device, per raccontare le loro vite, le loro esperienze di viaggio.
- * **“Big data is about people”**
- * I dispositivi mobile sono dei produttori naturali di big data, perché ognuno di noi è costantemente connesso. Ognuno di noi lascia delle tracce, volenti o meno, sui dispositivi digitali, come i pagamenti, i check in, le foto condivise durante una trasferta.

BTO 2014 Live – Big Data, Measurement & Analytics

- * Le persone sono non solo le prime protagoniste a produrre big data, ma saranno anche le prime beneficiarie dei big data.
- * Non sarebbe bello se l'albergatore riuscisse a sapere grazie a questi dati chi è la persona che deve accogliere prima ancora che raggiunga l'hotel o il sito?
- * ***“La personalizzazione dell'esperienza, il sacro graal del turismo di oggi, può essere raggiunta proprio grazie ai big data.”***

Anticipare il futuro con i big data: lo stiamo facendo nel modo sbagliato

Roberto Paura: italian Institute for the Future e direttore della rivista "Futuri".

- Una delle proprietà più affascinanti dei big data è la possibilità di usarli per “anticipare” il futuro. Non a caso uso la parola “anticipare”, piuttosto che “prevedere”, anche se alcuni strumenti di analytics dei big data promettono davvero di riuscire a prevedere i possibili futuri.
- Diverse aziende e start-up nel mondo hanno deciso di gettarsi in quest'affare. Il caso-scuola che le ha spinte a farlo è quello raccontato molto bene da Viktor Mayer-Schönberger e Kenneth Cukier nel loro libro *Big Data: il caso Walmart*.
- Nel 2004, grazie alla consulenza della compagnia Teradata, i dirigenti di Walmart scoprirono alcune correlazioni interessanti tra fenomeni esterni e picchi di vendita di taluni prodotti. In particolare scoprirono, incrociando dati di vendita con quelli forniti dai bollettini meteo, che all'approssimarsi di un uragano i clienti non solo acquistano più torce elettriche ma anche più merendine dolci.
- Una correlazione incomprensibile che però portò Walmart a organizzare i suoi punti vendita, uno o due giorni prima dell'arrivo di uragani nella zona, allestendo all'ingresso dei supermarket uno scaffale di torce elettriche e uno di merendine dolci, così da massimizzarne le vendite.

Anticipare il futuro con i big data: lo stiamo facendo nel modo sbagliato

Roberto Paura: italian Institute for the Future e direttore della rivista "Futuri".

- Questo è uno degli usi comuni delle proprietà anticipatrici dei big data. A nessuno importa, sostengono i *data scientist* (gli esperti di analisi dei big data), perché gli americani vogliono le merendine dolci prima di un uragano. È così e basta, tanto vale vedere se se ne può ricavare un guadagno.
- Il fatto che possa esserci un'altra correlazione che non abbia nulla a che vedere con gli uragani non li interessa. Sul dibattito riguardante la "fine delle teorie" che i big data annuncerebbero, si è detto già parecchio. Quello che qui preoccupa è piuttosto l'uso che viene fatto dei dati in questione. Davvero dovremmo usare la straordinaria capacità di calcolo dei supercomputer e i potenti algoritmi messi su da scienziati di primissimo livello per vendere più merendine dolci.

Anticipare il futuro con i big data: lo stiamo facendo nel modo sbagliato

Roberto Paura: italian Institute for the Future e direttore della rivista "Futuri".

- Dal caso Walmart hanno preso spunto diverse altre start-up. [Per esempio Blab, una compagnia fondata a Seattle due anni fa.](#) Blab sostiene di poter anticipare l'esplosione di un trend sociale di circa 40 ore, attraverso l'analisi delle conversazioni su Facebook e Twitter, le foto su Instagram, i video su YouTube.
- Anticipando l'esplosione del trend, un'azienda potrebbe progettare e lanciare la sua campagna pubblicitaria con qualche preziosissima ora di anticipo, sbaragliando la concorrenza e massimizzando i guadagni. «Quello che facciamo è dare ai brand oggi il giornale di domani», spiega Randy Browning, CEO di Blab.
- Il suo esempio è quello di Oreo, la famosa marca di biscotti che, durante la mezz'ora di blackout che bloccò il Super Bowl del 2013, twittò ai propri clienti lo slogan coniato sul momento: “Potete inzuppare i biscotti anche al buio”.
- Blab ha ottenuto a oggi 6 milioni di dollari di finanziamenti e lavora con almeno 10 grandi brand di primo piano a livello internazionale, sostenendo che le sue “predizioni” hanno un'affidabilità del 70% e possono trarre dati da conversazioni sui social network in qualsiasi lingua.

Anticipare il futuro con i big data: lo stiamo facendo nel modo sbagliato

Roberto Paura: italian Institute for the Future e direttore della rivista "Futuri".

- Un modello simile è alla base di EMBERS, un sofisticato programma messo a punto da un team di ricercatori diretti da Naren Ramakrishnan del Virginia Tech. Con EMBERS, Ramakrishnan e il suo gruppo hanno tenuto d'occhio per anni tutti i tweet, gli articoli online, i post sui blog e su Facebook prodotti in America Latina e hanno previsto correttamente con anticipo le sollevazioni in Venezuela e i tumulti in Brasile.
- EMBERS è il programma che ha vinto la competizione indetta dalla IARPA (Intelligence Advanced Research Projects Activity), agenzia governativa americana che lavora nello sviluppo di più sofisticati sistemi d'intelligence. L'interesse dei servizi segreti nelle capacità anticipatrici del web è evidente. Tuttavia, anche in questo caso, l'uso che si sta facendo dei big data è sbagliato. Che differenza c'è, dopotutto, tra l'utilizzo di questi dati per favorire una catena di supermercati o un'agenzia di intelligence piuttosto che un'altra?

Conclusioni

- * I Big Data sono già qui
- * Ampia disponibilità in diversi settori
- * La tecnologia per i Big Data esiste
- * Il problema è nell'uso che se ne fa
- * I Big Data possono creare valore
- * Gli aspetti metodologici sono importanti
- * Un'area in rapida evoluzione



Bibliografia sitografia

- * Wikipedia
- * "Big Data: The next frontier for innovation, competition, and productivity". Rapporto McKinsey&Company, 2012.
- * Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati 20 nov. 2013 di Alessandro Rezzani
- * "Taming The Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics". Bill Franks, John Wiley & Sons, 2012.
- * <http://www.instituteforthefuture.it/anticipare-il-futuro-con-big-data-lo-stiamo-facendo-nel-modo-sb>
- * Seminario di Riccardo Torlone Università Roma Tre