Complex and Neural Network

Lesson 3

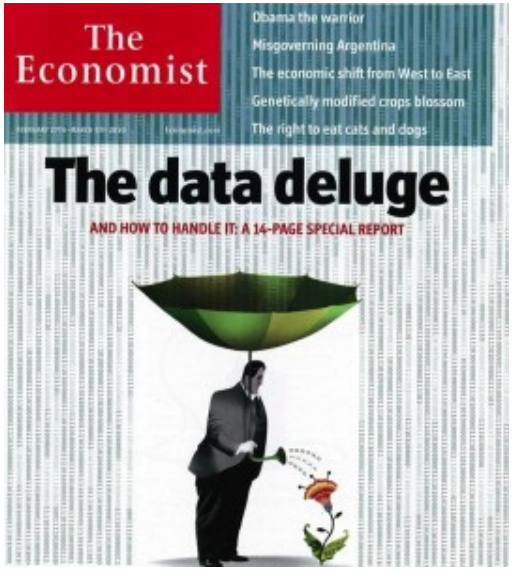
Gaetano Salina

"Big data is like teenage sex;
everyone talks about it,
nobody really knows how to do it,
everyone thinks everyone else is doing it,
so everyone claims they are doing it".

Dan Ariely, Duke University

An interesting metaphor

Is Big Data an important problem ?



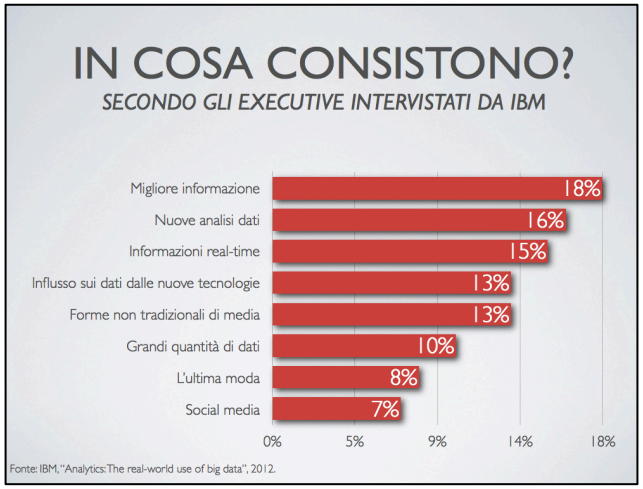
Is Big Data an *even more* important problem ?

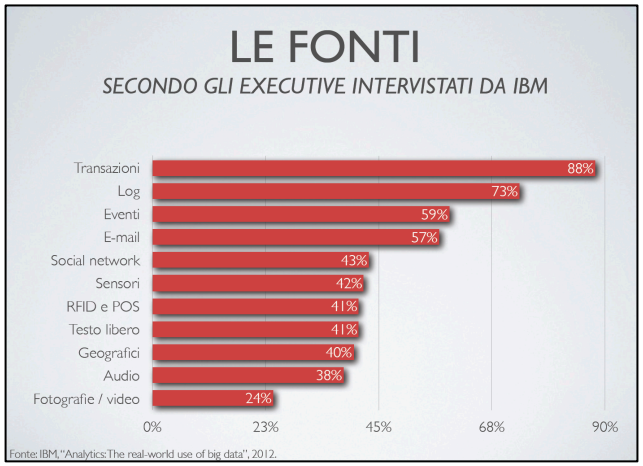
What is Big Data?

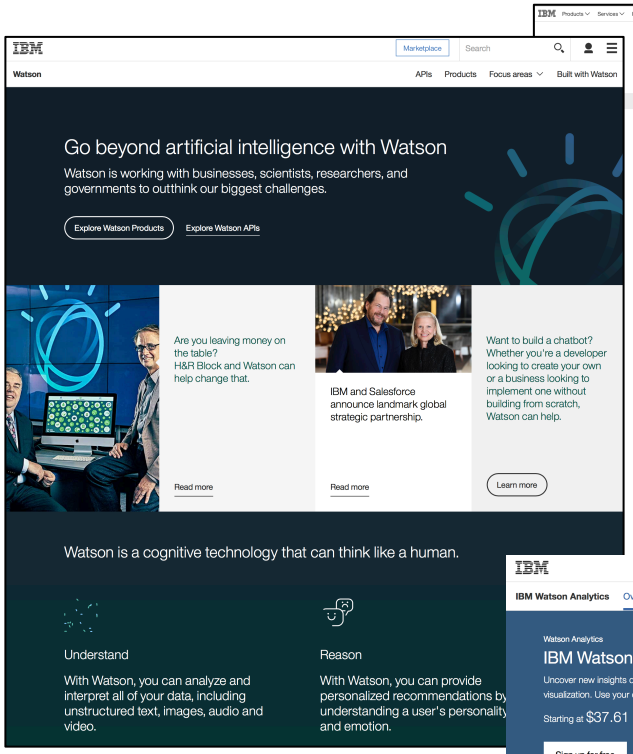
Definition:

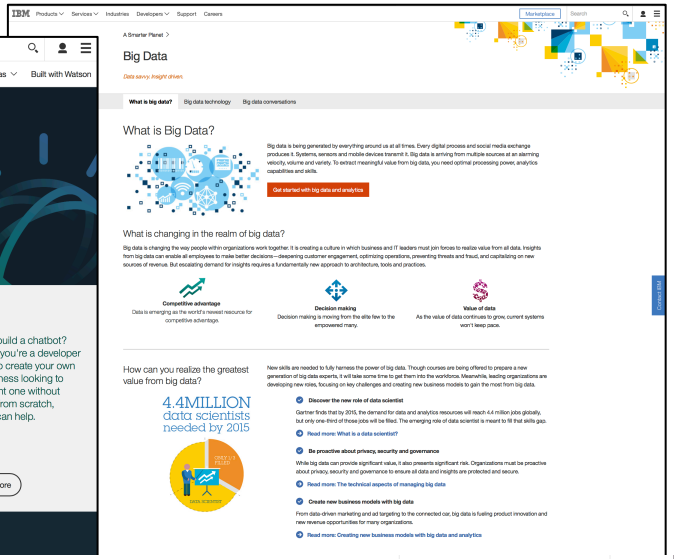
- Anything that Won't Fit in Excel!
- Big data is data defined by using the Vs
- Big data is data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it…
- Big data exceeds the reach of commonly used hardware environments and software tools to capture, manage, and process it with in a tolerable elapsed time for its user 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, 2014

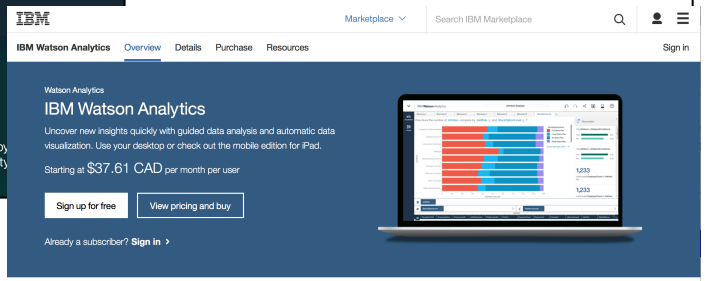
What is Big Data?

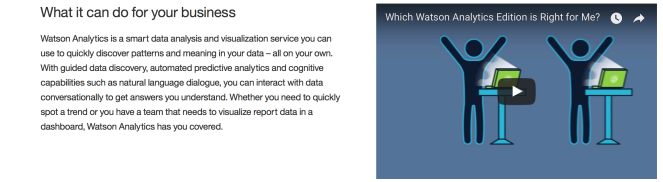








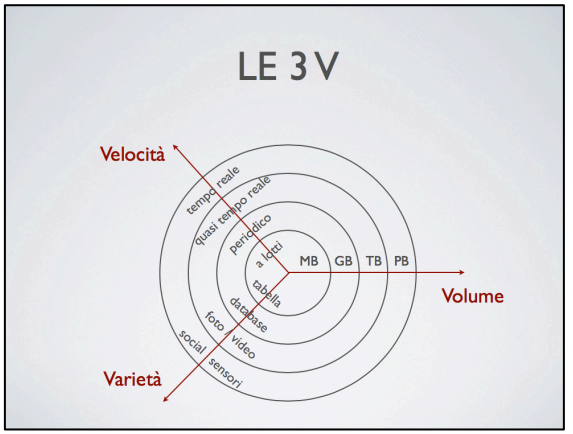


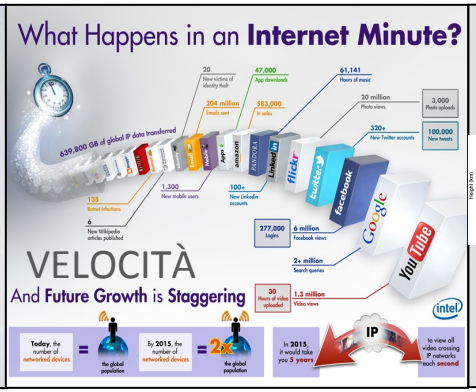


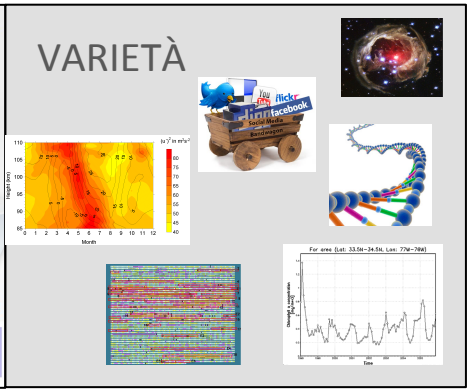
Big Data is now the core business of IBM

The three Vs

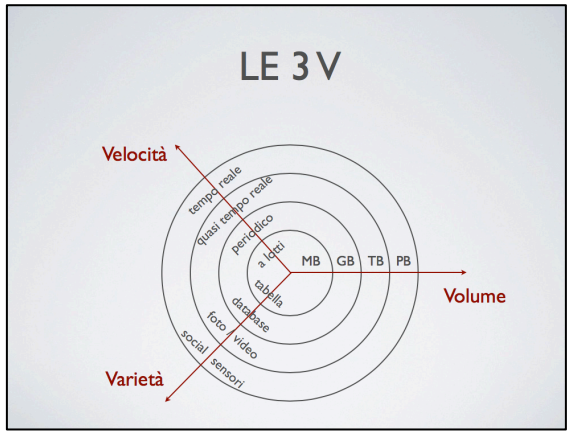
- **Volume:** size does matter!
- **Velocity**: data at speed, i.e., the data "fire-hose"
- **Variety:** heterogeneity is the rule

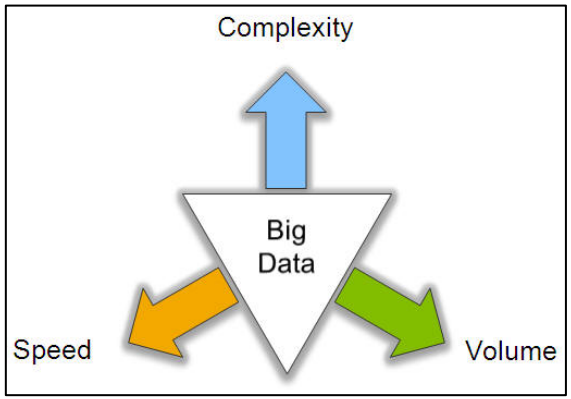


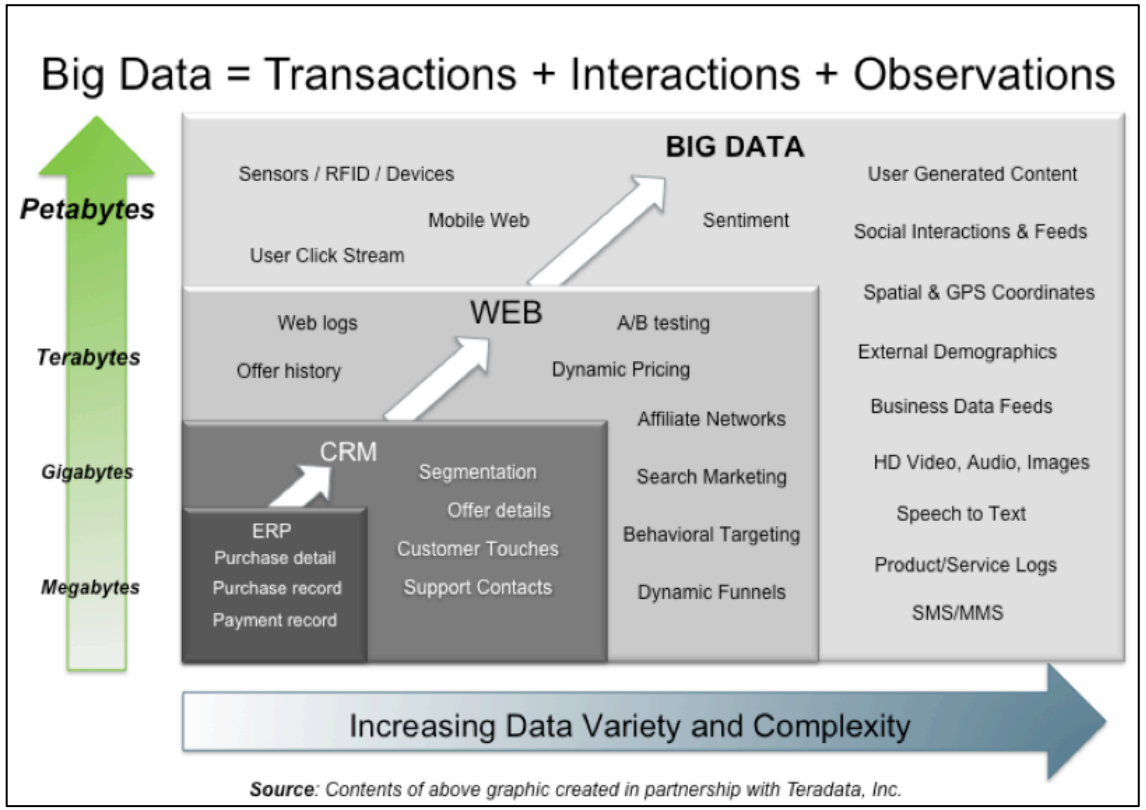




The three Vs

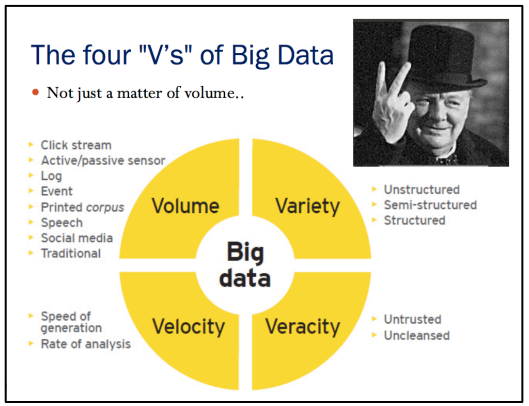

LE 3 V
Velocità
Volume
Varietà
tempo reale
quasi tempo reale
periodico
a lotti
MB
GB
TB
PB
tabella
database
foto
video
social
sensori


Complexity
Big
Data
Speed
Volume


Big Data = Transactions + Interactions + Observations
Petabytes
Terabytes
Gigabytes
Megabytes
BIG DATA
Sensors / RFID / Devices
Mobile Web
User Click Stream
Sentiment
WEB
Web logs
Offer history
A/B testing
Dynamic Pricing
CRM
ERP
Purchase detail
Purchase record
Payment record
Segmentation
Offer details
Customer Touches
Support Contacts
Affiliate Networks
Search Marketing
Behavioral Targeting
Dynamic Funnels
User Generated Content
Social Interactions & Feeds
Spatial & GPS Coordinates
External Demographics
Business Data Feeds
HD Video, Audio, Images
Speech to Text
Product/Service Logs
SMS/MMS
Increasing Data Variety and Complexity
Source: Contents of above graphic created in partnership with Teradata, Inc.

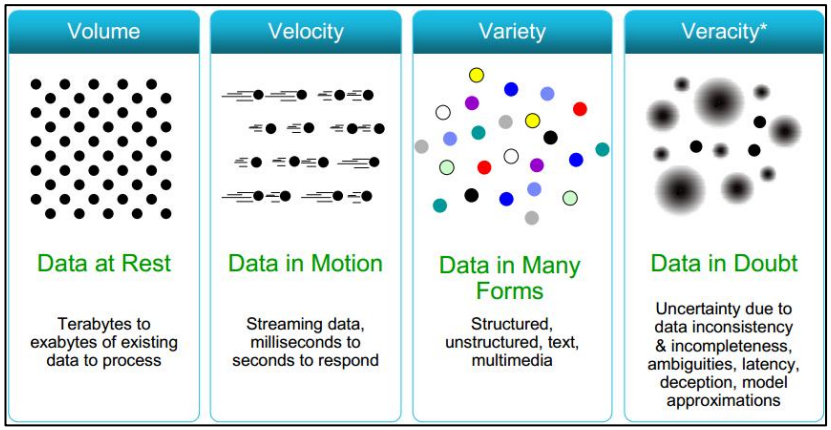
More Vs

- **Value:** usefulness & ability to find the right-needle in the stack
- **Veracity:** ability to handle uncertainty, inconsistency, etc
- **Variability:** rapid change of data characteristics over time
- **Visibility:** protect privacy and provide security
- **Voracity:** strong appetite for data!

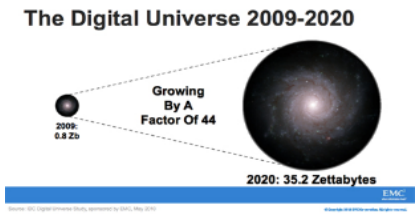


Big Data: V^4+VALUE

- Volume: Terabyte(10^{12}), Petabyte(10^{15}), Exabyte(10^{18}), Zettabyte (10^{21})
- Variety: Structured, semi-structured, unstructured; Text, image, audio, video, record
- Velocity: Periodic, Near Real Time, Real Time
- Veracity: Quality of the data can vary greatly
- Value: Big data can generate huge competitive advantages

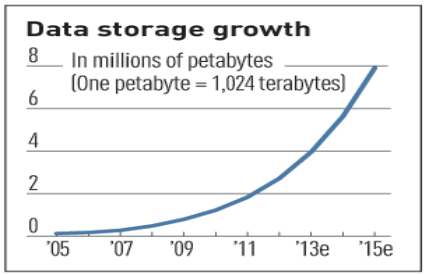


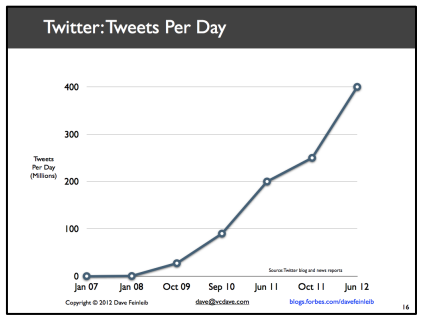
Characteristics of Big Data:
Scale (Volume)



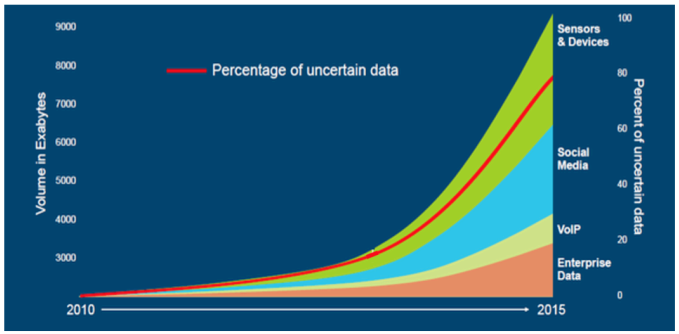
- Data Volume
  - 44x increase from 2009 2020
  - From 0.8 zettabytes to 35zb
- Data volume is increasing exponentially

Exponential increase in collected/generated data



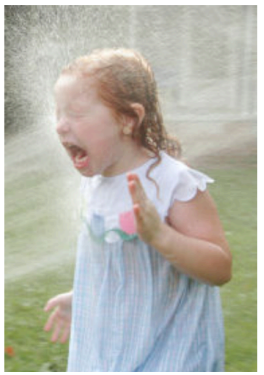


Growth rate



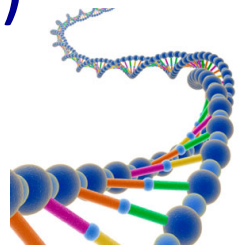
Some numbers

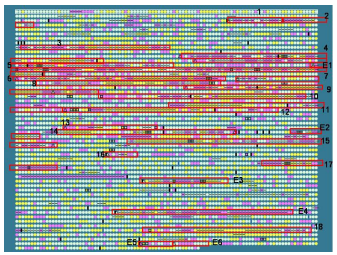
- How many data in the world?
  - 800 Terabytes, 2000
  - 160 Exabytes, 2006 (1EB = 10^{18}B)
  - 4.5 Zettabytes, 2013 (1ZB = 10^{21}B)
  - 44 Zettabytes by 2020
- How much is a zettabyte?
  - 1,000,000,000,000,000,000,000 bytes
  - A stack of 1TB hard disks that is 25,400 km high
- How many data in a day?
  - 2.5 Exabytes
  - 8 TB, Twitter
  - 50 TB, Facebook
- 90% of world's data:
  - generated over last two years!

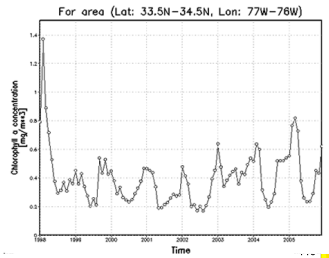


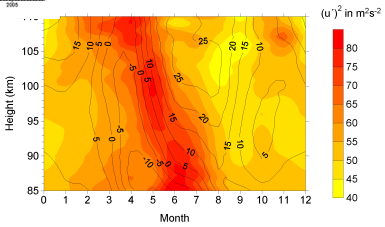
Characteristics of Big Data:
Complexity (Varity)

- Various formats, types, and structures
- Text, numerical, images, audio, video, sequences, time series, social media data, multi-dim arrays, etc…
- Static data vs. streaming data
- A single application can be generating/ collecting many types of data









To extract knowledge➔ all these types of data need to linked together.
Different Data Semantic

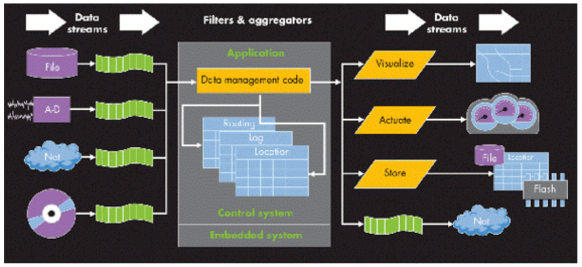
Characteristics of Big Data:
Speed (Velocity)

- Data is begin generated fast and need to be processed fast
- Online Data Analytics
- Late decisions ➔ missing opportunities
- Examples:

  E-Promotions: Based on your current location, your purchase history, what you like ➔ send promotions right now for store next to you

  Healthcare monitoring: sensors monitoring your activities and body ➔ any abnormal measurements require immediate reaction

Data-centric applications



The big data process

Who's Generating Big Data

Social media and networks

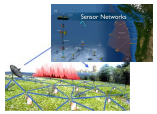
(all of us are generating data)

Scientific instructments

(collecting all sorts of data)

Mobile devices

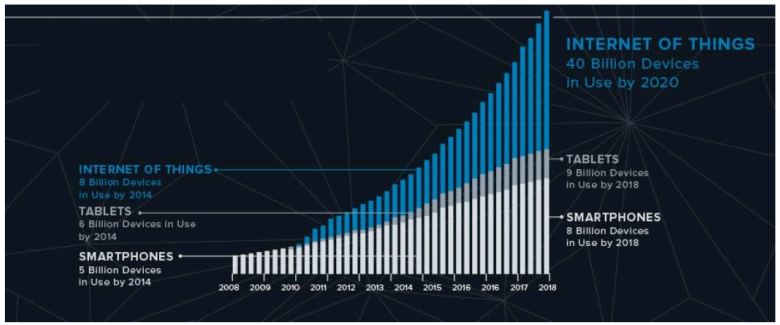
(tracking all objects all the time)



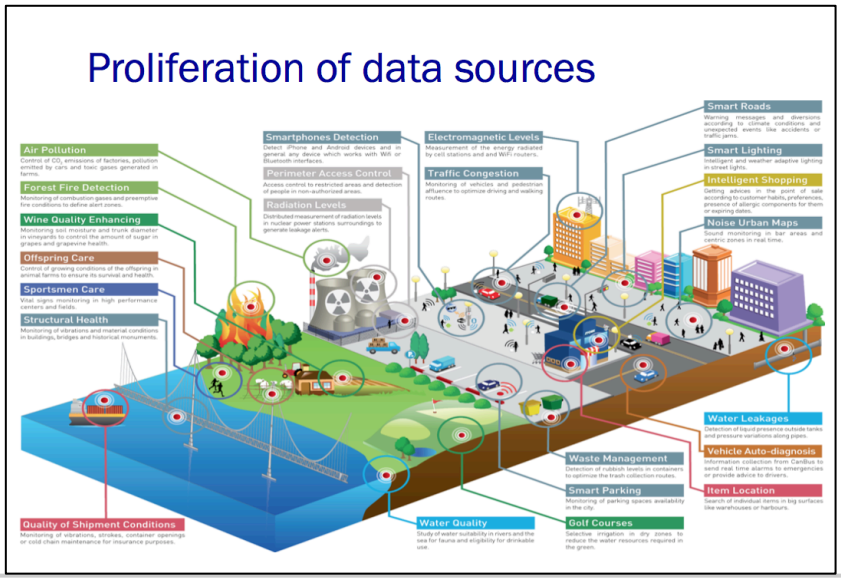
Sensor technology and networks

(measuring all kinds of data)

Global Internet device forecast



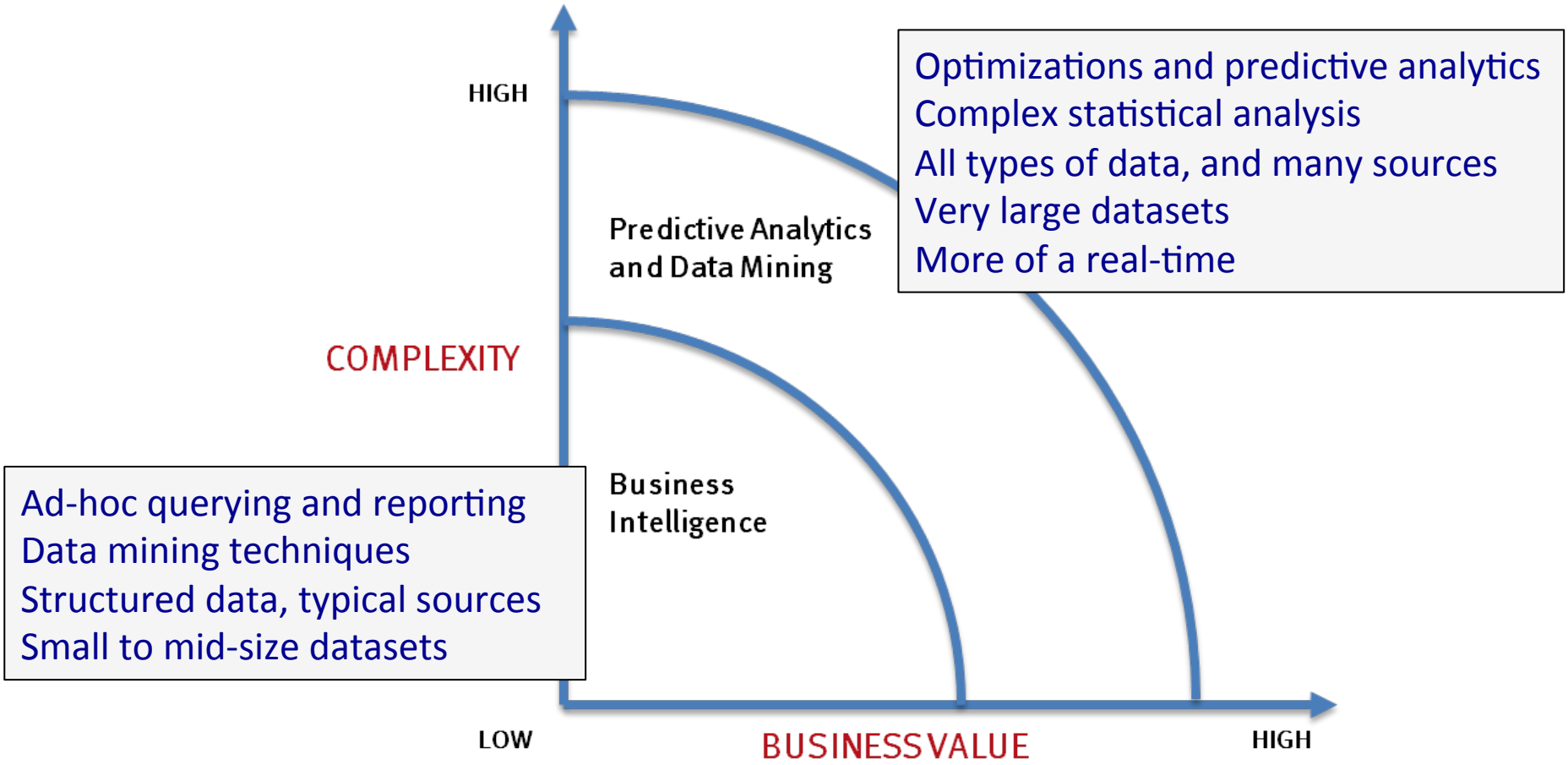
Proliferation of data sources



The progress and innovation is no longer hindered by the ability to collect data.

But, by the ability to manage, analyze, summarize, visualize, and discover knowledge from the collected data in a timely manner and in a scalable fashion.

What's driving Big Data

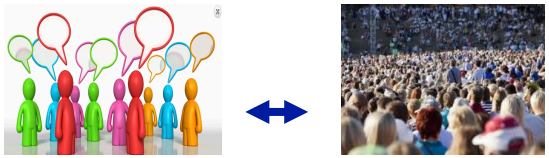


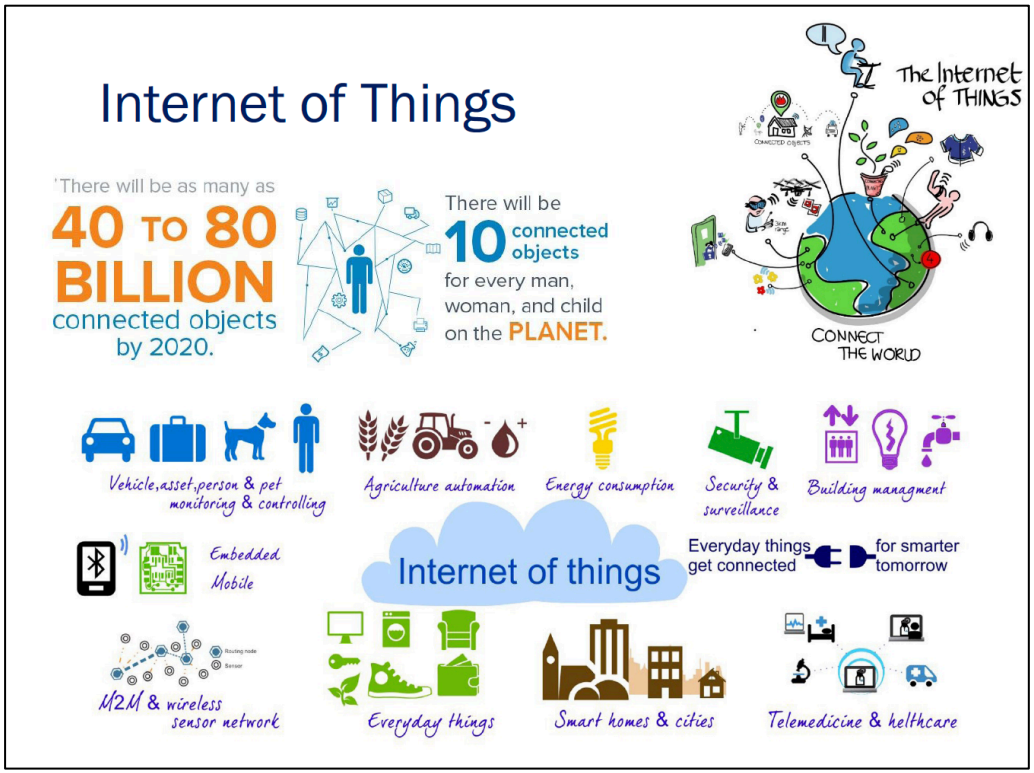
The Model Has Changed…

The Model of Generating/Consuming Data has Changed

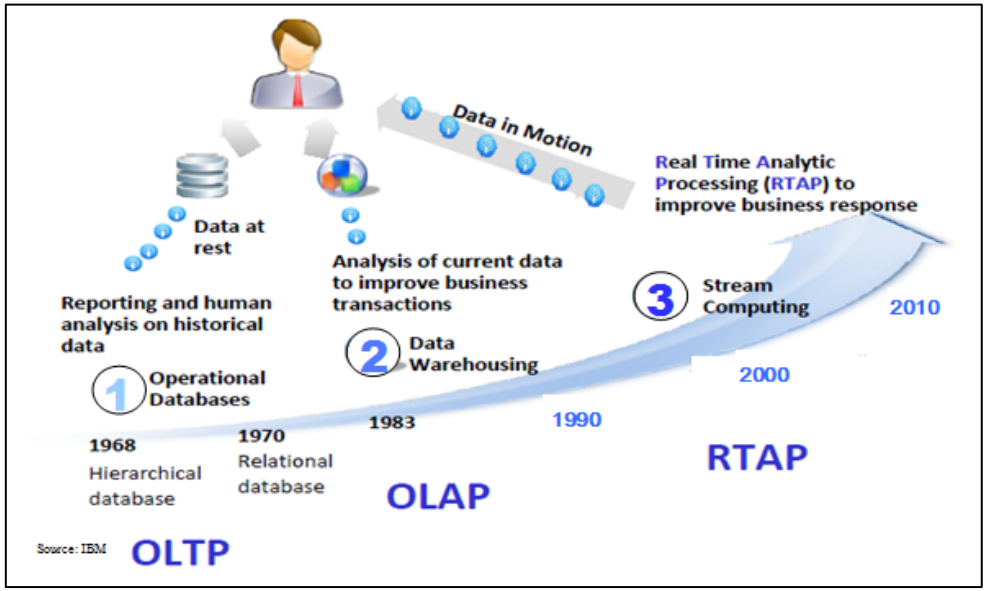
Old Model: Few companies are generating data, all others are consuming data

New Model: all of us are generating data, and all of us are consuming data





Harnessing Big Data

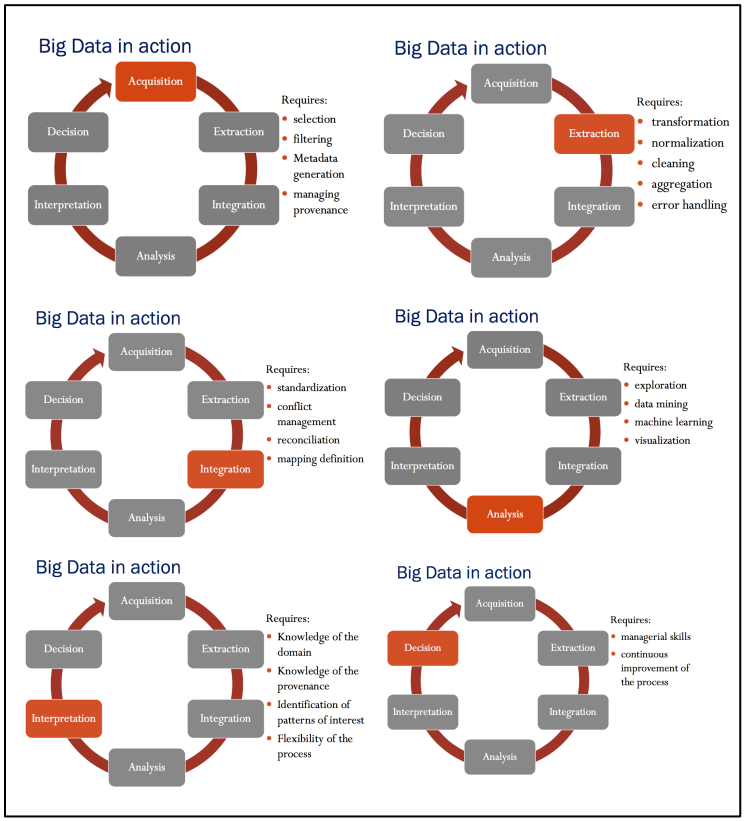


OLTP: Online Transaction Processing
(DBMSs)

OLAP: Online Analytical Processing
(Data Warehousing)

RTAP: Real-Time Analytics Processing
(Big Data Architecture & technology)

Real-Time Analytics Processing

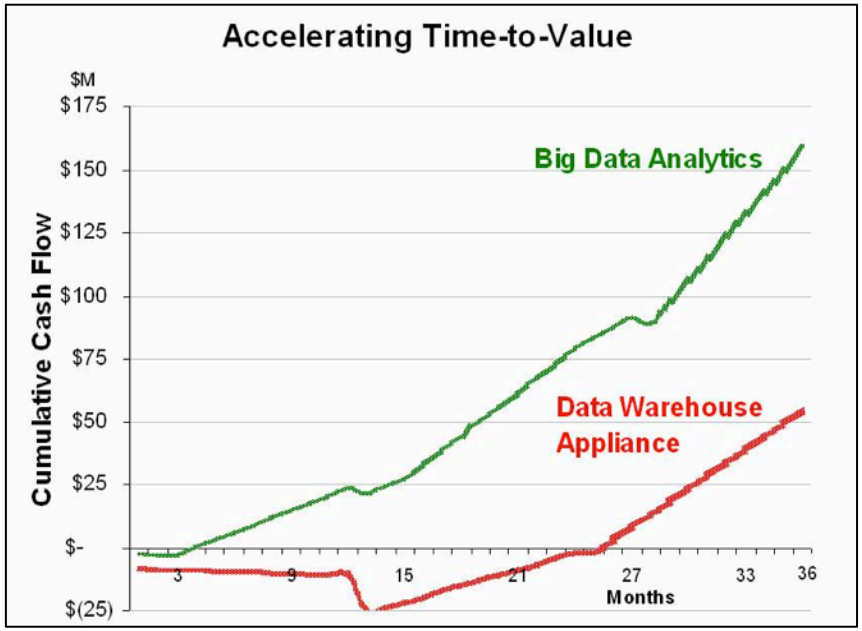


Value of Big Data Analytics

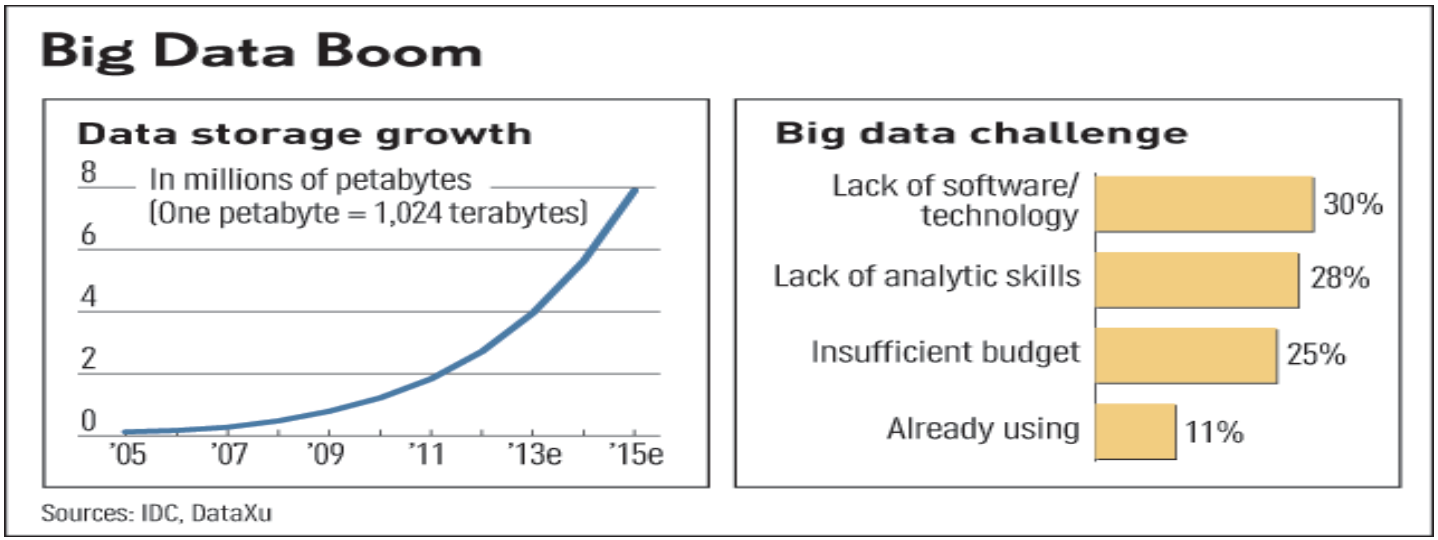
Big data is more real-time in nature than traditional DW applications

Traditional DW architectures (e.g. Exadata, Teradata) are not well-suited for big data apps

Shared nothing, massively parallel processing, scale out architectures are well-suited for big data apps



Challenges in Handling Big Data



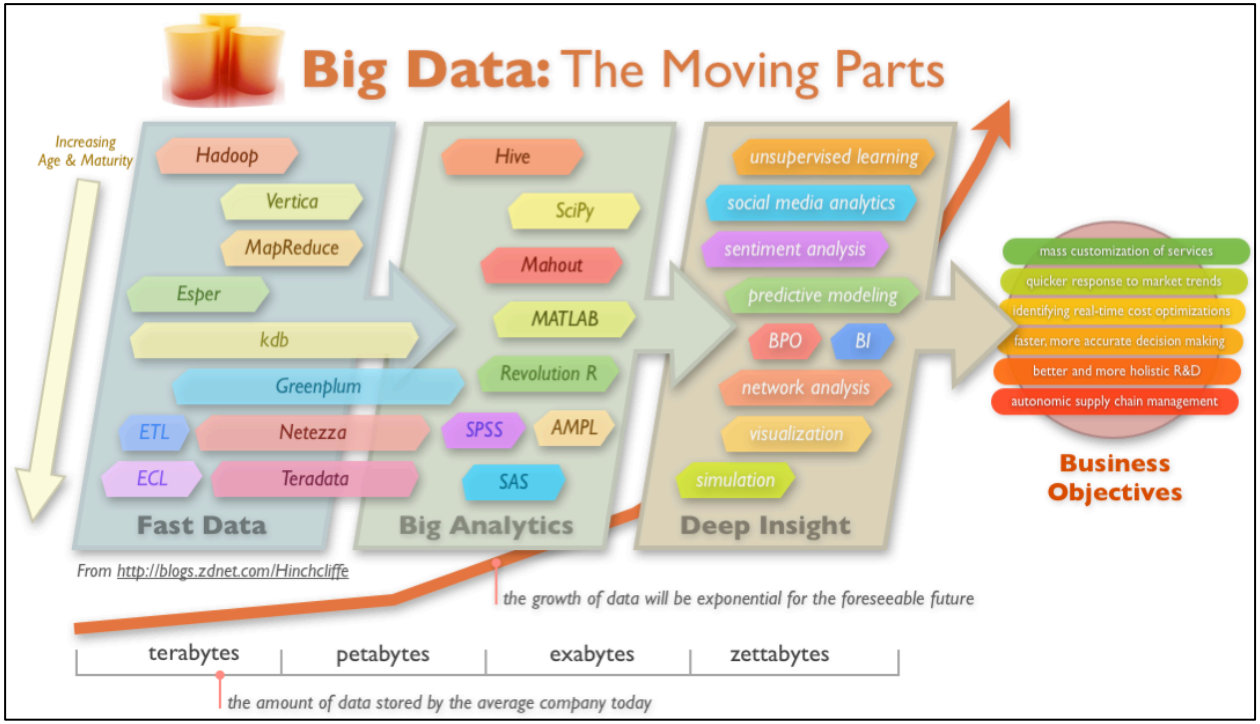
The Bottleneck is in technology

New architecture, algorithms, techniques are needed

Also in technical skills

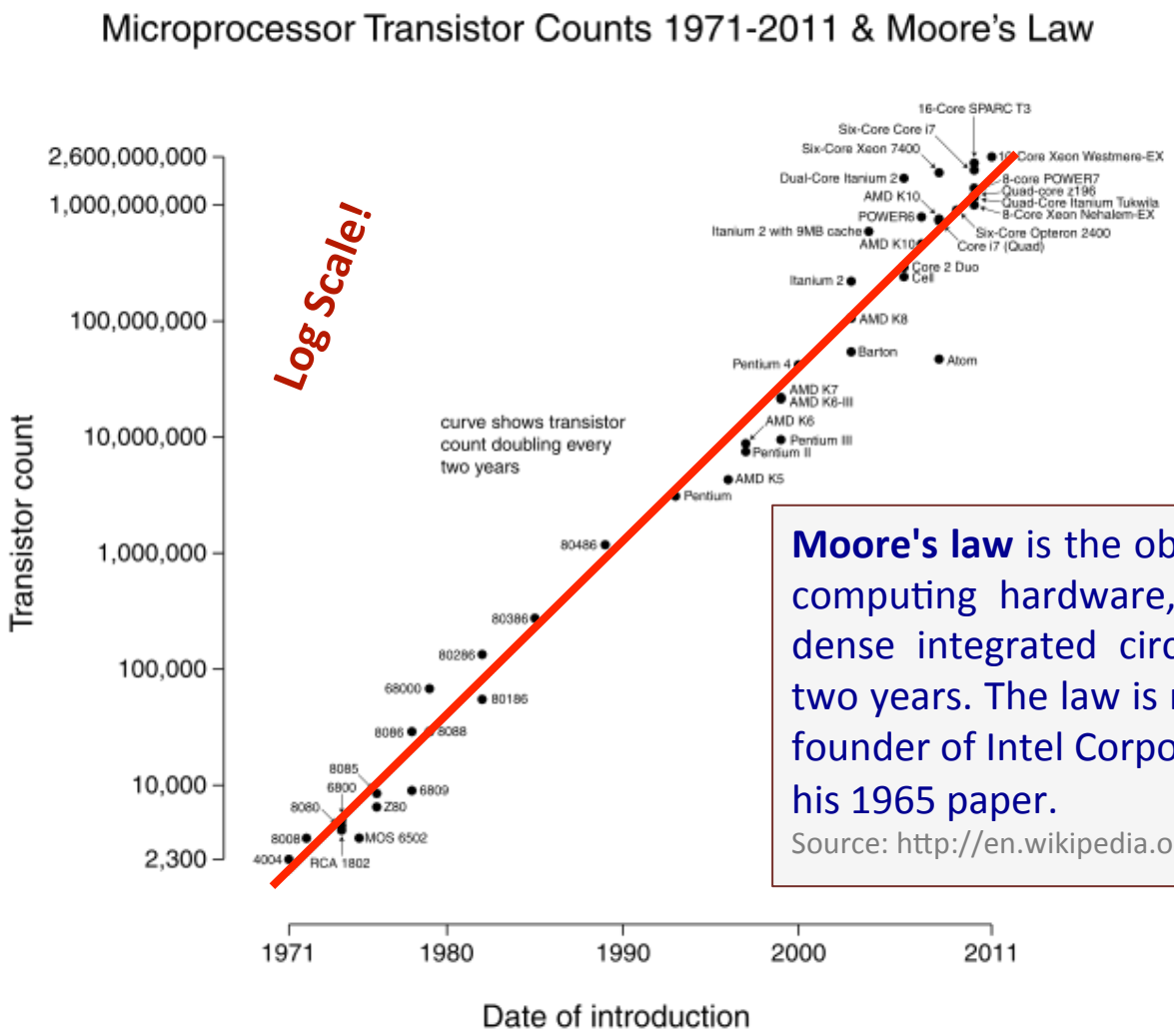
Experts in using the new technology and dealing with big data

What Technology Do We Have For Big Data ??



The Triumph of digital paradigm!

Enter Moore's Law



Moore's law is the observation that, over the history of computing hardware, the number of transistors in a dense integrated circuit doubles approximately every two years. The law is named after Gordon E. Moore, co-founder of Intel Corporation, who described the trend in his 1965 paper.

Source: http://en.wikipedia.org/wiki/Moore's_law

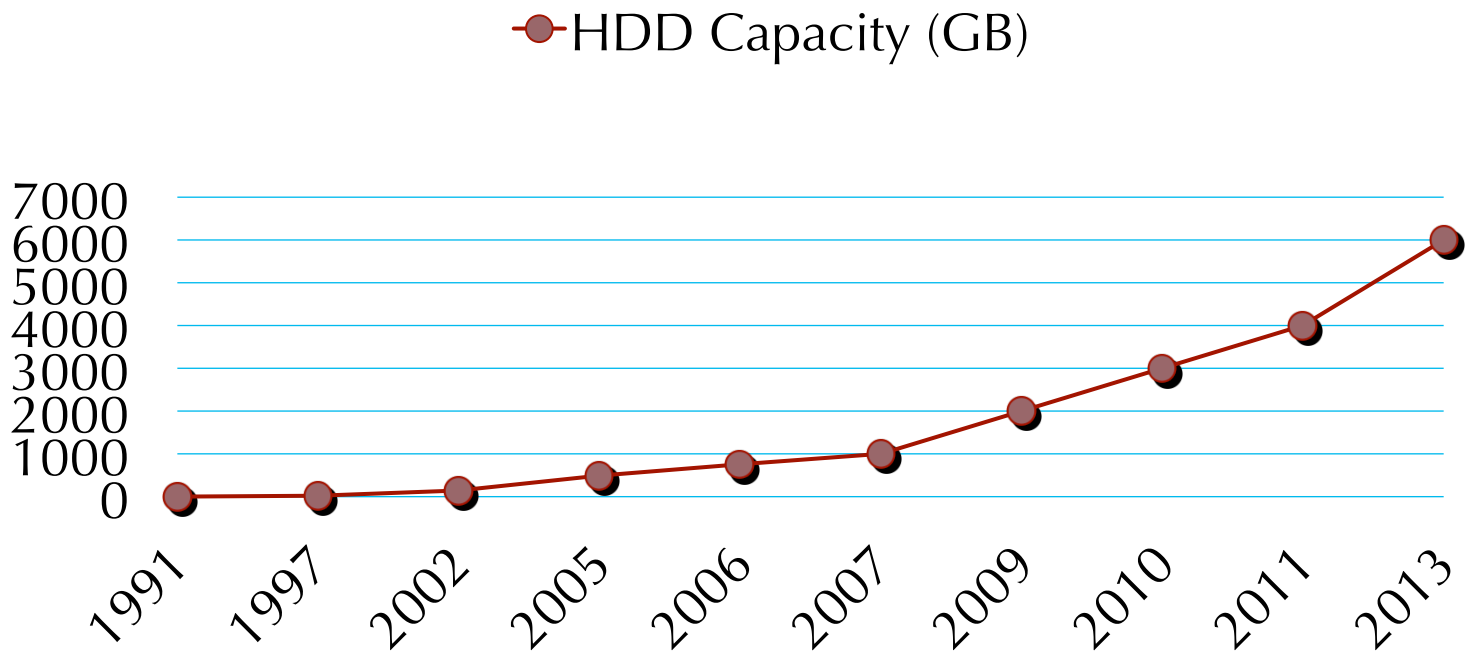
Enter Bezos' Law

Bezos' law is the observation that, over the history of cloud, a unit of computing power price is reduced by **50%** approximately every **3** years

Source: http://blog.appzero.com/blog/futureofcloud

Photo: http://www.slashgear.com/google-data-center-hd-photos-hit-where-the-internet-lives-gallery-17252451/

Enter Storage capacity increase



Insert other exponentially increasing graphs here

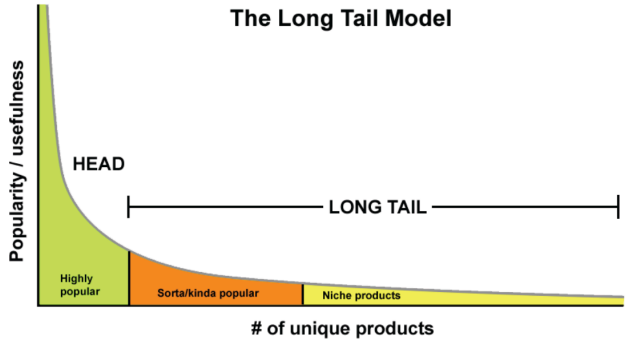
(e.g., data generation rates, world-wide smartphone access rates, Internet of Things, …)

Together with …

Bigger = Smarter?

- YES!
  - algorithms work much better
  - tolerate errors
  - discover the "long tail" and "corner cases"
- BUT:
  - more heterogeneity
  - data grows faster than energy on chip
  - still need humans to ask right questions

Big tail



"We sold more items today that didn't sell at all yesterday than we sold today of all the items that did sell yesterday" – Amazon employee.

Why now?

- Because we have data
  - Data born already in digital form
  - 40% of data growth per year
- Because we can

  - 400$ for a drive in which to store all the music of the world
  - >40 years of Moore's Law → large computational resources
  - 76% of organizations have invested in big data in 2016
  - 130 billions $ invested in big data in 2016
- "Because we reached dead end with logic"

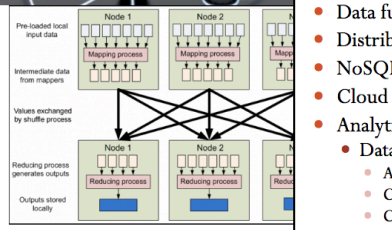
A simple example of bigger=smarter

- Google Translate
  - you collect snippets of translations
  - you match sentences to snippets
  - you continuously debug your system
- Why does it work?
  - there are tons of snippets on the Web
  - the accuracy improves as the training set grows

Together with …

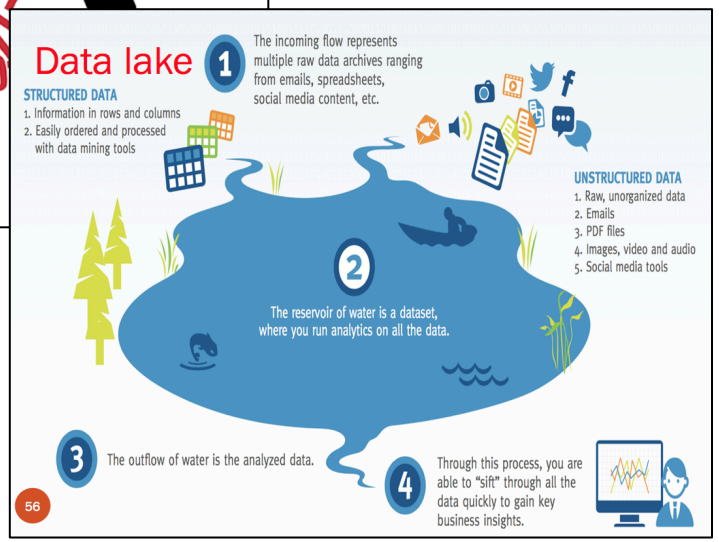
Distribution of resources and services

- Distributed Architecture
  - Clusters of computers that work together to a common goal
  - Scale out not up!
- Fault- tolerance
  - Resource replication
  - Eventual consistency
- Distributed processing
  - Shared-nothing model
  - New programming paradigms



Techniques for big data analysis

- Extract, transform, and load (ETL)
- Data fusion and data integration
- Distributed file system
- NoSQL database systems
- Cloud computing
- Analytics
  - Data mining
    - Association rule learning
    - Classification
    - Cluster analysis
    - Regression
  - Machine learning
    - Supervised learning
    - Unsupervised learning
- Crowdsourcing
- …



It is the golden age of the digital approach ?

A simple example of a big data process

- Problem: The sale of lollipops is going down!
- Acquisition:
  - Sales by customer, region and time
  - Surveys of users
  - Social networks
- Extraction:
  - Data loading from receipts
  - Automatic reading of questionnaires
  - Data extraction from twitter
- Integration:
  - On the basis of user types
- Analysis:
  - lollipops bought by people older than 25
  - lollipops preferred by people younger than 10
- Interpretation:
  - Moms believe: lollipops = bad teeth
  - Boys and girls believe that lollipops are for babies
- Decision:
  - We make lollipops without sugar
  - We ask dentists to advertise our lollipops
  - We make commercials targeted to boys and girls

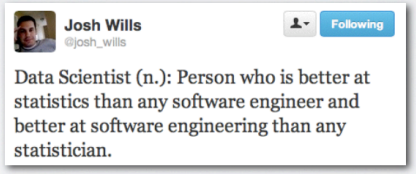
The New Opportunity …

LE OPPORTUNITÀ

1. I big data applicati alla sanità possono far risparmiare agli Stati Uniti **300 B$** in efficienza.
2. L'Europa può risparmiare **149 B$** in costi di amministrazione e governo.
3. Solo negli Stati Uniti serviranno nel breve periodo **1.5+ M** di *data scientist* e *data manager*.

NUOVE (?) PROFESSIONI

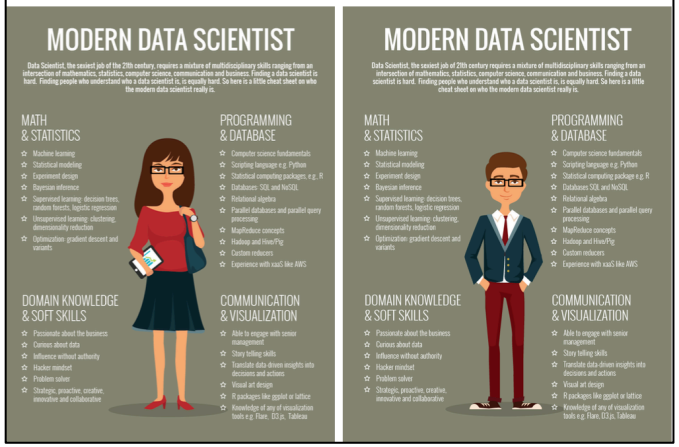
- I bit sono inutili senza qualcuno che li sappia interpretare!



Data scientist: a brand new profession

- Data Scientist: The Sexiest Job of the 21st Century [Harward Business Review 2013]
- Data scientist? A guide to 2015's hottest profession [Mashable 2015]
- "It's official – data scientist is the best job in America" [Forbes, 2016]

Skills of data scientists



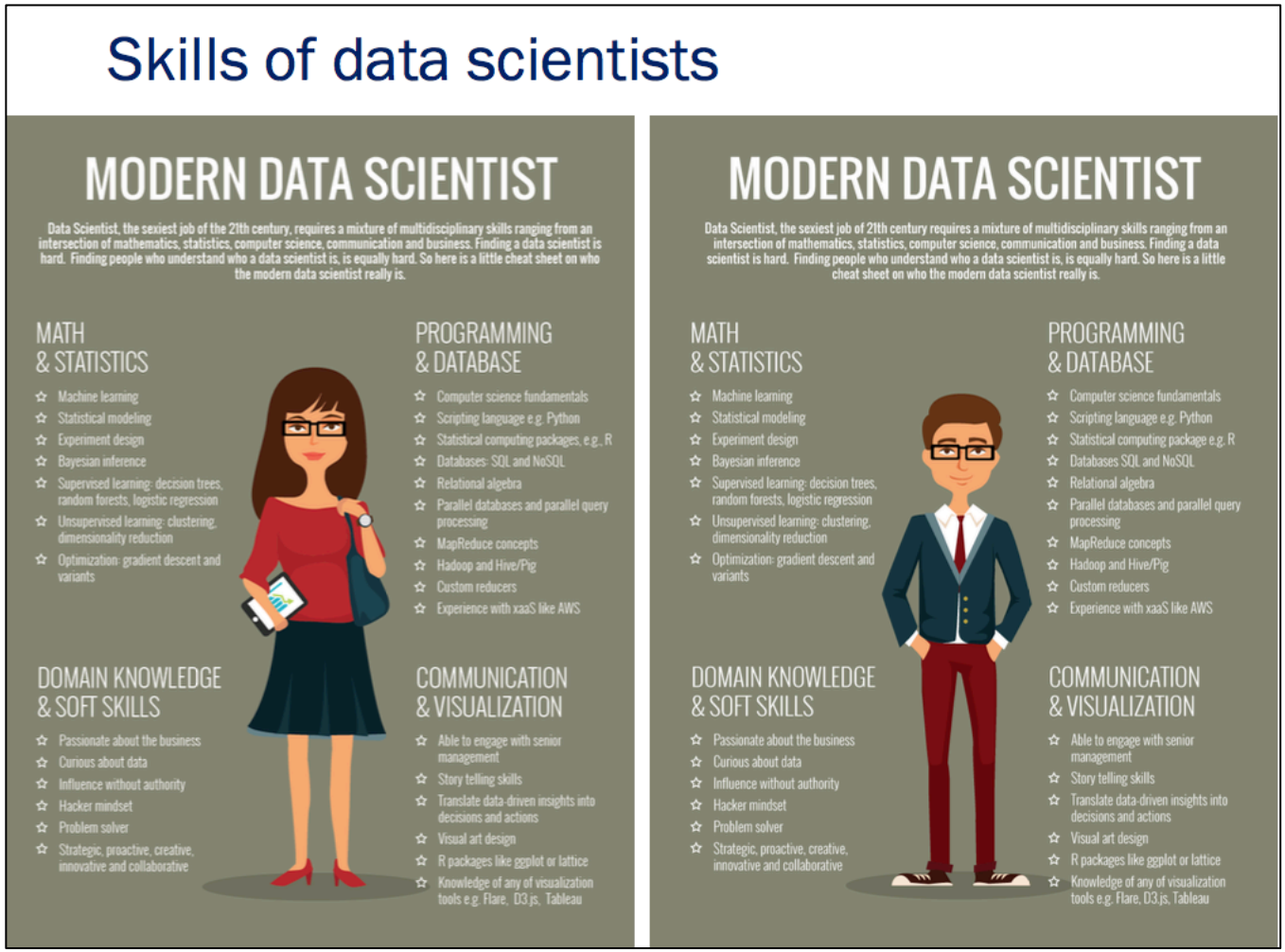
…

There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.

…

Chris Anderson

The New Opportunity …



…

There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.

…

Chris Anderson

Google's founding philosophy is that we don't know why this page is better than that one: If the statistics of incoming links say it is, that's good enough.

No semantic or causal analysis is required. That's why Google can translate languages without actually "knowing" them (given equal corpus data, Google can translate Klingon into Farsi as easily as it can translate French into German). And why it can match ads to content without any knowledge or assumptions about the ads or the content.

Speaking at the O'Reilly Emerging Technology Conference this past March, Peter Norvig, Google's research director, offered an update to George Box's maxim: "All models are wrong, and increasingly you can succeed without them."

This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.

https://www.wired.com/2008/06/pb-theory/

CHRIS ANDERSON SCIENCE 06.23.08 12:00 PM

THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE

Illustration: Marian Bantjes

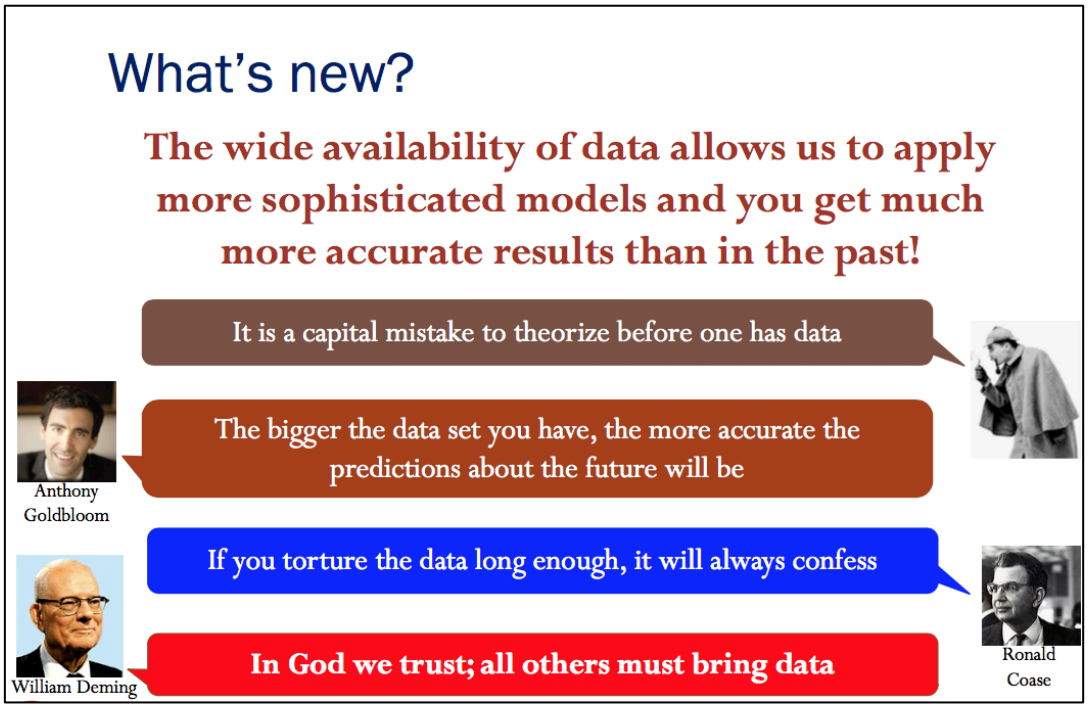
"All models are wrong, but some are useful."

This article has been reproduced in a new format and may be missing content or contain faulty links. Contact wiredlabs@wired.com to report an issue.

So proclaimed statistician George Box 30 years ago, and he was right. But what choice did we have? Only models, from cosmological equations to theories of human behavior, seemed to be able to consistently, if imperfectly, explain the world around us. Until now. Today companies like Google, which have grown up in an era of massively abundant data, don't have to settle for wrong models. Indeed, they don't have to settle for models at all.

It is all OK!


ft.com > life&arts >
FT Magazine
Home UK World Companies Markets Global Economy Lex C
Arts Magazine Food & Drink House & Home Lunch with the FT Style Books
March 28, 2014 11:38 am
Big data: are we making a big mistake?
By Tim Harford
Big data is a vague term for a massive phenomenon that has rapidly become an obsession with entrepreneurs, scientists, governments and the media
BIG DATA

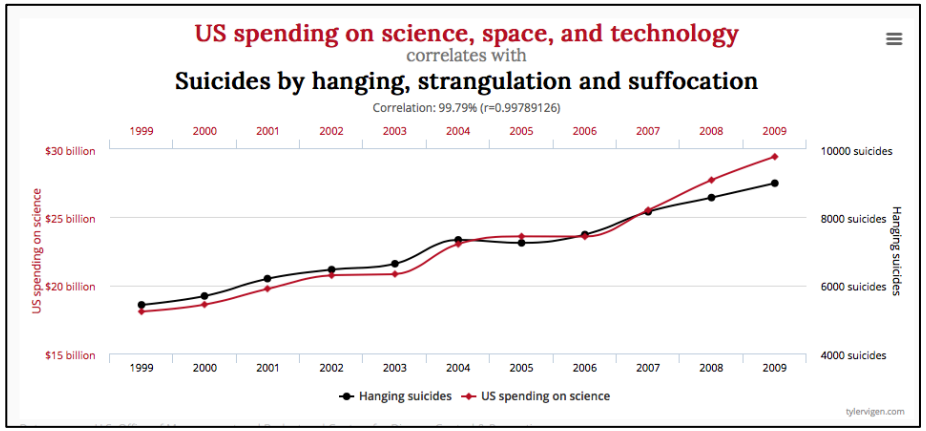

What's new?
The wide availability of data allows us to apply more sophisticated models and you get much more accurate results than in the past!
It is a capital mistake to theorize before one has data
The bigger the data set you have, the more accurate the predictions about the future will be
Anthony Goldbloom
If you torture the data long enough, it will always confess
Ronald Coase
In God we trust; all others must bring data
William Deming

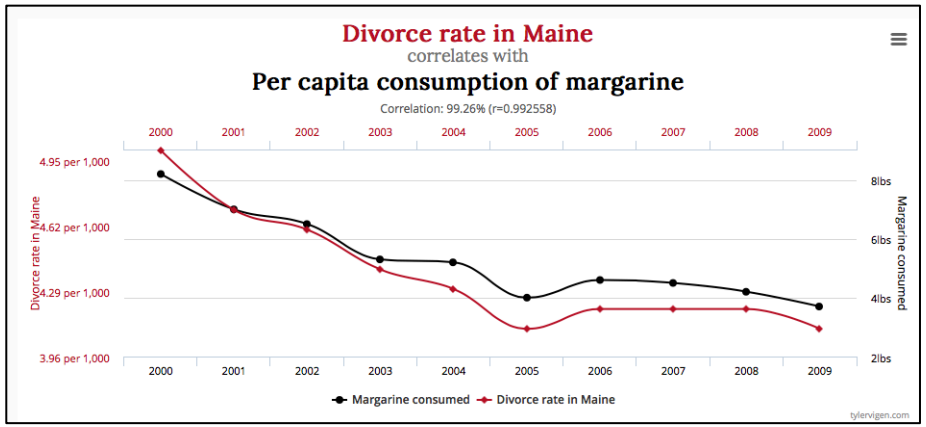
Big Data: una vera rivoluzione scientifica ?

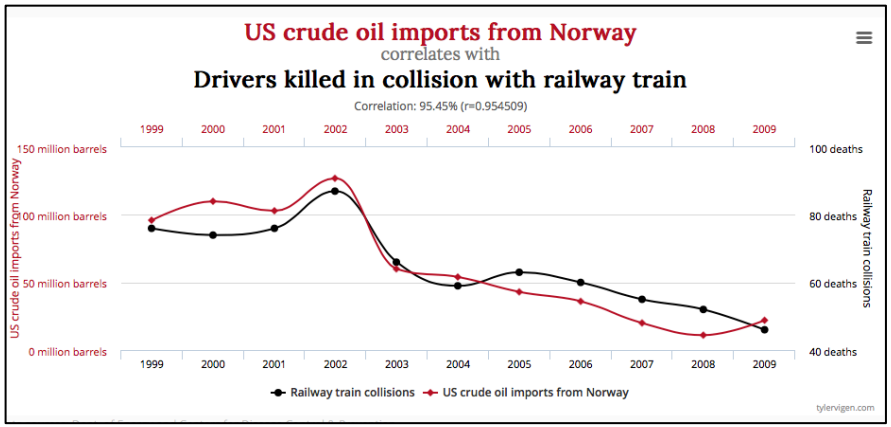
…

Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. …

Chris Anderson







http://www.tylervigen.com/spurious-correlations

Big Data: una vera rivoluzione scientifica ?

Viviamo in un mondo globalizzato con un'enorme produzione di dati, da anni ci sentiamo ripetere che siamo nell'era dei Big Data e che quest'abbondanza di informazioni non potrà che essere una risorsa importante in diversi ambiti, ad esempio per la sicurezza, per le assicurazioni o per aumentare l'efficienza della aziende.

...

Senza grande sorpresa, anche la politica ha scoperto i Big Data ed il loro ruolo potenziale nell'ambito scientifico e tecnologico. Ad esempio il governo italiano, sulla scia di Expo, intende lanciare il progetto Human Technopole, di cui ha recentemente discusso Francesco Sinopoli sul Menabò, che è incentrato in gran parte proprio sul trattamento di una grande mole di dati.

In questo articolo mi occuperò di un aspetto, sicuramente meno interessante per il grande pubblico, ma importante da un punto di vista culturale, in particolare per la ricerca scientifica, ovvero la possibilità, offerta dai Big Data, di realizzare una nuova rivoluzione scientifica che consenta di fondare una scienza senza basi teoriche.

http://www.eticaeconomia.it/big-data-una-vera-rivoluzione-scientifica/

eticaeconomia
menabò
fondato da Luciano Barca

Big Data: una vera rivoluzione scientifica?

Articoli | Istituzioni e regole | Angelo Vulpiani 31 gennaio 2017

Angelo Vulpiani si occupa dell'attesa "rivoluzione dei Big Data", partendo dalla diffusa opinione secondo cui grazie all'enorme mole di informazioni attualmente disponibili saremmo prossimi ad una nuova rivoluzione che consentirebbe di fondare una scienza senza basi teoriche. Vulpiani illustra le ragioni del suo scetticismo e sostiene che non è opportuno esagerare l'importanza dei Big Data, il cui impatto nell'ambito della ricerca, peraltro, non è stato finora sostanziale.

Viviamo in un mondo globalizzato con un'enorme produzione di dati, da anni ci sentiamo ripetere che siamo nell'era dei Big Data e che quest'abbondanza di informazioni non potrà che essere una risorsa importante in diversi ambiti, ad esempio per la sicurezza, per le assicurazioni o per aumentare l'efficienza della aziende.

Sicuramente avere a disposizione una grande mole di informazioni non può che essere utile: ad esempio, più dati sull'influenza vengono raccolti meglio sarà per la prevenzione. Ovviamente i Big Data fanno nascere una serie di problemi etici e legali: chi è il proprietario dei miei dati sanitari? Quale uso possono farne le compagnie di assicurazioni, i datori di lavoro, i tribunali?

Senza grande sorpresa, anche la politica ha scoperto i Big Data ed il loro ruolo potenziale nell'ambito scientifico e tecnologico. Ad esempio il governo italiano, sulla scia di Expo, intende lanciare il progetto Human Technopole, di cui ha recentemente discusso Francesco Sinopoli sul Menabò, che è incentrato in gran parte proprio sul trattamento di una grande mole di dati.

In questo articolo mi occuperò di un aspetto, sicuramente meno interessante per il grande pubblico, ma importante da un punto di vista culturale, in particolare per la ricerca scientifica, ovvero la possibilità, offerta dai Big Data, di realizzare una nuova rivoluzione scientifica che consenta di fondare una scienza senza basi teoriche.

A PENSARCI BENE...

Menabò n. 58/2017

Big Data: una vera rivoluzione scientifica ?

Secondo alcuni, con la disponibilità di grandi quantità di informazioni, saremmo di fronte ad una nuova rivoluzione scientifica: la possibilità di fare scienza attraverso l'analisi di dati avrebbe creato un quarto paradigma (T. Hey et al., "The Fourth Paradigm: Data Intensive Scientific Discovery", *Microsoft Research* 2009). Un nuovo approccio si aggiungerebbe, quindi, alle tre metodologie già esistenti: il metodo sperimentale, quello teorico matematico e quello computazionale (simulazioni numeriche).

Il guru informatico Chris Anderson è arrivato a sostenere, in un articolo dal titolo esplicitamente provocatorio "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", che "ormai la grande quantità di dati a disposizione rende il metodo scientifico obsoleto… i petabyte ci consentono di dire la correlazione è sufficiente, possiamo smettere di cercare modelli". Non sarebbe, quindi, più necessario studiare teorie generali, basterebbe scaricare i dati da Internet, trattarli al computer con opportuni algoritmi statistici ed avremmo tutto quello che serve. Uno degli slogan ricorrenti dei profeti dei Big Data è "basta la correlazione". Inutile insistere sul fatto che l'esistenza di una correlazione tra due quantità non dica molto, come mostrano alcuni esempi decisamente divertenti: la correlazione tra il numero di pirati e la temperatura media sulla terra, quella tra il consumo di cioccolata pro capite ed il numero di premi Nobel in un dato paese o quella tra il numero di affogati per caduta da un barca da pesca e il numero di matrimoni nel Kentucky.

Big Data: una vera rivoluzione scientifica ?

L'idea secondo la quale è sempre meglio avere più dettagli (o dati) è ingenua e fuorviante: quasi mai la scienza avanza per accumulo di dati, bensì per la capacità di eliminare gli aspetti secondari. Ovviamente fare questo non è semplice: più volte in fisica è stata sottolineata la difficoltà di individuare le "giuste variabili del sistema". In quasi ogni problema ci sono molti aspetti che sono irrilevanti e la prima cosa (forse la più difficile e importante) da fare è identificare la parte significativa del fenomeno, solo così si ha qualche speranza di capire.

Una descrizione molto dettagliata può avere conseguenze addirittura negative: Borges nel breve racconto "Funes, o della memoria" scrive di un personaggio che, in seguito ad un incidente, ricordava tutto di tutto, sin nei minimi dettagli della più comune delle situazioni. Questo, ben lungi dall'essere un fatto positivo, comportava la quasi incapacità di un pensiero astratto. Funes era infastidito che un cane visto di profilo alle 3:14 fosse lo stesso visto di fronte alle 3:15 e era quasi incapace di idee generali platoniche.

Per non rimanere troppo sul'astratto vale la pena discutere il caso delle previsioni meteo per mostrare chiaramente come, per un problema non banale, sia decisamente troppo ottimistico puntare solo sull'uso dei dati osservativi, ma sia necessaria una combinazione di tecniche matematiche, intuizione fisica e sviluppo tecnologico.
Assumiamo (cosa che non sempre è vera) di sapere che il fenomeno che vogliamo studiare è descritto da un set di variabili x(t) la cui evoluzione è deterministica. Per fare una previsione del futuro si potrebbe pensare di cercare nel passato una situazione "vicina" a quella di oggi, se la si trova al giorno k allora è sensato assumere che domani il sistema sarà "vicino" al giorno k+1 del passato.

Big Data: una vera rivoluzione scientifica ?

Sembrerebbe tutto facile, in particolare ora che siamo nell' era dei Big Data e, quindi, potremmo non perdere tempo con la teoria. Per prima cosa chiediamoci se sia sempre possibile individuare un analogo (cioè un giorno k nel passato in cui il sistema è "vicino" ad oggi).

Da un punto di vista matematico il problema è strettamente collegato ad un risultato classico della fine del diciannovesimo secolo (il teorema di ricorrenza di Poincaré): un sistema deterministico, in cui ogni variabile è contenuta in un intervallo limitato, dopo un certo tempo ritorna vicino alla sua condizione iniziale. Quindi l'analogo sicuramente esiste, c'è però un problema pratico: quanto indietro si deve andare per trovarlo? La risposta è un risultato ben noto della teoria matematica dell'ergodicità. La difficoltà di trovare un analogo dipende dalla dimensione "D" (in parole povere D è il numero minimo di variabili necessarie per descrivere il problema), per trovare un analogo con precisione percentuale "a" si deve andare indietro di un tempo ordine (1/a)D.

E facile convincersi che si può fare una previsione con l'idea degli analoghi, solo se la lunghezza della sequenza è di ordine almeno (1/a)D. Se D è grande (diciamo oltre 7−8) già per precisioni non enormi (ad esempio per a=0,05) in genere non si trova un analogo, basti notare che (1/a)10=2010 =1,024×1013. Da questo si capisce come la limitata lunghezza delle serie dei Big Data, per quanto grandi in situazioni non banali, non permette di usare per le previsioni un approccio puramente induttivo e senza teoria (Cecconi et al. , *American Journal of Physics 80*, 2012).

Nella realtà la situazione è più complicata di quella sopra descritta, infatti tipicamente non si conoscono nemmeno le "variabili giuste", e molto spesso non sappiamo neanche se il sistema evolve con regole deterministiche o stocastiche.

Beyond the Market !

Complex Networks
&
Big Data

A Smart System is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it.

Parafrasando l'aforisma di Dan Ariely sui Big Data

EPoSS – The European Technology Platform on Smart Systems Integration

Sistemi con un limitato numero di gradi di libertà, e leggi di evoluzione temporale note e limitate dimensioni spaziali portano sistemi di controllo efficienti, con algoritmi di ottimizzazione direttamente derivati dalla conoscenza del modello teorico.

Lo sviluppo delle reti di comunicazione (tecnologia wireless), la disponibilità di potenza di calcolo distribuita a basso costo e di tecniche software di tipo Cloud (che sono indipendenti dal Layer hardware) hanno permesso negli ultimi anni la realizzazione di sistemi per il monitoraggio e il controllo di sistemi aventi un numero elevato di gradi di libertà, utilizzando dati con semantiche diverse, di grande estensione spaziale e la cui evoluzione nello spazio e nel tempo è determinata da dinamiche solo parzialmente note.
Smart Systems generano Big Data

A seconda del campo di applicazione si parla di Smart City, Smart Building, Smart Transport. L'intelligenza di questi sistemi è limitata dalla capacità di estrazione delle caratteristiche della dinamica del sistema dall'analisi dei dati mediante tecniche di intelligenza artificiale. Questo limite fa sì che la maggior parte delle applicazioni si fermi alla funzione di monitoring o di alert, predominanti nella funzionalità dei sistemi, rispetto alle funzioni realmente intelligenti che non vengono affrontate.

Introduction

Smart Systems - defined as self-sufficient intelligent technical (sub-)systems with advanced functionality, enabled by underlying micro- nano- and bio-systems and components – present a crucial link in the innovation chain as they can provide heightened functionalities for upgraded and new industrial and consumer products. Because of their ability to enable and enhance products and services in every walk of life, Smart Systems are of strategic importance for the competitiveness of entire sectors and economies in Europe.

Smart Systems are able to sense, diagnose, describe, qualify and manage a given situation, their operation being further enhanced by their ability to mutually address, identify and work in consort with each other. They are able to interface, interact and communicate with users, their environment and with other Smart Systems.

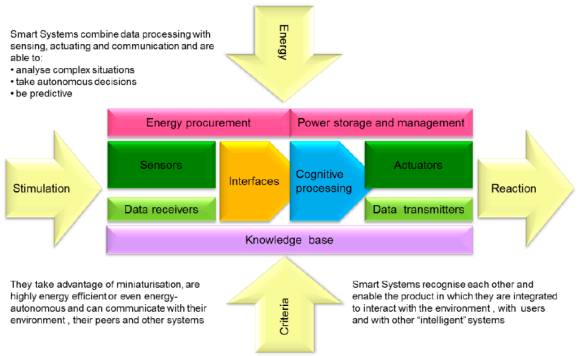


Figure 1: Structure and definition of a Smart System

Smart Systems make use of and integrate a multitude of Key Enabling Technologies (KETs), such as micro- and nanoelectronics, advanced materials, nanotechnologies, biotechnology, photonics and advanced manufacturing. That way they are mastering and widely introducing cross-cutting Key Enabling Technologies (Cross-KETs) such as Micro-Nano-Bio Systems (MNBS), bio-electronics, bio-photonics, Flexible, Organic and Large-Area Electronics (FOLAE), functionalised (nano) materials.

EPoSS is an industry-driven membership organisation, defining R&D and innovation needs as well as policy requirements related to Smart Systems Integration (SSI) and integrated Micro- and Nanosystems. EPoSS is contributing to EUROPE 2020, the EU's growth strategy for the coming decade, to become a smart, sustainable and inclusive economy.

Struttura, Controllo e Paradigmi Computazionali

ENEA
Agenzia Nazionale per le Nuove Tecnologie, l'Energia e lo Sviluppo Economico Sostenibile

Ministero dello Sviluppo Economico

RICERCA DI SISTEMA ELETTRICO

Modelli predittivi e soluzioni pilota per la diagnostica e controllo di smart buildings

M. Annunziato, M. Bosello, M. De Felice, C. Meloni, S. Pizzuti

SOMMARIO

Building Energy Management Systems

Attualmente non esistono dei veri e propri modellatori o sistemi in grado di effettuare una diagnostica avanzata in modo automatico.

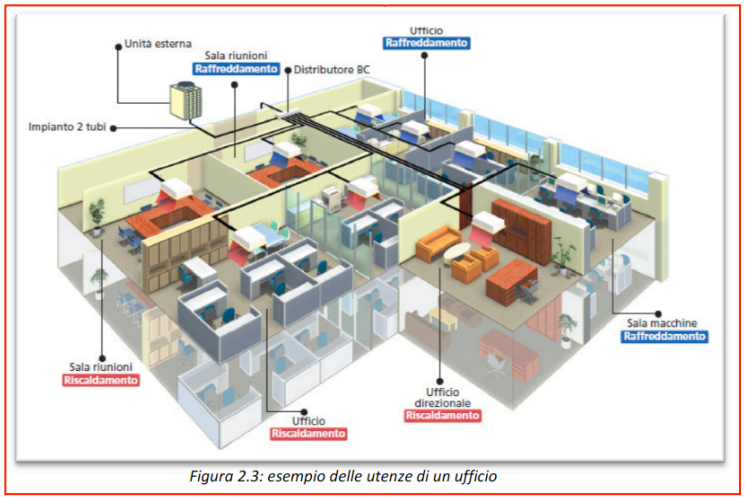


Figura 2.3: esempio delle utenze di un ufficio

Generalmente infatti ci si limita a delle semplici collezioni di dati mentre la diagnostica in senso stretto e` demandata ad operatori umani che analizzano i dati acquisiti.

Questo naturalmente causa un elevato costo operativo e la necessita` di un energy manager esperto il cui costo e` molto elevato.

Inoltre la competenza dell'energy manager e` tanto piu` elevata quanto piu` grande e` il parco di edifici che deve controllare (ad esempio esistono molte reti di edifici in Italia che superano i 3000 uffici) ma in questo caso e`praticamente impossibile visionare a fondo tutti i dati.

In definitiva l'intervento di diagnostica ed ottimizzazione si limita alla identificazione di cause macroscopiche senza disporre di un vero processo di elaborazione dei dati.

MetaDati

$$\left(y_1^{Type1}, y_2^{Type1}, y_3^{Type2}, \ldots, y_M^{TypeM_T}\right) \rightarrow M_D = \left\{y_i^k(t)\right\}_{M, M_T}$$

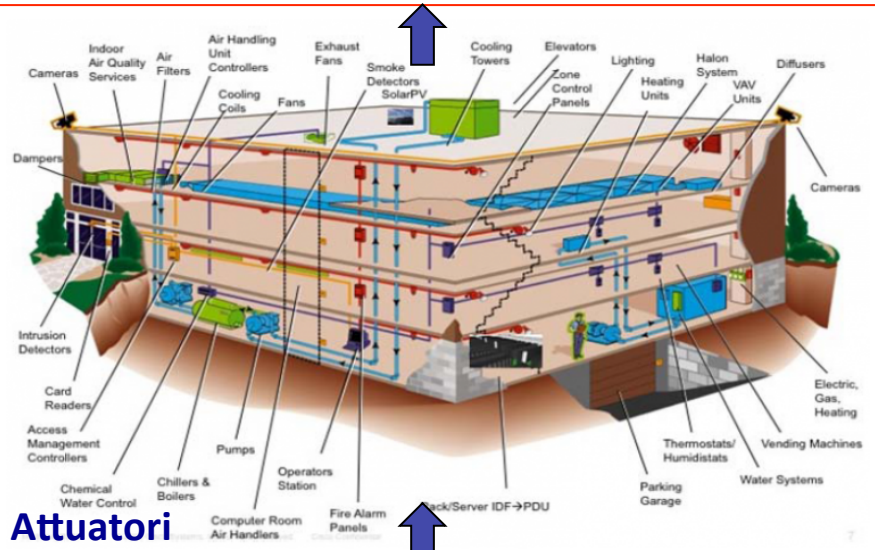
DataBase

$$Stor\left(\left\{y_i^k\right\}\right) = \left\{M_D = \left\{y_i^k\right\}_{M, M_T}\right\}$$

Sensori

$$\left(x_1^{Type1}, x_2^{Type1}, x_3^{Type2}, \ldots, x_N^{TypeN_T}\right) \rightarrow S = \left\{x_i^k(t)\right\}_{N, N_T}$$

$$Stor\left(\left\{x_i^k\right\}\right) = \left\{S = \left\{x_i^k\right\}_{N, N_T}\right\}$$



Monitor

$$Mon = \left\{M_\alpha\left(Stor\left\{x_i^k\right\}\right)\right\}$$

$$Mon = \left\{F_\beta\left(M_\alpha\left(Stor\left\{x_i^k\right\}\right)\right)\right\}$$

$$x_i^k \in S_A, S_B, \ldots, S_Z \subseteq S$$

Stati Locali o Globali

Indicazioni di modifiche strutturali e/o funzionali

Attuatori

$$\left(w_1^{Type1}, w_2^{Type1}, w_3^{Type2}, \ldots, w_{N_w}^{TypeN_{T_w}}\right) \rightarrow W = \left\{w_i^k\left(\left\{\lambda_k(t)\right\}, t\right)\right\}_{N_w, N_{T_w}}$$

Ottimizzazione

$$F_C^\alpha = F\left\{M_\alpha, \left\{\lambda_k(t)\right\}, \left\{x_i^k\right\}, t\right\}$$

$$Min = \frac{\partial F_C^\alpha}{\partial\left\{\lambda_k(t)\right\}} = 0 \rightarrow \left\{\lambda_k^*(t)\right\}$$

Minimizzazione globale di Funzioni Costo

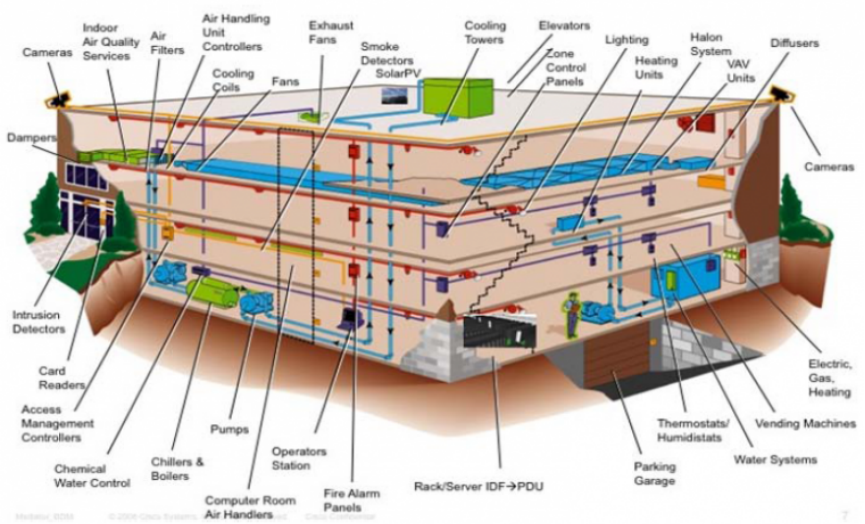
Ottimizzazione funzionale in Real Time.
Eventualmente supervisionata→ Sistemi Adattivi.

Modello Computazionale del Sistema

$$ModComp = \left\{S, M_D, Stor, Mon, Min, W\right\}$$

Struttura, Controllo e Paradigmi Computazionali

$$ModComp = \{S, M_D, Stor, Mon, Min, W\}$$



L'implementazione fisica del modello di calcolo, in termini di architettura, tecnologie e topologia della rete di comunicazione e dei livelli di gerarchia per gli elementi di processo diventa critica

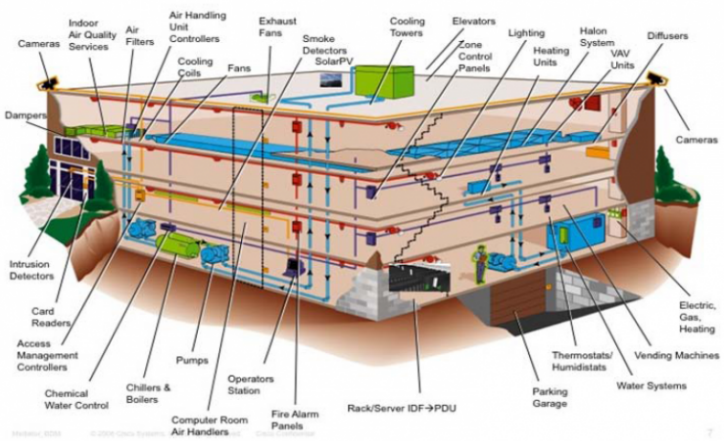
Nella situazione in cui:

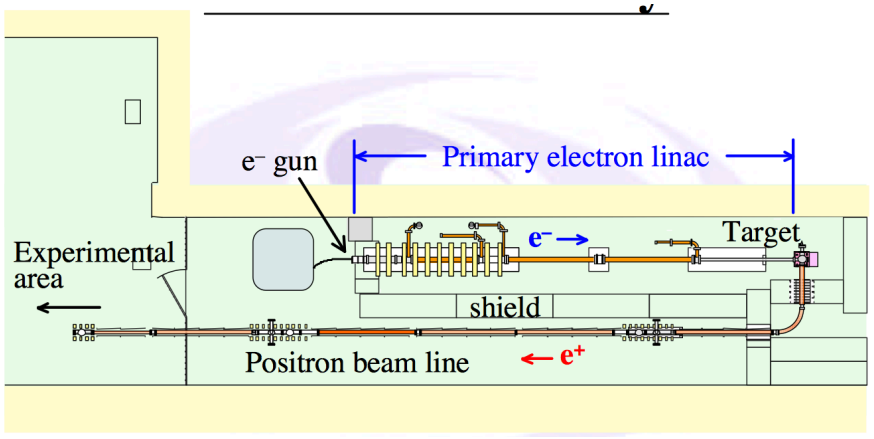
- Le sotto parti del sistema siano correlate (in maniera debole o forte e la distanza di correlazione e' confrontabile con le dimensioni del sistema stesso
- Il numero di sensori e attuatori diventa arbitrariamente elevato

Il monitoraggio ed l'ottimizzazione diventano task computazionali hard.

$$ModComp \xrightarrow[\xi >> 1]{N \to \infty} HARD$$

Use Cases Complexity: |*ModComp*|





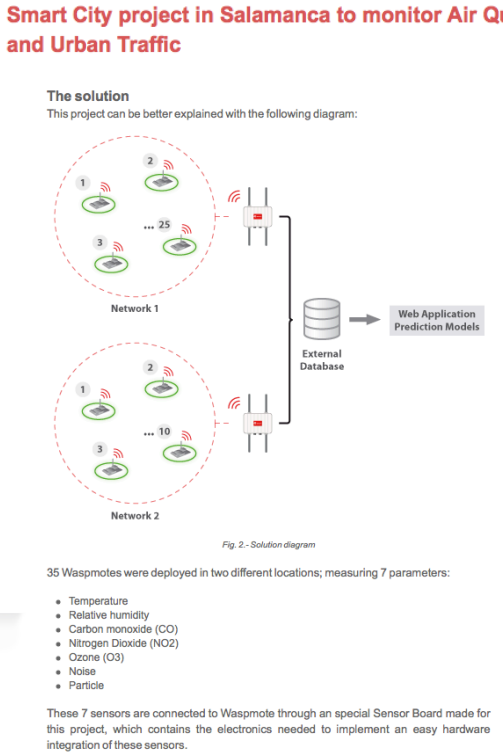
- Maggior numero di sensori ed attuatori e di natura diversa.
- Sistema difficilmente fattorizzabile in sotto sistemi non interagenti
- Funzioni costo complesse e algoritmi di ottimizzazione non noti e/o non efficaci. Possibile necessita' di modelli adattivi supervisionati iterativi.
- BD non sequenziali ?

- Numero di sensori ed attuatori limitato e omogeneo.
- Sistema facilmente fattorizzabile in sotto sistemi non interagenti
- Funzioni costo complesse e algoritmi di ottimizzazione noti e efficaci
- BD sequenziali !

$$|ModComp|_{SB} >> |ModComp|_{FP}$$

Piattaforme Hardware

- Mica Mote (Atmel,radio 32khz, Berkeley Univ.)
- EYES (MP430,TDA520, EU project)
- Btnodes (Atmel,Bluetooth e Chiocon 915MHz, ETH Zurigo)
- Scatterweb (MP430,Radio e Bluetooth,Univ. Berlino)
- FireFly(Atmel)
- TelosB(MSP430)
- Beagle Bone (Open hardware Open Software)
- Libelium (Open software)
- NI MyDaq
- ……..



Libelium (Open software)

Scegliere dipende dall'applicazione che si vuole creare, dal contesto in cui verrà adoperato, dalle dimensioni massime ammissibili, dal costo sostenibile, dalla loro efficienza energetica e dalla robustezza richiesta.

Non conviene sviluppare HW
Non conviene sviluppare SW

Definizione del Modello Computazionale per Smart System:

Layers Computazionale e Logico

$$ModComp = \{S, M_D, Stor, Mon, Min, W\}$$

Il Layer Computazionale:

Si individuano le osservabili fisiche in grado di descrivere e definire lo stato del sistema e si modellizza la loro interazione. In base alla funzionalità del sistema di controllo si determinano le funzioni per il monitoraggio e il *warning*. Si deve definire l'insieme dei *Meta-Data* che determinano e caratterizzano il sistema. La retroreazione sul sistema è effettuata mediante l'ottimizzazione di funzioni di costo rispetto a dei vincoli funzionali, ed è vista come l'introduzione di termini forzanti nell'evoluzione del sistema.

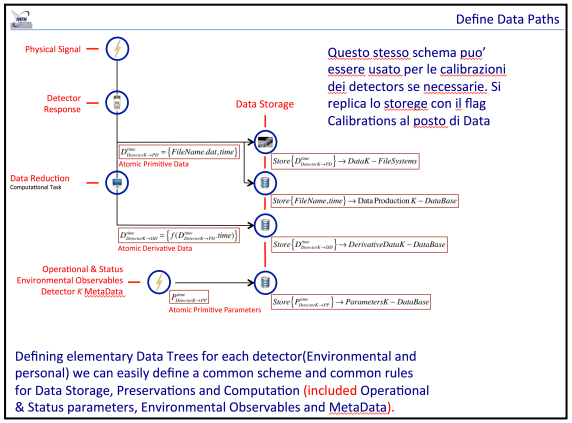
A seconda del livello di complessità, per il processo di ottimizzazione si possono contemplare algoritmi predittivi e deterministici, per le procedure più semplici, o algoritmi adattivi basati su tecniche di AI (*Neural Network*, Algoritmi Genetici o di *Deep Learning*), per le più complesse.

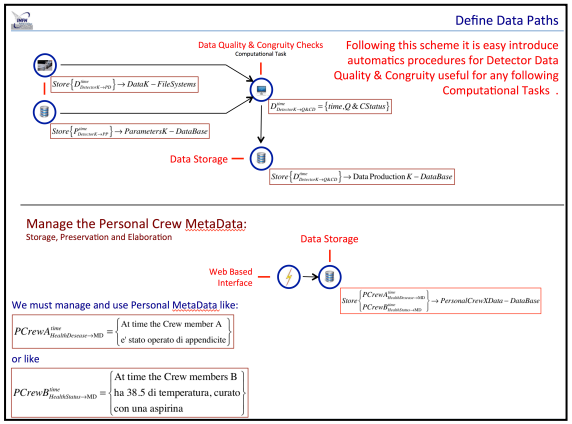
→ Agricoltori carcasi

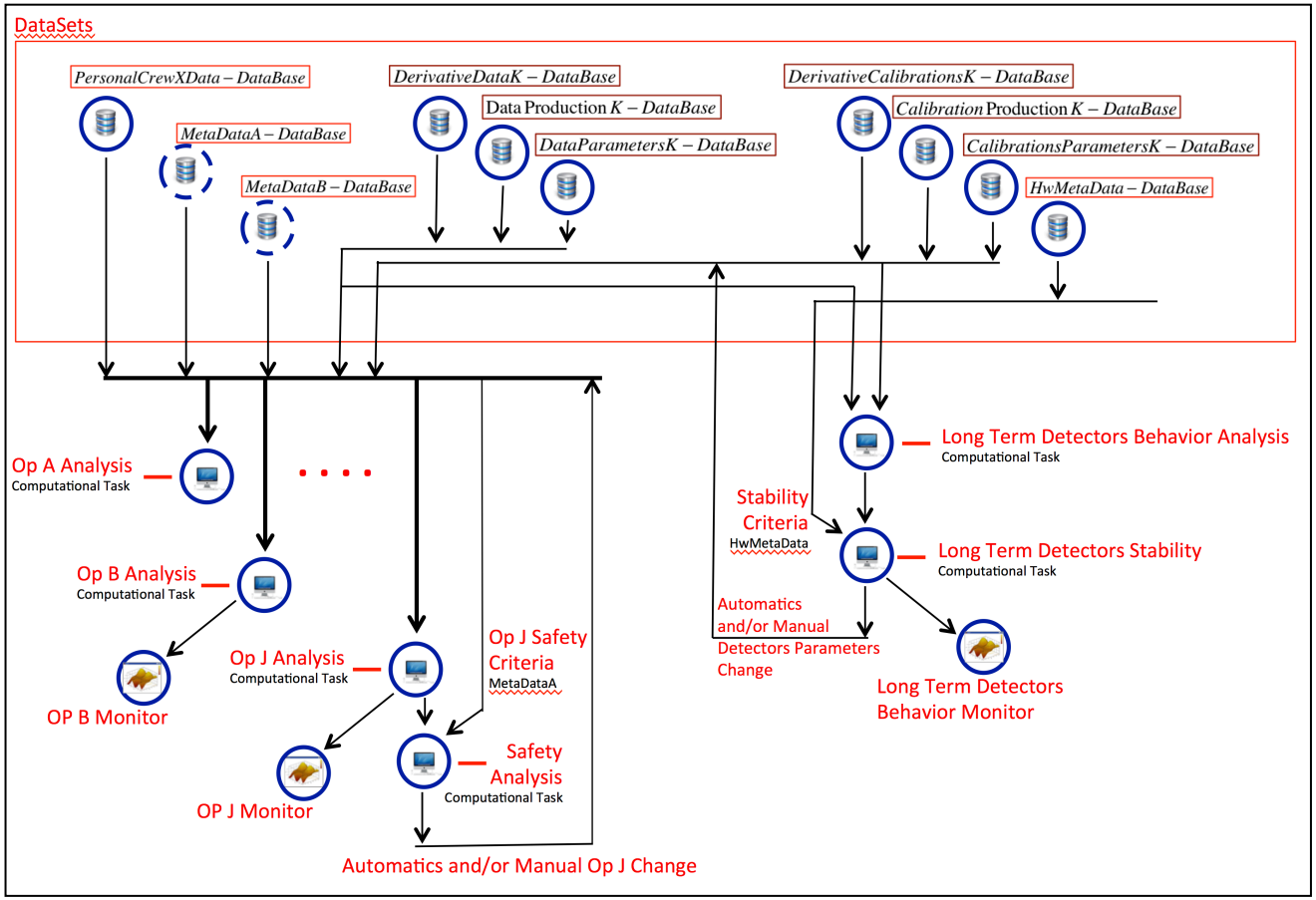
Il Layer Logico:

Per ogni osservabile fisica viene definito uno specifico *data path* definendo i compiti computazionali e le operazioni di immagazzinamento in memoria da eseguire sul data-set che la rappresenta. A questo livello devono essere adottate strategie di *Data Control* e *Data Quality* al fine di garantire consistenza e qualità dei dati stessi. I *Data-Set* delle osservabili fisiche sono, per loro natura, caratterizzati da una sintassi e una semantica comune che devono essere rese omogenee alla Sintassi e Semantica dei *Meta-Data*. È infine necessario definire strategie di *Data Preserving* sull'intero *Data-Set*.

→ Tecniche di Big Data







$$ModComp = \{S, M_D, Stor, Mon, Min, W\}$$

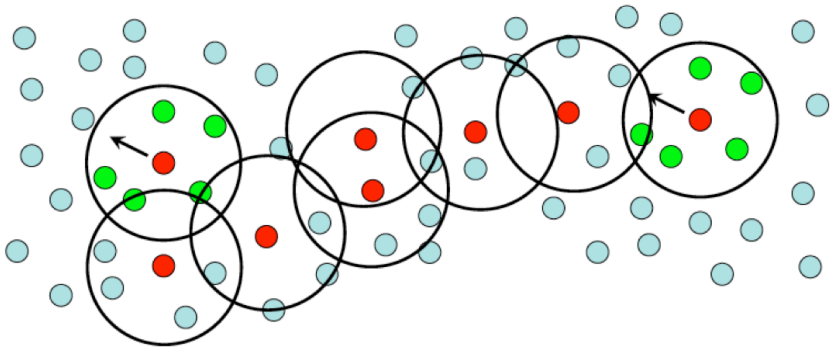
Definizione del Modello Computazionale per Smart System:

Layer Fisico

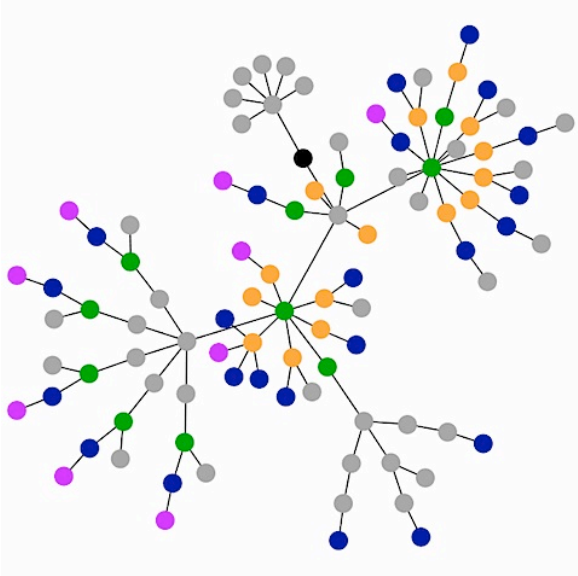
Il Layer Fisico:

Viene definita l'architettura hardware del sistema. Tale architettura deve ottimizzare i vari task computazionali sfruttando il loro naturale parallelismo in modo da ridurre i tempi necessari per l'elaborazione dei dati ed essere in grado di retroazioni sul sistema in *Real Time*. L'architettura software viene definita sfruttando le attuali tecniche di *Cloud Service* e *Cloud Computing* utilizzando dei *Data-Base* per immagazzinare i dati. Tali tecniche disaccoppiano l'architettura hardware da quella software e permettono un'evoluzione/aggiornamento del layer hardware senza una ridefinizione del layer software.

L'architettura hardware/software deve essere modulare in modo da permettere la gestione di nuovi elementi senza la necessità di una sua ridefinizione.



Wireless Sensor Network



Wired Hierarchical Network