

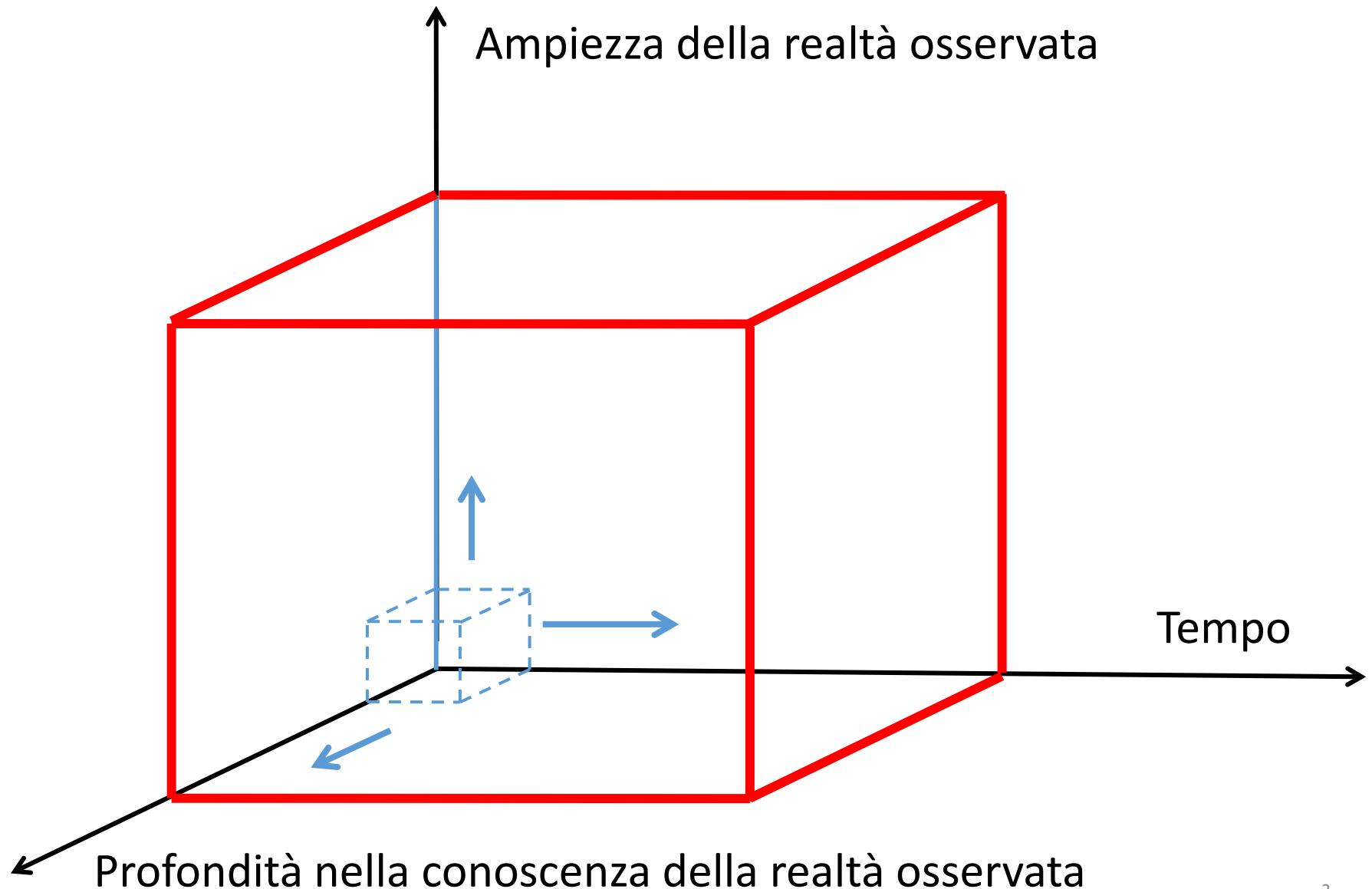
Laurea Magistrale in



DATASCIENCE

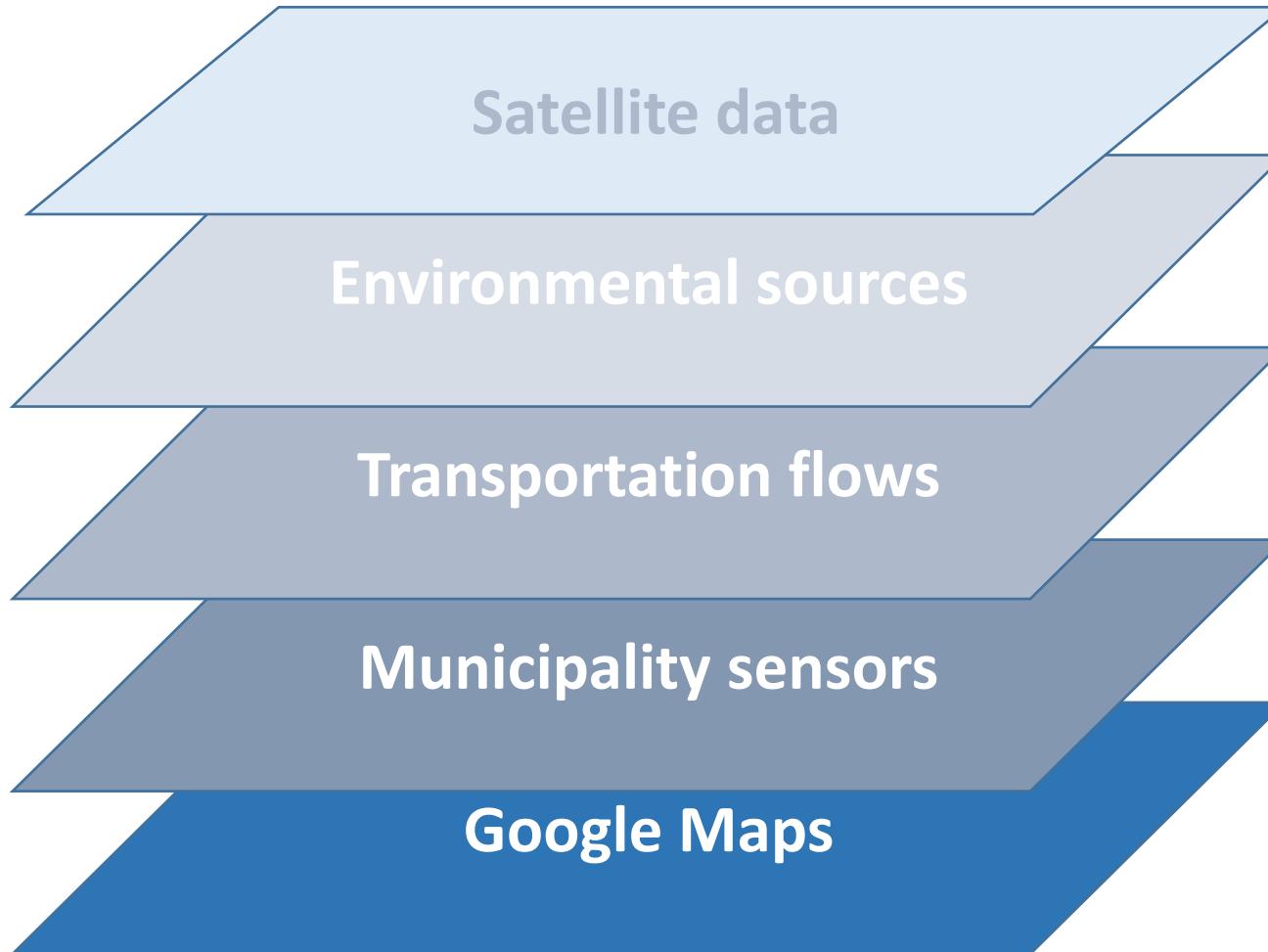
Prof. Carlo Batini
carlo.batini@unimib.it

Dagli small data ai big data

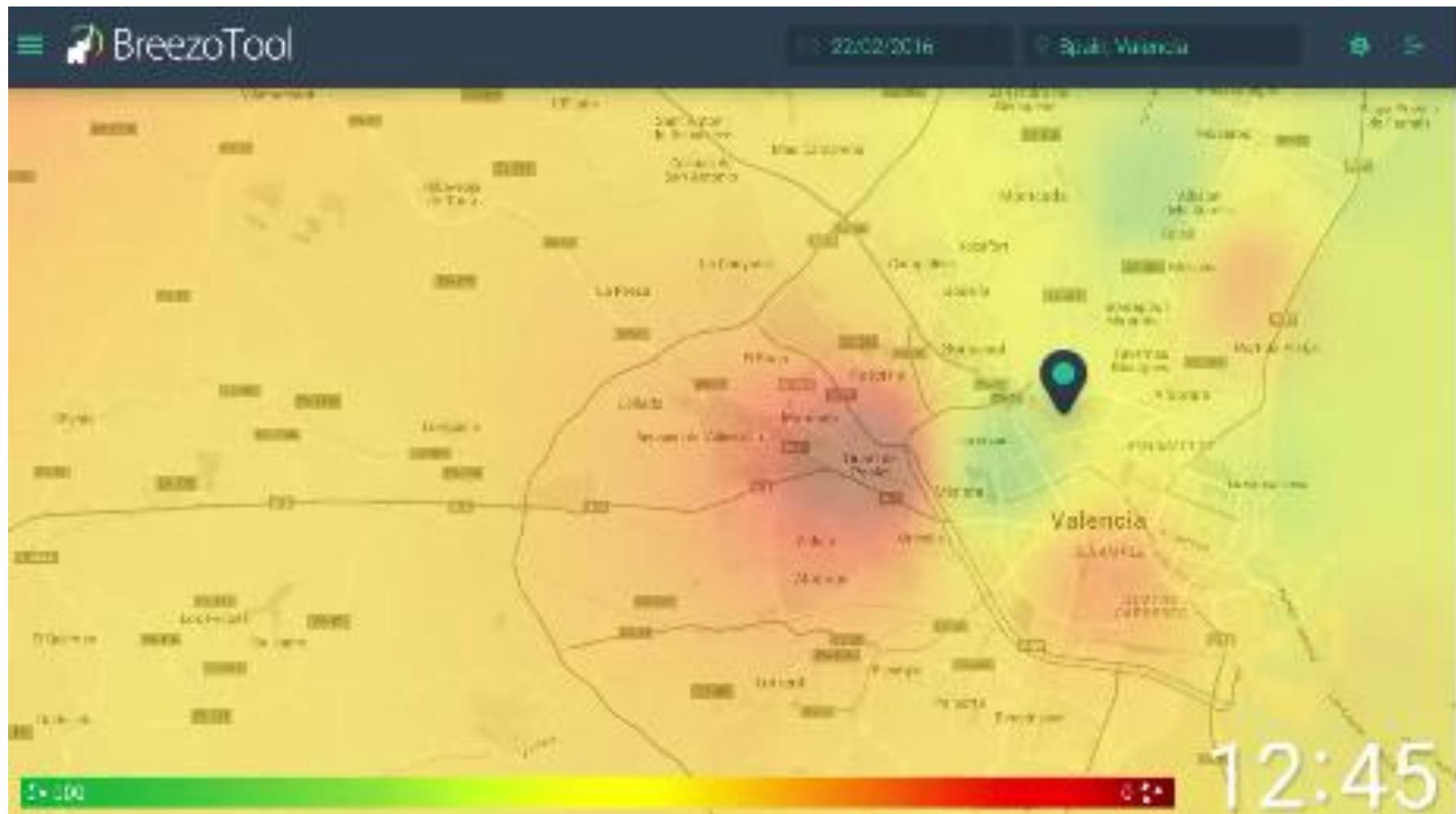


Innovazione abilitata dai big data

Breezometer: prevedere il livello di inquinamento



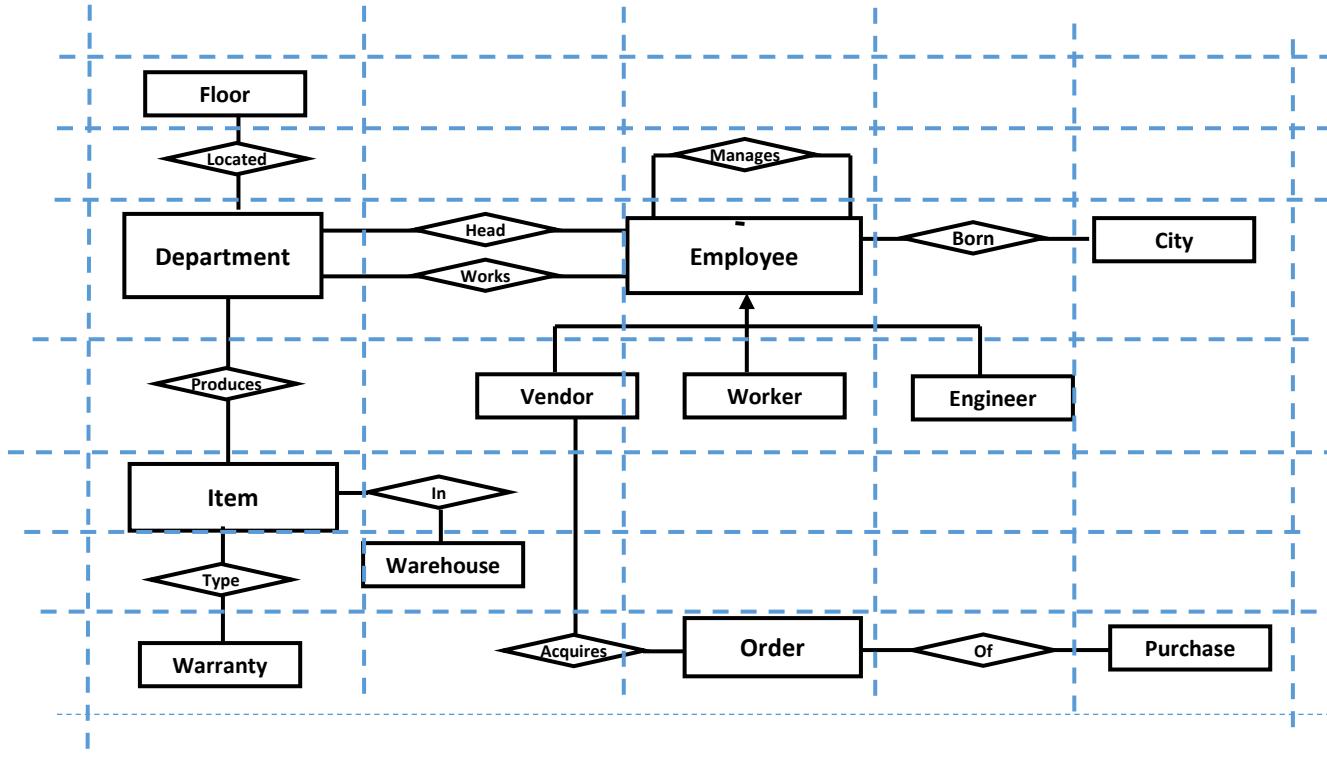
12.45



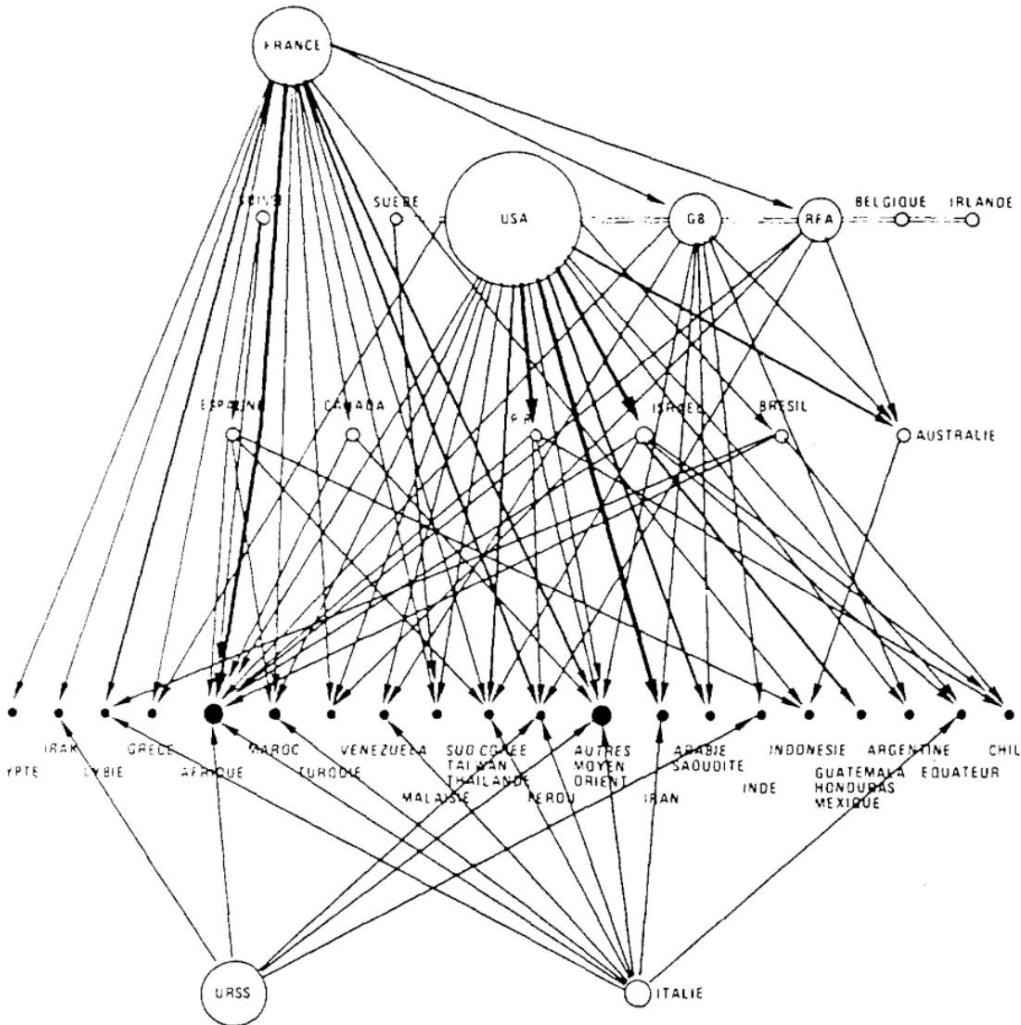
Complessità che va governata

Disegnare i grafi

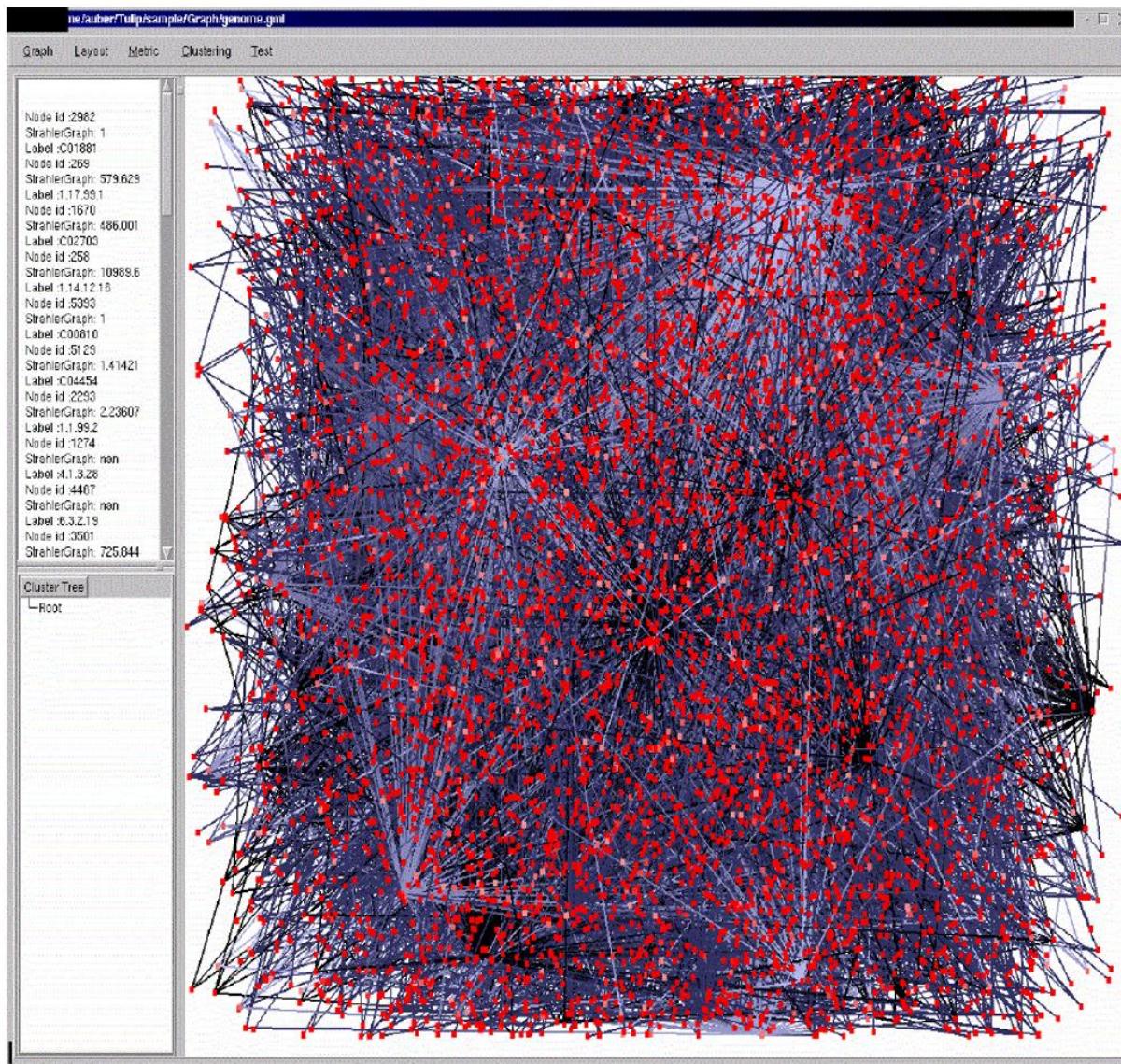
Grafo di 22 nodi + 23 archi



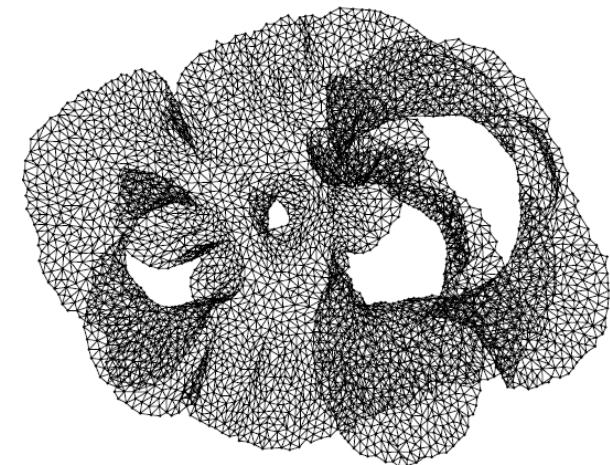
35 nodi +
80 archi



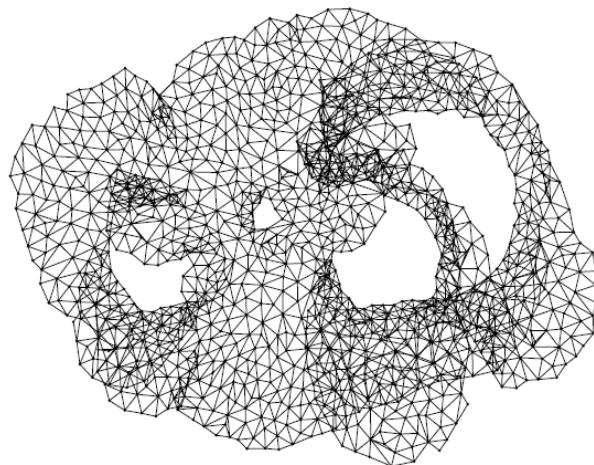
Raggiunto il limite – 6.000 nodi e 12.000 archi



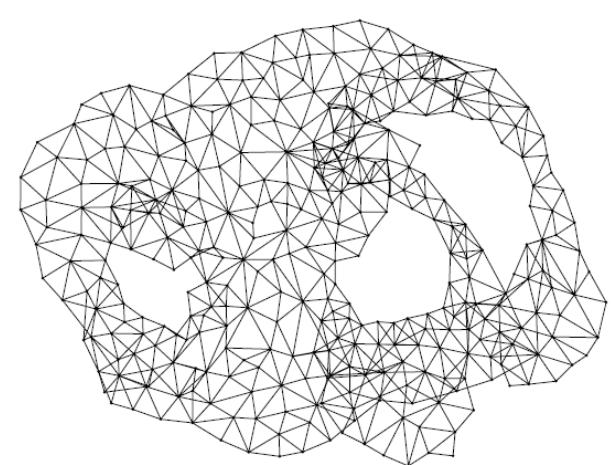
Possiamo costruire versioni sintetiche del grafo
a tre differenti scale di dimensione
e accuratezza



4394-node approximation



1223-node approximation



341-node approximation

Volume, Velocità, Varietà, Valore, Veridicità

Volume – la quantità di dati che possono essere acquisiti, elaborati e memorizzati

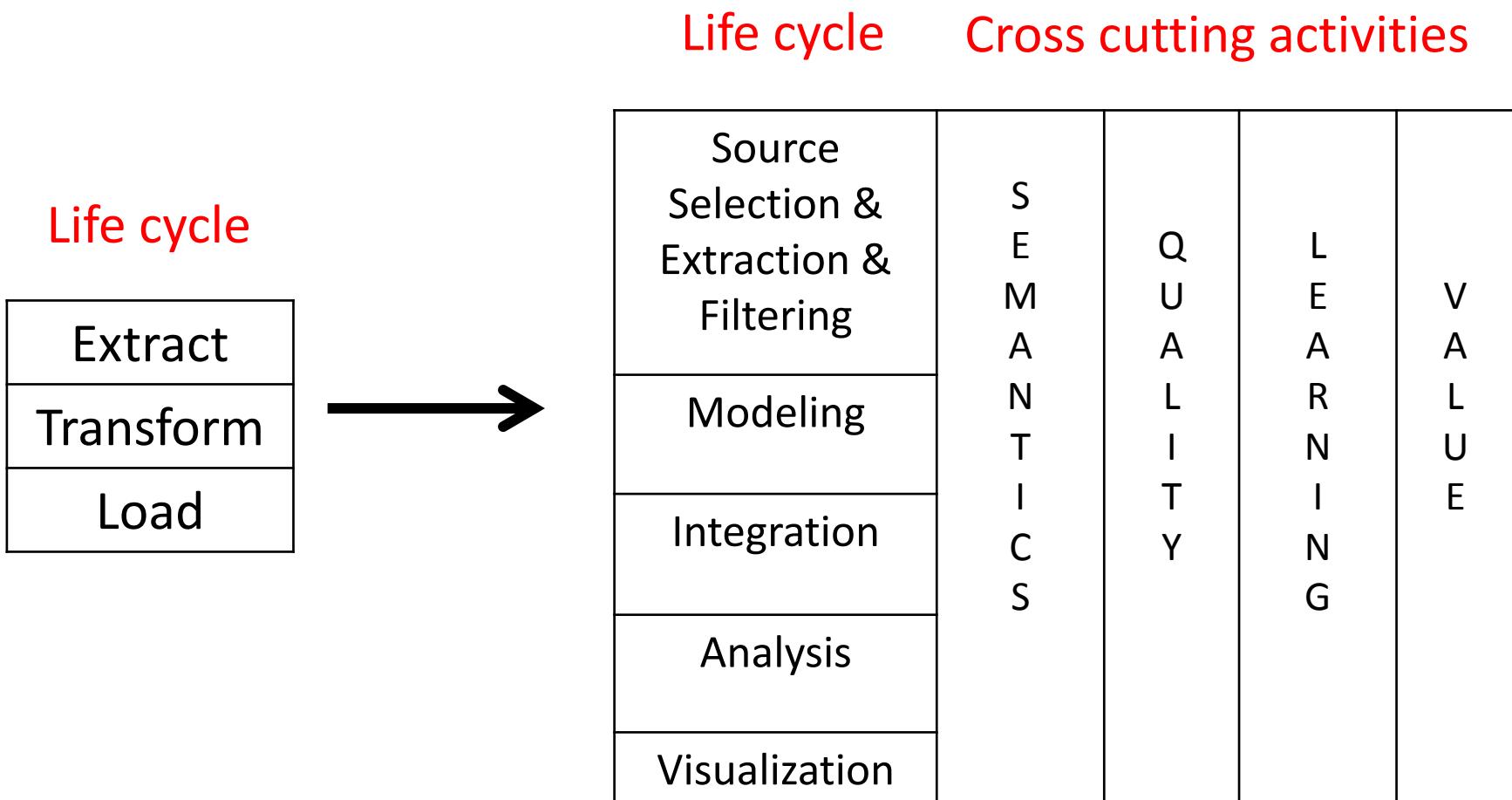
Velocità – la velocità a cui i dati possono essere acquisiti

Varietà – da dati strutturati (tavole relazionali) a dati debolmente strutturati (es. testi, immagini)

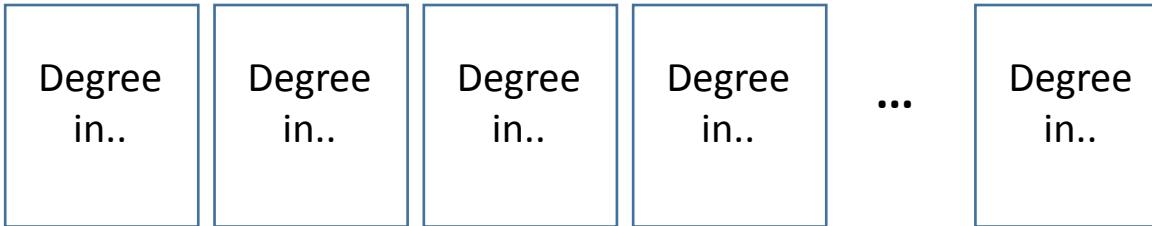
Valore – economico o sociale dei dati

Veridicità – Accuratezza, completezza e livello di aggiornamento dei dati

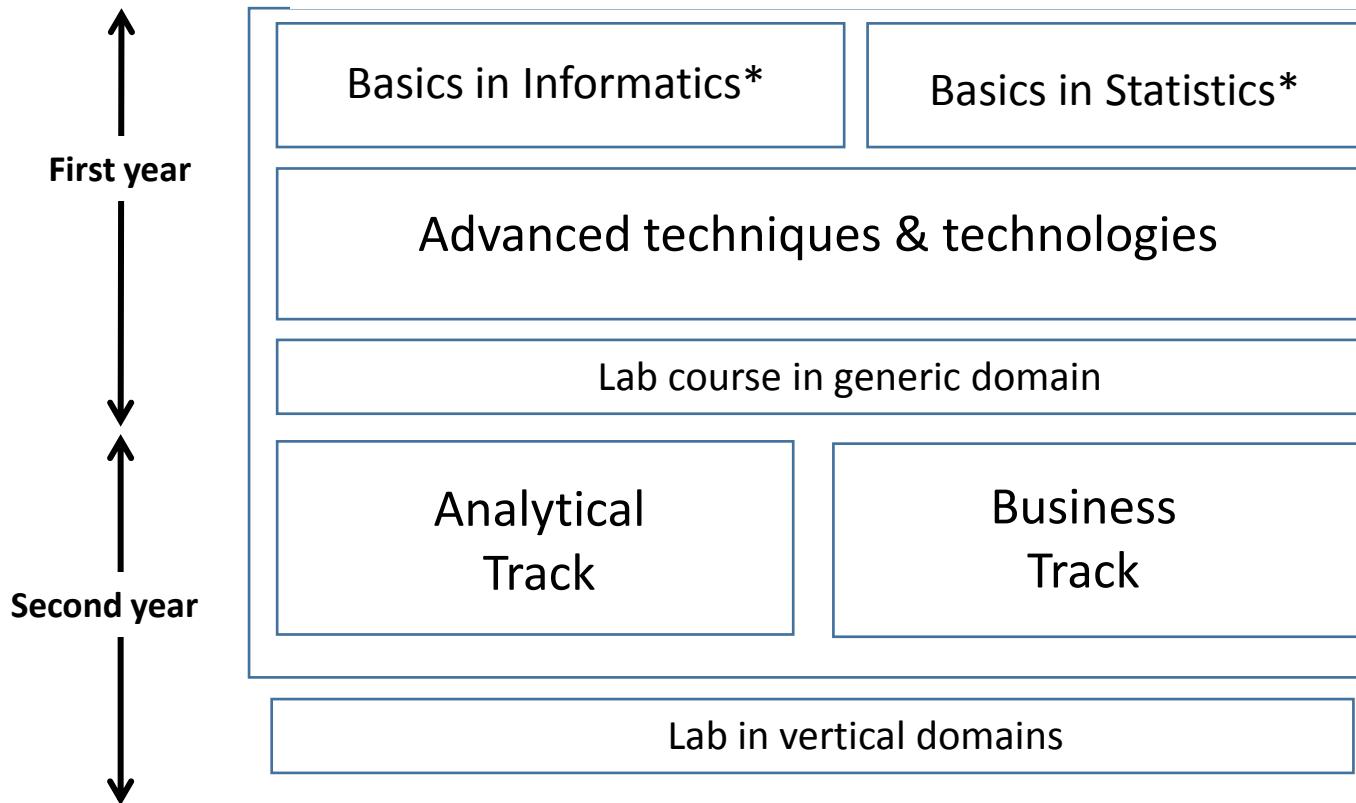
Come cambia il ciclo di vita del dato



Course organization and required skills



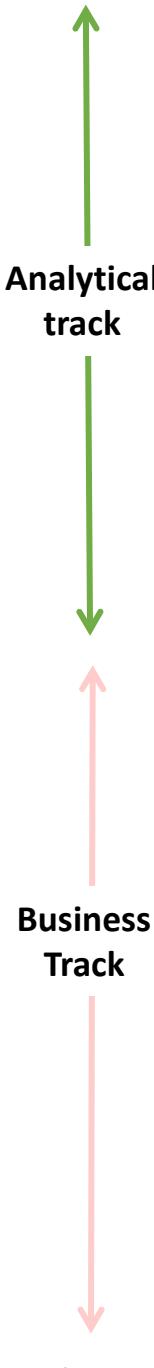
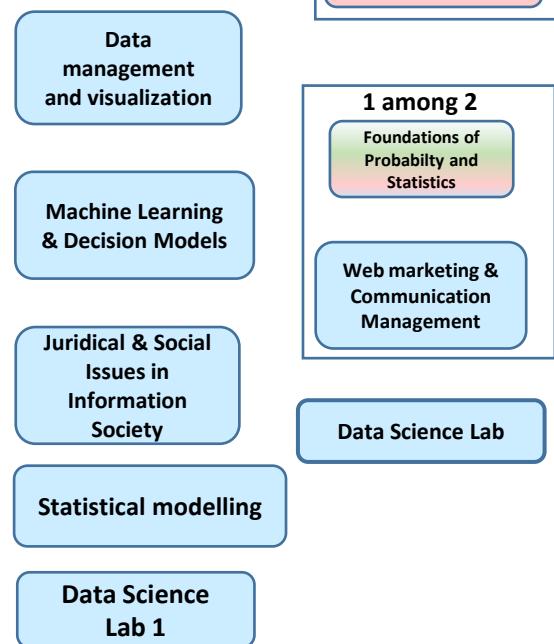
**At least 30 credits in informatics and/or
statistics and/or mathematics and/or physics
B2 English level certificate**



* If needed

Piano didattico

Programmi
dei Corsi e
Docenti sul
sito del
corso



Primo anno

Secondo anno

E in più..

- Stage obbligatorio in azienda o presso i laboratori di ricerca dell'ateneo
- Doppia laurea con Universidad de Barcelona
- Accordi Erasmus con Manchester e Stoccolma
- Ogni anno una Expert Week
- Progetto di coinvolgimento aziende con Assolombarda
- Corsi Mooc ed eLearning per non frequentanti

Per approfondimenti

Sito del Corso di Laurea Magistrale
datascience.disco.unimib.it

e-mail a orientamento.datascience@disco.unimib.it

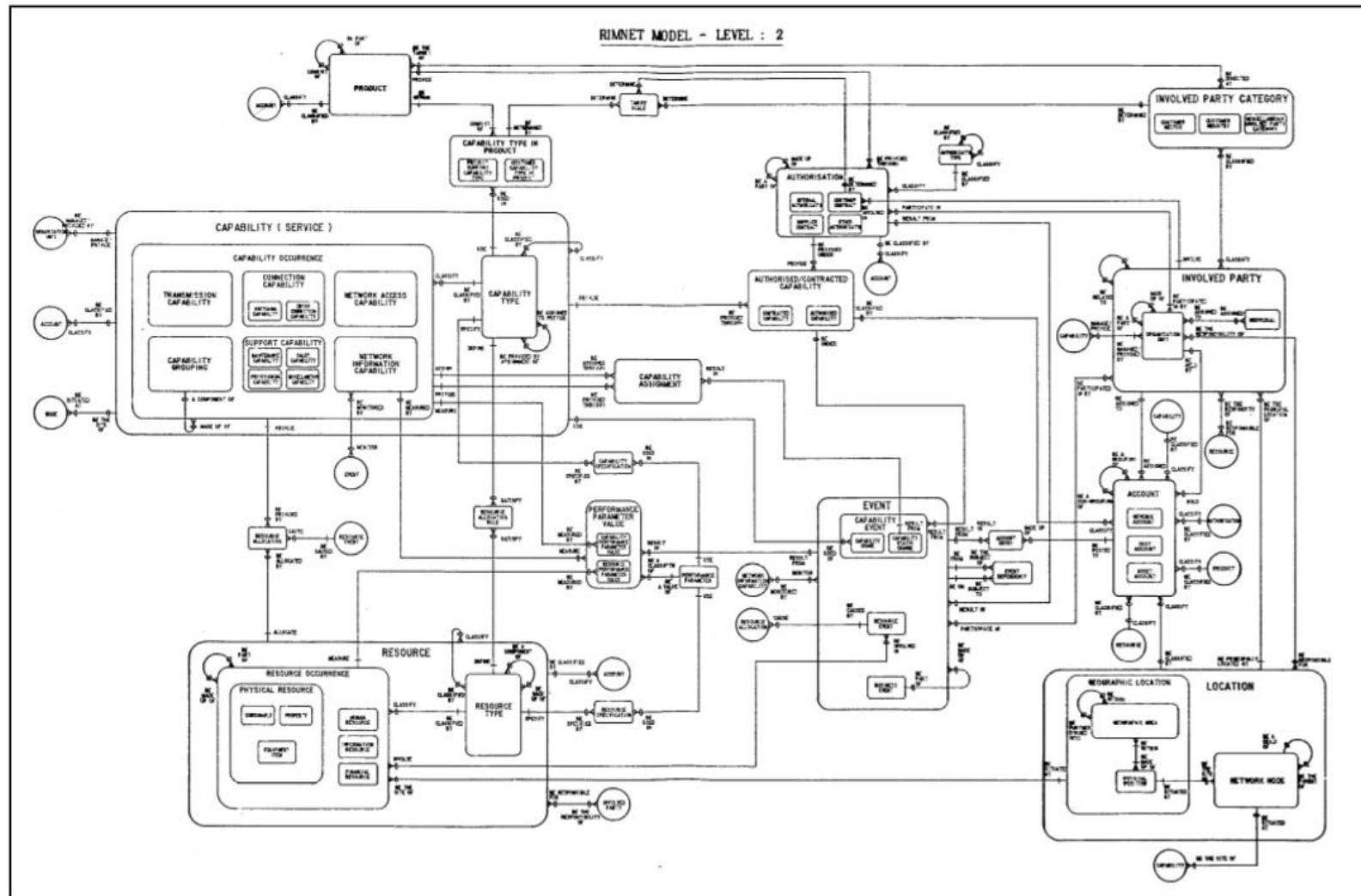
Facebook: @dataScienceBicocca
Twitter: @dsUniMib

Fine

Google trends per «Data Science»



Big schemas, big maps, big graphs



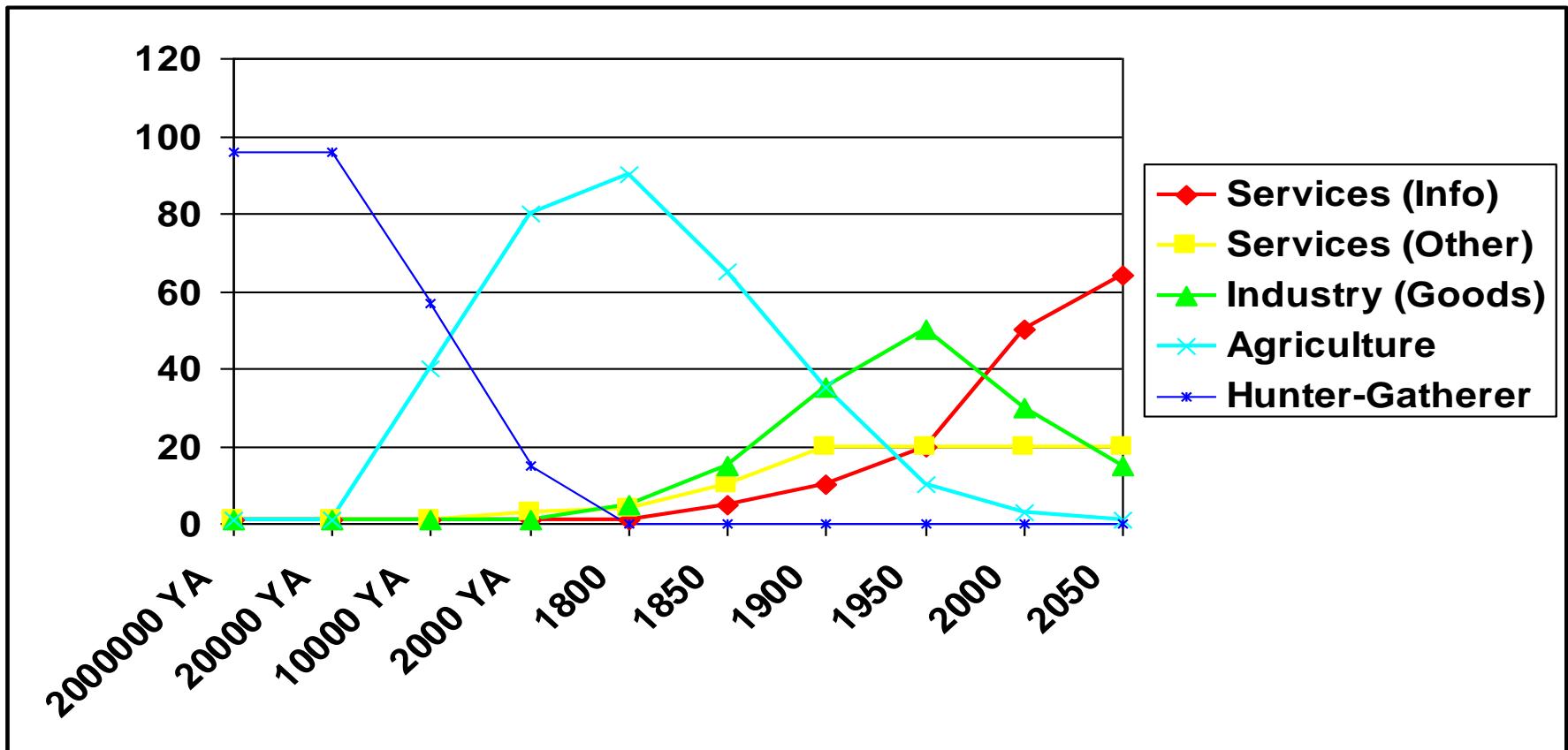
We are at the dawn of a new revolution
in the Information Age: data

Every animate and inanimate object on
Earth will soon be generating data”
(Smolan and Erwitt, 2013)

Relevance of Digital Data in Research, Economy, eGovernment and Society

- Data → Big Data: change of paradigm
- Fast growth of jobs in «informative services»,
services based on digital data (30% of overall jobs
in 2020, 65% in 2050)
- Digital economy
- Social and ethical issues in Data Science

Estimated world (pre-1800) and then U.S. Labor Percentages by Sector

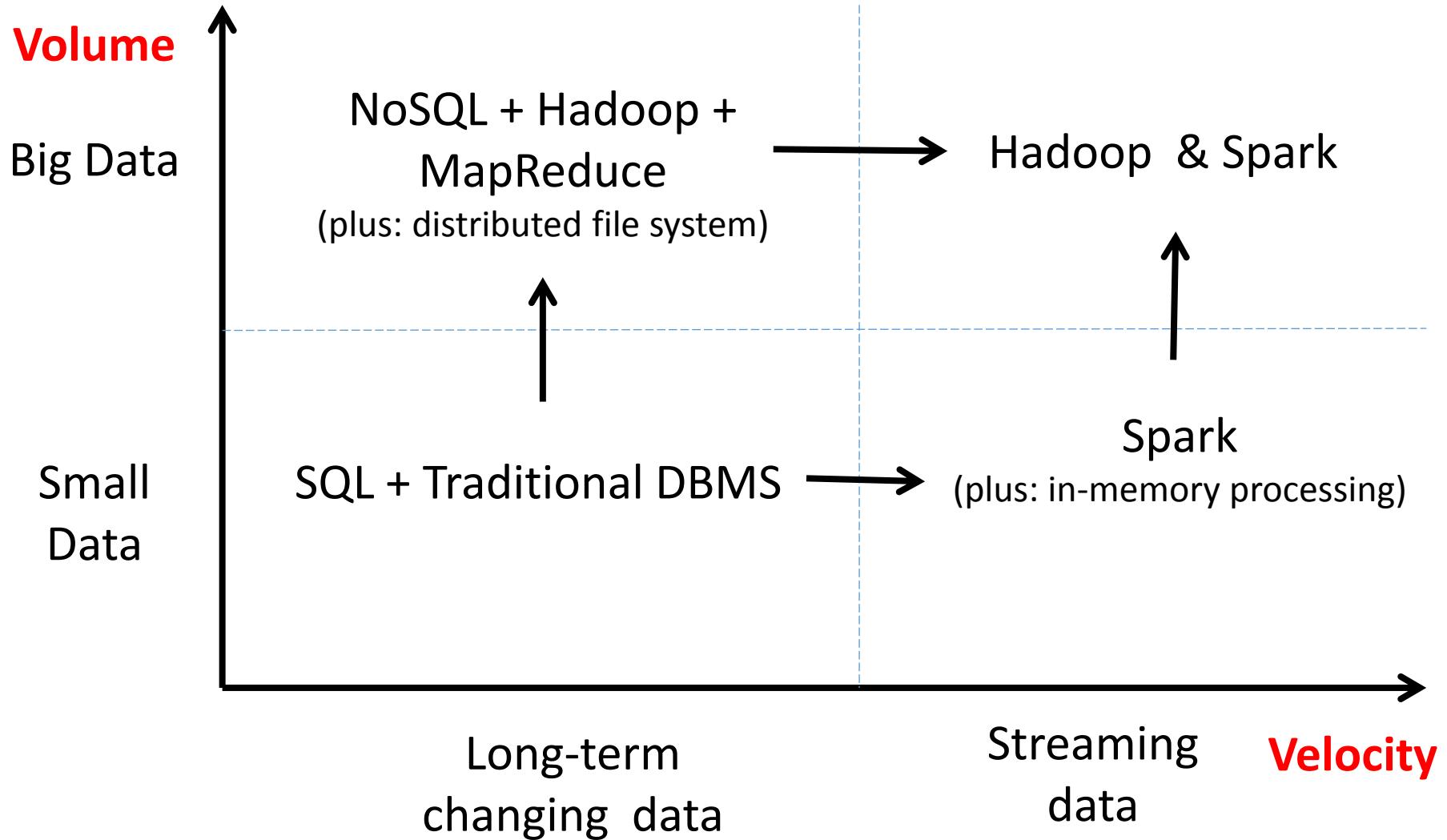


Estimations based on Porat, M. (1977) Info Economy:
Definitions and Measurement

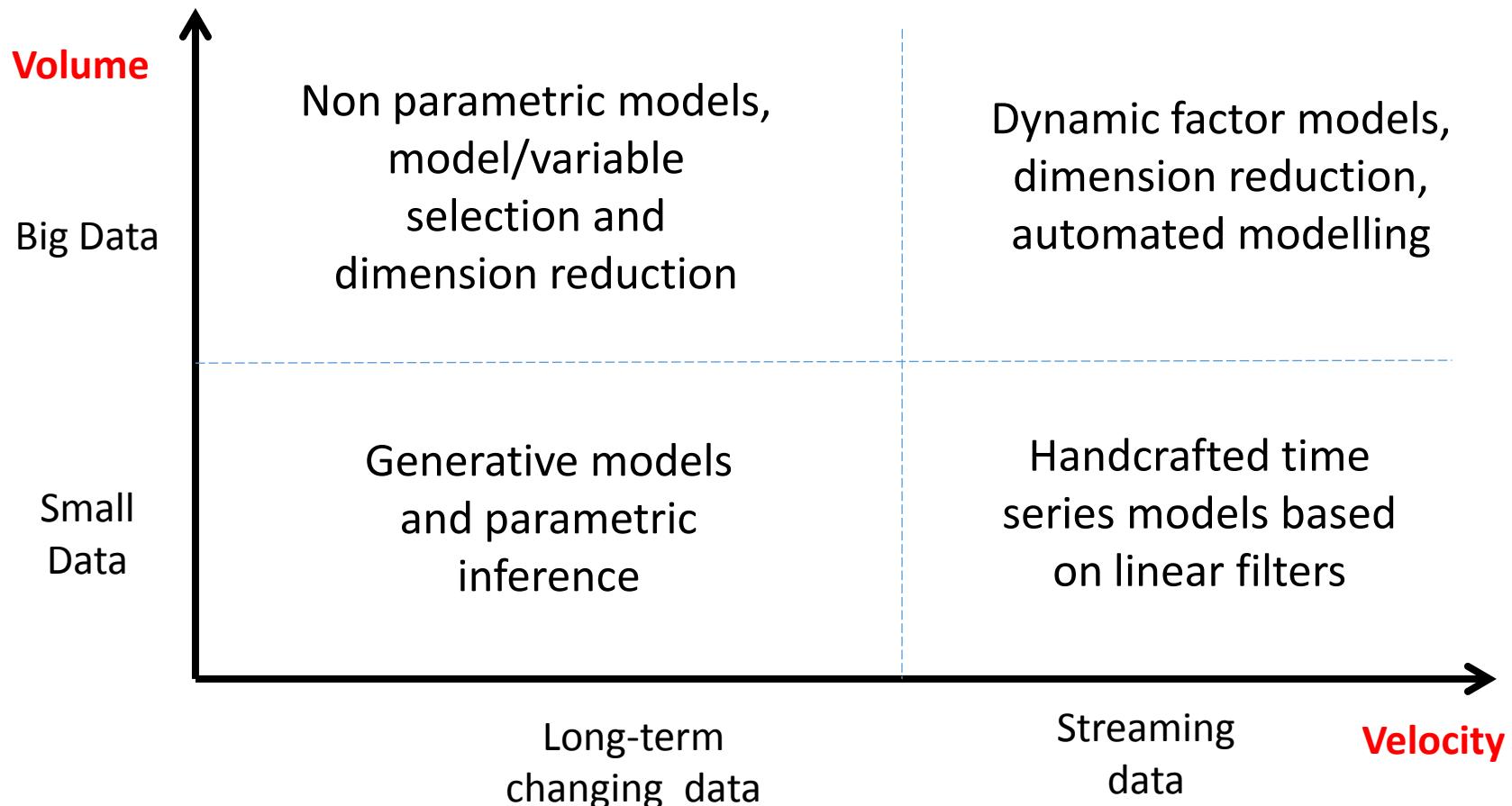
Four V's of Big Data

- Volume
- Velocity
- Variety
- Value

Change of Paradigm...in Data Management Systems



Change of Paradigm... in Machine Learning Techniques



Relevance of Data Science at University of Milano Bicocca

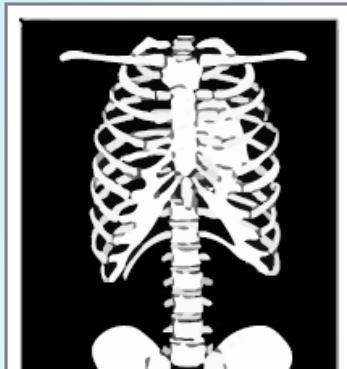
Many disciplines that make intensive usage of data and apply data management & analysis techniques and technologies

- Informatics and Statistics
- Sciences
 - Life sciences
 - Physics
 - Environment
 - Materials
 - Social
- Economics
 - Macro and Micro Economics
 - Finance
 - Marketing

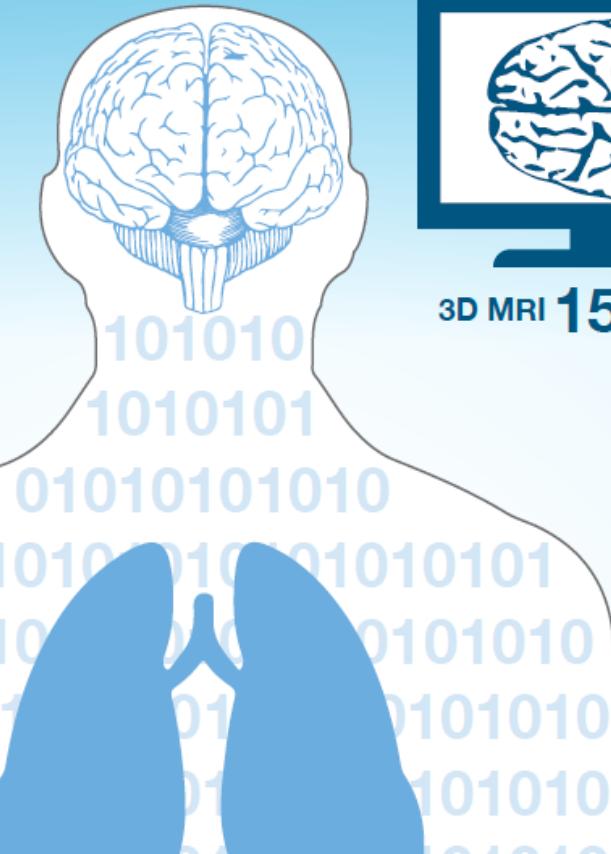
The Power of Healthcare Data

The Body as a Source of Big Data

Today data storage is essential for healthcare providers to see a patient's complete story of care, make the most informed decisions and enhance treatment and outcomes.



Access to electronic patient data beyond the desktop



3D MRI **150MB**



MAMMOGRAMS
120MB

What do Data Scientists do?

What do data scientist do?

Make discoveries while swimming in data. It's their preferred method of navigating the world around them.

At ease in the digital realm, they are able to **bring structure** to large quantities of formless data and **make analysis** possible.

They **identify rich data sources, join them** with other, **potentially incomplete** data sources, and **clean** the resulting set.

In a competitive landscape where challenges keep changing and data never stop flowing, data scientists **help decision makers** shift from ad hoc analysis to an ongoing conversation with data.

Professional profiles

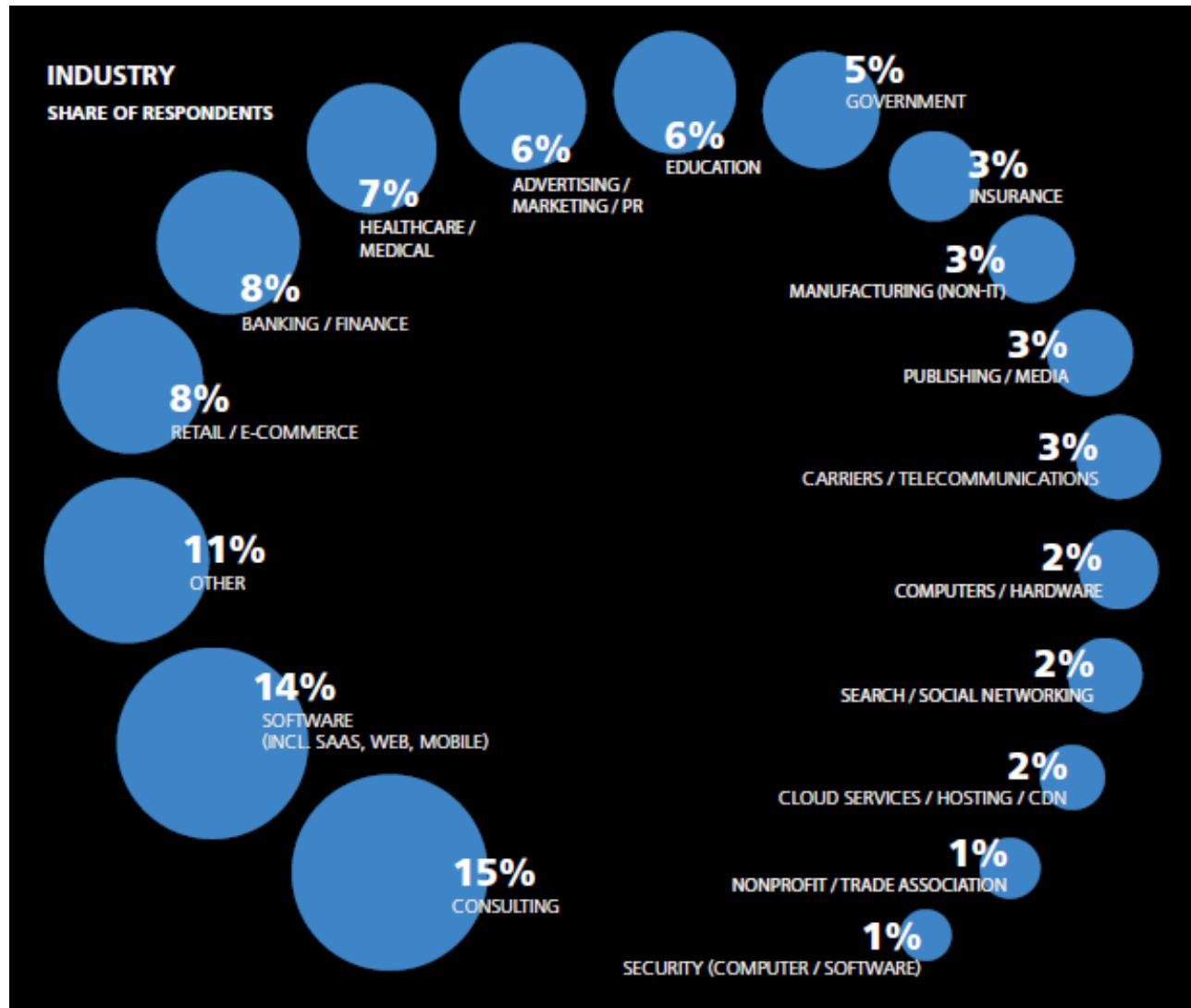
- Technological Data Scientist:
 - **Applies, adapts and extends statistical techniques** and computer science technologies providing effective analyses for decision, operational or research problems.
 - **Performs high level architectural design of services** based on digital data.
- Business Data Scientist:
 - **Finds solutions** based on statistical techniques and computer science technologies **to enhance value of decisions** and value of business processes in companies and public administrations
 - **Conceives new services based on digital data**, which optimize value in use for customers and value in exchange for service providers.

Data Scientist in the World

Source: Glassdoor 25 Best Jobs in America

1		Data Scientist	Job Openings 1,736	Median Base Salary \$116,840
		Career Opportunity	4.1	
		Job Score	4.7	
2		Tax Manager	Job Openings 1,574	Median Base Salary \$108,000
		Career Opportunity	3.9	
		Job Score	4.7	
3		Solutions Architect	Job Openings 2,906	Median Base Salary \$119,500
		Career Opportunity	3.5	
		Job Score	4.6	
4		Engagement Manager	Job Openings 1,356	Median Base Salary \$125,000
		Career Opportunity	3.8	
		Job Score	4.6	
5		Mobile Developer	Job Openings 2,251	Median Base Salary \$90,000
		Career Opportunity	3.8	
		Job Score	4.6	

Industry...



Source: O'Reilly Data Science Survey 2016

Cosa è accaduto
e cosa sta accadendo
sempre più velocemente

Rapida evoluzione tecnologica: es. evoluzione dei Samsung Galaxy Sensors

	Galaxy S	Galaxy SII	Galaxy SIII	Galaxy S4	Galaxy S5	Galaxy S6
Accelerometer	+	+	+	+	+	+
Light Meter	+	+	+	+	+	+
GPS	+	+	+	+	+	+
Magnetometer (Compass)	+	+	+	+	+	+
Microphone	+	+	+	+	+	+
Proximity	+	+	+	+	+	+
Battery Temp	+	+	+	+	+	+
Touchscreen	+	+	+	+	+	+
Camera	+	+	+	+	+	+
Cellular Radio	+	+	+	+	+	+
Wifi Radio	+	+	+	+	+	+
Bluetooth	+	+	+	+	+	+
Gyroscope	+	+	+	+	+	+
NFC		+	+	+	+	+
Barometer		+	+	+	+	+
Pedometer			+	+	+	+
Thermometer				+	-	-
Humidity				+	-	-
Gesture				+	+	+
Color Meter					+	+
Heart Rate					+	+
Fingerprint					+	+
Oxygen Saturation						+
Magnetic Secure Transmission						+

Secondo esempio: capire quando
conviene comprare un biglietto aereo

Comprare il biglietto che costa meno in un dato giorno? Basta consultare ad esempio eDreams..

Da Milano → A Melbourne Partenza Mer 14 Giu №. di passeggeri 1 adulto Cambiare ricerca ▾

Orario di partenza

Partenza Mer, 06:25 Mer, 22:20

L'opzione migliore!

664,11€ * Sola andata per passeggero Dettaglio ▾ Cancellazione gratuita entro 7 ore

10:40 17:50 (+ 1)
Malpensa, Milano Melbourne Airport, Melbo...
23h 10' → 1 scalo
7 posti disponibili!
Check-in bagaglio incluso

Selezione

Tutti i risultati della ricerca a partire da 665 €

Più brevi meno di 28h a partire da 665 €

Scali

Quantità

E' indifferente

Tratta diretta

Comprare un biglietto nel momento più favorevole?

Da Milano → A Melbourne Partenza Mer 14 Giu № di passeggeri 1 adulto Cambiare ricerca ▾

Orario di partenza

Partenza Mer, 06:25 Mer, 22:20

Scali

Quantità

E' indifferente

Tratta diretta

Tutti i risultati della ricerca a partire da 665 € a partire da 665 €

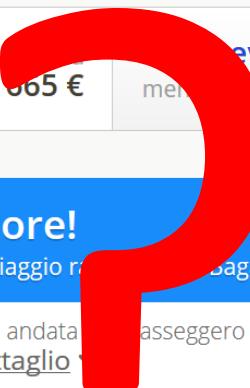
L'opzione migliore!

✓ Offerta migliore ✓ Viaggio non stop! Bagaglio da stiva incluso ✓ Cancellazione gratuita entro 7 ore

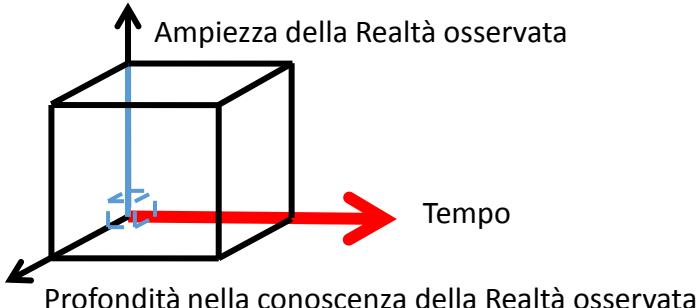
664,11€* Sola andata per passeggero Dettaglio

10:40 Malpensa, Milano 17:50 (+ 1) Melbourne Airport, Melbo... 23h 10' → 1 scalo 7 posti disponibili! Check-in bagaglio incluso

Seleziona



Dalla causalità (**Why**), cioè
la ricerca della politica di pricing...



.... alla correlazione
(What): Il modello
 predittivo di O. Etzioni

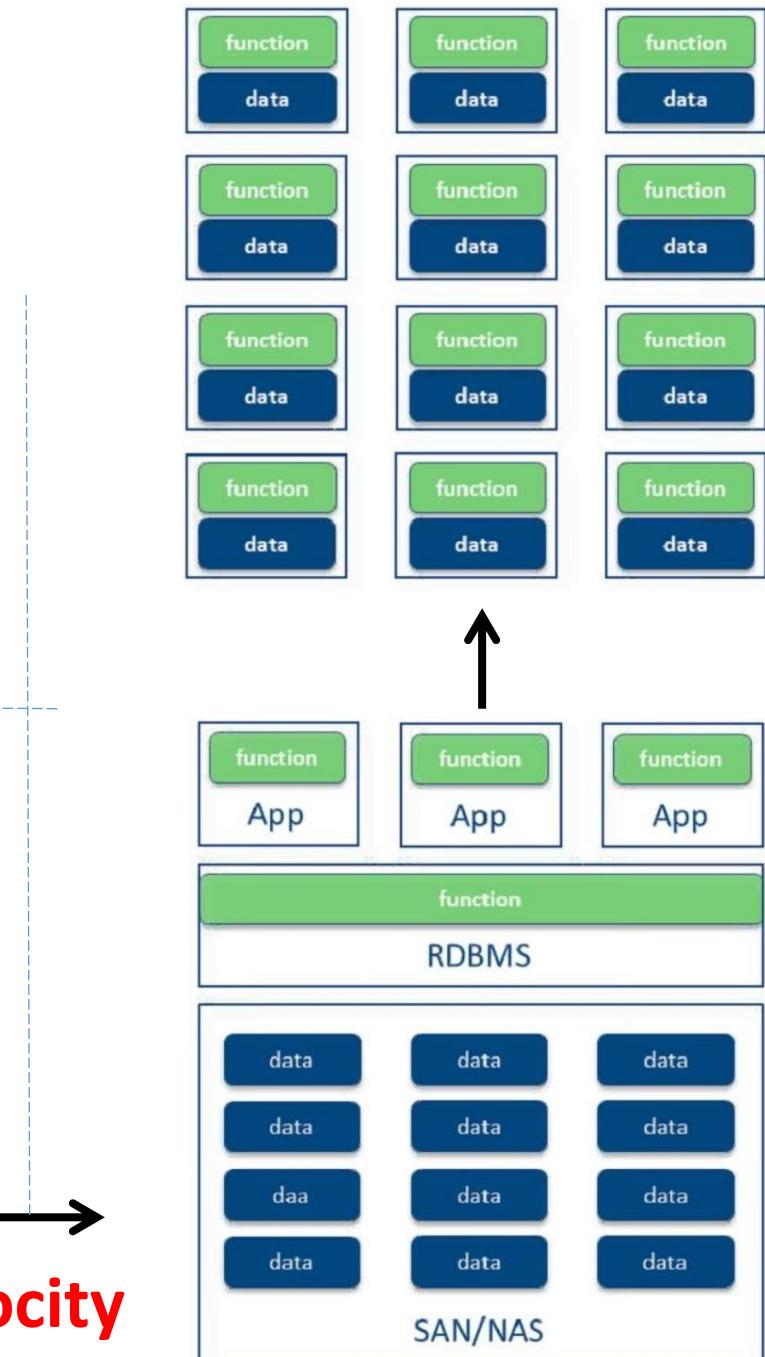
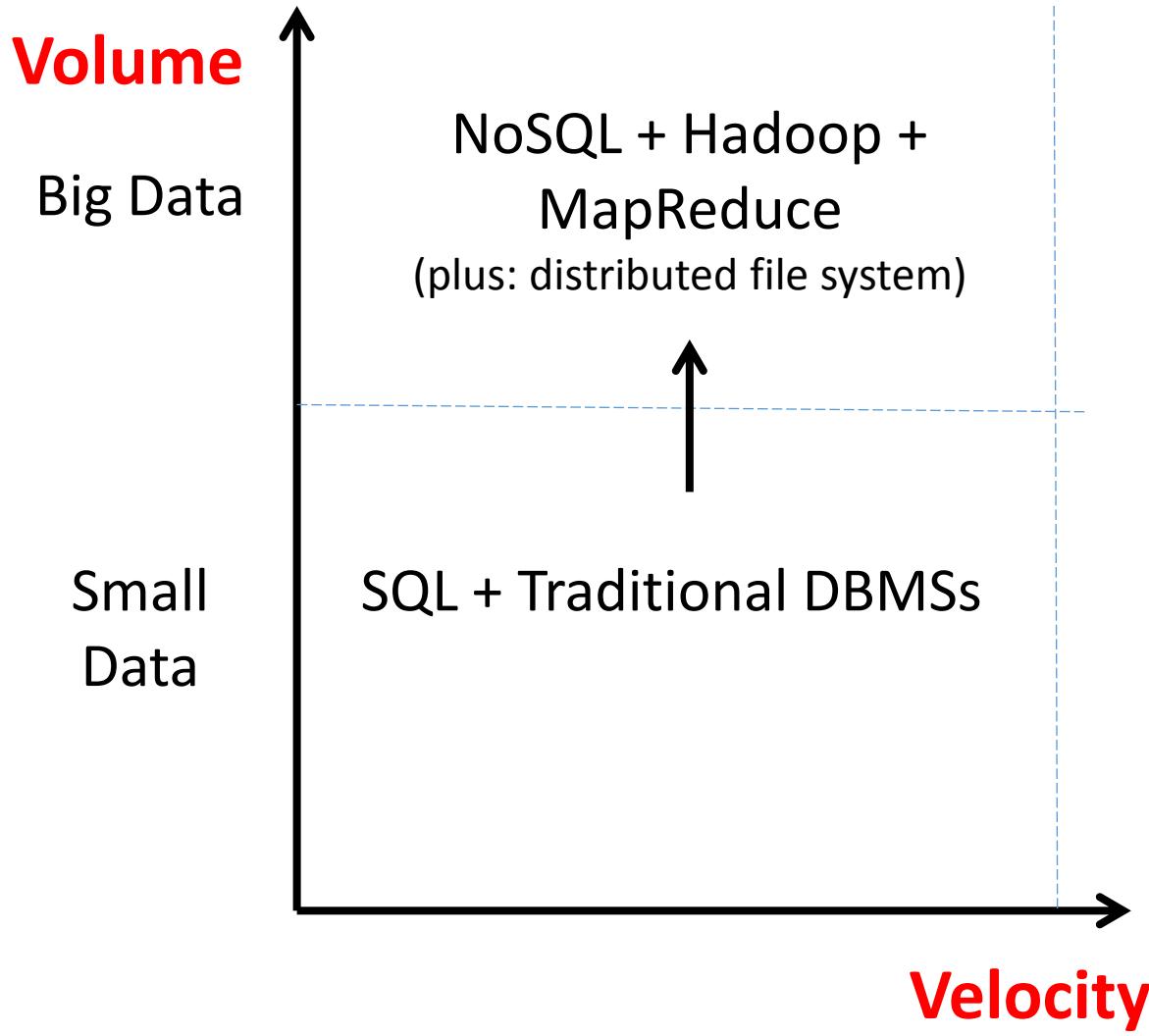
Campione
 di 12.000
 tickets



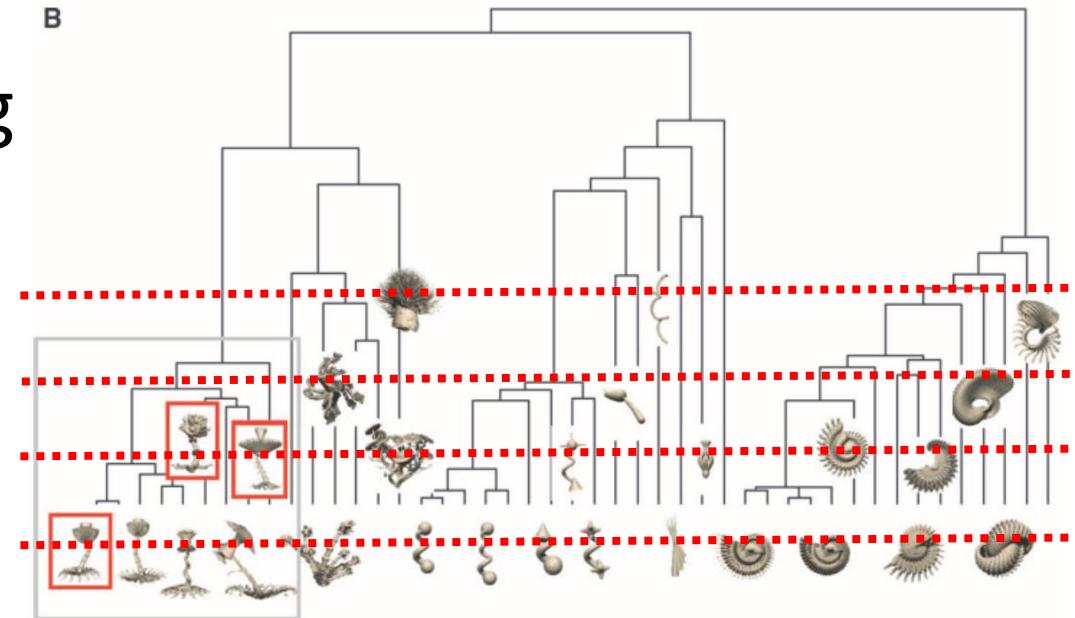
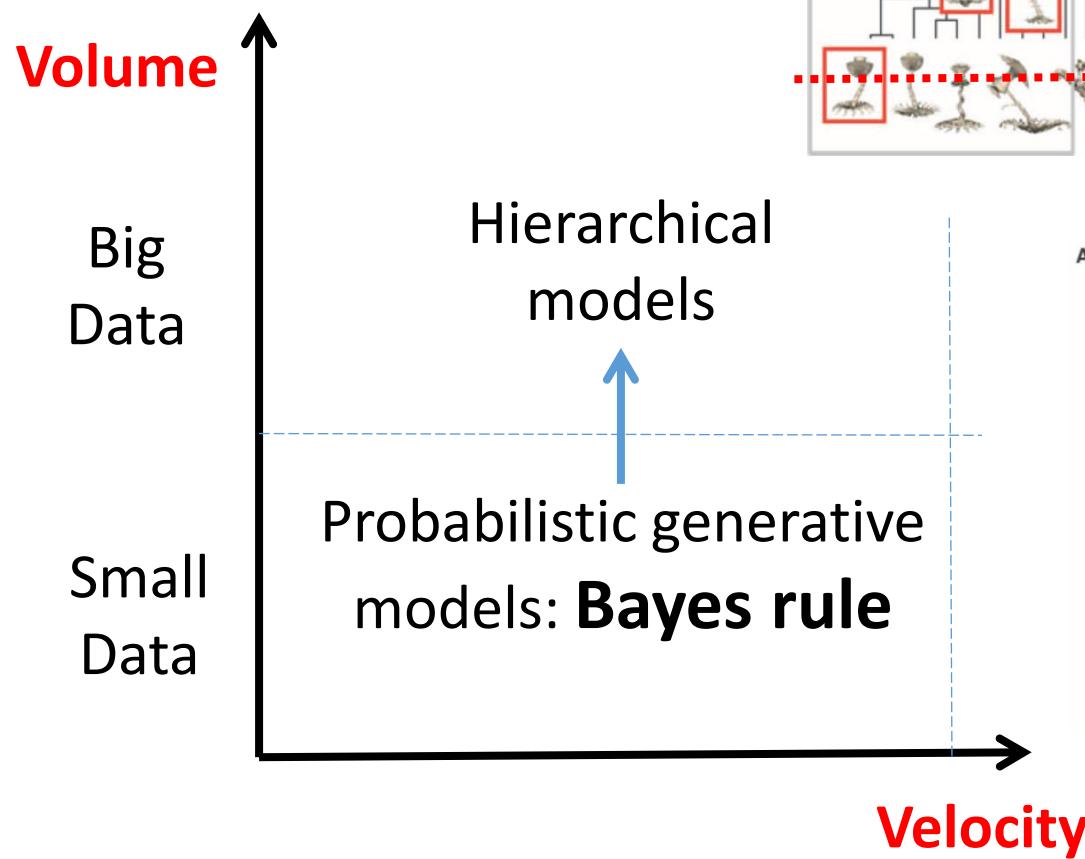
200 10^9

50 \$ medio di risparmio per biglietto
La start-up Farecast
venduta per $110 \cdot 10^6$ \$

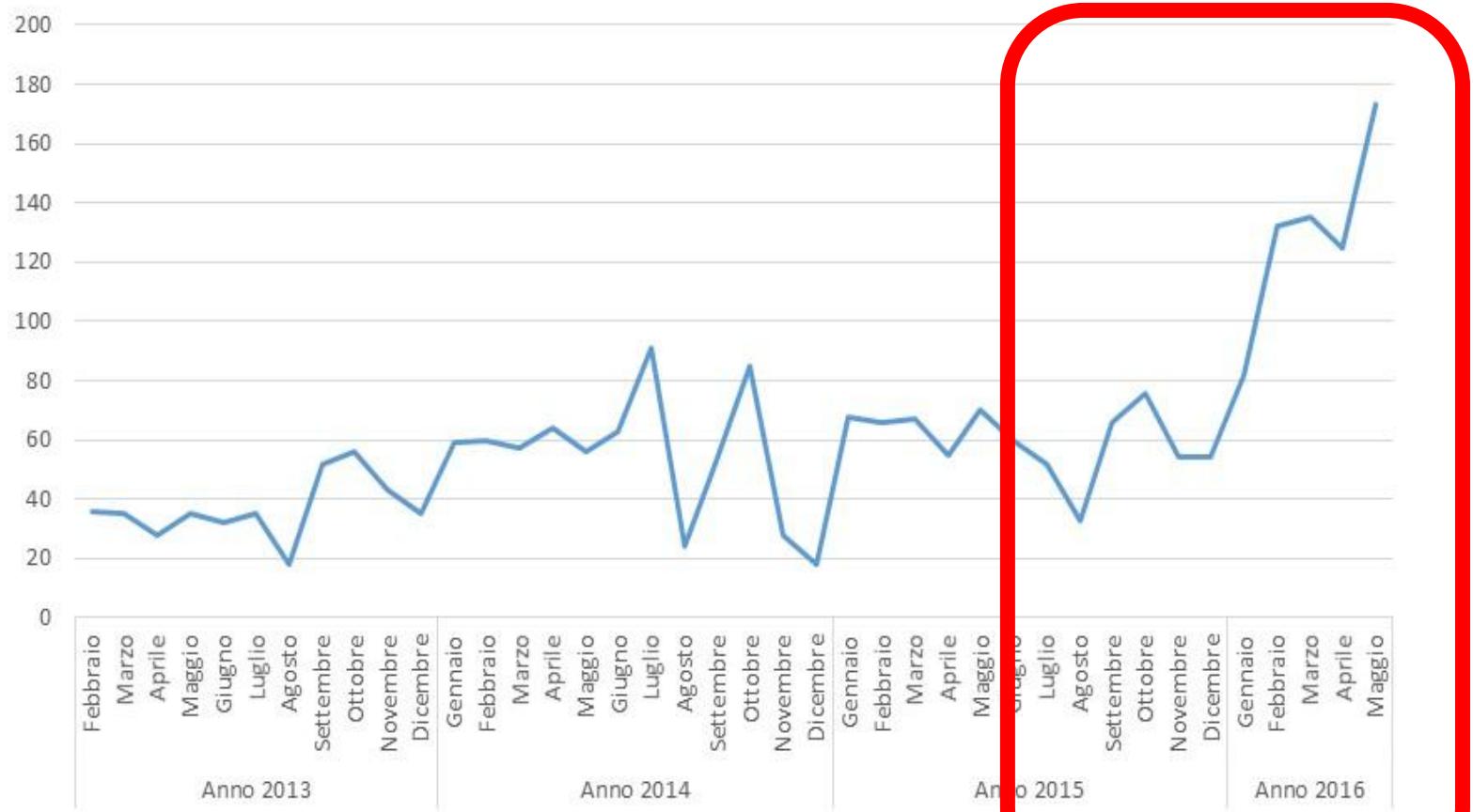
..... in Data Management Systems



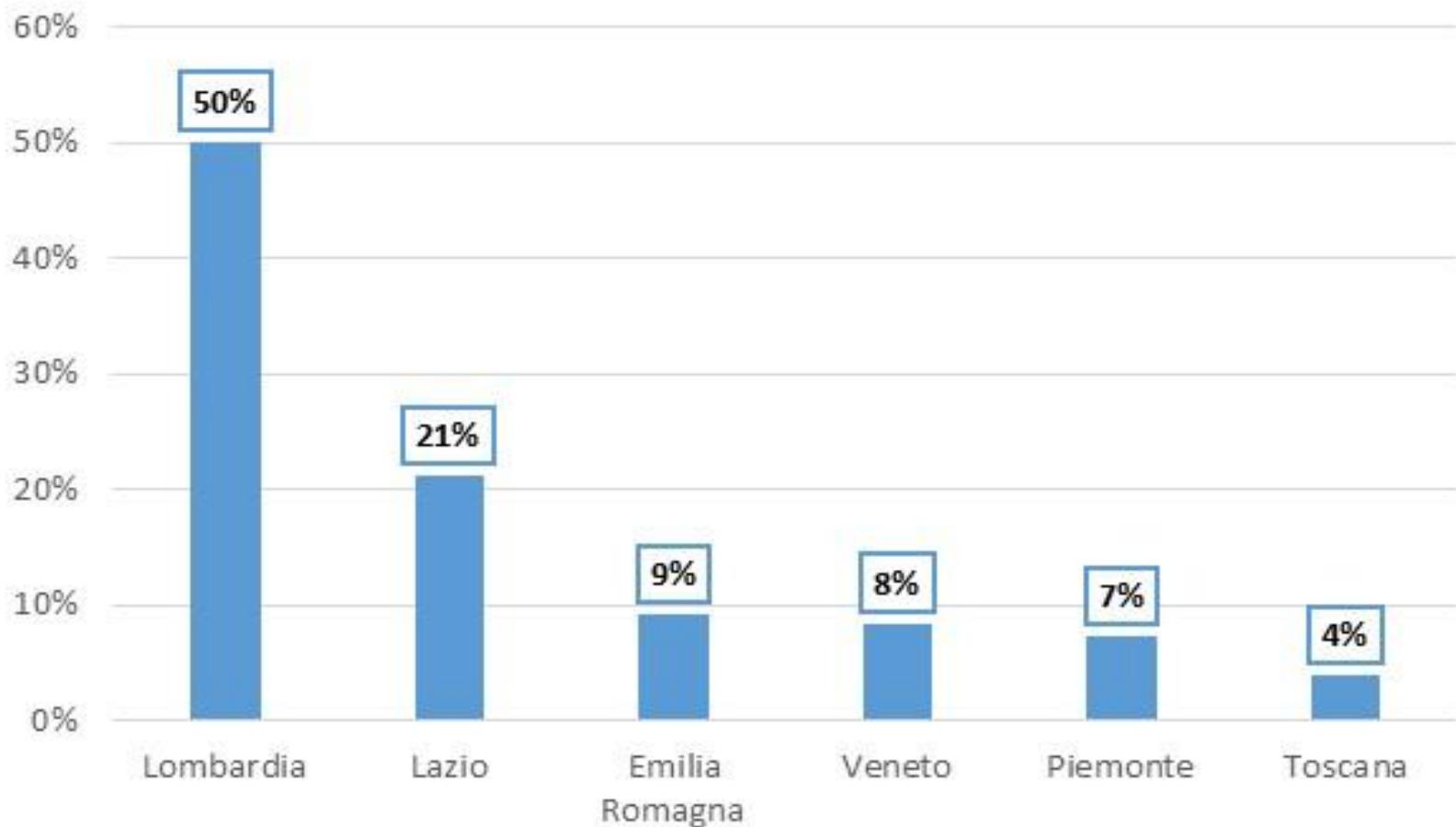
... in Machine Learning Techniques



Sbocchi professionali – Ricerca CRISP-Bicocca forte crescita delle richieste per Data Scientist – Big Data 50% in Lombardia



50% of job offers in Lombardy Region

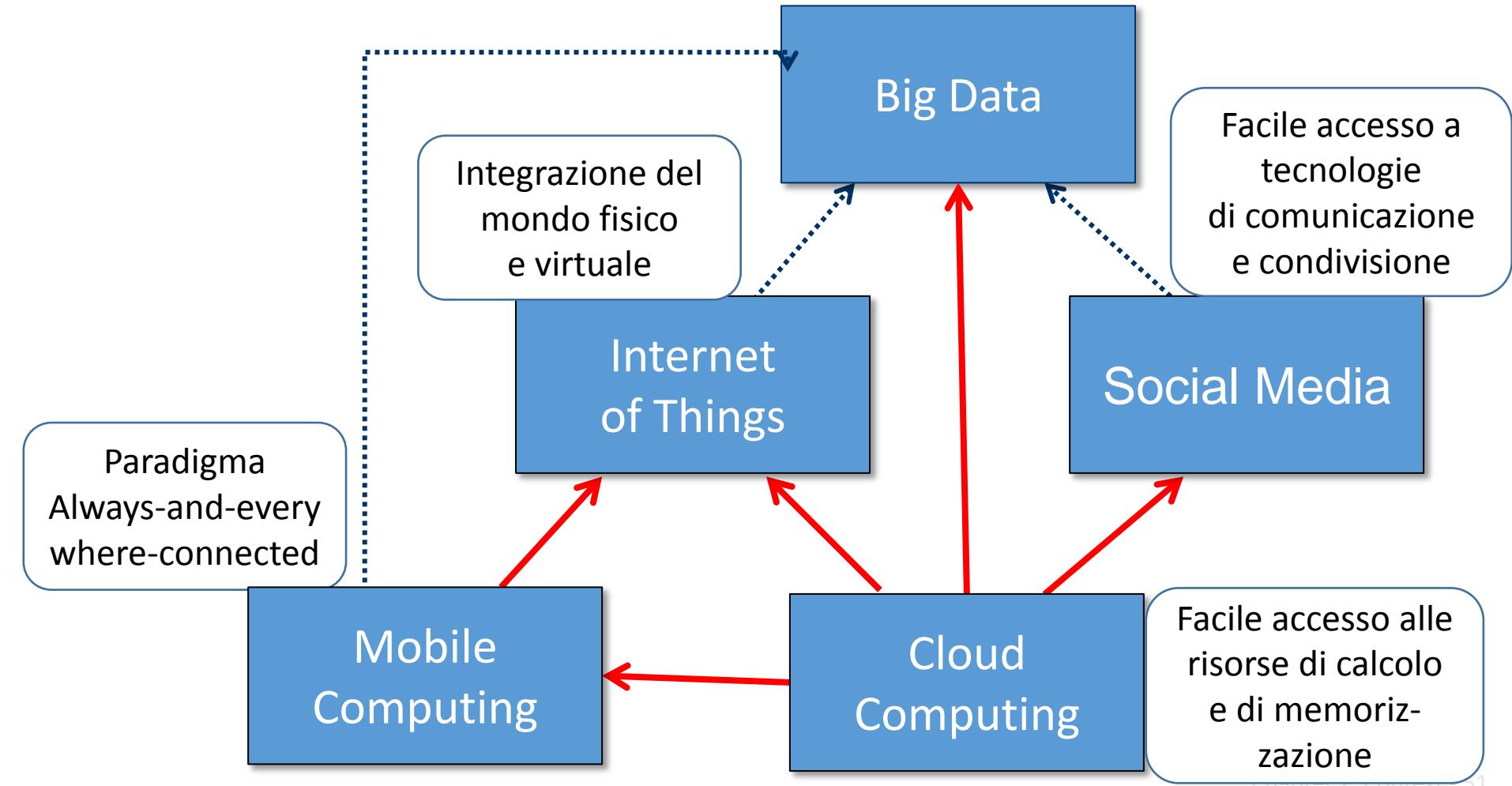


Dagli small data ai big data

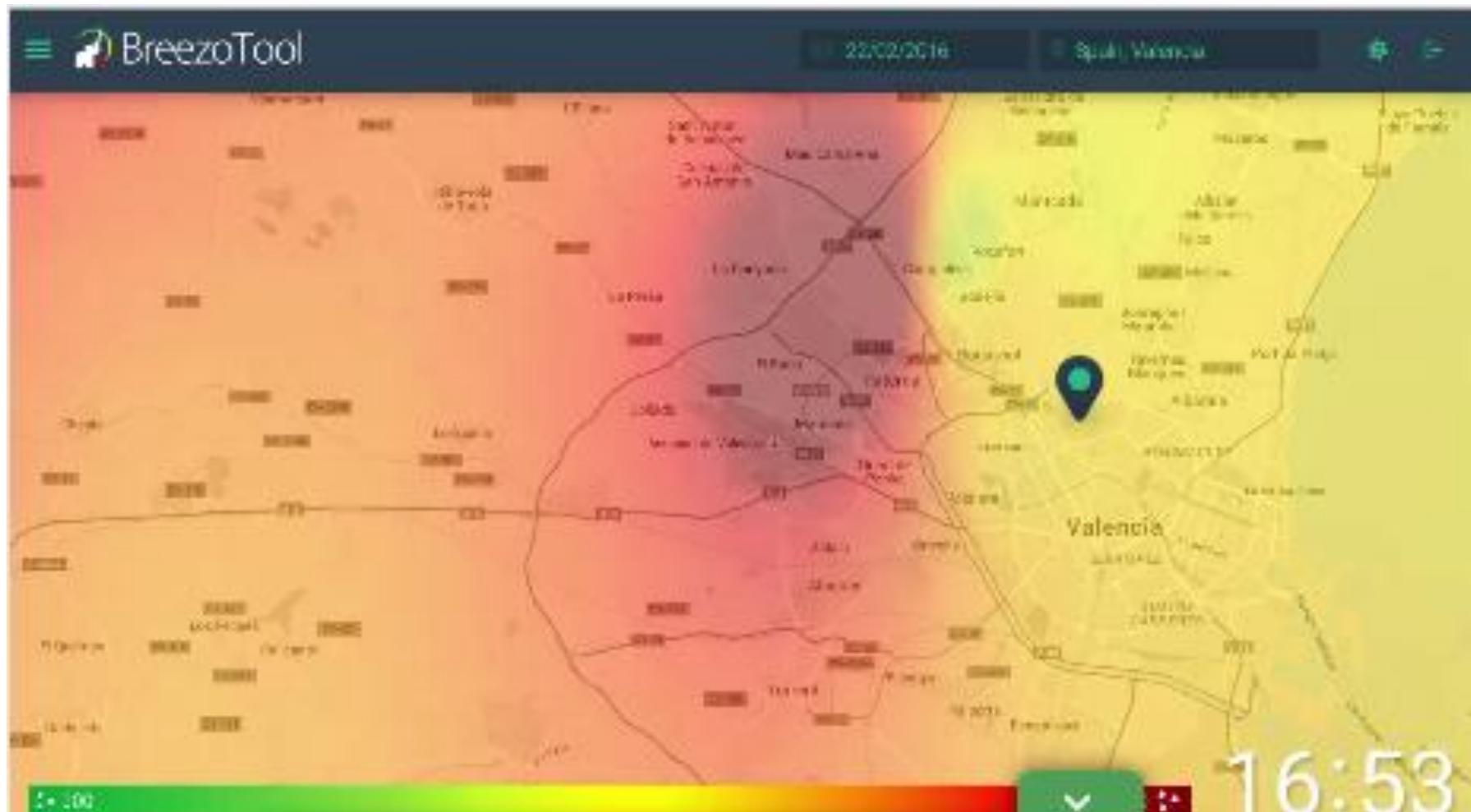
come cambiano
tecnologie informatiche
e tecniche statistiche

Le Big Five

← Abilita
← Alimenta



16.53



17:36

