

Corso di Laurea Magistrale in Economia

Data Science

A.A. 2018/2019

Lez. 8 – Big Data

Big Data

https://www.youtube.com/watch?v=A2pUx5B_C4A&feature=youtu.be

Introduzione

- Sarebbe carino pensare che ciascuna innovazione nella gestione dei dati è un nuovo inizio ed è completamente disconnessa dal passato. In realtà, sia se rivoluzionari sia se incrementali, gran parte dei nuovi stadi nella gestione dei dati sono stati costruiti sui loro predecessori
- Sebbene la gestione dei dati viene vista tipicamente attraverso la lente del software, in realtà essa deve essere considerata da una prospettiva olistica
- La gestione dei dati deve includere avanzamenti tecnologici nell'hardware, nella memorizzazione, nelle reti, nei modelli di calcolo quali la virtualizzazione e il *cloud computing*

Introduzione

- La convergenza di tecnologie emergenti e le riduzioni dei costi per qualunque aspetto, dalla memorizzazione ai cicli di calcolo, hanno trasformato il panorama dei dati e hanno reso possibili nuove opportunità
- La convergenza di tutti questi fattori tecnologici ha portato alla trasformazione del modo con cui si gestiscono e si manipolano i dati.
- I big data rappresentano l'ultimo trend emerso da questi fattori.

Cosa sono?

“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

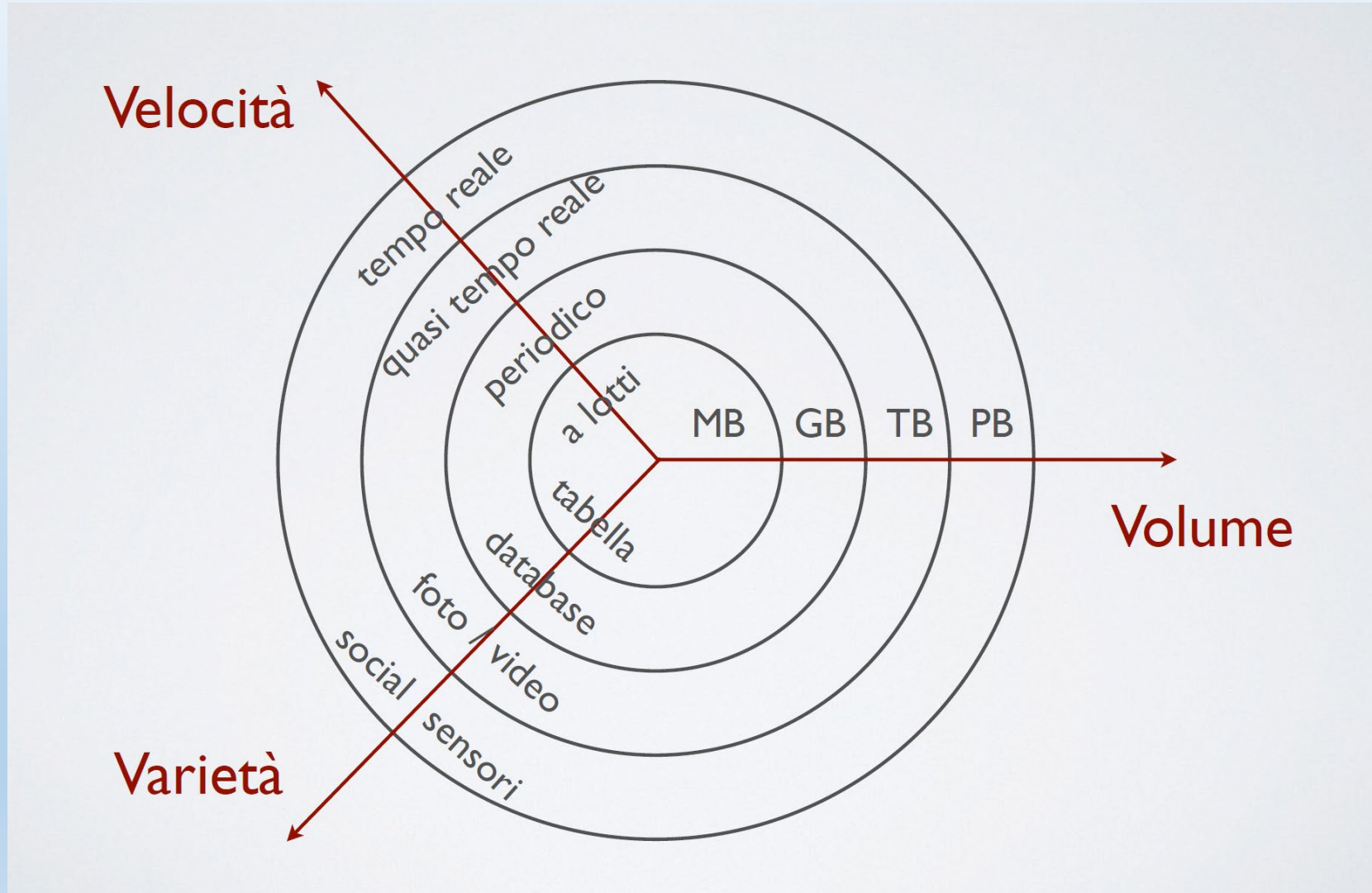
“Big data refers to data sets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze.” - The McKinsey Global Institute, 2012

“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

Definizione

- Per iniziare, possiamo dire che con il termine «big data» si intende una qualunque tipologia di sorgente di dati che possiede contemporaneamente almeno le seguenti tre caratteristiche:
 - Un **Volume** di dati estremamente grande
 - Una **Velocità** dei dati estremamente elevata
 - Una **Varietà** dei dati estremamente ampia
- I big data sono importanti perché consentono alle organizzazioni di collezionare, memorizzare, gestire e manipolare grandi quantità di dati alla giusta velocità, nel tempo giusto, per ottenere le giuste informazioni

Caratteristiche



Caratteristiche

- I big data non sono una tecnologia *stand-alone*; piuttosto, essi rappresentano la combinazione degli ultimi 50 anni di evoluzioni tecnologiche
- Sebbene sia conveniente semplificare i big data nelle tre «V», ciò può essere *misleading* e iper-semplificistico
- Per esempio, noi potremmo dover gestire una quantità relativamente piccola di dati molto eterogenei e complessi, oppure, al contrario, potremmo dover gestire un'enorme quantità di dati molto semplici. Questi dati semplici possono essere tutti strutturati o tutti non strutturati.

Quarta «V» - Veracità

- Quanto accurati sono i dati nel predire i valori di business? Hanno veramente senso i risultati di un'analisi che coinvolge i big data?
- E' importante non sottostimare quest'ultimo aspetto. I dati devono essere verificati sulla base dell'accuratezza e del contesto.
- Un'azienda innovativa può voler essere capace di analizzare un'enorme quantità di dati in *real time* per determinare velocemente il valore di un certo cliente e la potenzialità nel proporgli offerte aggiuntive
- E' necessario identificare la giusta quantità e i giusti tipi di dati che possono essere analizzati per soddisfare le esigenze di quell'azienda
- I big data includono tutti i dati, compresi quelli strutturati e non strutturati provenienti da e-mail, social media, flussi di testi, ed altro ancora. Questo tipo di gestione dei dati richiede che le compagnie siano capaci di manipolare tutti i loro dati, strutturati e non strutturati

Non bastano i dati!!!

What organizations do with big data

"Data is not information, information is not knowledge,
knowledge is not understanding, understanding is not wisdom"
Cliff Stoll

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



Anthony
Goldbloom

The bigger the data set you have, the more accurate the predictions about the future will be



William Deming

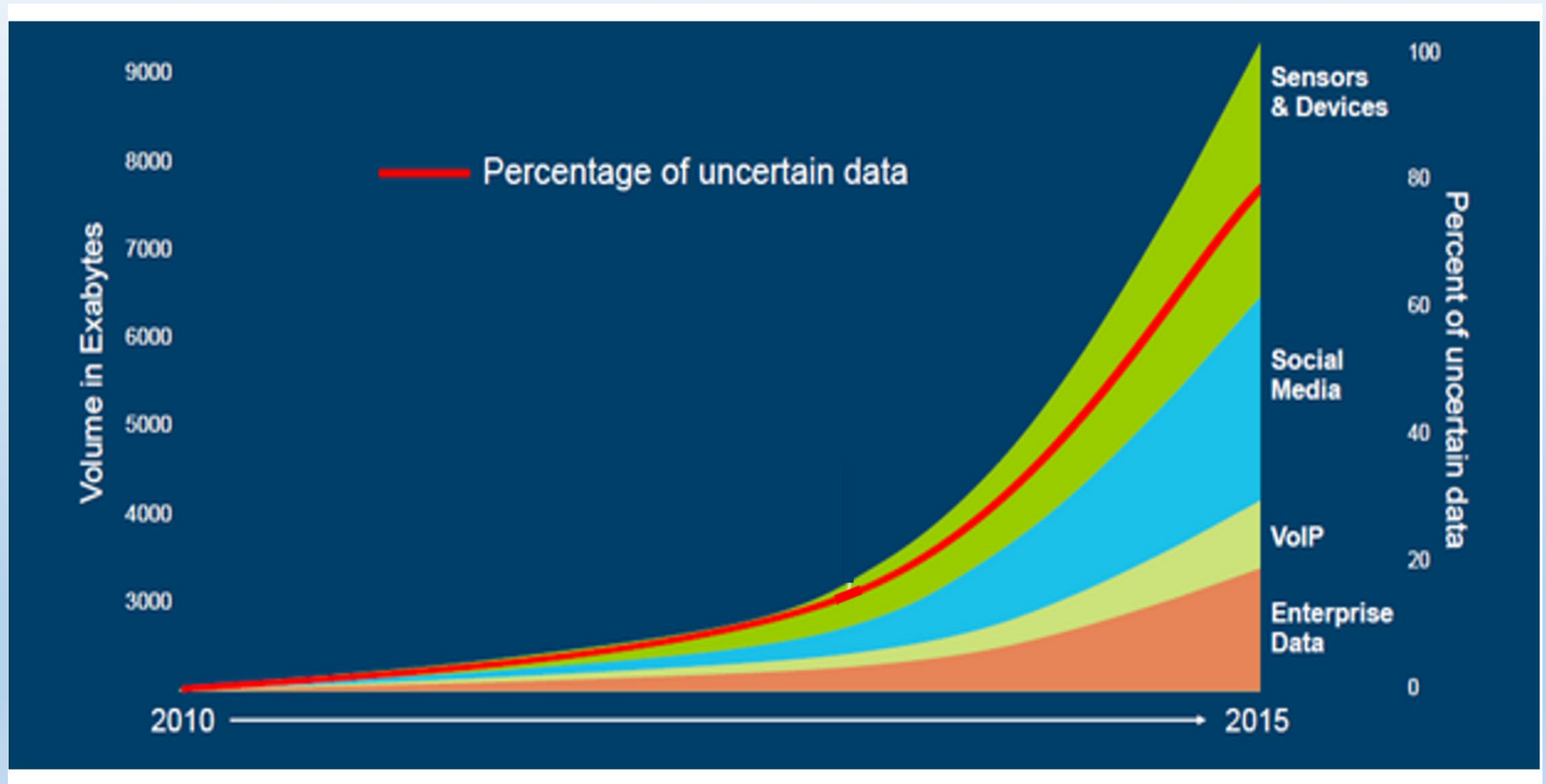
If you torture the data long enough, it will always confess



Ronald
Coase

In God we trust; all others must bring data

Tasso di crescita


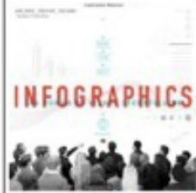



Opportunità

- E-commerce: sistema di *recommendation*

Altri prodotti da tenere presente




Hai visto Chi ha visto questo prodotto ha visto anche

 <p>Visualize This: The FlowingData Guide... Nathan Yau Brossura EUR 24,12</p>	 <p>L'arte funzionale. Infografica e... Alberto Cairo Brossura EUR 25,50</p>	 <p>Infographics: The Power of Visual... Jason Lankow, Josh Ritchie, Ross Crooks Brossura EUR 17,56</p>	 <p>Now You See It: Simple Visualization... Stephen Few Rilegato EUR 30,17</p>	 <p>Data Flow: Visualising Information... Robert Kintan, Nicolas Bourquin, ... Rilegato</p>	 <p><u>The Functional Art: An introduction...</u> Alberto Cairo Brossura EUR 23,47</p>
---	--	--	---	--	---

[Visualizza o modifica la tua cronologia di navigazione](#)

Simili a prodotti che hai già visto

Hai visto Chi ha visto questo prodotto ha visto anche

 <p>La società dei dati Vincenzo Cosenza Formato Kindle ★★★★☆ (6) EUR 0,99</p>	 <p>Social Media ROI (Guida completa) Vincenzo Cosenza Formato Kindle ★★★★☆ (6) EUR 14,99</p>	 <p>Big Data: A Revolution That Will... Kenneth Cukier, Viktor Mayer-Schonberger Formato Kindle ★★★★☆ (1) EUR 12,99</p>	 <p>Enciclopedia Hacker Marco Torrione Formato Kindle ★★★★☆ (2) EUR 3,96</p>	 <p>The Little Book of BIG DATA, 2012... Noreen Burlingame Formato Kindle EUR 4,02</p>	 <p>Big Data - An Introduction Subu Raj Formato Kindle EUR 2,68</p>
--	--	---	--	--	---

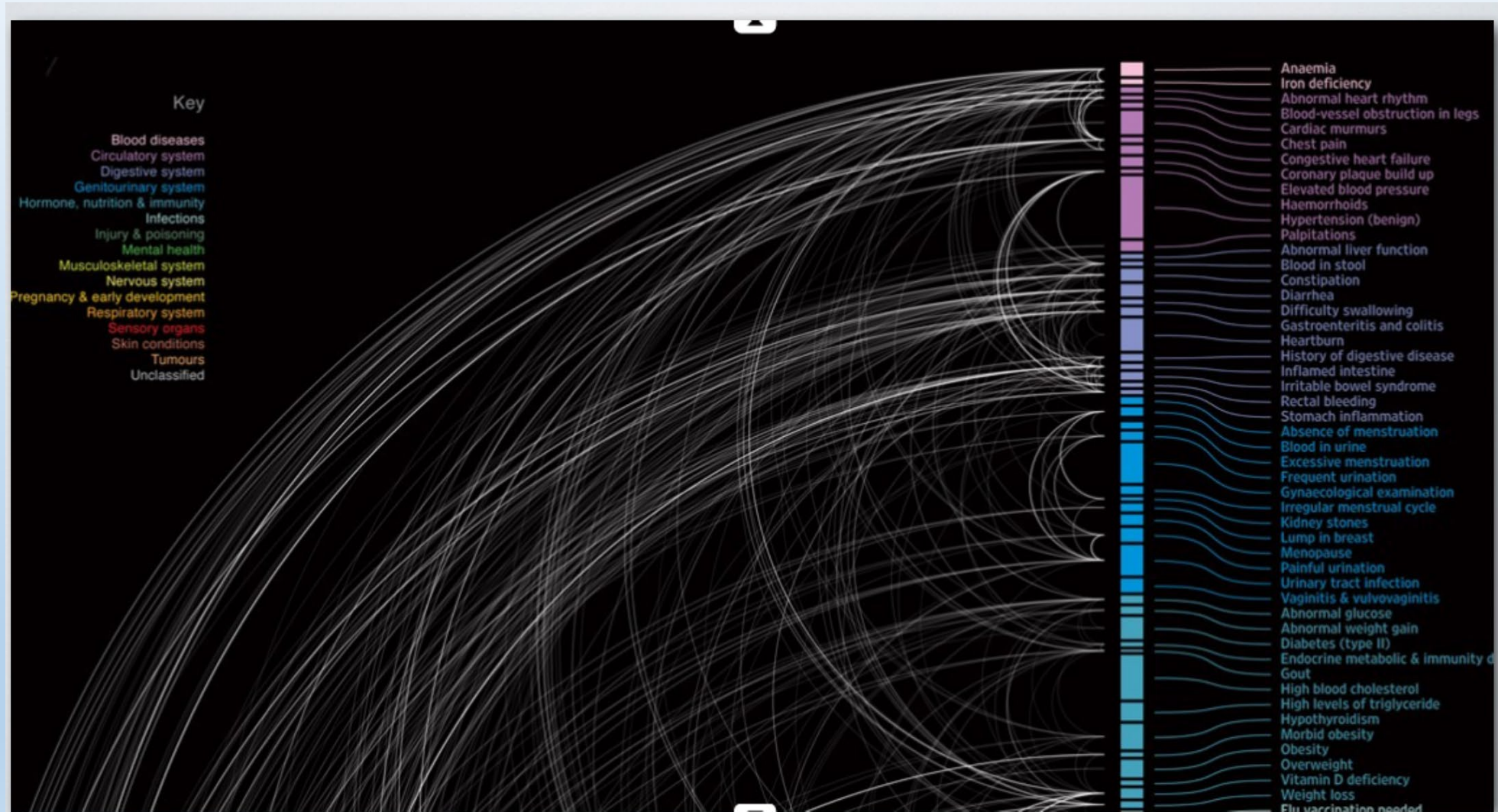
Opportunità

- Advertising



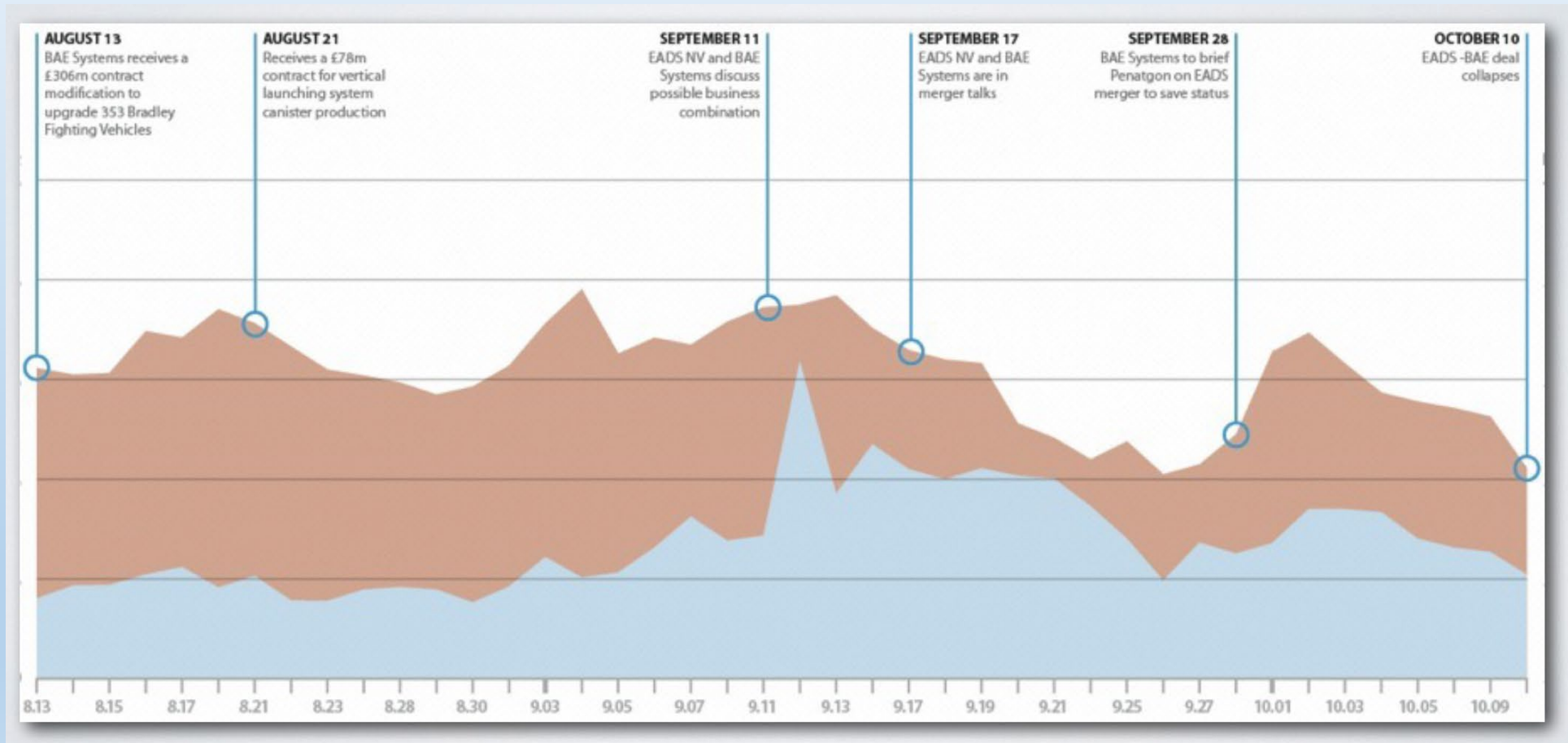
Opportunità

- Healthcare



Opportunità

- Finanza: *sentiment analysis*



Use cases

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

Potenziale



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

I casi di insuccesso

- Assenza di figure con la professionalità necessaria
- Uso di tecnologie grezze e immature
- Mancanza di un *business case* specifico

I casi di successo

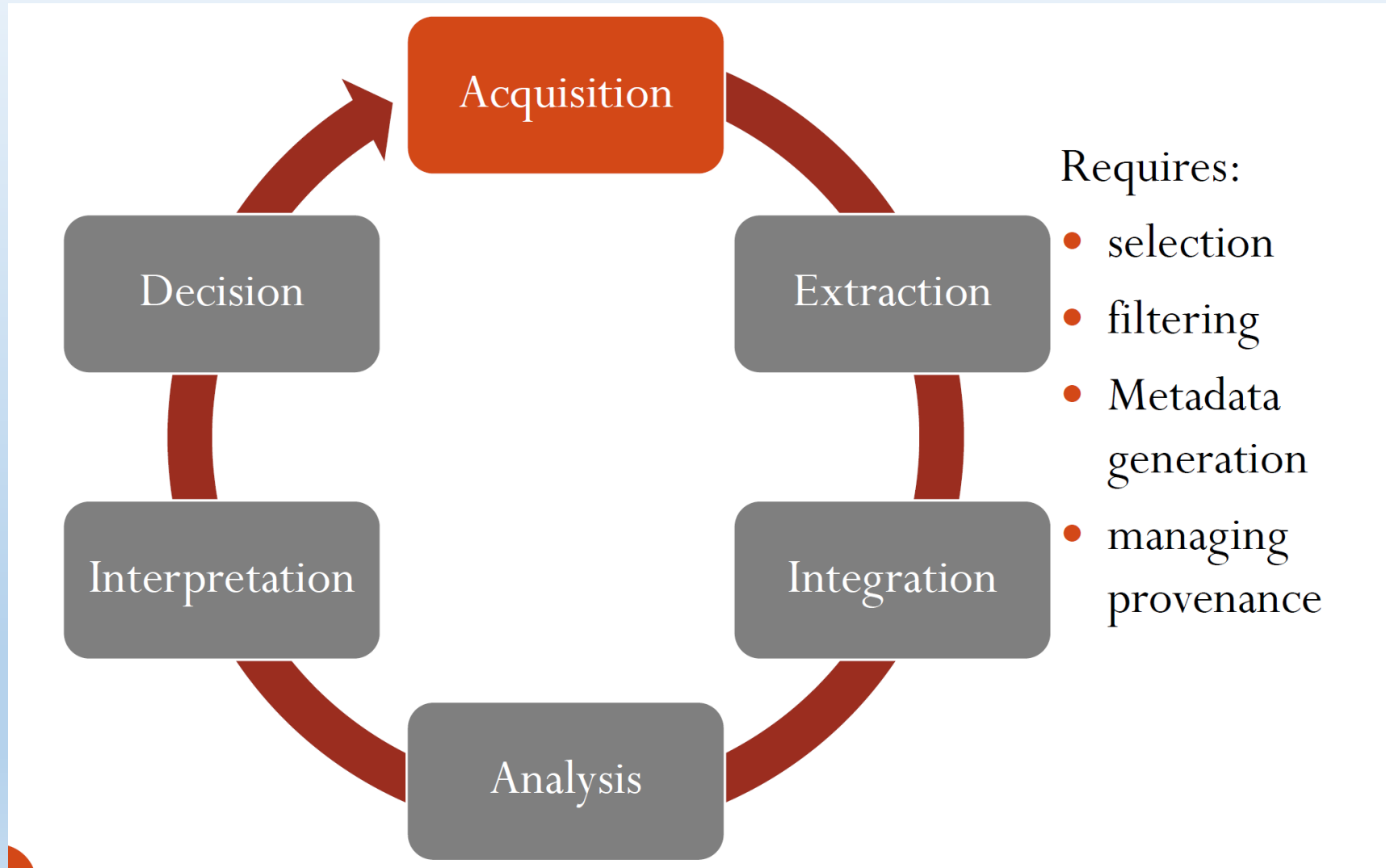
- Progetti non sponsorizzati di IT, ma da dipartimenti di *line-of-business*, come marketing o logistica
- Focus su un caso d'uso piccolo, ma strategico
- Iterare e crescere sulla base dei risultati precedenti

Il ciclo di gestione dei Big Data

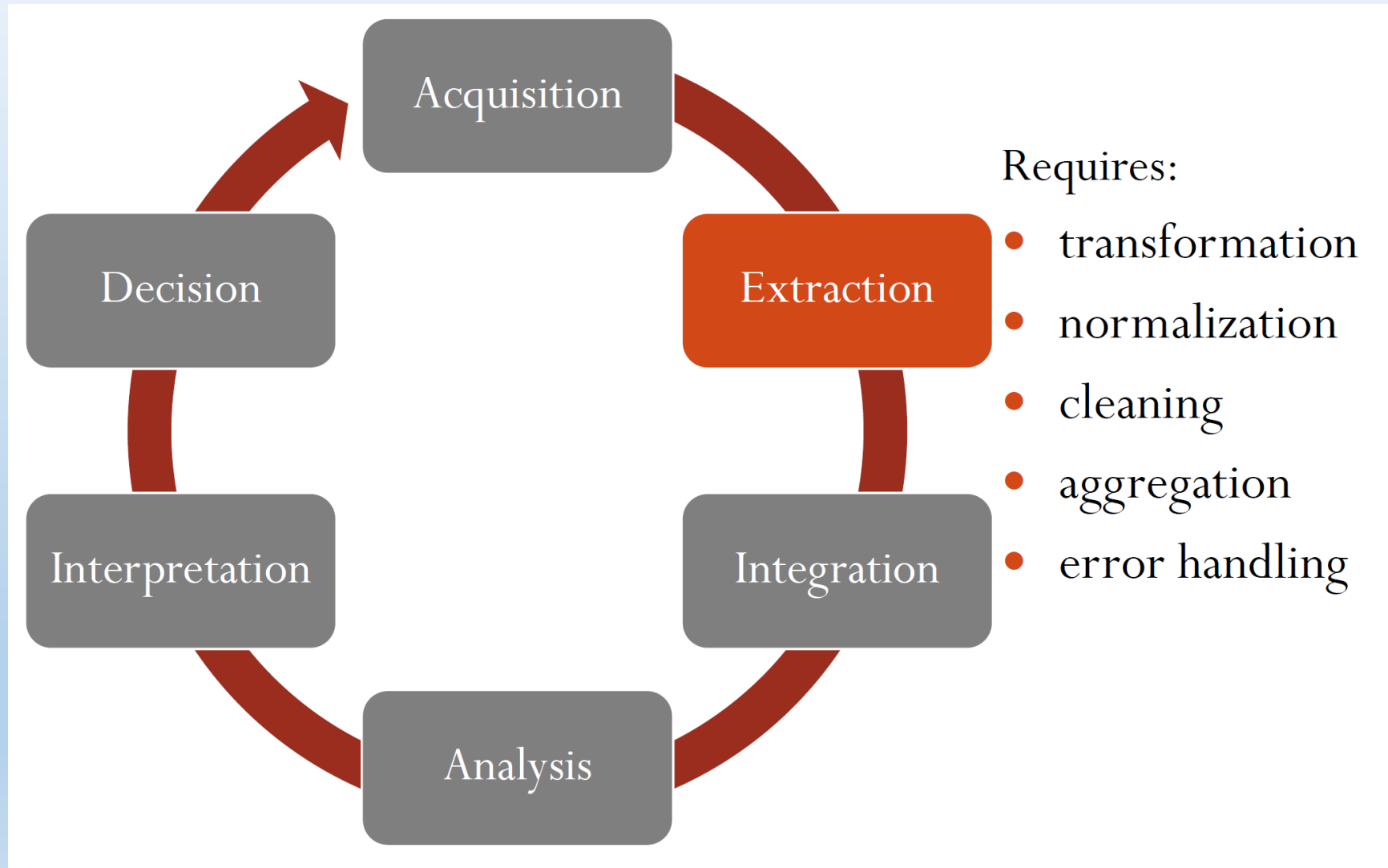
- I dati devono essere catturati, organizzati e integrati.
- Successivamente, possono essere analizzati sulla base del problema da risolvere
- Infine, possono essere effettuate azioni sulla base dei suggerimenti prodotti dalla fase di analisi.
- Per esempio, Amazon potrebbe raccomandare un libro sulla base degli acquisti passati o un cliente potrebbe ricevere un coupon per uno sconto per un acquisto futuro di un prodotto correlato ad uno già acquistato



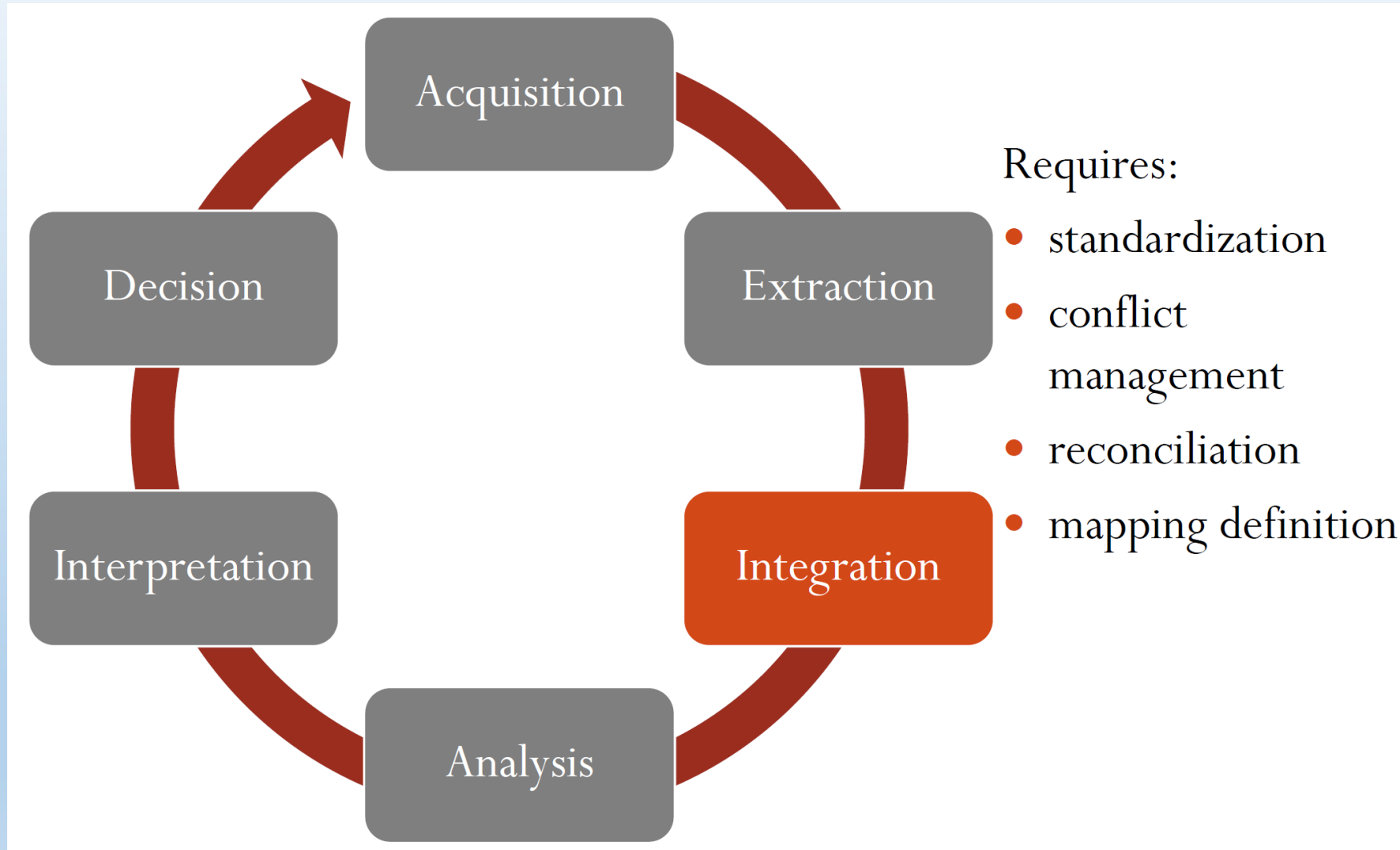
Il ciclo di gestione dei Big Data



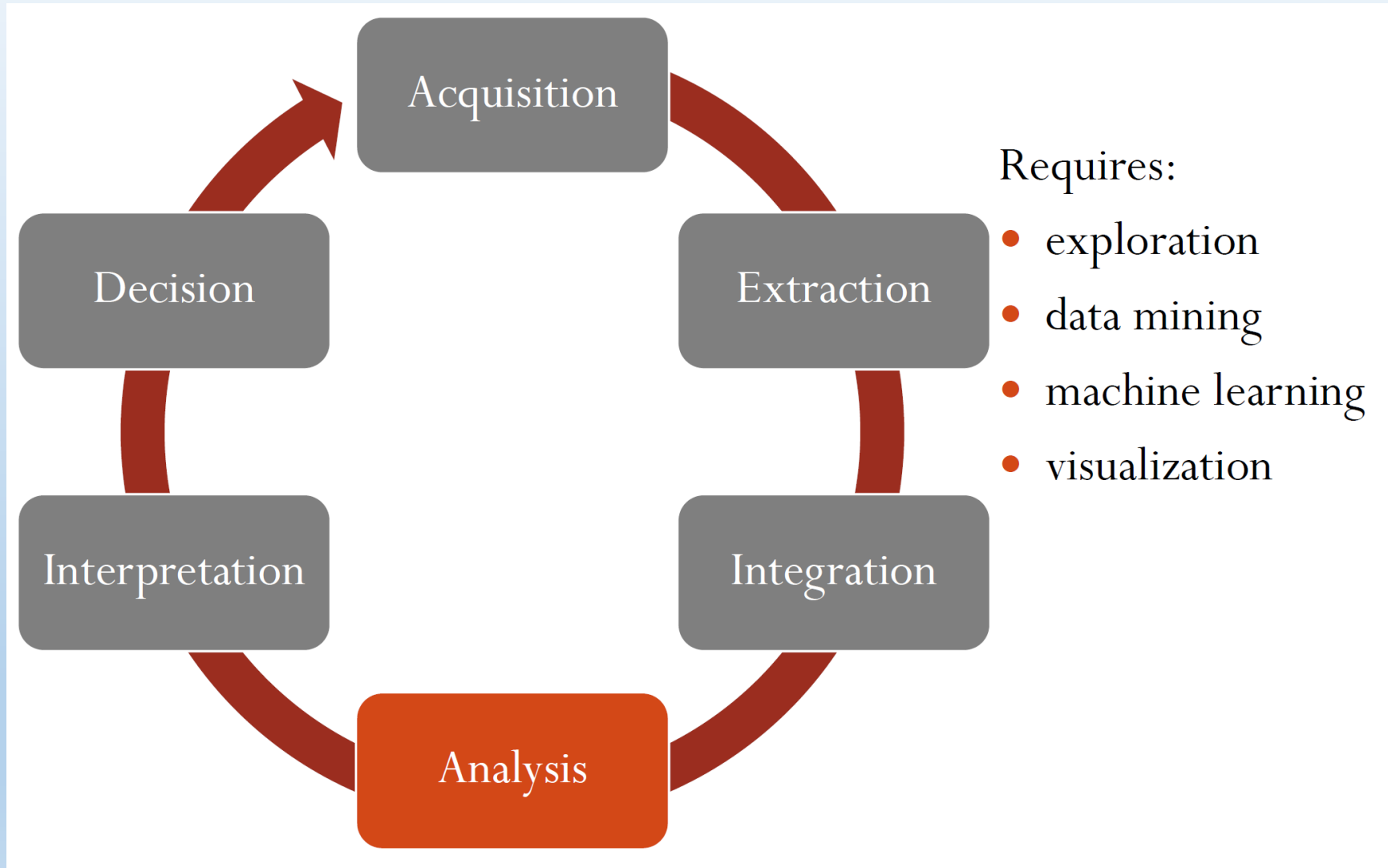
Il ciclo di gestione dei Big Data



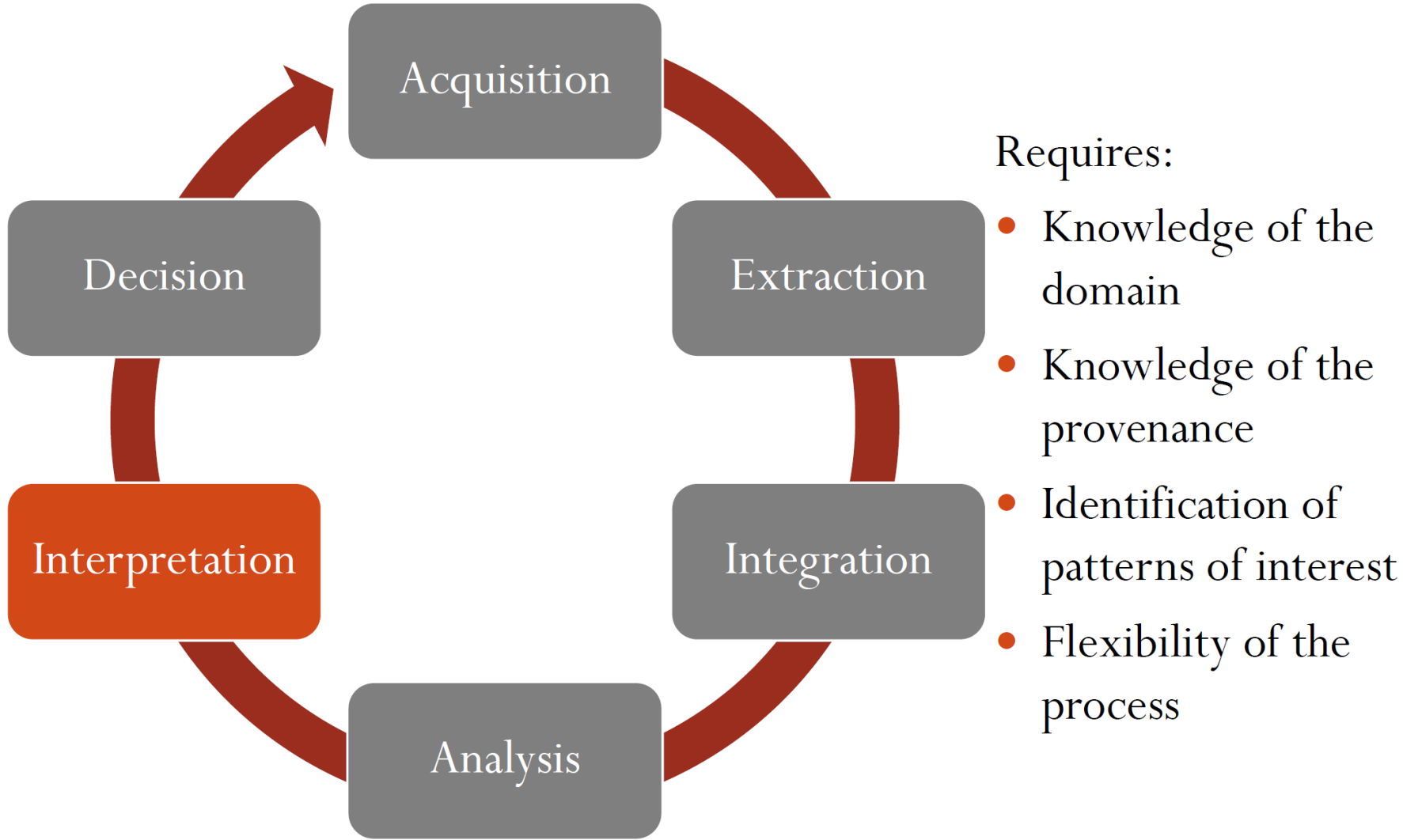
Il ciclo di gestione dei Big Data



Il ciclo di gestione dei Big Data



Il ciclo di gestione dei Big Data



Il ciclo di gestione dei Big Data

