

# Big Data e Business Intelligence

Marco Pozzan  
info@marcopozzan.it

# Chi sono

- Formazione & Divulgazione con [1nn0va](#)
- Consulente Business Intelligence
- MVP Sql Server 
  - [marco.pozzan@live.it](mailto:marco.pozzan@live.it)
  - [www.marcopozzan.it](http://www.marcopozzan.it)
  - [www.innovazionefvg.net](http://www.innovazionefvg.net)
  - @marcopozzan.it

# Agenda

- Business Intelligence e Business Analytics
- Introduzione ai Big Data
- Hadoop
- Analisi BigData a basso costo
- Strumento di analisi

# Business Intelligence



# Tanti Dati

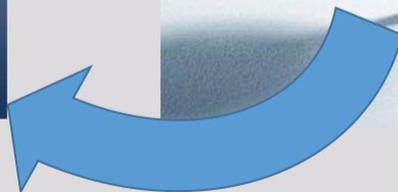
- Ogni giorno le organizzazioni devono prendere delle decisioni basandosi sulle informazioni che provengono dai **DATI**



# Cosa è la BI?



Molti dati disagregati  
(Gestionale, HR, CRM, Excel, Xml, ecc..) = **CAOS**



# Cosa vogliamo sapere?

- **Due caratteristiche di una azienda**
  - **Business is doing now**
    - Quello che si sta facendo in questo momento
  - **Perform in future**
    - Come sarà il futuro

# Che cosa è la BI?

Business Intelligence  
(Here & Now)

Quante transazioni  
sono rifiutate?

Quante transazioni al  
giorno?

Quale è il profitto per  
transazione?



Sistema di pagamenti con carta  
di credito e con milioni di  
transazioni al giorno

# Che cosa è la BA?

Quali tipi di transazioni saranno considerate fraudolente?

Dove si verificherà il maggior numero di transazioni fraudolente?



Business Analytics  
(Future)



Sistema di pagamenti con carta di credito e con milioni di transazioni al giorno

# Grande verità

**Tutto quello che in azienda non è  
misurato va male!!!**

# Big Data

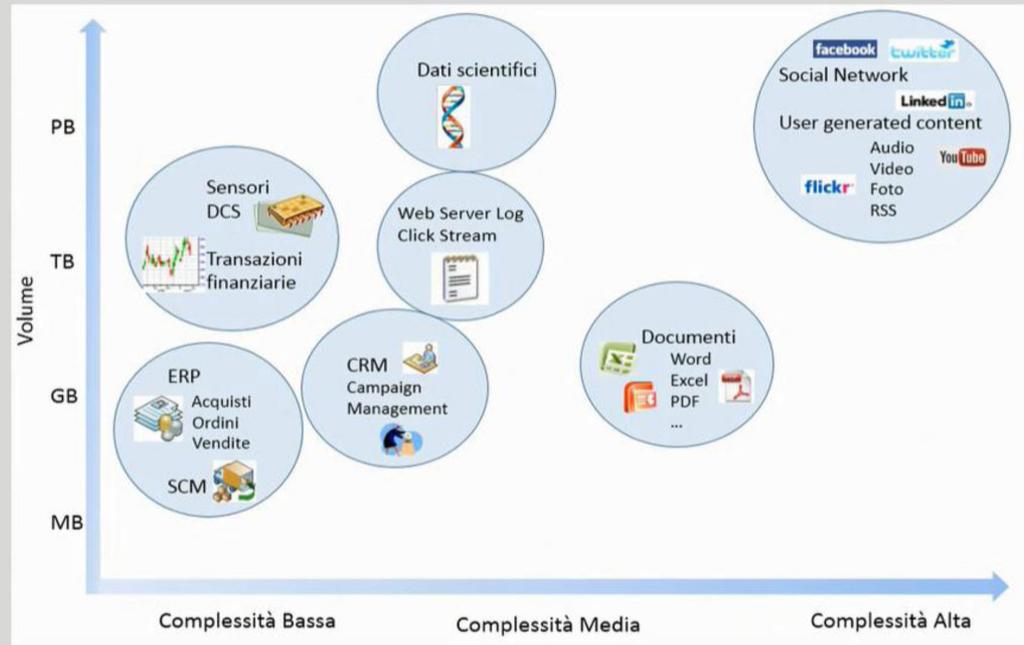
# Caratteristiche dei Big Data

- **Tre caratteristiche**
  - **Volume**
    - Grandi quantità di dati
  - **Varietà**
    - Varietà di strutture, tipo dati
    - Complessità delle strutture
      - Dati destrutturati o semi-strutturati
  - **Velocità**
    - Velocità con cui sono prodotti

# Che cosa sono i Big Data

- Sono i **dati non analizzabili** su una solo macchina o che non conviene analizzare con **tecnologie tradizionali**
- **Dati non strutturati** difficili da analizzare con un DB relazionale

# Esempio di Big Data



# Trend - Opportunità

- Dati digitali cresceranno del **44X** rispetto alla precedente decade
- Nel 2015 i servizi pubblici offerti dal cloud rappresenteranno il **46%** della crescita complessiva dei prodotti IT per la nuova infrastruttura di rete.

# Trend - Opportunità

- **40** Trillioni di Gigabytes sarà la dimensione dell'universo digitale nel 2020 rispetto ai 130 bilioni del 2005
- Il **90%** dei dati digitali nel mondo sono stati creati negli ultimi due anni

# Hadoop

# Hadoop

## Ecosistema Hadoop

- File system distribuito (HDFS)
- Sistema di calcolo distribuito
- Applicazioni di gestione
- Applicazioni varie

## Altre caratteristiche importanti

- Free / Open Source
- Scalabilità ordine dei PB

# Alcuni utilizzatori famosi

YAHOO!

AOL 

 ImageShack  
online media hosting

ebay™

facebook®

LinkedIn®



Spotify®

The New York Times

twitter 

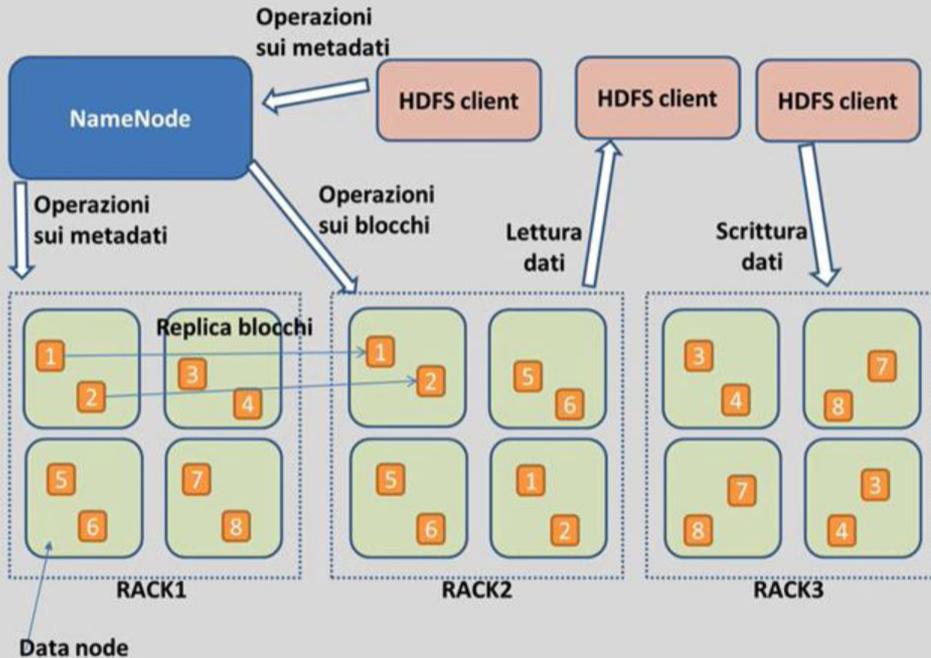
FONDAZIONE  
> ITS PORDENONE

INFORMATION COMMUNICATION TECHNOLOGY

Con la collaborazione di

innova

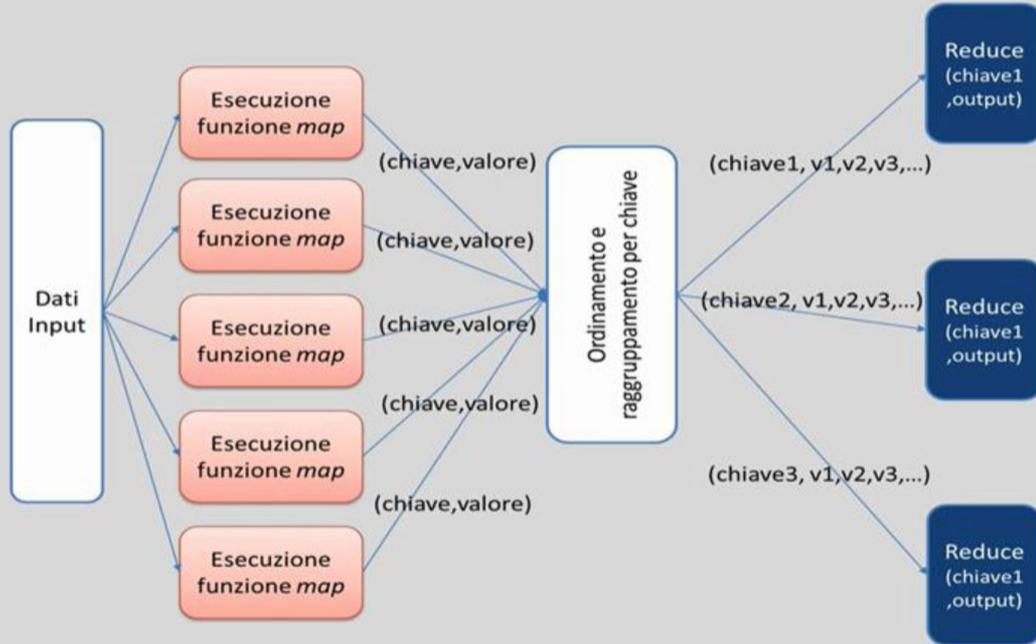
# HDFS



Hadoop Distributed File System

- Fault tolerant
- Distribuito e ridondato
- Scalabile (Migliaia di nodi)

# MAP Reduce



- Batch oriented
- Principio divide et impera
- Framework di calcolo
- Functional programming

# Conclusioni

- **Hadoop + Business Intelligence** = Analisi di grandi quantità di dati
- **analisi dati, big data**, sono tutti campi che – promettono i **bookmakers statunitensi** – daranno da mangiare a **sempre più giovani** anche nel resto del mondo non solo in Italia

# Analisi BigData a basso costo

# Self-Service BI

- Le organizzazioni hanno insufficienti capacità di andare in profondità nel dettaglio del dato aziendale
- La Business Intelligence **tradizionale** è **costosa**
- La Business Intelligence **tradizionale** è **lenta**
- Le **decisioni aziendali** possono aver bisogno di essere **fatte velocemente**

# Self-Service BI players



# Microsoft Excel for BI???

«...La soluzione di **self-service BI** per gli **utenti di business** passa attraverso un **tool familiare** pronto all'uso»

# Microsoft Excel for BI

- E' molto probabile che **Excel** sia già la linfa vitale della vostra organizzazione
- Le persone che **usano excel** saranno gli utenti primari di qualsiasi nuovo strumento self-service si decida di adottare
- **Excel è molto flessibile** rispetto altri tools
- «**Export in Excel**» esiste perché l'excel è flessibile non perché l'Excel è popolare

# Le figure richieste

- **Data Analyst (Analista)**

(**raccoglie, organizza e interpreta le informazioni statistiche** ) è come un motore di ricerca o un **radar umano capace** di orientarsi tra milioni di informazioni che riguardano un sistema aziendale

*«Diciamo che quella del **data scientist** di cui abbiamo parlato anche a Milano è una **figura codificata a livello professionale negli Stati Uniti, ma non in Italia**. Da noi è più corretto parlare di **analista dati più che di scienziato dei dati** proprio perché le mansioni più richieste ruotano, per ora, attorno alla capacità di **sintetizzare informazioni utili a chi poi prende le decisioni in azienda**».*

*Marco Russo*

# Le figure richieste

- **Data Steward (DBA e personale IT)**

uno steward di dati è una persona che è responsabile per il **mantenimento dei metadati e dei dati**. E' un ruolo ampio che incorpora i processi, le politiche, le linee guida e le responsabilità **per la gestione tutti i dati delle organizzazioni** in conformità con le imprese e / o di obblighi normativi. Un steward dati assicura che ci sono procedure documentate e linee guida per l'accesso e l'utilizzo dei dati.

# Conclusioni

«anche se le Facoltà italiane sono ancora sfornite di corsi appositi (**ad eccezione dell'Università di Tor Vergata di Roma che ha da pochissimo avviato il primo percorso formativo in Data Science**), all'estero l'offerta per chi vuole diventare scienziato dei dati e data steward inizia a diversificarsi»

<http://www.tecnicosuperiorekennedy.it/icorsi/i-corsi-2014/le-figure-professionali> (Intelligent Data Analysis)

# Grazie

> **FONDAZIONE**  
**ITS PORDENONE**

INFORMATION COMMUNICATION TECHNOLOGY

Con la collaborazione di  
**Innova**