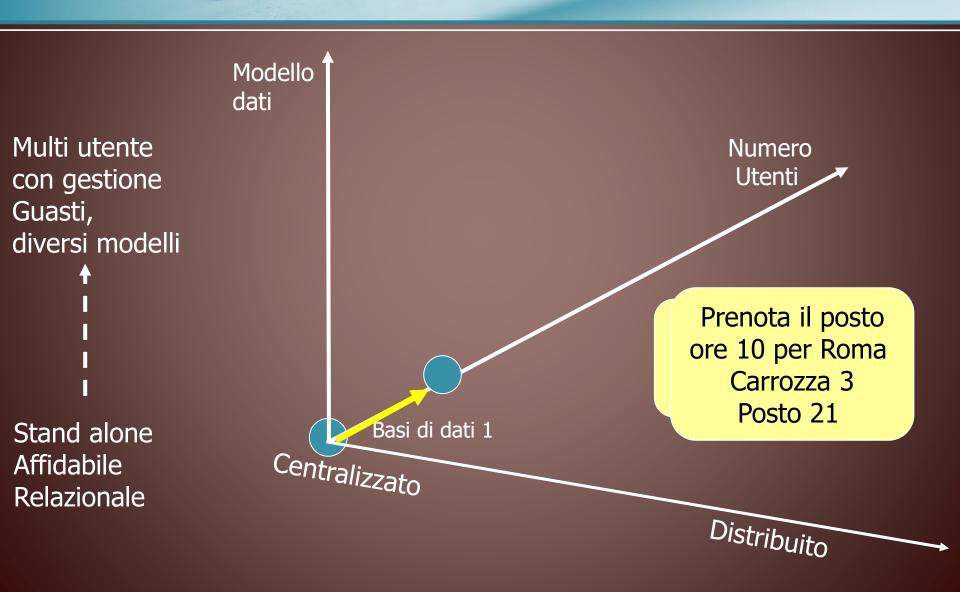
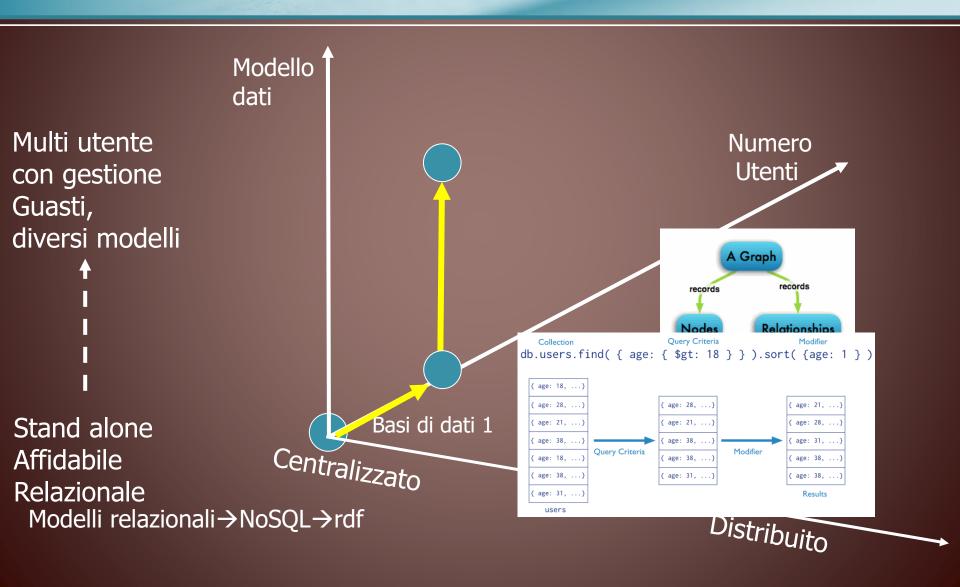


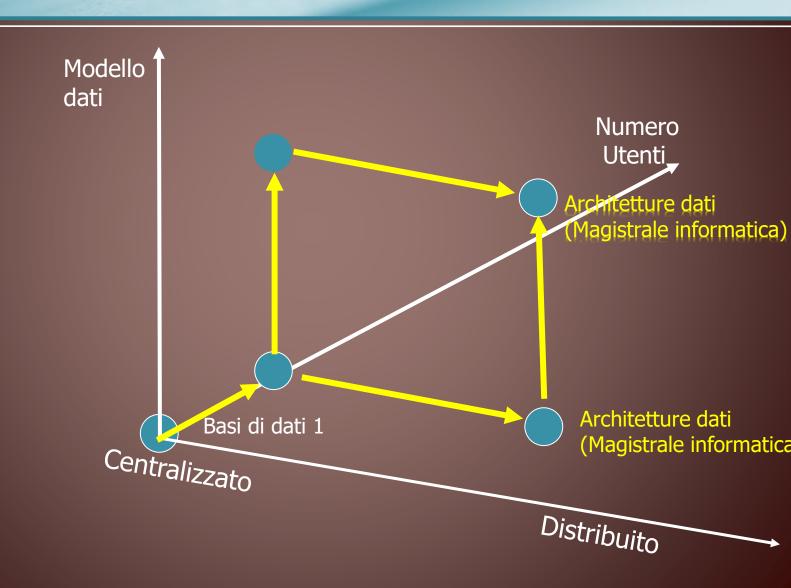
Architetture dati



Architetture dati



Architetture dati



Evolution of data technologies

Centralizzate Centralizzate

a. Gerarchiche Relazionali

b. linked

Anni 60 Anni 70

Evolution of data technologies

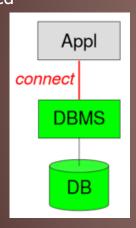
DB Machine
Parallele
Centralizzate Centralizzate

a. Gerarchiche Relazionali b. linked

Anni 60 Anni 70

Evolution of data technologies

Centralizzate Centralizzate a. Gerarchiche Relazionali b. linked

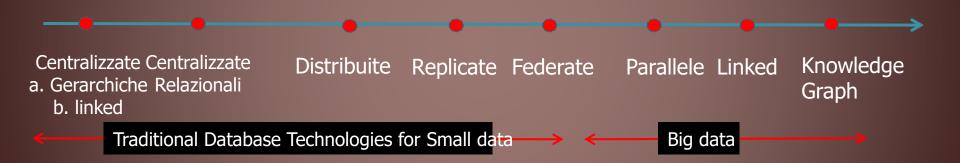


Distribuite Replicate Federate Parallele Linked Knowledge Graph

Data technologies and organization types



Data technologies and small vs big data

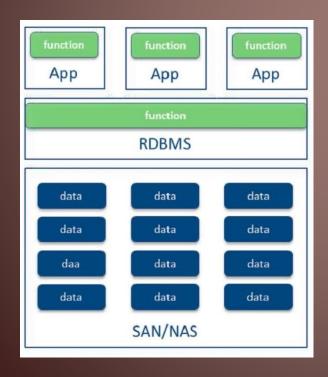


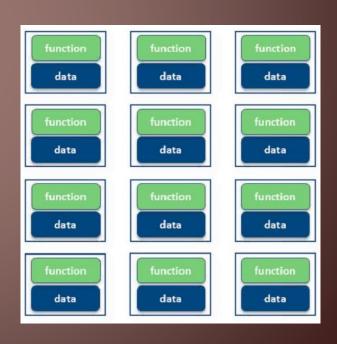
Evolution of data technologies: parallel processing

Centralizzate Centralizzate a. Gerarchiche Relazionali b. linked

Distribuite Replicate Federate Paralleleinked Kr

Knowledge Graph





Obiettivi del corso

- Vostri interessi:
 - Imparare/sapere
 - Utilizzare/Gestire
 - Progettare
- Cosa?
 - NoSQL
 - BigData
 - Architetture
 - Sistemi
 - Tipologie



Vostre curiosità

- le tecnologie di progettazione dei dati utilizzate oggi, con qualche esempio pratici
- No linguaggi/modelli obsoleti
- Spero di trattare modelli distribuiti su larga scala, quelli che oggi viviamo e mantengono i nostri dati.
- MongoDB, Grafi, RDF
- Non vorrei vedere argomenti già coperti nel corso complementi di basi di dati. Nello specifico l'aspetto relativo alle basi di dati relazionali come è stato trattato nel corso complementi di basi di dati.
 - 10/31 hanno già visto complementi basi di dati

Vostri interessi (35 votanti)

- Big data 86%
- NoSQL 69%
- Knowledge Graph 57%
- Blockchain 54%
- Qualità dei dati 51%
- Principi di progettazione di basi di dati distribuite 46%
- Linked data 40%
- Db Distribuiti 37%
- Data warehouse 31%
- Integrazione dati 20%

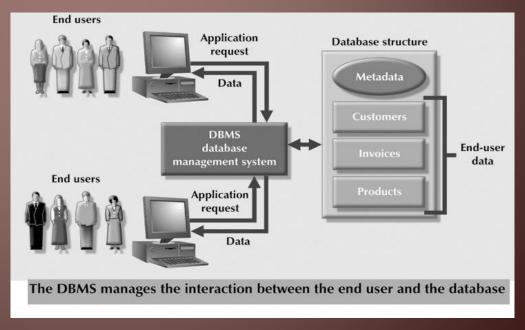
Argomenti del corso

- Richiami sulle tecnologie delle basi di dati relazionali centralizzate
- Tecnologie delle basi di dati relazionali distribuite
- Blockchain
- Progettazione delle basi di dati relazionali distribuiti
- NoSQL
- Integrazione dati
- Qualità dati
- Big data

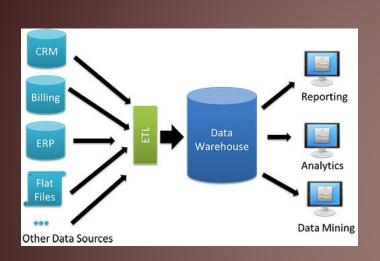
Architetture: obiettivi

Utente/programmi Schema logico Schema fisico

- Database di tipo transazionale
- OLTP



Architetture: obiettivi



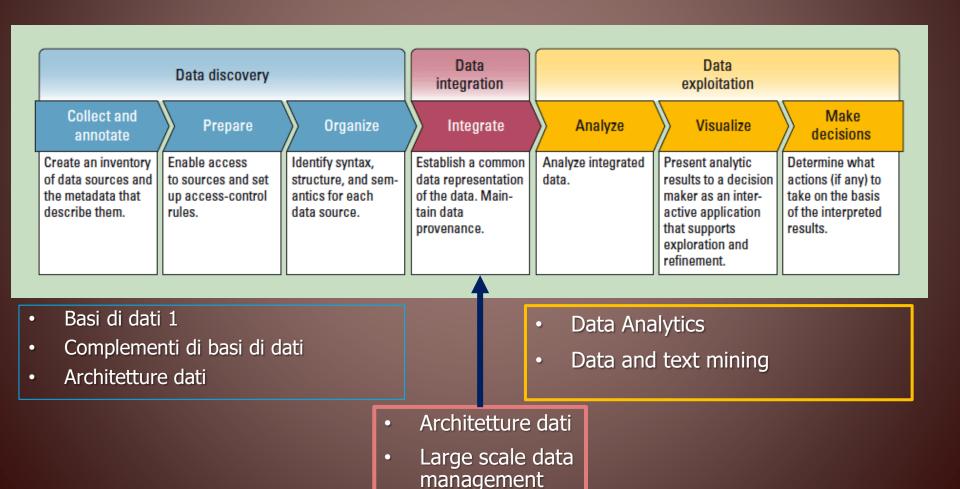
Datawarehouse



- Database di tipo analitico
- OLAP

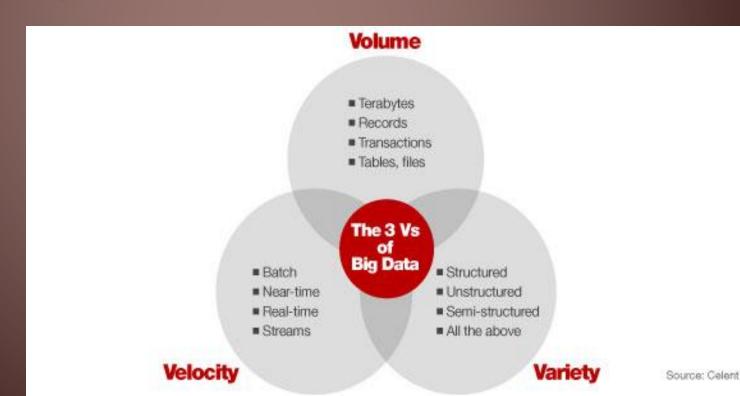
Data Value chain

https://www.computer.org/csdl/mags/it/2013/01/mit2013010057-abs.html



Big Data

- 2012 Gartner
- "Big data is high volume, high velocity, and/or high variety information assets"



Altra definizione informale



captures

1 TB OF TRADE INFORMATION during each trading session

Velocity

that monitor items such as

uel level and tire pressure

ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be

18.9 BILLION NETWORK CONNECTIONS

- almost 2.5 connections per person on earth



The FOUR V's of Big **Data**

break big data into four dimensions: Volume, Velocity, Variety and Veracity

4.4 MILLION IT JOBS



As of 2011, the global size of data in healthcare was estimated to be

[161 BILLION GIGABYTES]



DIFFERENT

Variety FORMS OF DATA

By 2014, it's anticipated there will be 420 MILLION WEARABLE, WIRELESS

HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO

are watched on YouTube each month



30 BILLION PIECES OF CONTENT

are shared on Facebook



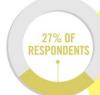




million monthly active users

1 IN 3 BUSINESS

don't trust the information they use to make decisions



in one survey were unsure of how much of their data was inaccurate



Poor data quality costs the US economy around



Veracity UNCERTAINTY

OF DATA



L'insegnamento

- I sistemi di gestione di basi di dati nei sistemi informativi tradizionali,
 - evoluzione dai sistemi centralizzati verso i sistemi distribuiti e federati.
- data governance,
 - integrazione dati, fusione dati e assessment di qualità dei dati.
- tecnologie di gestione per big data,
 - NoSQL e architetture big data

Riferimenti

- Slides
- Next Generation Databases: NoSQL, NewSQL, and Big Data by Guy Harrison Publisher: Apress Release Date: January 2016 ISBN 9781484213292
- Basi di Dati (Atzeni, Paraboschi, Torlone,) parte
 3
- Lemahieu, Broucke, Baesens; Principles of database management, Cambridge university press 2018

Comunicazioni

 La distribuzione del materiale didattico e le comunicazioni avverranno tramite il sito del corso, accessibile tramite Moodle http://elearning.unimib.it/

• E' strettamente richiesta l'iscrizione al corso.

Lezioni / Esercitazioni

- Lezioni / Esercitazioni :
 - Mercoledi 13:30 15:30 \rightarrow 14.00-15.30 streaming
 - Giovedì 8:30 10:30 → video lezioni registrate
 - Venerdì 8:30 −10:30 → video lezioni registrate
 - Test di autoverifica (non valido per il superamento dell'esame)
 - Risposte a domande chiuse
 - Risposte a domande aperte da discutere il mercoledì successivo

Modalità d'esame

Scritto

- Domande a risposte aperte sulla teoria/esercizi di progettazione di architetture o sistemi di gestione dati/esercizi sui linguaggi di interrogazioni NoSQL
- Prova orale (opzionale)
 - Eventuali domande sui contenuti comuni del corso.
- In alternativa è possibile realizzare un elaborato che approfondisca un particolare tema mostrato a lezione
 - Da concordare con il docente
- Architettura dati equivale al vecchio data technology