

Business Intelligence per i Big Data



Tania Cerquitelli

AA. 2019-2020

1



Docenti

- **Docente: Prof. Tania Cerquitelli**
Dip. Automatica e Informatica
Tel: 011 090-7178
e-mail: tania.cerquitelli@polito.it
- **Esercitatore:**
 - Giuseppe Attanasio
Dip. Automatica e Informatica
e-mail: giuseppe.attanasio@polito.it
- **Ricevimento: su appuntamento e in remoto**



2

2



Organizzazione del corso

- Inizio lezioni: 09/03/2020
- Fine lezioni: 13/06/2020

- Orario delle lezioni ed esercitazioni

Giorno	Orario	Aula
Lunedì	11:30-14:30	Virtual Classroom
Giovedì	11:30-14:30	Virtual Classroom



Obiettivi

- Business Intelligence: attività di analisi dei dati per il supporto delle decisioni aziendali
 - Data analytics
 - Data warehousing and visualization
 - No-SQL databases



Obiettivi

- Business Intelligence: attività di analisi dei dati per il supporto delle decisioni aziendali
- Studio delle basi di dati destinate all'analisi dei dati (*data warehouse*)
 - Architettura dei sistemi di data warehouse e delle metodologie di progettazione e sviluppo
 - Introduzione alle basi di dati non relazionali (no-relational databases) a supporto dei Big Data
- Tecniche di analisi
 - OLAP (On Line Analytical Processing) dei dati
 - Tecniche ed algoritmi di data mining
- Sviluppo di casi applicativi mediante strumenti open



Lezioni

- Data analytics
 - Tecniche di preparazione dei dati
 - Tecniche di esplorazione dei dati
 - Processo di estrazione della conoscenza
 - Algoritmi di classificazione e tecniche di validazione
 - Algoritmi di clustering ed indici di qualità
 - Algoritmi di estrazione di regole di associazione ed indici di qualità
 - Analisi di serie temporali e regressione



Lezioni

- Data warehouse
 - architetture
 - progettazione e sviluppo
 - analisi OLAP (On Line Analytical Processing) dei dati
 - estensioni del linguaggio SQL per interrogare i data warehouse
- Basi di dati NO-SQL
 - Tipi di basi di dati NO-SQL
 - Concetti di base
 - Esempi di interrogazioni



Lezioni: casi di studio

- Studio di casi applicativi di ampio interesse scientifico
 - Analisi di requisiti per una corretta progettazione
 - Processo di estrazione della conoscenza su dati open
 - Processo di estrazione della conoscenza nel contesto Industry 4.0
 - Progettazione di data warehouse
 - interrogazioni di sistemi no relazionali



Esercitazioni in aula

- Discussione di come progettare un workflow di data analytics
- Progettazione concettuale, logica e fisica di un data warehouse
- Interrogazione di un data warehouse mediante il linguaggio SQL esteso
- Interrogazione di un sistema NO-SQL mediante paradigma Map-Reduce



Esercitazioni in laboratorio

- Applicazioni di tecniche di data analytics
 - Tool open source (i.e., RapidMiner)
- Interrogazione di una base di dati relazionale mediante il linguaggio SQL e le sue estensioni
 - DBMS: Oracle
- Progettazione di un data warehouse e analisi OLAP
 - Tool open source (i.e., Pentaho) che si interfaccia con il DBMS Oracle
- Progettazione di un sistema NO-SQL e analisi
 - Tool open source: Mongo DB



Esame

- Progetto
 - Utilizzo di una tecnica di data mining per analizzare un dataset reale mediante RapidMiner o altri tools di data analytics noti allo studente
- Prova scritta (durata: 2 ore)
 - Progettazione concettuale, logica e fisica di un data warehouse
 - Interrogazione del data warehouse progettato mediante il linguaggio SQL
- Consegna di 2 esercitazioni di laboratorio
 - Uso di algoritmi di classificazione in un caso di studio
 - Esercitazione con un sistema di basi di dati no relazionale



11

11



Materiale

- Sito web del corso
 - <http://dbdmq.polito.it/wordpress/teaching/business-intelligence/>
 - Copia dei lucidi utilizzati a lezione
 - Testi e risoluzioni di esercizi proposti in aula
 - Testi e materiale utile per le esercitazioni in laboratorio
 - Materiale utile per la risoluzione dei progetti
 - Materiale integrativo
 - Modello relazionale e linguaggio SQL



12

12



Testi

- Data warehousing
 - Golfarelli, Rizzi. Data warehouse: teoria e pratica della progettazione, McGraw-Hill 2006
 - Kimball e altri, numerosi testi su metodologia e casi di studio, Wiley
- Data mining
 - Han, Kamber. Data mining: concepts and techniques, Morgan Kaufmann 2006
 - Tan, Steinbach, Kumar. Introduction to data mining, Pearson 2006
- Basi di dati no relazionali
 - K. Chodorow and M. Dirolf. MongoDB: the definitive guide. O'Reilly Media, 2010



13

13



Link utili

- Data warehouse
 - <http://www.dwinfocenter.org>
 - <http://www.dwreview.com>
 - <http://kimballuniversity.com>
- Data mining
 - <http://www.kdnuggets.com>
- Tool open source
 - Pentaho: <http://www.pentaho.com/>
 - RapidMiner: <http://rapid-i.com/>
 - Mongo-DB: <https://www.mongodb.com/it>



14

14