

Business Intelligence per Big Data



Data Base and Data Mining Group of Politecnico di Torino

Tania Cerquitelli

AA. 2022-2023



Docenti

- **Docente: Prof. Tania Cerquitelli**
Dip. Automatica e Informatica
Tel: 011 090-7178
e-mail: tania.cerquitelli@polito.it
- **Esercitatore: Daniele Rege Cambrin**
Dip. Automatica e Informatica
e-mail: daniele.regecambrin@polito.it
- **Esercitatrice: Eliana Pastor**
Dip. Automatica e Informatica
e-mail: eliana.pastor@polito.it



Organizzazione del corso

- Inizio lezioni: 27/02/2023
- Fine lezioni: 09/06/2023

- Orario delle lezioni ed esercitazioni

Giorno	Orario	Aula
Lunedì	13:00-16:00	Aula 4P
Giovedì	13:00-16:00	Aula R3



Obiettivi

- Business Intelligence: attività di gestione e analisi dei dati per il supporto delle decisioni aziendali
 - Data analytics
 - Data warehousing
 - Dashboard informative e tecniche di visualizzazione
 - No-SQL databases



Obiettivi

- Business Intelligence: attività di gestione e analisi dei dati per il supporto delle decisioni aziendali
- Tecniche di analisi di dati
 - Tecniche ed algoritmi di data science e machine learning per trasformare la conoscenza a supporto del business
 - Tecniche di visualizzazione
- Studio delle basi di dati per il supporto decisionale (*data warehouse*)
 - Architettura dei sistemi di data warehouse e delle metodologie di progettazione e sviluppo
 - OLAP (On Line Analytical Processing) dei dati
 - Introduzione alle basi di dati non relazionali (no-relational databases) a supporto dei Big Data
- Sviluppo di casi applicativi mediante strumenti di data science e data warehouse



Lezioni

- Business Intelligence: attività di gestione e analisi dei dati per il supporto delle decisioni aziendali
- Data analytics
 - Tecniche di preparazione dei dati
 - Tecniche di esplorazione dei dati
 - Processo di estrazione della conoscenza (*data science pipeline*)
 - Algoritmi di clustering ed indici di qualità
 - Algoritmi di estrazione di regole di associazione ed indici di qualità
 - Analisi di serie temporali e regressione
 - Algoritmi di classificazione e tecniche di validazione
 - Introduzione agli algoritmi di deep learning



Lezioni

- Data warehouse
 - architetture
 - progettazione e sviluppo
 - analisi OLAP (On Line Analytical Processing) dei dati
 - estensioni del linguaggio SQL per interrogare i data warehouse
 - Dashboard informative (basate su *KPIs – Key Performance Indicators*)
 - Tecniche di data visualization
- 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
 - Processo di estrazione della conoscenza mediante pipeline di data science
 - Dai dati alla conoscenza per trasformare il business (data-driven strategies & actions)
 - Presentazione di progetti di ricerca di data science
 - Progettazione di data warehouse
 - Analisi di requisiti per una corretta progettazione
 - Dashboard informative e infografiche
 - Interrogazioni di sistemi no relazionali



Esercitazioni in aula

- Discussione di come progettare un workflow di data science
- 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 di laboratorio in Aula

- 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
 - Google Data Studio
- Progettazione di un sistema NO-SQL e analisi
 - Tool open source: Mongo DB



Esame

- Progetto
 - Caso di studio da analizzare in gruppo (5 persone)
 - Consegna di una relazione scritta (con descrizione del contributo individuale)
 - Compilazione di 1-2 questionari
- Prova scritta
 - Progettazione concettuale, logica e fisica di un data warehouse
 - Interrogazione del data warehouse progettato mediante il linguaggio SQL e definizione di viste materializzate
 - 3 domande teoria/esercizi su data analytics
- Consegna di 2 homework
 - Data science
 - Basi di dati no relazionale



Materiale

- Sito web del corso

- https://dbdmg.polito.it/dbdmg_web/index.php/2023/02/22/business-intelligence-per-big-data/
- 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



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

DBMG guide. O'Reilly Media, 2010



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>