

INGEGNERIA INFORMATICA (LM75)

(Lecce - Università degli Studi)

Insegnamento **GESTIONE DI BIG DATA**

GenCod A006801

Docente titolare MARCO SALVATORE ZAPPATORE

Docenti responsabili dell'erogazione ANTONELLA LONGO, MARCO SALVATORE ZAPPATORE

Insegnamento GESTIONE DI BIG DATA

Anno di corso 1

Insegnamento in inglese BIG DATA MANAGEMENT

Lingua ITALIANO

Settore disciplinare ING-INF/05

Percorso PERCORSO COMUNE

Corso di studi di riferimento INGEGNERIA INFORMATICA

Tipo corso di studi Laurea Magistrale

Sede Lecce

Crediti 6.0

Periodo Primo Semestre

Ripartizione oraria Ore Attività frontale: 54.0

Tipo esame Orale

Per immatricolati nel 2022/2023

Valutazione Voto Finale

Erogato nel 2022/2023

Orario dell'insegnamento

<https://easyroom.unisalento.it/Orario>

BREVE DESCRIZIONE DEL CORSO

Il corso ha l'obiettivo di fornire agli studenti la possibilità di apprendere i concetti e i paradigmi principali della generazione, gestione e archiviazione di big data, con particolare riferimento ad architetture, modelli e applicazioni.

PREREQUISITI

Buona conoscenza di linguaggi orientati agli oggetti (almeno 1), tecniche e strumenti. Elementi di reti informatiche e tecnologie per il Web.

OBIETTIVI FORMATIVI

- Comprendere i principi alla base della gestione dei Big Data, della loro governance e della loro generazione
 - Conoscere i modelli dati e le architetture per i Big Data
 - Apprendere i criteri alla base della Big Data Quality
 - Conoscere gli approcci di analisi esplorativa dei Big Data
 - Comprendere definizioni, caratteristiche e differenze di Data Warehouse, Data Lake e Data Lakehouse
 - Conoscere i sistemi e le piattaforme a supporto della gestione dei Big Data
 - Conoscere i DBMS NoSQL e NewSQL

METODI DIDATTICI

Le lezioni frontali mirano a migliorare la conoscenza e la comprensione degli studenti attraverso la presentazione di teorie, modelli e metodi; gli studenti sono invitati a partecipare alla lezione con autonomia di giudizio, ponendo domande e presentando esempi. Le esercitazioni di laboratorio sono finalizzate all'utilizzo di strumenti che supportano i modelli e gli approcci presentati.

MODALITA' D'ESAME

1. Prova di ammissione al progetto (prova scritta obbligatoria)
2. Progetto (prova obbligatoria che consta di implementazione e successiva discussione orale, su argomento concordato con i docenti)
3. Data hackathon (prova opzionale implementazione e successiva discussione orale)

PROGRAMMA ESTESO

- Presentazione del corso
- Database relazionali ed SQL
- Introduzione alla gestione dei Big Data
- Modelli dati, Big Data Quality
- Exploratory Data Analysis (EDA)
- Architetture per Big Data
- Distributed database
- Data Warehouse
- Data Lake, Data Lakehouse
- Introduzione alla virtualizzazione, Map Reduce
- Apache Flask, Apache Hop
- PostgreSQL
- Introduzione alla teoria dei grafi
- NoSQL: MongoDB, Hbase, Neo4j

TESTI DI RIFERIMENTO

- R. Elmasri, S. Navathe, Fundamental of Database Systems, 7th Ed., Pearson
- M. Golfarelli, S. Rizzi, Data Warehouse Design, Mc Graw Hill
- Balamurugan Balusamy, Nandhini Abirami R, Amir H. Gandomi, Big Data: Concepts, Technology, and Architecture, John Wiley & Sons Inc.; 1st Ed.
- Materiale fornito durante il corso