

Course: **Big Data**

Speaker: **Prof. Alfredo Cuzzocrea**

(16 hours)

Syllabus (ITA)

1. Big Data: Principi, Caratteristiche, Fondamenti. Introduzione ai big data. Principi dei big data. Caratteristiche dei big data. Evoluzione dei big data. Big data analytics. Big data technology. Cloud computing. Applicazioni dei big data.

2. Hadoop. Introduzione ad Apache Hadoop. L'ecosistema Hadoop. Vantaggi di Hadoop. Architettura di Hadoop. Hadoop Distributed File System (HDFS). Distribuzione di processi Hadoop su cluster. Rappresentazione e memorizzazione dei dati con HDFS. Gestione di dati relazionale su Hadoop. Impala e Hive.

3. MapReduce. Il paradigma di programmazione MapReduce. Esempi di MapReduce job. Partizioni e combinazioni in MapReduce. Implementazioni di MapReduce. Applicazioni basate su MapReduce.

4. Spark. Apache Spark. Resilient Distributed Datasets (RDDs). Programmazione funzionale in Spark. Applicazioni RDDs con Spark. Operazioni su RDDs. Programmazione parallela con Spark.

5. OLAP. OLAP: principi, definizioni e metodologia. Modelli dati multidimensionali. Analisi OLAP. Tecnologie OLAP. Progettazione di data cube OLAP. Calcolo di data cube OLAP. Aggregazione di data cube OLAP. Interrogazione di data cube OLAP. OLAP mining. Scoperta di conoscenza e analisi multidimensionale da data cube OLAP.

Syllabus (EN)

1. Big Data: Principles, Features, Fundamentals. Introduction to big data. Big-data's features. Big-data evolution. Big data analytics. Big data technology. Cloud computing. Big data applications.

2. Hadoop. Introduction to Apache Hadoop. The Hadoop ecosystem. Advantages of Hadoop. Hadoop architecture. Hadoop Distributed File System (HDFS). Hadoop process distribution over clusters. HDFS-based data representation and storage. Relational data management in Hadoop. Impala and Hive.

3. MapReduce. MapReduce programming paradigm. MapReduce job examples. Partitions and combinations in MapReduce. MapReduce implementations. MapReduce-based applications.

4. Spark. Apache Spark. Resilient Distributed Datasets (RDDs). Spark functional programming. RDDs-based applications with Spark. Operations over RDDs. Parallel programming with Spark.

5. OLAP. OLAP: principles, definitions and methodologies. Multidimensional data models. OLAP analysis. OLAP technology. OLAP data cube design. Computing OLAP data cubes. Aggregating OLAP data cubes. Querying OLAP data cubes. OLAP mining. Knowledge discovery and multidimensional analysis over OLAP data cubes.