

Big Data course

Period: 2nd semester 2015-2016

Lecturer: Alessandro Rezzani

Syllabus of the course

Lecture	Topics
1	Big Data introduction <ul style="list-style-type: none">- Big data: definition and taxonomy- Big data value for the enterprise- Setting up the demo environment- First steps with the Hadoop “ecosystem” <p>Exercises</p>
2	The Hadoop ecosystem <ul style="list-style-type: none">- Introduction to Hadoop- Hadoop components: MapReduce/Pig/Hive/HBase- Loading data into Hadoop- Handling files in Hadoop- Getting data from Hadoop <p>Exercises</p>
3	Querying big data with Hive <ul style="list-style-type: none">- Introduction to the SQL Language- From SQL to HiveQL <p>Exercises</p>
4	Querying big data with Hive <ul style="list-style-type: none">- Introduction to HIVE e HIVEQL- Using Hive to query Hadoop files <p>Exercises</p>

5 Big data & Machine learning

- Quick intro to Machine learning
- Big Data & Machine Learning
- Machine learning tools
 - Spark & SparkML
 - H2O
 - Azure ML

Exercises

6 Big data & Machine learning

- Big Data & Machine Learning (continued)
- Next steps in the big data world

Exercises

- A case study
-

Software used

Apache Hadoop

Suggested bibliography

Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati, A. Rezzani, Apogeo Education, 2013

Hadoop For Dummies, Dirk deRoos, For Dummies, 2014