## Big Data course

**Period:** 2\(^{nd}\) semester 2015-2016  
**Lecturer:** Alessandro Rezzani

### Syllabus of the course

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Topics</th>
</tr>
</thead>
</table>
| 1 | **Big Data introduction**  
- Big data: definition and taxonomy  
- Big data value for the enterprise  
- Setting up the demo environment  
- First steps with the Hadoop “ecosystem”  
  
  **Exercises** |
| 2 | **The Hadoop ecosystem**  
- Introduction to Hadoop  
- Hadoop components: MapReduce/Pig/Hive/HBase  
- Loading data into Hadoop  
- Handling files in Hadoop  
- Getting data from Hadoop  
  
  **Exercises** |
| 3 | **Querying big data with Hive**  
- Introduction to the SQL Language  
- From SQL to HiveQL  
  
  **Exercises** |
| 4 | **Querying big data with Hive**  
- Introduction to HIVE e HIVEQL  
- Using Hive to query Hadoop files  
  
  **Exercises** |
5 Big data & Machine learning
- Quick into 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