

Sistemi e Architetture per Big Data

A.A. 2017/18

Valeria Cardellini, Matteo Nardelli

Laurea Magistrale in
Ingegneria Informatica - Il anno

Staff

- Valeria Cardellini
 - Tel: 06 72597388, office: Ing. Informazione, room D1-17
 - Email: cardellini@ing.uniroma2.it
 - www.ce.uniroma2.it/~valeria
- Matteo Nardelli
 - Office: Ing. Informazione, room D1-19
 - Email: nardelli@ing.uniroma2.it
 - www.ce.uniroma2.it/~nardelli
- Email: use [SABD] in the subject line
- Office hours:
 - When: Tuesday 9:00-11:00
 - Where: room D1-17

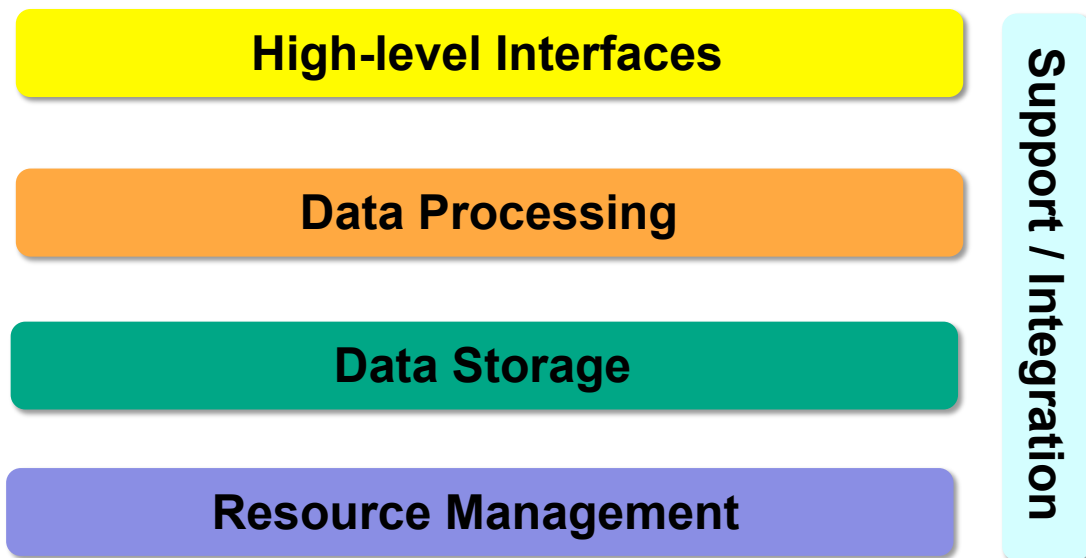
General information

- Web site of the course
www.ce.uniroma2.it/courses/sabd1718/
- Number of credits: 6 CFU
 - 60 hours of lessons (each lesson of 105 minutes)
- Class period: 2nd semester
 - From 8/3/2018 to 7/6/2018
- Class schedule
 - Monday 9:30-11:15, room B12
 - Thursday 11:30-13:15, room B12
- Register to the course through Delphi

Educational objectives

- Principles, paradigms, tools and technologies to design and manage distributed **systems** and **architectures** for **big data analytics** services and applications

The Big Data stack we will consider



Course program at-a-glance

- Frameworks for **resource management**
- Tools and frameworks for collecting and ingesting data from various sources into the big data analytics infrastructure
- Systems and frameworks for storing data either temporary or permanently, including distributed file systems and non-relational (NoSQL) databases for **data storage**
- **Processing** frameworks for batch and real-time analytics, including architectural and programming aspects of these frameworks

Course program in details

- Introduction to Big Data: issues and challenges
- Data storage: distributed file systems and NoSQL databases
 - Case studies: HDFS, HBase, MongoDB, DynamoDB, Neo4j
 - Lab: HDFS and NoSQL databases
- Systems for data acquisition
 - Pub/sub, message queues, collection systems
 - Lab: Kafka
- Systems for batch processing
 - Case studies: Hadoop, Pig, Hive, Spark
 - Batch processing in the Cloud
 - Lab: Hadoop and Spark

Course program in details (2)

- Systems for stream processing
 - Case studies: Spark Streaming, Storm, Flink, Heron
 - Stream processing in the Cloud
 - Lab: Storm and Spark Streaming
- Frameworks for cluster resource management
 - Case studies: Mesos, YARN, Kubernetes
- Future reference infrastructure: Fog computing

Teaching material

- Your notes
- Lesson slides on the course web site (after the lesson!)
- Scientific papers, articles, etc. on the course web site
- Suggested textbooks:



- A. Bahga, V. Madisetti, [Big Data Science and Analytics: A Hands-On Approach](#), VPT, 2016.



- R. Buyya, R. Calheiros, A. V. Dastjerdi (eds.), [Big Data: Principles and Paradigms](#), 1st Edition, Morgan Kaufmann, 2016.

Exam

- a) 2 programming projects assigned during the course
 - [Programming project #1](#): assigned at the end of April 2018, due at the end of May 2018
 - [Programming project #2](#): assigned at the end of May 2018, due at the end of June 2018
 - Students in groups of 2
- b) [Final oral exam](#) on the entire course program
 - When:
 - 2 dates in the period 18/6-28/7 2018
 - 2 dates in the period 28/8-22/9 2018
 - 2 dates in January-February 2019

Grading

- Programming project #1: 25%
- Programming project #2: 25%
- Final oral exam: 50%

- Participation during class will also be taken into account