

# Sistemi e Architetture per Big Data

## A.A. 2016/17

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](mailto:cardellini@ing.uniroma2.it)
  - [www.ce.uniroma2.it/~valeria](http://www.ce.uniroma2.it/~valeria)
- Matteo Nardelli
  - Office: Ing. Informazione, room D1-19
  - Email: [nardelli@ing.uniroma2.it](mailto:nardelli@ing.uniroma2.it)
  - [www.ce.uniroma2.it/~nardelli](http://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/sabd1617/](http://www.ce.uniroma2.it/courses/sabd1617/)
- Number of credits: 6 CFU
  - 60 hours of lessons (4 hours per week)
- Class period: 2° semester
  - From 6/3/2017 to 22/6/2017
- Class schedule
  - Monday 9:30-11:15, room B12
  - Thursday 9:30-11:15, room C2
- Register to the course using Delphi (or send an email)



Valeria Cardellini - SABD 2016/17

2

## Educational objectives

---

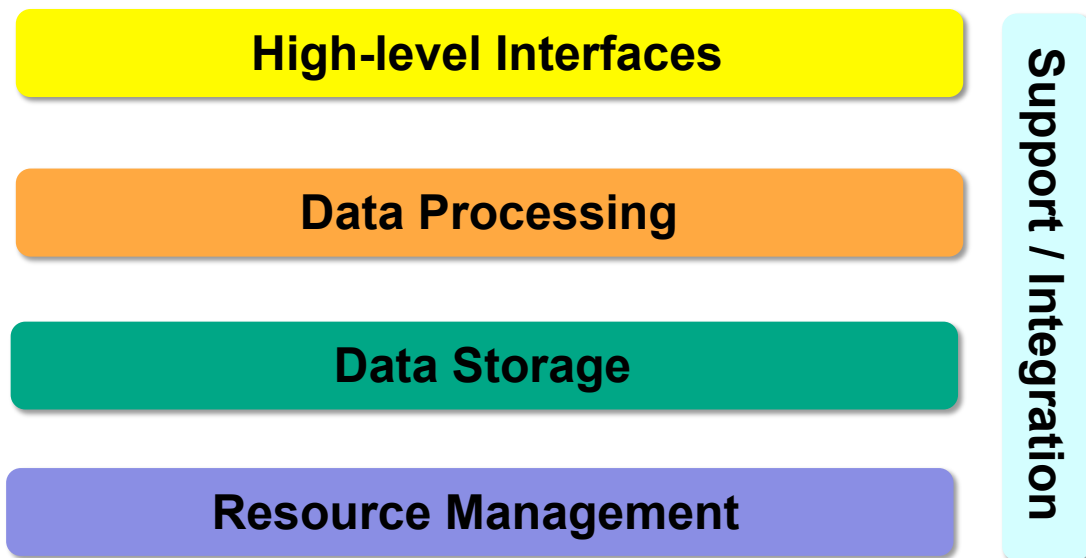
- Principles, paradigms, tools and technologies to design and manage distributed **systems** and **architectures** for **big data analytics** services and applications

Valeria Cardellini - SABD 2016/17

3

## 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, incorporating 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
- Future reference infrastructure: Fog computing
- Frameworks for cluster resource management
  - Case studies: Mesos, Kubernetes
- 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
- Systems for batch processing
  - Case studies: Hadoop, Pig, Hive, Spark
  - Batch processing in the Cloud
  - Lab: Spark

## Course program in details (2)

---

- Systems for stream processing
  - Case studies: Spark Streaming, Storm, Flink, Heron
  - Stream processing in the Cloud
  - Lab: Storm

## 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 beginning of May 2017, due at the end of May 2017
  - [Programming project #2](#): assigned at the beginning of June 2017, due at the end of June 2017
  - Students in groups of 2
- b) [Final oral exam](#) on the entire course program
  - When:
    - 2 dates in the period 26/6-28/7 2017
    - 2 dates in the period 4/9-30/9 2017
    - 2 dates in the period in February 2018

# Grading

---

- Programming project #1: 25%
- Programming project #2: 25%
- Final oral exam: 50%
  
- Participation during class will also be taken into account