



Lecture Presentation

CSC5003

The Teaching Team



Amel Bouzeghoub
4A 308 (Palaiseau)

amel.bouzeghoub [at] telecom-sudparis.eu



Bruno Defude
DIR 221 (Evry)

bruno.defude [at] telecom-sudparis.eu



Julien Romero
D 307 (Evry)

julien.romero [at] telecom-sudparis.eu

Goals of the First Part of the Lecture

- Understand the problematics of big data
- Acquire experience with big data tools, in particular the Hadoop ecosystem
- Find and present relevant tools for your application
- Learn to process big data in the form of massive datasets or data streaming

Evaluation

The evaluation of this module will combine:

- The grade from the written exam (9 points)
- The grade from the short presentation (1 point)
- The grade from the final projects (5 + 5 = 10 points)

Short Presentations

You will be given a tool from the Hadoop ecosystem that you will have to present to the class in 5 minutes. In details, you need to:

- Describe its purpose, i.e. when it should be used
- Show examples of applications
- Show an short example of how it works in practice (a short demo)
- Some tools work without Hadoop. Give examples when Hadoop is involved.

List Short Presentations

- HBase - Lancelot
- Hive - Matthieu
- Pig -
- Apache ORC - Louis
- Avro - Corentin
- Solr -
- Lucene - Chenwei
- Flink - Aziz
- Ambari -
- Cassandra -
- ZooKeeper -
- Impala - Charles
- Storm - Paul G.
- Parquet - Sabine
- Kudu - Paul B.
- Beam - Sri Raam Appakutti Palani

Written Exam

In this exam, we will evaluate that you know:

- The basic concepts presented in this lecture
- What each tool of the short presentations does

Duration: 2h

Final Big Data Project

Find a real world dataset and use the relevant tools we learned about during this lecture to exploit it.

In group of 2-3 people.

Final presentation in front of the class.

More about this later.