INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

DPT INFORMATIQUE
COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
INSA Campus LyonTech - 7 Avenue Jean Capelle
Batiment Blaise Pascal - 69621 VILLEURBANNE
Phone 0472438892
Web site : http://if.insa-lyon.fr

Conferences and Seminars

Cloud Computing for Distributed Big Data Applications

IDENTIFICATION

CODE :	IF-5-S1-EC-OT6
ECTS :	6.0

HOURS

Lectures :	0.0 h
Seminars :	64.0 h
Laboratory :	0.0 h
Project :	0.0 h
Teacher-student contact :	64.0 h
Personal work :	64.0 h
Total :	128.0 h

ASSESSMENT METHOD

- Active participation

TEACHING AIDS

- Lecture slides

TEACHING LANGUAGE

English

CONTACT

M. BELAL Yacine
yacine.belal@insa-lyon.fr
MME BEN MOKHTAR Sonia
sonia.benmokhtar@insa-lyon.fr
MME BOUCHENAK KHELLADI Sahra
sara.bouchenak@insa-lyon.fr
M. EL HATTAB Fatma
@

AIMS

The objective of this course is to present the paradigms, algorithms and mechanisms underlying distributed computing systems and cloud computing for building large-scale big data applications.

Its content consists of the following three modules:
- Project « Large-scale cloud services for big data management »
- Project « Mobile and distributed social networks »
- Seminar « Distributed systems and algorithms »

CONTENT

The three modules are organized as follows.

Project « Large-scale cloud services for big data management »
- 100% practical work
- Deploy distributed big data analytics and learning framework on a real distributed infrastructure (e.g., Apache Spark / Spark ML, cloud instances, Grid 5000 instances, etc.)
- Deploy real datasets
- Evaluate scalability with regard to data, with regard to distributed infrastructure

Project « Mobile and distributed social networks »
- 100% practical work
- Use real user mobility datasets
- Two use cases to implement: novel useful mobility data-based applications, inference of sensitive information from anonymized mobility data

Seminar « Distributed systems and algorithms »
- Third of time dedicated to technological case studies (e.g., Apache Spark, Spark ML lib, Hadoop MapReduce, Hadoop Yarn, etc.)
- Third of time dedicated to research& innovation solutions (e.g., research papers on « PCC Vivace: Online-Learning Congestion Control » in NSDI 2018, « Improving spark application throughput via
memory aware task co-location » in Middleware 2017, etc.)
- Third of time dedicated to lectures on distributed algorithms(e.g., fault tolerance algorithms, etc.)

BIBLIOGRAPHY

- George Coulouris, Jean Dollimore, Tim Kindberg. Distributed Systems: Concepts and Design (4th Edition).
Addison Wesley, 2005.
- Consensus: How to enable a set of processes to agree on a given value (in presence of failures)?
- Andrew S. Tanenbaum, Maarten van Steen. Distributed Systems: Principles and Paradigms (2nd Edition). Prentice Hall, 2006.
- Nancy A. Lynch. Distributed Algorithms. Morgan Kaufmann Publishers, 2001.

PRE-REQUISITE

Programming skills