

EIIN943	Large Scale Distributed Systems	CM 12h	TD 14h	HNE 24h
---------	---------------------------------	-----------	-----------	------------

Cours proposé dans la mineure / Course offered in the minor :

AL	CyberSec	IA-ID	IHM	IoT-CPS	Ubinet	IF	M1 EIT DSC	M2 EIT DSC	M2 Fintech
x					x	x	x	x	x

Responsable / In charge of : **Huet Fabrice** (Fabrice.HUET@univ-cotedazur.fr)

Résumé / Abstract :

Large-scale distributed infrastructures leverage high-performance networks to federate distributed resources, whether they are co-located or geographically dispersed. The purpose of this course is to provide an in-depth overview of such architectures, with a specific focus on their underlying motivation and the various layers involved, spanning from hardware to software. Throughout the lectures, we will primarily concentrate on large-scale data processing as a key exemplar. Additionally, the lab sessions will offer hands-on experience in deploying and utilizing real-world systems.

Prérequis / Prerequisite :

Knowledge of Java and/or Python and basic knowledge of Command Line Interface (Unix) is required. A laptop with VM support is required for the lab sessions.

Objectifs / Objectives :

- Become familiar with large-scale distributed infrastructures
- Learn distributed computing principles and underlying technologies
- Design performing distributed applications
- Be alert to emerging technologies and research trends

Contenu / Contents :

- Principles behind distributed systems
 - o Is many better than one ?
 - o CAP Theorem
 - o Latency and Scalability
- Storing large data
 - o From RAID to distributed File Systems
 - o Introduction to NoSQL
 - o Data consistency
 - o Case studies : Facebook's Cassandra, HDFS
- Processing Big Data
 - o Map-Reduce
 - o Case studies : Hadoop, Spark
- Managing Fast Data
 - o Data streams, real-time processing
 - o Case studies : Spark Streaming, Apache Flink, Apache Kafka

Références / References :

- Ali Davoudian, Liu Chen, and Mengchi Liu. 2018. A Survey on NoSQL Stores. ACM Comput. Surv. 51, 2, Article 40 (March 2019), 43 pages. <https://doi.org/10.1145/3158661>
- Jeffrey Dean and Sanjay Ghemawat. 2008. MapReduce: simplified data processing on large clusters. Commun. ACM 51, 1 (January 2008), 107–113. <https://doi.org/10.1145/1327452.1327492>
- Navigating the 8 fallacies of distributed computing, <https://ably.com/blog/8-fallacies-of-distributed-computing>

Acquis / Knowledge :

Evaluation / Assessment :

Lab session exercises, group project with presentation at the end of the course.