

# UE Large Scale Processing @ ESIGELEC

## 2019/2020

Abdel Dadouche  
DJZ Consulting

[adadouche@hotmail.com](mailto:adadouche@hotmail.com)  
@adadouche

# Quand allons nous passer du temps ensemble?

- 2019-11-04 – 08:30 – 10:30 (2h) - Cours
- 2019-11-04 – 10:30 – 12:30 (2h) - TP
- 2019-11-04 – 13:30 – 15:30 (2h) - Cours
- 2019-11-04 – 15:30 – 17:30 (2h) – TP
- 2019-11-08 – 08:30 – 10:30 (2h) - Cours
- 2019-11-08 – 10:30 – 12:30 (2h) – TP
- 2019-11-18 – 13:30 – 15:30 (2h) - Cours
- 2019-11-18 – 15:30 – 17:30 (2h) – TP
- 2019-12-02 – 13:30 – 15:30 (2h) - Cours
- 2019-12-02 – 15:30 – 17:30 (2h) – TP
- 2019-12-04 – 08:30 – 10:30 (2h) - Cours
- 2019-12-04 – 10:30 – 12:30 (2h) - TP
- 2019-12-04 – 13:30 – 15:30 (2h) - Cours
- 2019-12-04 – 15:30 – 17:30 (2h) – TP
- 2019-01-06 – 08:30 – 10:30 (2h) - Cours
- 2019-01-06 – 10:30 – 12:30 (2h) – TP
- 2020-01-06 – 13:30 – 15:30 (2h) - Cours
- 2020-01-06 – 15:30 – 17:30 (2h) – TP
- 2020-01-08 – 13:30 – 15:30 (2h) - Cours
- 2020-01-08 – 15:30 – 17:30 (2h) – TP
- 2020-01-10 – 08:30 – 12:30 (4h) – TP
- 2020-01-20 – 13:30 – 17:30 (4h) – TP
- **2020-02-04 – 08:30 – 10:30 (2h) - Examen**

# Modalités de Validation

- Contrôle continu (restitution de TP, assiduité, implication et collaboration)
- Note de Projet
- Examen Final (QCM et question ouvertes)

# Petit sondage avant de (bien) commencer

- Qui a déjà suivi un cours ou travaillé avec une base de données?
- Qui a déjà suivi un cours ou travaillé sur HADOOP?
- Qui a déjà travaillé sur des « gros » volumes de données? (gros, c'est-à-dire?)
- Qui a déjà utilisé SQL? Java? Linux? Machines virtuelles?

# Qui suis-je?

Abdel Dadouche

Consultant Freelance



## Dernier Jobs:

- 2016-2019: SAP Developer Relation
- 2013-2016: SAP Predictive Go To Market and CoE
- 2010-2013: Solution Architect @ KXEN (acquired by SAP)

## Langage:

- Français & English
- Java, SQL, “predictive” & “machine learning”

## Hobby:

- Hackathon (InnoJam ESIGELEC)
- Bricolage



# Objectifs de ce cours

- "Large Scale Processing", c'est quoi et pourquoi?
- Comprendre l'évolution de la « data » dans le temps
- Définir ce qu'est et n'est pas le « Big Data »
- Un peu d'histoire autour de « Big Data » et « Hadoop »
- Comprendre le rôle de « Hadoop » dans l'évolution du « Big Data »

# Objectifs de ce cours

- Appréhender l'architecture de « Hadoop » et son évolution
  - HDFS, YARN, MapReduce, ZooKeeper, Hive, Pig, Tez, HBase
- Comprendre les distributions et mode d'installation de « Hadoop »
  - On-premise ou Cloud, Bare Metal, Cloudera, HortonWorks, MapR, AWS, Azure, GCP...
- Comprendre les cas d'utilisation de « Hadoop »
  - Archivage, Analytics, Machine Learning, Streaming, IoT



# Et dans la pratique

- Installer votre premier cluster « Hadoop »
- Interagir avec HDFS pour importer des données
- Construire votre premier programme « MapReduce »
- Construire votre premier programme « Spark »
- Explorer Hive, Pig, HBase, etc.
- *Démarrer votre projet! (après l'avoir fait valider!)*

# Pourquoi cette UE?

- Depuis quelques années, de nouveaux types de compétences sont de plus en plus recherchées:
  - Business Intelligence
    - PowerBI, Qlik, Tableau, SAP Analytics Cloud
  - Data Scientist
    - SAS, R, Python, Neural Network, Python, Sci-Kit...
  - Data Engineer
    - ETL, Data Cleansing, Management & Governance, Performances...
  - Big Data Architect & Engineer
    - Hadoop & Distributed systems
- And some more functional roles as well:
  - Master Data Manager
    - Ensure a “single version of truth” with the organization
  - Data Officer
    - Define the company overall “Data” strategy
  - Data Protection Officer
    - Guarantees that legislation and regulations are properly implemented and enforced

## Slack - Support de cours & discussion



[esigelec-ue-lsp.slack.com](https://esigelec-ue-lsp.slack.com)

# Recapitulatif

- Agenda
  - périodes creuse pour travailler sur votre projet!
- Modalité de validation
  - Pas juste un examen à la fin!
- Il y a plein de Job dans le domaine!
- Ne soyez pas timide et posez vos questions!