

# Large Scale Data Processing

Sylvain Tenier, Vincent Derrien

Département TIC - Esigelec - Dominante BDTN

Année 2016/2017 - Semestre 9

À l'issue de ce module, vous serez en mesure de :

- expliquer les différents composants d'un système de traitement Big Data,
- définir une architecture répondant à un besoin métier,
- créer un environnement « docker »,
- déployer sur un cluster de machines,
- sécuriser et monitorer le cluster
- écrire et exécuter des programmes sur le cluster

Première partie (V Derrien) Ligne de commande, VMs, Docker

12/12 Choix du sujet, état de l'art, choix des technologies.

14/12 Présentation par équipes des choix effectués.

janvier Mise en place de l'architecture, développement de l'application. Cours de restructuration à votre demande.

lundi 6/02 Soutenance et démonstration

mardi 7/02 DS

# Modalités d'évaluation

CCTP - partie 1 25% de la note

CCTP - partie 2 25% de la note

DS 50% de la note

## Sujets

- 1 Sujet Innojam
- 2 Traitement IoT : où va mon chat ?
- 3 Traitement Réseaux Sociaux : que dit-on de votre Ecole ?

## Équipes

- 4 équipes de 6
- Auto-répartition - priorité équipes innojam pour sujet 1
- Répartition des rôles, méthode de travail et planning prévisionnel pour mercredi 14/12

- Moteur de traitement Spark
  - <https://spark.apache.org/>
- Architecture batch + temps réel
  - Lambda, Kappa...
- Déploiement avec Docker
  - VM puis en natif sous Linux