

EE-IRM83B1

Technologies pour le Big Data

Système d'évaluation

CC+Examen

Crédits

2 ECTS

Liste des compétences visées : CG2, CG5, CS2, CS4, CS6, CS8.

Pré-requis: Algorithmique & structures de données (EE-IRM511), Base de données réparties (EE-IRM831).

Mots clés: Big Data, Hadoop, HDFS, MapReduce.

Objectifs de l'enseignement : Ce cours présente les différents aspects de la mouvance Big Data, en présentant son origine, ses principales applications, ses outils technologiques spécifiques et quelques méthodes statistiques et analytiques mises en oeuvre.

Contenu de l'enseignement :

1. Leçon 1 : Introduction au Big Data

- (a) Section 1 : Big Data : Faits, Intérêts, Sources et Challenges
- (b) Section 2 : Qu'est-ce que le Big Data ?
- (c) Section 3 : Problématique et termes clés
- (d) Section 4 : Forces majeurs
- (e) Section 5 : Importance et défis du Big Data
- (f) Section 6 : Big Data et décisionnel
- (g) Section 7 : Approche traditionnelle versus approche Big Data
- (h) Section 8 : Mise en place du Big Data : des chantiers de grande ampleur
 - (i) Section 9 : Enjeux stratégiques
 - (j) Section 11 : Enjeux technologiques

2. Leçon 2 : Hadoop, MapReduce et le Big Data

- (a) Section 1 : Hadoop : Un standard de gestion des données et applications distri- buées
 - i. Section 1.1 : Problématiques du calcul distribué
 - ii. Section 1.2 : La solution : Apache Hadoop Single-sign-on)
 - iii. Section 1.3 : L'écosystème Hadoop
- (b) Section 2 : HDFS : Hadoop Distributed File System
 - i. Section 2.1 : Présentation
 - ii. Section 2.2 : Architecture
 - iii. Section 2.3 : Ecriture / Lecture d'un fichier

- (c) Section 3 : MapReduce
 - i. Section 3.1 : Présentation
 - ii. Section 3.2 : Modèle de programmation
 - iii. Section 3.3 : Exemples
- 3. Leçon 3 : Programmation MapReduce sous Hadoop
 - (a) Section 1 : Architecture fonctionnelle
 - (b) Section 2 : MapReduce du point de vue du développeur Java : Les entrées / Sorties
 - (c) Section 3 : Création d'un projet Java Wordcount sous Eclipse
 - (d) Section 4 : Hadoop Streaming (Python)
- 4. Leçon 4 : Traitement de données et requêtage sous Hadoop
 - (a) Section 1 : Traitement des données
 - i. Section 1.1 : Traitement par lot : Batch processing
 - ii. Section 1.2 : Traitement par streaming : Stream processing
 - (b) Section 2 : Comment réconcilier les Big Data avec le temps réel ?
Lambda- architecture
 - i. Section 2.1 : Big Data + Temps Réel : Pour quels use-cases ?
 - ii. Section 2.2 : Le système Big Data à construire
 - iii. Section 2.3 : Architecture Lambda
 - (c) Section 3 : Requêtage des données Hadoop
 - i. Section 3.1 : Pig
 - ii. Section 3.2 : Hive
 - iii. Section 3.3 : Spark et SparkQL

Eventuels projets à réaliser: (à suivre ...)