

SYLLABUS

Domaine : Maths et Informatique

Filière : Informatique

Spécialité : Gestion et Analyse des Données Massives : Big Data (GADM)

Semestre : 3

Année scolaire : 2022/2023

Identification de la matière d'enseignement

Intitulé : Hadoop MapReduce

Unité d'enseignement: UEF2

Nombre de Crédits: 2

Coefficient : 2

Volume horaire hebdomadaire total :

- Cours (nombre d'heures par semaine) : 1h30
- Travaux pratiques (nombre d'heures par semaine) : 1h30

Responsable de la matière d'enseignement

Nom, Prénom, Grade : Dr KLAI SIHEM, MCA

Localisation du bureau (Bloc, Bureau) : Bureau 10

Email : klai@labged.net; klai_univ_annab@yahoo.com

Tel (Optionnel) :

Horaire du cours et lieu du cours : Mardi de 13h15mn à 16h30 à la Salle GADM

Description de la matière d'enseignement

Pré-requis :

- Des connaissances sur le système de gestion de fichier classique ;
- Des connaissances générales sur l'architecture maître-esclave classique;
- Le paradigme de la programmation fonctionnelle et l'algorithmique en général ;
- Le langage Java.

Objectif général de la matière d'enseignement : Dans ce module, nous rappelons les problèmes causés par l'importance et la rapidité de l'évolution des données massives. Nous introduisons par la suite le framework Hadoop qui apporte des solutions à ces problèmes voire de nouvelles manières de traitement, de nouveaux ordres de grandeur concernant la capture, le stockage, la recherche, le partage, l'analyse et la visualisation des données tout en présentant les limites des approches classiques.

Nous étudions en détail les composants principaux de Hadoop : Hadoop Distributed File System (HDFS) qui gère le stockage des données dans les clusters et MapReduce qui traite les données sur les clusters.

Nous aborderons par la suite l'exécution de programme MapReduce dont les étapes sont les suivantes :

- Le découpage des données d'entrée (opération *split*).
- L'opération *map*, une fonction exécutée indépendamment sur chacun de ces fragments, et qui génère une série de paires (clé;valeur) ;
- Le regroupement de tous les paires (clé;valeur) ainsi générées par clé distincte (opération *shuffle*).
- L'opération *reduce*, une fonction exécutée indépendamment sur chacun de ces groupes associés à une clé distincte, et produisant elle aussi une série de paires (clé;valeur).

Objectifs d'apprentissage : (de 3 à 6 objectifs, n'inclure que les objectifs que vous pouvez évaluer)

- Maîtriser les concepts de Hadoop, HDFS et MapReduce ;
- Installer Hadoop sur Windows et exécuter les commandes HDFS via commande et via des programmes java;
- Modéliser les problèmes selon le paradigme Mapreduce;
- Développer et exécuter un programme Mapreduce java sous Eclipse.

Contenu de la matière d'enseignement

Chapitre 1 : Généralité sur le Big Data

Chapitre 2 : **Hadoop** (présentation du framework)

Chapitre 3 : HDFS (fonctionnement)

Chapitre 4 : MapReduce : Présentation et Modélisation

Chapitre 5 : Mise en oeuvre de MapReduce dans Hadoop : API Java

Chapitre 6 : **Spark**

Annexes : Fiches TP

Modalités d'évaluation

Nature du contrôle	Pondération en %
Examen	60%
Micro – interrogation	
Travaux dirigés	
Travaux pratiques	30%
Projet personnel	
Travaux en groupe	10%
Sorties sur terrains	
Assiduité (Présence /Absence)	
Autres (à préciser)	
Total	100%

Références & Bibliographie

Textbook (Référence principale) :		
Titre de l'ouvrage	Auteur	Éditeur et année d'édition
HADOOP the définitive guide	Tom White	
Hadoop MapReduce Cookbook	Srinath Parera	2009
.....	Thilina Gunarathne	2013
.....		

Planning du déroulement du cours

Semaine	Titre du Cours	Date
1	<ul style="list-style-type: none"> • Présentation du Syllabus • Chapitre 1 : Introduction, spécificités de Big Data, technologies liées au Big Data. • Conduite des mini-projets • TP1 : TPN°1 : L'installation et la vérification d'un ensemble de pré-requis 	27/09/2022
2	Présentation de Hadoop : <ul style="list-style-type: none"> • Fonctionnement et architecture de Hadoop ; • Les composants principaux de Hadoop ; • Les modes de fonctionnement de Hadoop. • TP2 : TPN°2 : Télécharger, copier et installer Hadoop 	04/10/2022
3	HDFS : <ul style="list-style-type: none"> • Organisation des fichiers dans HDFS ; • Organisation des machines et Fonctionnement de HDFS ; • Manipuler HDFS ; • Ecriture et Lecture des données dans HDFS. • TPN°3 : Se familiariser avec les commandes HDFS 	11/10/2022
4	MapReduce : Présentation et Modélisation <ul style="list-style-type: none"> • Présentation de MapReduce ; • Modélisation des problèmes selon le paradigme MapReduce ; • MapReduce et Hadoop ; • Exercices ; 	18/10/2022 25/10/2022 01/11/2022

	<ul style="list-style-type: none"> • Solution des exercices. 	
5	<p>Mise en oeuvre de MapReduce dans Hadoop : API Java</p> <ul style="list-style-type: none"> • Structure d'un programme MapReduce ; • Les types de données de MapReduce ; • Spécification des Entrées/Sorties dans un programme Mapreduce; • Spécification des résultats intermédiaires (les sorties du Mapper, qui sont aussi les entrées du Reducer) ; • Spécification des résultats (données de sortie) ; • Compilation et exécution des programmes Mapreduce ; • Exercice ; • Solution des exercices. • TPN° 4 : Implémenter des programmes Mapreduce sous Eclipse 	<p>08/11/2022</p> <p>15/11/2022</p> <p>22/11/2022</p> <p>29/11/2022</p>
6	SPARK	<p>06/12/2022</p> <p>13/12/2022</p>
7	<ul style="list-style-type: none"> • Travail en groupe sous forme d'exposés sur les cas d'utilisation de Mapreduce (domaine de recherche) • Validation des mini-projets 	<p>A partir du</p> <p>13/12/2022</p>
	Examen de fin de semestre	
	Examen de rattrapage	

Liste des étudiants (indiquer le délégué de groupe et ses coordonnées)

Nom	Prénom	Emargement
AFIFI	HOUNAIDA	
AISSOU	MOUNDER	
ALOUYERE	YOUCEF	
AOUAMRI	ACHOUAK	
AYAD	TAKIEDDINE	
BENDJAH	ZAKARIA	
BOUCHMELLA	MOHAMED WASSIM	
BOUDJADJA	LINA NADA	
BOUKHEMIS	AYMEN	
BOUTABIA	NADJETTE	
CHAOLI	AIMEN	
CHAREF	MERIEM	
CHETTOUH	NASSIM	
CHOUAF	LAMIA	
DEBBAH	MOUFIDA	
DJENDLI	ABDELHAKIM	
FERFAR	INAAM	
FERRADJ	ZAHRA	
GUERGAH	KHALED	
HARBI	FARAH	
HAROUN	HICHAM	
KABAR	INES	
LAIB	AKRAM	
LAKEHAL	ABDELOUAHEB	
LOUCIFI	ISKANDER ABD ELDJALIL	
MENEI	ANOUAR	
MESSAOUD	BOUTHAINA	
MESSAOUDI	AYA	
MOULFI	HANANE	
NABTI	ZINEB	
NAIT ATIA	MOHAMED LAMINE	
OUAFFAI	SABRI	
REFFAD	SERINE	
ROUABHI	ROUMAISSA	
SOUALEM	OUSSAMA	
TLEMSANI	HILLEL	
TORCHE	SALMA	