

Module : BigData 1				Code	
				ING-4-SDIA-S8-P2	
Période	S8-P1	Volume horaire	21h	ECTS	2

Responsable	Aicha Ben Jrad	Email	Aicha.benjrad@tek-up.tn
Equipe pédagogique	Aicha Ben Jrad		

1. Objectifs de Module (*Savoirs, aptitudes et compétences*)

Ce module vise à fournir une compréhension approfondie des concepts fondamentaux du Big Data, des outils et des techniques de traitement des données volumineuses, en mettant l'accent sur l'écosystème Hadoop et Apache Spark.

Acquis d'apprentissage :

A la fin de cet enseignement, l'élève sera capable de :

- Comprendre les concepts fondamentaux du big data (**C1.1**)
- Comprendre comment les données hors ligne sont stockées, distribuées et traitées à l'aide des outils des écosystèmes Hadoop et Spark (**C1.2**)
- Maîtriser les bases de la programmation parallèle avec Spark (**C5.1**)

2. Pré-requis (*autres UE et compétences indispensables pour suivre l'UE concernée*)

- Python & Java
- Virtualisation
- Docker et Kubernetes

3. Répartition d'Horaires de Module

Intitulé de l'élément d'enseignement	Total	Cours	TD	Atelier	PR
Module : BigData 1	21	9		12	

4. Méthodes pédagogiques et moyens spécifiques au Module

(*pédagogie d'enseignement, ouvrages de références, outils matériels et logiciels*)

- Supports de Cours
- Projecteur et Tableau
- Travaux pratiques

Bibliographie

Titre	Auteur(s)	Edition
Big Data Analytics	Francisco Herrera	

HCIA BigData	Huawei	
Social media big data analytics: A survey	Norjihan AbdulGhani	

5. Contenu (<i>Descriptifs et plans des cours / Déroulement / Détail de l'évaluation de l'activité pratique</i>)	Durée allouée	
Séance 1 <ul style="list-style-type: none">● Introduction au BigData<ul style="list-style-type: none">1- Définition2- Impact du Big Data3- Traitement parallèle, mise à l'échelle et parallélisme des données4- Outils du Big Data5- Cas d'utilisation● Ecosystème Hadoop<ul style="list-style-type: none">1- Qu'est-ce que Hadoop ?2- Introduction à MapReduce3- Composants de l'écosystème Hadoop4- HDFS et ZooKeeper5- Hive	Cours	3H
Séance 2 <ul style="list-style-type: none">● Atelier HDFS, Hive et MapReduce	Atelier	3H
Séance 3 <ul style="list-style-type: none">● Apache Spark<ul style="list-style-type: none">1- Pourquoi utiliser Spark ?2- Bases de la programmation fonctionnelle3- Programmation parallèle avec les RDD4- Parallélisme des données● DataFrames et SparkSQL<ul style="list-style-type: none">1- Introduction aux DataFrames et SparkSQL2- DataFrames et Datasets3- Catalyst et Tungsten4- ETL avec DataFrames	Cours	3H
Séance 4 <ul style="list-style-type: none">● Atelier PySpark● Atelier ETL avec DataFrames et SparkSQL	Atelier	3H
Séance 5 <ul style="list-style-type: none">● Développement d'applications Spark<ul style="list-style-type: none">1- Architecture de Spark2- Modes de cluster de Spark3- Exécution d'une application Spark4- Surveillance des jobs5- Débogage des jobs parallèles	Cours	3H

Séance 6 <ul style="list-style-type: none"> • Atelier Spark sur Kubernetes 	Atelier 	3H
Séance 7 <ul style="list-style-type: none"> • Examen TP 	Atelier 	3H

6. Mode d'évaluation de Module(*nombre, types et pondération des contrôles*)

<i>Eléments d'enseignement</i>	Coeff	DS	EX	TP	PR
Module – BigData 1	1		60%	40%	

Pour valider le module, les étudiants passeront un examen dont le coefficient est de 60 %, un TP dont le coefficient est de 40% .

La durée de tous les examens (Examen, DS...) est de 1h30.

Quant à l'examen, il est planifié après l'écoulement des 7 semaines et portera sur toutes les thématiques enseignées tout au long des 21 heures.

Concernant le TP, il est planifié une semaine avant l'examen et testera les connaissances acquises tout au long du module.

Le module est validé si l'étudiant obtient une moyenne supérieure ou égal à 10 sur 20.

Dans l'ensemble, l'examen vise à évaluer les étudiants à la fois sur leurs capacités à comprendre les concepts fondamentaux du Big Data, à expliquer comment les données hors ligne sont stockées distribuées et traitées à l'aide des outils des écosystème Hadoop et Spark, et à développer des programmes parallèles avec Spark.