	<b>Fiche module</b>		Mise à jour :
	Cycle de formation d'ingénieurs en Informatique		Page 1 sur 3

<b>Module : BigData 2</b>				Code	
				<b>ING-5-SDIA-S9-P4</b>	
<i>Période</i>	<b>S9</b>	<i>Volume horaire</i>	<b>21h</b>	<i>ECTS</i>	<b>2</b>

<i>Responsable</i>	<b>Aicha Ben Jrad</b>	<i>email</i>	<a href="mailto:Aicha.benjrad@tek-up.tn">Aicha.benjrad@tek-up.tn</a>
<i>Equipe pédagogique</i>	Aicha Ben Jrad		

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

Ce module porte sur les techniques avancées en Big Data avec un accent sur Apache Spark, Apache Airflow et Apache Kafka.

#### Acquis d'apprentissage :

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

- Maîtriser les techniques avancées de traitement des données en utilisant Apache Spark. **(C5.1)**
- Concevoir, gérer et automatiser des workflows de données avec Apache Airflow. **(C5.2)**
- Intégrer et configurer Apache Kafka pour le streaming de données en temps réel. **(C5.2)**

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

- Python & Java
- Virtualisation
- Docker
- Big Data 1

### 3. Répartition d'Horaire de Module


<i>Intitulé de l'élément d'enseignement</i>	<i>Total</i>	<i>Cours</i>	<i>TD</i>	<i>Atelier</i>	<i>PR</i>
Module : Big Data 2	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

	<b>Fiche module</b>		Mise à jour :
	Cycle de formation d'ingénieurs en Informatique		Page 2 sur 3

Big Data Analytics	Francisco Herrera	
HCIA BigData	Huawei	
Social media big data analytics: A survey	Norjihan AbdulGhani	

<b>5. Contenu</b> <i>(Descriptifs et plans des cours / Déroulement / Détail de l'évaluation de l'activité pratique)</i>		Durée allouée	
<b>Séance 1</b> <ul style="list-style-type: none"> <li>Rappel au BigData</li> <li>Spark pour l'ingénierie des données               <ol style="list-style-type: none"> <li>Spark Streaming</li> <li>GraphFrames</li> <li>ELT</li> </ol> </li> </ul>		Cours	3H
<b>Séance 2</b> <ul style="list-style-type: none"> <li>Atelier Spark</li> </ul>		Atelier	3H
<b>Séance 3</b> <ul style="list-style-type: none"> <li>Spark ML pour le Machine Learning               <ol style="list-style-type: none"> <li>Les bases de Spark ML</li> <li>Modèles de régression et de classification avec Spark ML</li> <li>Clustering avec Spark ML</li> </ol> </li> </ul>		Cours	3H
<b>Séance 4</b> <ul style="list-style-type: none"> <li>Atelier Spark ML</li> </ul>		Atelier	3H
<b>Séance 5</b> <ul style="list-style-type: none"> <li>Intégration d'Apache Airflow pour les workflows ELT               <ol style="list-style-type: none"> <li>Introduction à Apache Airflow</li> <li>Architecture d'Airflow</li> <li>Création et gestion de workflows ELT avec Airflow</li> <li>Intégration avec Spark</li> </ol> </li> <li>Introduction à Apache Kafka pour le Streaming de données               <ol style="list-style-type: none"> <li>Introduction</li> <li>Architecture de Kafka</li> <li>Concepts de base</li> <li>Gestion des topics</li> <li>Intégration de Kafka avec Spark</li> </ol> </li> </ul>		Cours	3H
<b>Séance 6</b> <ul style="list-style-type: none"> <li>Atelier Airflow</li> <li>Atelier Kafka</li> </ul>		Atelier	3H
<b>Séance 7</b>		Atelier	3H

	<b>Fiche module</b>		Mise à jour :
	Cycle de formation d'ingénieurs en Informatique		Page 3 sur 3

<ul style="list-style-type: none"> <li>Examen TP</li> </ul>		
---	--	--

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

<i>Eléments d'enseignement</i>	Coeff	DS	EX	TP	PR
Module – Big Data 2	1		60%	40%	
<p>Pour valider le module, les étudiants passeront un examen dont le coefficient est de 60%, et un TP dont le coefficient est de 40%.</p> <p>La durée de tous les examens (Examen, DS...) est de 1h30.</p> <p>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.</p> <p>Concernant le TP, il est planifié une semaine avant l'examen et testera les connaissances acquises tout au long du module.</p> <p>Le module est validé si l'étudiant obtient une moyenne supérieure ou égal à 10 sur 20.</p> <p>Dans l'ensemble, l'examen vise à évaluer les étudiants sur leurs capacités à utiliser les techniques avancées de traitement des données volumineuses, à concevoir, gérer et automatiser des workflows de données avec Apache Airflow, et à intégrer Apache Kafka pour le streaming de données en temps réel.</p>					