



## Plan du cours

### UIT et Ecole Supérieure Multinationale des Télécommunications (ESMT)

Titre	Hadoop pour le traitement distribué des Big Data
Modalité de la formation	Formation en présentiel par un formateur
Dates	Du 19 au 22 juillet 2022
Durée	4 jours
Date limite d'inscription	12 juillet 2022
Frais d'inscription	453 USD
Description	<p>Les entreprises produisent d'énormes quantités de données chaque jour. Ces données sont stockées puis traitées et analysées pour en tirer de la valeur. Grâce aux plateformes de stockage et de traitements distribués de type Hadoop, il est devenu plus facile pour les ingénieurs de répondre aux problématiques du Big Data avec une grande efficacité et à un coût réduit.</p> <p>Ce séminaire vous offre l'occasion de vous essayer au traitement distribué de données massives via la plateforme Hadoop et ses outils comme Hive et Spark.</p>
Code de la formation	22WS28051AFR-F

#### 1.OBJECTIFS D'APPRENTISSAGE

---

Vous apprendrez le traitement distribué de données massives via la plateforme Hadoop et ses outils comme Hive et Spark.

#### 2.RESULTATS ATTENDUS

---

A la fin de la formation, les participants seront capables de :

1. Stocker des données sur Hadoop
2. Décrire les différentes étapes du modèle de programmation MapReduce
3. Charger des données sur Hive et les traiter à l'aide du langage HiveQL
4. Charger des données dans Spark et les traiter à l'aide du langage Spark SQL

#### 3.PUBLIC CIBLE

---



Cette formation s'adresse aux :

- Développeurs
- Experts en business intelligence
- Analystes de données

#### 4.CONDITIONS D'ENTREE

Pour suivre ce cours, les participants doivent avoir des :

- Notions en administration système sous Linux.
- Notions en programmation orientée objet
- Connaissances de base du langage SQL

#### 5.FORMATEURS/INSTRUCTEURS

Nom du (des) formateur /instructeur(s)	Contacts
M. Jean-Marie PREIRA, Enseignant-Chercheur en Informatique à l'ESMT	Email : <a href="mailto:jeanmarie.preira@esmt.sn">jeanmarie.preira@esmt.sn</a> Tél. : +221 70 637 03 15

#### 6.CONTENU DU COURS

Date	Horaire	Contenus et activités
19 juillet	08:30 – 09:00	Accueil et installation des participants
	09:00 – 09:30	Ouverture officielle
	09:30 – 10:00	Introduction de la formation
	10:00 – 10:30	Pause-café
	10:30 – 11:30	<b>L'écosystème du Big Data</b> <ul style="list-style-type: none"><li>• Les origines et caractéristiques du Big Data</li><li>• Les avancées technologiques</li><li>• L'informatique parallèle</li><li>• Aperçu des plateformes Big Data</li></ul>
	11:30-13:00	<b>Amazon Athena</b> <ul style="list-style-type: none"><li>• Présentation de Amazon Web Services (AWS)</li><li>• Amazon EC2</li><li>• Amazon S3</li><li>• Présentation de Amazon Athena</li><li>• <b>Démonstration</b> : Requêtes SQL sur des données massives avec Amazon Athena</li></ul>
	13:00 – 14:30	Pause-déjeuner
	14:30 – 15:30	<b>Le paradigme MapReduce</b> <ul style="list-style-type: none"><li>• Historique</li><li>• Le pattern MapReduce</li><li>• Cas d'utilisation de MapReduce</li><li>• Rappel de quelques commandes Linux</li></ul>
	15:30 – 16:00	<b>Travaux pratiques</b> : programmation MapReduce
	16:00 – 16:30	Pause-café



Date	Horaire	Contenus et activités
	16:30 – 17:30	<b>Travaux pratiques</b> : programmation MapReduce
20 juillet	08:30 – 09:00	Accueil et installation des participants
	09:00 – 10:00	<b>Hadoop</b> <ul style="list-style-type: none"> <li>• Historique</li> <li>• Socle technique</li> </ul>
	10:00 – 10:30	Pause-café
	10:30 – 11:30	<b>Hadoop</b> <ul style="list-style-type: none"> <li>• HDFS</li> <li>• Hadoop – YARN</li> </ul>
	11:30 – 13:00	<b>Travaux pratiques</b> : les commandes HDFS
	13:00 – 14:30	Pause-déjeuner
	14:30 – 16:00	<b>Hadoop</b> <ul style="list-style-type: none"> <li>• Ecosystème Hadoop</li> <li>• Les distributions Hadoop</li> <li>• Installation de Cloudera Quickstart VM</li> </ul>
	16:00 – 16:30	Pause-café
	16:30 – 17:30	<b>Travaux pratiques</b> : programmation Hadoop MapReduce
Date	Horaire	Contenus et activités
21 juillet	08:30 – 09:00	Accueil et installation des participants
	09:00 – 10:00	<b>Apache Hive</b> <ul style="list-style-type: none"> <li>• Présentation</li> <li>• Création d'une base de données sous Hive</li> <li>• Commandes HQL</li> </ul>
	10:00 – 10:30	Pause-café
	10:30 – 13:00	<b>Travaux pratiques</b> : analyse de données avec HiveQL
	13:00 – 14:30	Pause-déjeuner
	14:30 – 16:00	<b>Spark SQL</b> <ul style="list-style-type: none"> <li>• Apache Spark</li> <li>• Les RDD</li> </ul>
	16:00 – 16:30	Pause-café
	16:30 – 17:30	<b>Spark SQL</b> <ul style="list-style-type: none"> <li>• Installation de Apache Spark</li> </ul>
Date	Horaire	Contenus et activités
22 juillet	08:30 – 09:00	Accueil et installation des participants
	09:00 – 10:00	<b>Travaux pratiques</b> : calcul distribué avec PySpark
	10:00 – 10:30	Pause-café
	10:30 – 11:30	<b>Travaux pratiques</b> : calcul distribué avec PySpark
	11:30 – 13:00	<b>Spark SQL</b> <ul style="list-style-type: none"> <li>• Spark SQL et DataFrames</li> </ul>
	13:00 – 14:30	Pause-déjeuner
	14:30 – 16:00	<b>Travaux pratiques</b> : analyse de données avec Spark SQL
	16:00 – 16:30	<b>Test de connaissance</b> : Quiz
	16:30 – 17:30	<b>Clôture de la formation</b> <ul style="list-style-type: none"> <li>• Evaluation à froid et à chaud</li> <li>• Remise des attestations</li> </ul>



## 8. AGENDA DU COURS

---

Jour	Activités	Travaux pratiques
Jour 1	<ul style="list-style-type: none"><li>• Cérémonie d'ouverture</li><li>• L'écosystème du Big Data</li><li>• Amazon Athena</li><li>• Le paradigme MapReduce</li></ul>	<ul style="list-style-type: none"><li>• Programmation MapReduce</li></ul>
Jour 2	<ul style="list-style-type: none"><li>• Hadoop</li></ul>	<ul style="list-style-type: none"><li>• Les commandes HDFS</li><li>• Programmation Hadoop MapReduce</li></ul>
Jour 3	<ul style="list-style-type: none"><li>• Apache Hive</li><li>• Spark SQL</li></ul>	<ul style="list-style-type: none"><li>• Analyse de données avec HiveQL</li></ul>
Jour 4	<ul style="list-style-type: none"><li>• Spark SQL (suite et fin)</li><li>• Test des connaissances</li><li>• Evaluation de la formation</li><li>• Remise des attestations et clôture de la formation</li></ul>	<ul style="list-style-type: none"><li>• Calcul distribué avec PySpark</li><li>• Analyse de données avec Spark SQL</li></ul>

## 9. MODE D'ANIMATION PEDAGOGIQUE

---

La formation sera animée en mobilisant des les activités pédagogiques suivantes :

Exposés théoriques

Questions-réponses

Exercices d'application

Démonstrations

Travaux pratiques

---

## 10. EVALUATION ET NOTATION

---

Ce cours est évalué via l'assiduité des participants (1/3) d'une part et par un examen final (2/3) d'autre part. Seuls les candidats ayant totalisé un score global de 63% minimum se verront délivrer un certificat codéveloppé par l'ESMT et l'UIT.

## 11. COORDINATION DU COURS

---



**Coordonnateur du centre  
d'excellence :**

Nom : Dr André ONANA  
Titre : Coordonnateur du Centre  
d'Excellence UIT à l'ESMT  
Tel : +221 70 637 03 04  
Email : [andre.onana@esmt.sn](mailto:andre.onana@esmt.sn)

**Coordinateur UIT:**

Nom: M Emmanuel NIYIKORA  
Email : [emmanuel.niyikora@itu.int](mailto:emmanuel.niyikora@itu.int)