

IOT & BIG DATA For Supporting and Developing Business

COURSE DESCRIPTION

Title	IOT & BIG DATA For Supporting and Developing Business
Period	17-20/12/2017
Objectives	<p>Le développeur Big Data après ce cours peut comprendre le concept de Big Data, créer des requêtes orientées des méga-données structurées, semi-structurées et non-structurées. Comprendre Hadoop et les projets créés la dessus.</p> <p><u>Pre-requisites :</u></p> <p>Connaissance generale de programmation et aussi des des systems de gestion de base de données classiques.</p>
Duration	4 jour

LEARNING OUTCOMES

A la fin de cette formation vous devez être capable de :

- Comprendre le concept de Hadoope écosystème et son implémentation
- Ecrire des programmes de MapReduces.
- Comprendre le concept de système de fichiers HDFS
- Comprendre les projets de moteurs de recherché orientés Hadoop comme HIVE, HBASE, BIG SQL
- Comprendre l'analyse des textes non-structurés
- Les langages de requêtages orientés des données non-structurées
-

TARGET AUDIENCE

Ingénieurs, IT admistrateurs, Techniciens en informatique, des consultants et des adminstrateurs de bases de données

TRAINER

BIOGRAPHY	CONTACT
<p>Actuellement, Maher HENI enseignant (Maître assistant à ISI KEF) chercheur au laboratoire Innov'COM@ l'école Supérieur de Communication SUP'COM, à TUNIS. Docteur ingénieur en Télécom depuis Juin 2013.</p> <p>Mon diplôme d'ingénieur est obtenu en 2008, de l'école nationale d'ingénieurs de Tunis(ENIT) suite à un PFE de recherche effectué au sein du laboratoire CITI @ INSA-Lyon sous la direction de Mme Isabelle AUGÉ-BLUM portant comme thème l'évaluation des performances des protocoles de routages pour les réseaux de capteurs sans fil.</p>	

Mon diplôme de Mastère est obtenu en 2009, de l'école nationale d'ingénieurs de Tunis(ENIT) suite à un stage de Mastère 2 effectué au sein de l'Institut d'Electronique Fondamentale à l'Université de Paris-SUD sous la tutelle de Mme Véronique VEQUE autour de l'optimisation énergétique dans les réseaux Ad Hoc. Simulations effectués avec JIST-SWANS utilisant le langage de programmation JAVA.

Ma thèse de doctorat a été une continuation sur le même thème de Mastère2, la proposition de nouvelles approches pour l'optimisation de la consommation énergétique dans les réseaux sans fil soutenu en JUIN 2013 avec la mention "très honorable". Mes travaux de thèse ont fait l'objet de publications dans des revues et conférences internationales avec comité de lectures. J'ai aussi publié le rapport de thèse en tant que livre dans l'édition de publication européenne.

TRAINING SCHEDULE

jour 1	Course's Content (Time)
17/ 12 / 2018	<ul style="list-style-type: none"> ○ Présentation d'Hadoop ○ Évaluer la valeur que peut apporter Hadoop à l'entreprise ○ Examiner l'écosystème d'Hadoop ○ Choisir un modèle de distribution adapté ○ Défier la complexité de la programmation parallèle ○ Examiner les difficultés liées à l'exécution de programmes parallèles : algorithmes, échange des données ○ Évaluer le mode de stockage et la complexité du Big Data ○ Programmation parallèle avec MapReduce ○ Fragmenter et résoudre les problèmes à grande échelle ○ Découvrir les tâches compatibles avec MapReduce <ul style="list-style-type: none"> ○ Résoudre des problèmes métier courants
jour 2	<ul style="list-style-type: none"> ○ Course's Content (Time)
18 /12 / 2017	<ul style="list-style-type: none"> ○ Appliquer le paradigme Hadoop MapReduce ○ Configurer l'environnement de développement ○ Examiner la distribution Hadoop ○ Étudier les démons Hadoop ○ Créer les différents composants des tâches MapReduce ○ Analyser les différentes étapes de traitement MapReduce : fractionnement, mappage, lecture aléatoire et réduction ○ Créer des tâches MapReduce complexes ○ Choisir et utiliser plusieurs outils de mappage et de réduction : exploiter les partitionneurs et les fonctions map et reduce intégrées, coordonner les tâches avec le planificateur de workflow Oozie, rationaliser les tâches dans différents langages de programmation

	<ul style="list-style-type: none"> ○ Résoudre les problèmes de manipulation des données ○ Exécuter les algorithmes : tris, jointures et recherches pour analyser les fichiers journaux, les données des médias sociaux, courriels ○ Mise en œuvre des partitionneurs et des comparateurs ○ Identifier les algorithmes parallèles liés au réseau, au processeur et aux disques ○ Répartir la charge de travail avec les partitionneurs ○ Contrôler l'ordre de groupement et de tri avec les comparateurs ○ Mesurer les performances avec les compteurs ○ Bien-fondé des données distribuées ○ Optimiser les performances du débit des données ○ Utiliser la redondance pour récupérer les données ○ Interfacier avec le système de fichiers distribué Hadoop ○ Analyser la structure et l'organisation du HDFS ○ Charger des données brutes et récupérer le résultat ○ Lire et écrire des données avec un programme ○ Manipuler les types SequenceFile d'Hadoop ○ Partager des données de référence avec DistributedCache ○
jour 3	Course's Content (Time)
19 /12 / 2017	<ul style="list-style-type: none"> ○ Structurer les données avec HBase ○ Passer du stockage structuré au stockage non structuré ○ Appliquer les principes NoSQL avec une application de modèle de lecture, se connecter à HBase à partir des tâches MapReduce ○ Comparer HBase avec d'autres types de magasins de données NoSQL ○ Exploiter la puissance de SQL avec Hive ○ Structurer bases de données, les tables, les vues et les partitions ○ Extraire, transformer et charger les données ○ Lancer des requêtes avec HiveQL ○ Accéder aux servers Hive via JDBC, ajouter des fonctionnalités HiveQL avec les fonctions définies par l'utilisateur
jour 4	Course's Content (Time)
20/12 / 2017	<ul style="list-style-type: none"> ○ Exécuter des workflows avec Pig ○ Développer des scripts Pig Latin pour consolider les workflows et intégrer des requêtes Pig à Java ○ Interagir avec les données par le biais de la console Grunt ○ Étendre Pig avec les fonctions définies par l'utilisateur ○ Tester et déboguer le code Hadoop ○ Enregistrer des événements importants à auditer et à déboguer ○ Valider les spécifications avec MRUnit ○ Déboguer en mode local

- | | |
|--|--|
| | <ul style="list-style-type: none">○ Déployer, surveiller et affiner les performances○ Déployer la solution sur un cluster de production, utiliser des d'administration pour optimiser les performances, sur l'exécution des tâches via les interfaces utilisateur web |
|--|--|