

Réaliser de l'ingénierie de données avec Microsoft HDInsight

Référence : **MS20775**

Durée : **5 jours**

Certification : **70-775**

Eligible CPF : **236056**

CONNAISSANCES PREALABLES

- 1- Avoir une expérience dans la programmation via R.
- 2- Posséder des connaissances sur les méthodes statistiques courantes et les meilleures pratiques d'analyse des données.
- 3- La connaissance du système d'exploitation Windows et ses fonctionnalités principales ainsi que des bases de données relationnelles est recommandée.

PROFIL DES STAGIAIRES

- Architectes de données.
- Data Scientists.
- Développeurs.

OBJECTIFS

- Acquérir les connaissances et compétences nécessaires pour planifier et mettre en œuvre les Workflows Big Data sur Microsoft HD Insight.

METHODES PEDAGOGIQUES

- Mise à disposition d'un poste de travail par stagiaire
- Remise d'une documentation pédagogique papier ou numérique pendant le stage
- La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience
- Le suivi de cette formation donne lieu à la signature d'une feuille d'émargement

FORMATEUR

Consultant-Formateur expert SQL Server 2016

CERTIFICATION PREPAREE

Ce test prépare en partie le titre de MCSA Ingénierie des données avec Azure et le titre de MCSE Gestion des données et analyse

METHODE D'EVALUATION DES ACQUIS

- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation de fin de stage adressée avec la facture

CONTENU DU COURS

Introduction à HD Insight

- Qu'est-ce que la Big Data ?
- Introduction à Hadoop
- Travailler avec la fonction MapReduce
- Introduction à HDInsight

Déploiement des clusters HDInsight

- Identifier les types de clusters HDInsight
- Gérer les clusters HDInsight via le portail Azure
- Gérer les clusters HDInsight via Azure Powershell

Autorisation des utilisateurs pour l'accès aux ressources

- Clusters non joints aux domaines
- Configurer les clusters HDInsight joints aux domaines
- Gérer les clusters HDInsight joints aux domaines

Chargement des données dans HDInsight

- Stockage des données pour le traitement HDInsight
- Utiliser les outils de chargement des données
- Valoriser les données stockées

Dépannage de HDInsight

- Analyser les journaux HDInsight
- Journaux YARN
- Dump de tas (Heap Dump)

- Operations Management Suite

Mise en œuvre des solutions Batch

- Stockage Apache Hive
- Requêtes de données HDInsight via Hive et Pig
- Mise en œuvre de HDInsight

Conception des solutions Batch ETL

- Qu'est-ce que SPARK ?
- ETL avec SPARK
- Performance de SPARK

Analyse des données avec SPARK SQL

- Mettre en œuvre des requêtes itératives et interactives
- Réaliser des analyses de données exploratoires

Analyse des données avec Hive et Phoenix

- Mettre en œuvre les requêtes interactives pour la Big Data avec Hive interactive
- Réaliser des analyses de données exploratoires via Hive
- Réaliser un traitement interactif via Apache Phoenix

Analyse de flux

- Analyse de flux
- Processus de diffusion des données à partir de l'analyse de flux
- Gérer les travaux d'analyse de flux

Mettre en œuvre les solutions de diffusion avec Kafka et HBase

- Créer et déployer un cluster Kafka
- Publier, consommer et traiter les données via le cluster Kafka
- Utiliser HBase pour stocker des données et effectuer des requêtes

Développement des solutions de traitement Big Data en temps réel

- Travailler avec Spark Streaming
- Créer des applications structurées Spark Streaming
- Stockage et Visualisation

Création des applications Spark Streaming

- Utilisation de Spark Streaming
- Création d'applications Spark structurée pour le Streaming
- Persistance et visualisation

Certification Microsoft - Perform Data Engineering on Azure HD Insight

- Cette formation prépare la certification Microsoft Perform Data Engineering on Azure HD Insight