

Hadoop, stocker et traiter des données pour le big data

Tutorat en option

Formation en ligne - 2h

Réf : 4ER - Prix 2024 : 95CHF HT

Ce cours en ligne a pour objectif de vous faire découvrir la plateforme Hadoop, son écosystème et comprendre ses concepts clés, son fonctionnement et ses possibilités. Il s'adresse aux décideurs et développeurs intéressés par le big data. La pédagogie s'appuie sur un auto-apprentissage séquentiel par actions de l'utilisateur sur l'environnement à maîtriser. Une option de tutorat vient renforcer l'apprentissage.

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

- Définir les notions de NoSQL et de big data
- Installer Hadoop et étudier les différents outils au cœur de la plateforme
- Déployer un cluster sur des machines virtuelles dans le cloud grâce à la distribution Cloudera Data Platform
- Stocker des données dans Hadoop
- Appréhender le paradigme de traitement distribué des données avec MapReduce
- Développer des traitements avec Spark ou en SQL avec Hive

LE PROGRAMME

dernière mise à jour : 10/2023

1) Présentation des origines de Hadoop

- Définir le NoSQL et le big data.
- Découvrir l'histoire de Hadoop.
- Découvrir l'écosystème.
- Appréhender les distributions Hadoop.

2) Installation d'un environnement Hadoop avec Cloudera

- Provisionner des machines virtuelles.
- Installer un serveur Cloudera Manager.
- Installer les agents sur les serveurs.
- Installer un service Hadoop.

3) Gestion du stockage de données

- Interagir avec n'importe quel stockage grâce à HadoopFS.
- Mettre en place et exploiter le HDFS.
- Stocker des objets avec Apache Ozone.
- Exploiter HBase, la base de données de Hadoop.

PARTICIPANTS

Décideurs et développeurs intéressés par le big data.

PRÉREQUIS

Aucun.

COMPÉTENCES DU FORMATEUR

Les experts qui ont conçu la formation et qui accompagnent les apprenants dans le cadre d'un tutorat sont des spécialistes des sujets traités. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

La progression de l'apprenant est évaluée tout au long de sa formation au moyen de QCM, d'exercices pratiques, de tests ou d'échanges pédagogiques. Sa satisfaction est aussi évaluée à l'issue de sa formation grâce à un questionnaire.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : documentation et support de cours, exercices pratiques d'application et corrigés des exercices, études de cas ou présentation de cas réels. ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques. Une attestation de fin de formation est fournie si l'apprenant a bien suivi la totalité de la formation.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

4) Réalisation de traitements sur les données

- Comprendre et développer un MapReduce avec Hadoop Streaming.
- Réaliser un MapReduce Java.
- Optimiser la vitesse des traitements avec Apache TEZ®.
- Traiter les données avec Spark.
- Interroger les données en SQL avec Hive.
- Utiliser Sqoop pour échanger des données avec un SGBDR.