

Hadoop - HBase, mise en œuvre et administration

Cours Pratique de 2 jours - 14h

Réf : HBA - Prix 2024 : 1 620€ HT

Le prix pour les dates de sessions 2025 pourra être révisé

A l'issue de la formation, le stagiaire sera capable d'installer et de mettre en œuvre une configuration distribuée de données sous la solution HBase.

OBJECTIFS PÉDAGOGIQUES

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

Appréhender l'écosystème Hadoop

Comprendre l'architecture et le fonctionnement de HBase

Identifier les apports d'HBase en termes de stockage distribué des données

Mener à bien l'installation du système

Savoir mettre en place une configuration distribuée

Ce cours, interactif, s'appuie sur de nombreux exercices et travaux pratiques

LE PROGRAMME

dernière mise à jour : 11/2022

1) Présentation de l'écosystème d'Hadoop

- L'architecture Hadoop.
- Description des principaux composants de la plateforme Hadoop.
- Rappels MapReduce.
- Le système de fichiers distribué HDFS (Hadoop Distributed File System) d'Hadoop.
- Présentation de HBase, ses apports, lien avec HDFS.
- Format des données dans HBase.
- Fonctionnalités : failover automatique, sharding, interface avec des jobs MapReduce.

Echanges : Le stockage distribué de données et les apports d'HBase.

2) Architecture HBase

- Terminologie de HBase : table, région, ligne, famille de colonnes, cellules, espace de nommage...
- Les 3 composants principaux : HMaster, Region Server, Zookeeper.
- Rôle de Zookeeper.
- Le schéma des données. Les différents scénarios de schémas.

3) Installation

- Installation et configuration dans le fichier conf/hbase-site.xml.
- Installation en mode distribué.
- Gestion des connexions avec hbase shell.

Travaux pratiques : Installation, configuration, test de connexion et interrogations depuis le serveur http intégré.

4) Base utilisation : Shell

- Les différentes interfaces disponibles.

FINANCEMENT

Ce cours fait partie des actions collectives Atlas.

PARTICIPANTS

Chefs de projet, administrateurs

PRÉREQUIS

Avoir des connaissances des principes de base Hadoop et des bases de données.

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. 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 qu'ils enseignent. 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

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

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.

- Commandes de base, syntaxe, variables, manipulation des données : create, list, put, scan, get.
- Programmation de scripts.
- Activer/désactiver une table, effacer une table : enable/disable, drop...
- Principe des filtres. Mise en œuvre de filtres de recherche.
- Paramètres des tables.
- Les espaces de nommage.

Travaux pratiques : Ecriture de scripts Hase Schell.

5) Cluster HBase

- Rappels : préparation et configuration du cluster Hadoop, principe de fonctionnement de HDFS.
- Fonctionnement en mode distribué.
- Fonctionnement indépendant des démons (HMaster, HRegionServer, Zookeeper).
- Mise en œuvre avec HDFS dans un environnement distribué.

Travaux pratiques : Mise en œuvre des splits dans des tables réparties.

6) Développement

- Les APIs (REST, Avro, Thrift, Java, Ruby, Spark, ...).
- La surveillance des applications.
- Exemple de programmation d'un client.
- Gestion des tables.
- Map/reduce.

Démonstration : Détail de programmes et lancement

LES DATES

CLASSE À DISTANCE

2024 : 14 nov.

2025 : 27 févr., 12 mai, 21 juil., 16 oct.

PARIS

2025 : 20 févr., 05 mai, 07 juil., 09 oct.