

Hadoop : Stockage Hbase

Référence : PYCB034

Durée : 2 jours

Certification : **Aucune**

CONNAISSANCES PREALABLES

- Connaissance des principes de base Hadoop et des bases de données.

PROFIL DES STAGIAIRES

- Chefs de projet, administrateurs et toute personne souhaitant stocker des données avec Hbase.

OBJECTIFS

- Comprendre le fonctionnement de HBase, savoir mettre en place une configuration distribuée.

CERTIFICATION PREPAREE

Aucune

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 Bigdata

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

- Rappels rapides sur l'écosystème Hadoop
- Les fonctionnalités du framework Hadoop
- Le projet et les modules : Hadoop Common, HDFS, YARN, Spark, MapReduce
- Présentation HBase. Historique. Lien avec HDFS
- Format des données dans HBase
- Définitions : table, région, ligne, famille de colonnes, cellules, espace de nommage, ...
- Fonctionnalités : failover automatique, sharding, interface avec des jobs MapReduce

Architecture

- HBase master node, Region Master, liens avec les clients HBase
- Présentation du rôle de Zookeeper

Installation

- Choix des packages

- Installation et configuration dans le fichier conf/hbase-site.xml
- Démarrage en mode standalone start-hbase
- Test de connexion avec hbase shell
- Installation en mode distribué
- Travaux pratiques : Interrogations depuis le serveur http intégré

HBase utilisation : shell

- Présentation des différentes interfaces disponibles
- Travaux pratiques avec hbase shell
- Commandes de base, syntaxe, variables, manipulation des données : create, list, put, scan, get, désactiver une table ou l'effacer : disable (enable), drop, ...
- Programmation de scripts
- Gestion des tables : principe des filtres
- Mise en oeuvre de filtres de recherche, paramètres des tables
- Présentation des espaces de nommage

Cluster HBase

- Fonctionnement en mode distribué
- Première étape : fonctionnement indépendant des démons (HMaster, HRegionServer, Zookeeper)
- Passage au mode distribué : mise en œuvre avec HDFS dans un environnement distribué
- Travaux pratiques : sur un exemple de tables réparties : mise en œuvre des splits

Programmation

- Introduction, les APIs (REST, Avro, Thrift, Java, Ruby, ...)
- Utilisation d'un client Java
- Gestion des tables. Lien avec MapReduce
- Principe des accès JMX
- Travaux pratiques : création d'un client JMX