



Hadoop : stockage avec HBase

VOUS ÊTES

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

VOS OBJECTIFS

Comprendre le fonctionnement de HBase
Mettre en place une configuration distribuée

VOTRE FORMATION



DURÉE : 2 jours
14 heures



PROCHAINE SESSION :
14 et 15 mai 2020



LIEU : En distanciel



PRIX : 1 470 €
net de taxes

PRÉ-REQUIS

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

MODALITÉS

La formation est accessible à distance en classe virtuelle : accès à l'infrastructure de travaux pratiques, machines physiques, outils pédagogiques, échanges avec le formateur

De 4 à 12 participants

Financement éligible au FNE Formation pour tout salarié d'entreprise en activité partielle

VOTRE CONTACT :



Andrea FALLOURD

Conseillère en formation

06 74 51 44 97

afalourd@itescia.fr

ITESCIA - Campus de Pontoise

8 rue Pierre de Coubertin

95300 PONTOISE

www.itescia.fr



VOTRE PROGRAMME

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

Installation en mode distribué.

Test de connexion avec hbase shell.

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 oeuvre avec HDFS dans un environnement distribué.

Travaux pratiques : sur un exemple de tables réparties, mise en oeuvre des splits.

Programmation

Introduction, les APIs (REST, Avro, Thrift, Java, Ruby, Spark...)

Principe des accès JMX. Démonstration avec Prométheus.

Exemple de programmation d'un client Java.

Gestion des tables. Lien avec map/reduce.

Travaux pratiques : lancement et écriture de programmes Spark

Rejoignez nos réseaux sociaux

