

Hadoop : stockage avec HBase

Durée: 2 jours

1 130 €

6 au 7 mars
12 au 13 juin

25 au 26 septembre
20 au 21 novembre

Public:

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.

Connaissances préalables nécessaires:

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

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
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.

Hadoop : stockage avec HBase

- 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, ...)
Utilisation d'un client Java.
Gestion des tables. Lien avec MapReduce.
Principe des accès JMX.
Travaux pratiques :
création d'un client JMX