

BigData avec Cassandra

Durée: 3 jours

1710 €

12 au 14 mars

11 au 13 juin

17 au 19 septembre

26 au 28 novembre

Public:

Chefs de projet, gestionnaires de bases de données.

Objectifs:

Connaître les apports de Cassandra, savoir l'installer et le configurer et maîtriser le CQL. Gérer Cassandra en production avec OpsCenter. Savoir interfacier avec Hadoop, Spark.

Connaissances préalables nécessaires:

Connaissances générales sur les bases de données.

Programme:

- Introduction** : Historique, fonctionnalités de Cassandra, licence
Format des données,"key-value", traitement de volumes importants,
haute disponibilité, système réparti de base de données, ...
- Installation et configuration** : Prérequis. Plate-formes supportées. Etude du fichier de configuration : conf/cassandra.yaml
Répertoire de travail, de stockage des données, gestion de la mémoire.
Démarrage d'un noeud et test de l'interface cliente cqlsh.
- CQL** : Commandes de base : connexion au système de base de données,
création de colonnes,insertion, modification recherche,
Le CQL : Cassandra Query Language. Exécution de scripts.
Comment écrire des requêtes? Approches.
- Gestion de la grappe** : Principe.Préparation du premier noeud : adresse d'écoute.
Configuration de nouveaux noeuds.Notion de bootstrapping et de token.
Paramètres listen_address et rpc_address.
Réplication : topologie du réseau et EndpointSnitch.Stratégie de réplication. Ajout de noeuds, suppression.
Cassandra dans un cloud. Mise en oeuvre avec OpenStack.

BigData avec Cassandra

- Supervision : OpsCenter : installation, lancement. Utilisation de base.
Supervision avec nodetool cfstats, ou export JMX vers des outils de supervision comme Nagios.
- Exploitation : Sauvegardes. Import/export au format JSON.
- Support Hadoop : Principe de MapReduce. Implémentation Hadoop. Mise en oeuvre depuis Cassandra.
- Support Spark : Description rapide de l'architecture spark. Mise en oeuvre depuis Cassandra.
Execution de travaux Spark s'appuyant sur une grappe Cassandra.