

Formation Apache Cassandra : Utilisation

■ Durée :	3 jours (21 heures)
■ Tarifs inter-entreprise :	2 070,00 € HT (standard) 1 656,00 € HT (remisé)
■ Public :	Responsables des bases de données, techniciens, scientifiques des données, analystes, commerciaux cherchant à intégrer Cassandra dans leur environnement actuel
■ Pré-requis :	Connaissances des principes fondamentaux des bases de données, de SQL et de langage de programmation Java, sont fortement conseillées
■ Objectifs :	- Structurer et concevoir des bases de données Cassandra pour garder une longueur d'avance sur vos concurrents - Appliquer des modèles de requêtes pour modéliser les données de vos bases Cassandra - Accéder aux bases de données Cassandra avec CQL et Java - Trouver le juste équilibre entre le débit de lecture / écriture et la cohérence des données - Intégrer Cassandra à Hadoop, Pig et Hive - Mettre en oeuvre les design patterns Cassandra les plus courants
■ Modalités pédagogiques, techniques et d'encadrement :	• Formation synchrone en présentiel et distanciel. • Méthodologie basée sur l'Active Learning : 75 % de pratique minimum. • Un PC par participant en présentiel, possibilité de mettre à disposition en bureau à distance un PC et l'environnement adéquat. • Un formateur expert.
■ Modalités d'évaluation :	• Définition des besoins et attentes des apprenants en amont de la formation. • Auto-positionnement à l'entrée et la sortie de la formation. • Suivi continu par les formateurs durant les ateliers pratiques. • Évaluation à chaud de l'adéquation au besoin professionnel des apprenants le dernier jour de formation.

■ Sanction :	Attestation de fin de formation mentionnant le résultat des acquis
■ Référence :	BUS100296-F
■ Note de satisfaction des participants:	Pas de données disponibles
■ Contacts :	commercial@dawan.fr - 09 72 37 73 73
■ Modalités d'accès :	Possibilité de faire un devis en ligne (www.dawan.fr, moncompteformation.gouv.fr, maformation.fr, etc.) ou en appelant au standard.
■ Délais d'accès :	Variable selon le type de financement.
■ Accessibilité :	Si vous êtes en situation de handicap, nous sommes en mesure de vous accueillir, n'hésitez pas à nous contacter à referenthandicap@dawan.fr, nous étudierons ensemble vos besoins

NoSQL : Vue d'ensemble

Pourquoi utiliser des entrepôts de données non relationnels

Les différentes catégories d'entrepôts de données NoSQL

Structure des bases Cassandra

Définir des entrepôts de données avec des familles de colonnes

Interroger Cassandra

Examiner les principaux composants de l'architecture de Cassandra

Interrogation des bases Cassandra

Définir le langage CQL (Cassandra Query Language)

Énumérer les différents types de données CQL

Manipuler les données depuis l'interface cqlsh

Structures et types de données Cassandra

Faire un parallèle avec le modèle relationnel

Organiser les données avec les keyspaces, les tables et les colonnes

Créer des collections et des compteurs

Modélisation des données avec les requêtes

Créer des tables axées sur les modèles (patterns) d'accès
Créer des clusters avec des clés primaires composites
Améliorer la distribution des données avec les clés de partition composites

Optimisation de la cohérence

Identifier les différents niveaux de cohérence
Choisir les niveaux de cohérence en lecture / écriture des données Différencier les fonctionnalités d'ajustement des niveaux de cohérence

Équilibre entre cohérence et performances

Comprendre le lien entre cohérence et facteurs de réplication
Sacrifier la cohérence au profit de la disponibilité
Développer la cohérence linéaire avec Compare-And-Set

Les types de collections Cassandra

Regrouper les éléments dans des ensembles
Classer les éléments dans des listes
Cartographier les relations
Imbriquer les collections

Stockage des données pour faciliter leur récupération

Mapper les données avec des tuples et des types définis par l'utilisateur Comprendre le frozen keyword
Appliquer le pattern Valueless Columns
Mise en œuvre stratégique des colonnes de cluster

Contrôle de la durée de vie des données

Expiration des données temporelles avec time-to-live
Utiliser des tombstones pour les suppressions distribuées
Exécuter ultérieurement des instructions DELETE et UPDATE

Création de vues matérialisées et de séries chronologiques

Modéliser les séries chronologiques
Améliorer les requêtes avec les vues matérialisées
Maintenance des vues matérialisées dans l'application
Analyser les données à partir des vues matérialisées

Gestion des triggers

- Créer des triggers avec ITrigger
- Associer des triggers à des tables
- Gérer les vues matérialisées avec des triggers

Interrogation des données Cassandra avec le pilote Java Datastax

- Se connecter à un cluster Cassandra
- Exécuter des instructions CQL via le pilote Java
- Traiter les instructions préparées par lots
- Paginer les requêtes volumineuses

Persistance des objets Java avec Kundera

- Définir la JPA (Java Persistence Architecture)
- Configurer Kundera pour Cassandra
- Générer des schémas automatiquement
- Gérer les transactions JPA dans Kundera

Connecteurs Cassandra intégrés

- Charger les données dans Hadoop MapReduce avec la fonction Cassandra InputFormat
- Utiliser l'outil de chargement Cassandra pour créer des relations avec Pig
- Convertir une table Cassandra en table Hive avec la sérialisation / désérialisation Cassandra