

Certification Data Analyst : Utiliser Hive et Impala avec Hadoop de Cloudera (CCA Data Analyst)

<https://training.xebia.fr/formation-data/certification-data-analyst-utiliser-hive-et-impala-avec-hadoop-de-cloudera-cca-data-analyst/>



Agile



Agilité à l'échelle



DevOps



Product
Management



Data



JVM/Scala

Certification Data Analyst : Utiliser Hive et Impala avec Hadoop de Cloudera (CCA Data Analyst)

Avec *Bruno Bouchahoua*

3 jours, soit 21 heures

Passez au niveau supérieur en termes de connaissance grâce à la formation Data Analyst de Cloudera.

Cette formation Data Analyst « Hands-On » de trois jours, axée sur **Hive** et **Cloudera Impala**, vous apprendra à appliquer l'analyse traditionnelle des données et les compétences de « business intelligence » aux **Big Data**.

Xebia, partenaire privilégié de Cloudera en France, vous présentera les outils permettant aux participants de manipuler et analyser des ensembles complexes de données en utilisant **SQL** et des langages de script familiers.

Apache Hive rend accessibles les données multi-structurées pour les analystes, administrateurs de bases de données et d'autres profils n'ayant pas d'expertise en **programmation Java**. Cloudera Impala permet une analyse interactive en temps réel des données stockées dans Hadoop via un environnement **SQL** natif.

Au travers de discussions dirigées par le formateur et d'exercices « Hands-On » interactifs, les participants vont naviguer dans l'écosystème Hadoop et adresser des sujets tels que :

- Notions fondamentales de Apache Hadoop et de l'ETL (extract, transform, load), ingestion et traitement avec les outils Hadoop.
- Organiser des données dans un tableau, effectuer des transformations et simplifier les requêtes complexes avec Hive.
- Effectuer des analyses interactives en temps réel sur un ensemble important de données stockées dans HDFS ou HBase en utilisant SQL avec Impala.
- Comment choisir le meilleur outil pour une tâche donnée dans Hadoop, atteindre l'interopérabilité et manager les workflows récurrents.

Programme

Introduction à la formation Hadoop

Les participants navigueront à travers l'écosystème **Hadoop** et aborderont les points suivants :

- Au sujet de cette formation
- Au sujet de Xebia et Cloudera
- Logistique de la formation
- Introductions

Fondamentaux d'Hadoop

- L'intérêt d'Hadoop
- Vue globale d'Hadoop
- HDFS
- MapReduce
- L'écosystème Hadoop
- Explication de scénarios de laboratoire
- Exercices « Hands-On » : ingestion de données avec les outils Hadoop

Introduction à Hive

- Qu'est-ce que Hive ?
- Schema Hive et stockage de données
- Comparer Hive aux bases de données traditionnelles
- Hive vs. Pig
- Cas d'utilisation de Hive
- Interagir avec Hive

Analyse de données relation avec Hive

- Bases de données et tableaux Hive
- Syntaxe HiveQL basique
- Types de données
- Assembler des ensembles de données
- Fonctions communes de Built-in
- Exercice « Hands-On » : « Running Hive Queries on the Shell, Scripts and Hue »

Management de données avec Hive

- Formats de données Hive
- Créer des bases de données et tableaux de management Hive
- « Altering Databases and Tables »
- Tableaux auto-managés
- Simplifier les requêtes avec Views
- Stocker les résultats de requêtes
- Contrôler l'accès aux données
- Exercice « Hands-On » : management des données avec Hive

Traitement de texte avec Hive

- Vue d'ensemble du traitement de texte
- Fonctions String importantes
- Utiliser des expressions habituelles dans Hive
- « Sentiment Analysis » et « N-Grams »
- Exercices « Hands-On » (optionnels) : se faire une idée de l'analyse de sentiment

Optimisation d'Hive

- Comprendre la performance de requête
- Contrôler le plan d'exécution des tâches
- Partitionner
- « Bucketing »
- Indexer les données

Etendre Hive

- SerDes
- Transformation de données avec des Scripts personnalisés
- Fonctions définies par l'utilisateur
- Paramétrer les requêtes
- Exercices « Hands-On » : transformation de données avec Hive

Introduction à Impala

- Qu'est-ce qu'Impala ?
- En quoi Impala diffère d'Hive
- Comment Impala diffère des bases de données relationnelles
- Limitations et directions futures
- Utiliser le Shell Impala

Analyser les données avec Impala

- Syntaxe basique
- Types de données
- Filtrer, trier et restreindre les résultats
- Assembler et grouper les données
- Augmenter les performances d'Impala
- Exercices « Hands-On » : analyse interactive avec Impala

Choisir le meilleur outil pour la tâche

- Comparer MapReduce, Hive, Impala et les bases de données relationnelles
- Lequel choisir ?

À l'issue de cette formation, les stagiaires seront en mesure :

- D'appliquer l'analyse traditionnelle des données et les compétences de « business intelligence » aux Big Data.
- D'utiliser les outils permettant de manipuler et analyser des ensembles complexes de données en utilisant SQL et des langages de script familiers.

Méthodes pédagogiques

50% de travaux pratiques et 50% de théorie.

Les participants apprendront comment naviguer dans l'écosystème Hadoop au travers d'échanges avec le formateur ainsi qu'en réalisant des exercices pratiques sur les sujets suivants :

- Comment les données sont distribuées, stockées et enregistrées dans un cluster Hadoop ?
- Comment utiliser Sqoop et Flume pour intégrer les données ?
- Comment enregistrer des données distribuées avec Apache Spark ?
- Comment modéliser des données structurées en tant que tableau dans Impala et Hive ?
- Comment choisir le meilleur format de stockage de données pour différents patterns d'utilisation de données ?
- Les meilleures pratiques pour le stockage de données.

Tarif HT

Inter : 2 445 €

Prochaines dates de formation

11 -13 juillet

19 – 21 décembre

[Je m'inscris](#)

info@xebia-training.fr

[Programme PDF](#)

[Imprimer le programme](#)

Public visé

Cette formation est parfaitement adaptée aux analystes de données, business analysts, développeurs et administrateurs qui ont de l'expérience avec SQL et les commandes

UNIX ou Linux basiques.

Prérequis

Une connaissance préalable de **Java** et **Apache Hadoop** n'est pas requise.

Les postes de travail et les logiciels nécessaires au bon déroulement de la formation sont fournis par Xebia.

Validation

À la fin de cette formation, les stagiaires recevront une attestation de présence.



SOFTWARE TRAINING DONE RIGHT