

Big Data : intégration SQL, Hive, Spark/Dataframe

Comprendre les connexions et l'intégration de bases de données NoSQL dans un projet Big Data

Filière : **Gestion de données** Sous-filière : **NoSql & Big Data**

RÉFÉRENCE	DURÉE	PRIX UNITAIRE HT
DB-BDI	2 JOURS (14H)	1 850 €

Objectifs pédagogiques

- Comprendre les connexions existantes entre les mondes relationnels et NoSQL en environnement Big Data
- Savoir mettre en oeuvre Hive et Pig, Impala, les Spark Dataframes

Public cible

- Chefs de projets
- Experts en bases de données relationnelles
- Toute personne souhaitant comprendre les connexions existantes entre les mondes relationnels et nosql en environnement big data

Pré-requis

Connaissance générale des systèmes d'informations et des bases de données

Programme de la formation

Présentation

- Besoin
- Adéquation entre les objectifs et les outils
- Faciliter la manipulation de gros volumes de données en conservant une approche utilisateurs
- Rappels sur le stockage : HDFS, Cassandra, HBase et les formats de données : parquet, orc, raw, clés/valeurs
- Les outils : Hive, Impala, Tez, Presto, Drill, Pig, Spark/QL

Hive et Pig

OXiane Institut

98 avenue du général Leclerc
92100 Boulogne-Billancourt

RCS Nanterre 430 112 250 000 21 / Code NAF 6202A
Organisme de formation N° 11 92 16 52 492

- Présentation
- Mode de fonctionnement. Rappel sur map/reduce
- Hive : le langage HiveQL. Exemples
- Pig : le langage pig/latin. Exemples

Impala

- Présentation
- Cadre d'utilisation
- Contraintes
- Liaison avec le métastore Hive

Travaux pratiques : mise en évidence des performances

Presto

- Cadre d'utilisation
- Sources de données utilisables

Travaux pratiques : mise en œuvre d'une requête s'appuyant sur Cassandra et PostgreSQL

Spark DataFrame

- Les différentes approches
- Syntaxe Spark/SQL
- APIs QL
- Compilation catalyst
- Syntaxe, opérateurs

Travaux pratiques : mise en œuvre d'une requête s'appuyant sur HBase et HDFS

Drill

- Utilisation d'APIs JDBC, ODBC
- Indépendance Hadoop
- Contraintes d'utilisation
- Performances

Comparatifs

- Compatibilité ANSI/SQL
- Approches des différents produits
- Critères de choix

OXiane Institut



Qualité

Cette formation est accessible aux personnes en situation de handicap, nous contacter en cas de besoin d'informations complémentaires.



Programme mis à jour le **3 novembre 2023**