EduGroupe
Accompagner pour réussir

Big Data : intégration SQL, hive, SparkDataframes

Référence : **PYCB045** Durée : **2 jours (14 heures)** Certification : **Aucune**

Connaissances préalables

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

Profil des stagiaires

- Experts en bases de données, chefs de projet
- Toute personne souhaitant comprendre le fonctionnement et les apports des bases NoSQL

Objectifs

- Connaître les caractéristiques techniques des bases de données NoSQL, les différentes solutions disponibles
- Identifier les critères de choix

Certification préparée

- Aucune

Méthodes pédagogiques

- 6 à 12 personnes maximum par cours, 1 poste de travail par stagiaire
- Remise d'une documentation pédagogique papier ou numérique pendant le stage
- La formation est constituée d'apports théoriques, d'exercices pratiques et de réflexions

Formateur

- Consultant-Formateur expert Bigdata

Méthodes d'évaluation des acquis

- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation des compétences acquises envoyée au stagiaire
- Attestation de fin de stage adressée avec la facture

Contenu du cours

1. Présentation

- Besoin
- Comment concilier le quoi et le comment ?
- 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

2. Hive et Pig

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

3. Impala

- Présentation
- Cadre d'utilisation
- Contraintes
- Liaison avec le métastore Hive
- Travail Pratique : Mise en évidence des performances

4. Presto

- Cadre d'utilisation
- Sources de données utilisables
- Travail Pratique : Mise en oeuvre d'une requête s'appuyant sur Cassandra, HDFS et PostgreSQL

5. Spark DataFrame

- Les différentes approches
- Syntaxe Spark/SQl
- APIs QL
- Compilation catalyst
- Syntaxe, opérateurs
- Comparaison avec Presto
- Travail Pratique : Mise en oeuvre d'une requête s'appusant sur HBase et HDFS

6. Drill/sqoop

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

7. Comparatifs

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

Notre référent handicap se tient à votre disposition au 01.71.19.70.30 ou par mail à referent.handicap@edugroupe.com pour recueillir vos éventuels besoins d'aménagements, afin de vous offrir la meilleure expérience possible.