

## Big Data : intégration SQL, Hive, Spark/Dataframe

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

### Détails

- Code : DB-BDI
- Durée : 2 jours ( 14 heures )

#### Public

- 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

#### Objectifs

- 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

### Programme

#### 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

- 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 oeuvre 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 oeuvre 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

### Modalité

- Stage pratique en présentiel
- Stage pratique en distanciel
- Nombre de stagiaires minimum : 4
- Nombre de stagiaires maximum : 10

### Méthodes pédagogiques

- Exposés
- Cas pratiques

### Profils des intervenants

- Toutes nos formations sont animées par des consultants-formateurs expérimentés et reconnus par leurs pairs.

### Modalités d'évaluation

- Evaluation des acquis de la formation par le biais de cas pratiques et/ou mises en situation.
- Attestation de formation remise à chaque participant.

### Démarche qualité

- Questionnaire d'évaluation de satisfaction à chaud complété par chaque participant à l'issue de la formation.

### Moyens pédagogiques

- Salle équipée de PC (1 poste par stagiaire), vidéo-projecteur.
- Espace de pause.

Dernière mise à jour le 19/01/2022