



# ORGANISME DE FORMATION AUX TECHNOLOGIES ET METIERS DE L'INFORMATIQUE

## Formation Big Data - Le SQL distribué

N° ACTIVITÉ : 11 92 18558 92

TÉLÉPHONE : 01 85 77 07 07

E-MAIL : inscription@hubformation.com

Langage historique né avec les systèmes de gestion de bases de données relationnelles au début des années 2000, le SQL a été tellement utilisé par les développeurs et autres utilisateurs de solutions de Business Intelligence qu'il fait aujourd'hui figure de standard. Mais l'émergence des bases de données distribuées qui reposent sur des architectures machines en cluster vient changer la donne en rendant l'utilisation traditionnelle du SQL inopérante. C'est pourquoi la communauté Open Source a créé différents outils capables d'exécuter des requêtes SQL sur ces nouveaux systèmes distribués (également appelés bases NoSQL). Les participants à cette formation de 2 jours découvriront les principales solutions aujourd'hui disponibles et seront à même de choisir la plus adaptée à leur problématique spécifique.

### Objectifs

- | Identifier 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

- | Experts en bases de données relationnelles, chefs de projet

### Prérequis

- | 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

- | 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

Référence	BIG552
Durée	2 jours (14h)
Tarif	1 510 €HT

### PROCHAINES SESSIONS

Pour connaître les prochaines dates ou organiser un intra-entreprise, contactez-nous, nous vous répondrons sous 72 heures.

## **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

## **Méthode pédagogique**

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

## **Méthode d'évaluation**

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

---

## **Accessibilité**



Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.  
Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.