

# BigData: intégration SQL, Hive, SparkDataFrames

Action collective: Non

# INFORMATIONS GÉNÉRALES

Type de formation : Formation continue Éligible au CPF : Non

**Domaine :** IA, Big Data et Bases de données **Filière :** Big Data

Rubrique: Hive - Spark

**€** Tarifs

Prix public : 1390 €

Code de formation: BD045

Tarif & financement:

Nous vous accompagnons pour trouver la meilleure solution de financement parmi les suivantes :

Le plan de développement des compétences de votre entreprise : rapprochez-vous de votre service RH.

Le dispositif FNE-Formation.

**L'OPCO** (opérateurs de compétences) de votre entreprise.

**France Travail:** sous réserve de l'acceptation de votre dossier par votre conseiller Pôle Emploi.

**CPF** -MonCompteFormation Contactez nous pour plus d'information : contact@aston-institut.com

# **PRÉSENTATION**

# **Objectifs & compétences**

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 visé

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.

### Pré-requis

Experts en bases de données relationnelles, chefs de projet.

#### **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, SparkQL

#### **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 metastore Hive.

Atelier: mise en évidence des performances.

#### **Presto**

Cadre d'utilisation. Sources de données utilisables.

Atelier: mise en oeuvre d'une requête s'appuyant sur Cassandra et PostgreSQL.

#### Spark DataFrame

Les différentes approches. Syntaxe SparkQL. APIs QL. Compilation catalyst. Syntaxe, opérateurs.

Atelier: mise en oeuvre d'une requête s'appuyant sur HBase et HDFS.

#### Drill

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

#### **© Lieux & Horaires**

**Durée:** 14 heures

**Délai d'accès :** Jusqu'a 8 jours avant le début de la formation, sous condition d'un dossier d'insciption complet

### # Prochaines sessions

Consultez-nous pour les prochaines sessions



#### **Comparatifs**

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

# **MODALITÉS**

#### Modalités

**Modalités :** en présentiel, distanciel ou mixte . Toutes les formations sont en présentiel par défaut mais les salles sont équipées pour faire de l'hybride. – Horaires de 9H à 12H30 et de 14H à 17H30 soit 7H – Intra et Inter entreprise.

**Pédagogie :** essentiellement participative et ludique, centrée sur l'expérience, l'immersion et la mise en pratique. Alternance d'apports théoriques et d'outils pratiques. **Ressources techniques et pédagogiques :** Support de formation au format PDF ou PPT Ordinateur, vidéoprojecteur, Tableau blanc, Visioconférence : Cisco Webex / Teams / Zoom.

**Pendant la formation :** mises en situation, autodiagnostics, travail individuel ou en sous-groupe sur des cas réels.

#### Méthode

Fin de formation : entretien individuel.

Satisfaction des participants : questionnaire de satisfaction réalisé en fin de formation

Assiduité : certificat de réalisation.

**Validations des acquis** : grille d'evalution des acquis établie par le formateur en fin de formation.