Formation Hadoop : développement

N° ACTIVITÉ : 11 92 18558 92 | TÉLÉPHONE : 01 85 77 07 07 | E-MAIL : inscription@hubformation.com

Référence	HCB033
Durée	2 jours (14h)
Tarif	1 670 €HT
Repas	46 €HT(en option)

Objectifs

| Identifier l'écosystème Hadoop Cloudera/Hortonworks
| Présenter les principes du Framework Hadoop
| Mettre en oeuvre des tâches Hadoop pour extraire des éléments pertinents d'ensembles de données volumineux et variés
| Développer des algorithmes parallèles efficaces avec MapReduce
| Charger des données non structurées des systèmes HDFS et HBase

Public

| Développeurs, Chefs de projets, data-scientists, architectes, ...

Prérequis

| Avoir la connaissance d'un langage de programmation objet comme Java et du scripting

Programme de la formation

Comprendre l'écosystème Hadoop
| Les fonctionnalités du framework Hadoop. Les différentes versions.
| Distributions : Apache, Cloudera, Hortonworks, EMR, MapR, DSE.
| Spécificités de chaque distribution.
| Architecture et principe de fonctionnement. Zoom sur la distribution Cloudera/Hortonworks
| Terminologie : NameNode, DataNode, ResourceManager, NodeManager. Rôle des différents composants. Le projet et les modules : Hadoop Common, HDFS, YARN, Spark, MapReduce, Hue, Oozie, Pig, Hive, HBase, Zeppelin, ...
| Atelier : Manipulations de base sur la console Hadoop

Présenter les principes du Framework Hadoop
| Le projet et les modules : Hadoop Common, HDFS, YARN, Spark, MapReduce
| Utilisation de yarn pour piloter les jobs map/reduce.
| Infrastructure/mise en oeuvre : Avro, Ambari, Zookeeper, Pig, Tez, Oozie.
| Vue d'ensemble.
| Gestion des données. Exemple de sqoop.
| Restitution : webhdfs, hive, Hawq, Mahout, ElasticSearch, ...
| Outils complémentaires de traitement : Spark, SparkQL, SparkR, Spark/ML, Storm, BigTop ; outils de développement : Cascading, Scalding, Flink; outils d'analyse : RHadoop, Hama, Chukwa, kafka
| Atelier : exécution de jobs sur la ferme Hadoop

Mettre en oeuvre des tâches Hadoop pour extraire des éléments pertinents d'ensembles de données volumineux et variés
| Lac de données. Construction et utilisation. Exploitation des données du lac.
| Les différents outils : Yarn, MapReduce, Spark, Hive, Pig
| Différentes solutions : calculs en mode batch, ou en temps réel,sur des flux de données ou des données statiques.
| Principe de map/reduce et exemples d'implémentations, langages et sur-couches.
| Découpage des travaux (jobs) avec stockage intermédiaire. Le format parquet.
| Atelier : développement d'un extracteur de données et qualification de la donnée.

SESSIONS PROGRAMMÉES

A DISTANCE (FRA)

du 4 au 5 septembre 2025
du 11 au 12 décembre 2025

PARIS

du 4 au 5 septembre 2025
du 11 au 12 décembre 2025

VOIR TOUTES LES DATES

HUB Formation - 14 avenue de l'Opéra - 75001 Paris
01 85 77 07 07 - www.hubformation.com - SIREN 525 287 785

Développer des algorithmes parallèles efficaces avec MapReduce
| Principe et objectifs du modèle de programmation map/reduce.
| Configuration des jobs, notion de configuration.
| Les interfaces principales : mapper, reducer, fonctions map() et reduce(). Couples (clés, valeurs).
| Implémentation par le framework Hadoop.
| Etude de la collection d'exemples.
| Atelier : Réduction de la donnée extraite précedemment. Recherche et scores.

Charger des données non structurées des systèmes HDFS et HBase
| Format des données : texte, json, csv, parquet, ...
| Format des entrées et sorties d'un job map/reduce : InputFormat et OutputFormat.
| Atelier : type personnalisés : création d'un writable spécifique. Utilisation. Contraintes.
| Accès à des systèmes externes : S3, hdfs, har, hbase, ...
| Outils d'interfaçage entre les différents composants
| Atelier : Ecriture d'un ETL HDFS vers HBase

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.

Suivre cette formation à distance

Voici les prérequis techniques pour pouvoir suivre le cours à distance :

| Un ordinateur avec webcam, micro, haut-parleur et un navigateur (de préférence Chrome ou Firefox). Un casque n'est pas nécessaire suivant l'environnement.
| Une connexion Internet de type ADSL ou supérieure. Attention, une connexion Internet ne permettant pas, par exemple, de recevoir la télévision par Internet, ne sera pas suffisante, cela engendrera des déconnexions intempestives du stagiaire et dérangera toute la classe.
| Privilégier une connexion filaire plutôt que le Wifi.
| Avoir accès au poste depuis lequel vous suivrez le cours à distance au moins 2 jours avant la formation pour effectuer les tests de connexion préalables.
| Votre numéro de téléphone portable (pour l'envoi du mot de passe d'accès aux supports de cours et pour une messagerie instantanée autre que celle intégrée à la classe virtuelle).
| Selon la formation, une configuration spécifique de votre machine peut être attendue, merci de nous contacter.
| Pour les formations incluant le passage d'une certification la dernière journée, un voucher vous est fourni pour passer l'examen en ligne.
| Pour les formations logiciel (Adobe, Microsoft Office...), il est nécessaire d'avoir le logiciel installé sur votre machine, nous ne fournissons pas de licence ou de version test.
| Horaires identiques au présentiel.

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.