

ECOSYSTÈME HADOOP & DISTRIBUTIONS

Hadoop - Cloudera / Hortonworks pour développeurs

REF : SIHA002

DUREE : 21h

Mixte Classe virtuelle

PUBLIC

Cette formation Hadoop - Cloudera / Hortonworks pour développeurs est destinée aux développeurs qui développeront, maintiendront et optimiseront des applications pour Hadoop.

Modalités et délais d'accès : les inscriptions sont fermées 24h avant la 1ère journée de formation.

Accessibilité : Si vous avez des contraintes particulières liées à une situation de handicap, veuillez nous contacter au préalable afin que nous puissions, dans la mesure du possible, adapter l'action de formation.

PREREQUIS

Cette formation Hadoop - Cloudera / Hortonworks pour développeurs nécessite une expérience dans le développement, si possible avec Java. Une connaissance préliminaire d'Hadoop n'est pas exigée.

MODALITES PEDAGOGIQUES

1 poste et 1 support par stagiaire

8 à 10 stagiaires par salle

Remise d'une documentation pédagogique papier ou numérique pendant le stage

La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience

MODALITES D'EVALUATION

Auto-évaluation des acquis par le stagiaire via un questionnaire en ligne

Attestation de fin de stage remise au stagiaire

OBJECTIFS PEDAGOGIQUES

Cette formation Hadoop - Cloudera / Hortonworks pour développeurs vous permettra :

- Réaliser une étude en profondeur dans les développements des applications Hadoop 2.0
- Définir et à développer des applications MapReduce(Yarn), Hive et Pig efficaces et pertinentes avec Hadoop 2.0
- Exploiter toute la puissance d'Hadoop 2.0 pour manipulation, analyse et exécution des calculs sur leur cluster Hadoop

PROGRAMME

Introduction

- Présentation générale d'**Hadoop**
- Exemples d'utilisation dans différents secteurs
- Historique et chiffres clés : Quand parle-t-on de **Big Data** ?

HDFS : interfacement avec Java

- Le système de fichier Hadoop Distributed File System (HDFS)
- Lecture et écriture sur le HDFS
- Manipulation des types de fichier SequenceFile
- Qu'est-ce que le DistributedCache et comment partager les références ?

MapReduce/YARN : principe et développement Java

- Le paradigme **MapReduce** et l'utilisation à travers YARN
- Implémentation de problématiques réelles avec le paradigme **MapReduce**
- Implémentation de problématiques plus complexes
- Utilisation des combinateurs et partitionners
- Les bonnes pratiques pour le développement des applications **MapReduce**

Manipulation des données avec un Query language et Java

- Présentation et fonctionnement de HBase, Hive et PIG
- HBase et Hive : JDBC et requêtes
- Extension du HiveQL avec des fonctionnalités développées : UDF (User-Defined Functions)
- Pig : Développement et intégration avec JAVA
- Extension de PIG avec des fonctionnalités développées : UDF (User-Defined Functions)

Version du : 29/11/2021



