Contactez-nous Tél. : +33 (0)1 83 35 34 40
Email inscription@lepont-learning.com
Site web www.lepont-learning.com

SPARK

Certification Hadoop avec Spark pour Développeurs de Cloudera

REF : SICL003

DUREE : 28h

Présentiel Classe virtuelle

PUBLIC

Cette formation Certification Spark avec Hadoop pour Développeurs de Cloudera est destinée aux développeurs et ingénieurs qui développeront, maintiendront et optimiseront des tâches Hadoop.

Modalités et délais d'accés : les inscriptions sont fernées 24h avant la 1ére 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 Certification Spark avec Hadoop pour Développeurs de Cloudera nécessite d'avoir une expérience dans le développement, si possible avec Scala ou Python.

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

OBJECTIFS PEDAGOGIQUES

Cette formation Certification Spark avec Hadoop pour Développeurs de Cloudera vous permettra de :

- Comment les données sont distribuées, stockées et enregistrées dans un cluster Hadoop ?
- Comment utiliser Sqoop et Flume pour intégrer les données ?
- Comment enregistrer des données distribuées avec Apache Spark ?
- Comment modéliser des données structurées en tant que tableau dans Impala et Hive ?
- Comment choisir le meilleur format de stockage de données pour différents patterns d'utilisation de données ?
- Les meilleures pratiques pour le stockage de données.

PROGRAMME

- Introduction
- Présentation de **Hadoop** et de son écosystème
- L'architecture **Hadoop** et HDFS
- Importer des données relationnelles avec **Apache Sqoop**
- Introduction à **Impala** et **Hive**
- Les formats de données
- La partition des données
- Les données **Apache** Flume
- Les bases de **Spark**
- Rédiger et déployer des applications **Spark**
- La programmation parallèle avec **Spark**
- **Spark** : mise en cache et persistence
- Le traitement de données avec **Spark** et « Common patterns »
- Spark SQL et DataFrames
- Conclusion

Version du : 03/02/2020

stagiaire