

Data Analyse Pig, Hive et Spark

Formation Informatique / Bureautique / Sécurité



Cette formation permet aux participants d'acquérir les compétences et connaissances nécessaires pour utiliser les outils permettant de traiter et d'analyser des données sur Hadoop. Elle leur permettra de développer des compétences en analyse de données en se focalisant sur Pig, Hive et Spark.

OBJECTIFS

- Comprendre ce que sont Hadoop et YARN
- Pouvoir manipuler des données sous Hadoop
- Savoir manipuler les données PIG
- Savoir analyser les données avec HIVE

PUBLIC

Cette formation s'adresse à toute personne souhaitant manipuler et analyser des données dans un système Hadoop.

PRE-REQUIS

- Avoir des connaissances sur les systèmes d'information, les bases de données et les concepts de programmation.

PROGRAMME

Introduction

- Introduction au Big Data - Comprendre les concepts clés et les enjeux du Big Data
- Introduction à Hadoop – Principales distributions de Hadoop
- La plateforme Hadoop

Architecture et composants de la plateforme Hadoop

- HDFS
- NameNode / DataNode / ResourceManager
- Paradigme MapReduce et YARN
- Les technologies émergentes

Traitement des données avec Pig

- Description et caractéristiques de Pig : Présentation Pig, Différence entre Pig et MapReduce, Cas d'utilisation de Pig
- Traitement des données : Modélisation des données, Programmation avec Pig Latin, Transformations dans la syntaxe Pig Latin, Fonctions de chargement et de stockage
- Travaux pratiques

Requêtage des données avec Hive

- Description et caractéristiques de Hive
- Utilisation de Hcatalog
- Analyse des données avec Hive
- Management des données Hive : Formats de données Hive, Création des bases de données et des tableaux de management, Tableaux auto-managés, Simplification des requêtes avec Views, Stockage des résultats de requêtes, Contrôle 'accès aux données
- Traitement de texte avec Hive : Fonctions String, Utilisation des expressions habituelles dans Hive

Apache Spark SQL

- Présentation générale
- Caractéristiques – Architecture
- Les bases de Spark
- DataFrame et DataSets
- Les RDD
- Le SQL Contexte
- Opérations sur le DataFrames et les DataSets
- Comparaison entre Spark SQL et Hive

Horaires, Planning et Déroulement de cette formation

Horaires

- Formation de 9h00 (9h30 le premier jour) à 17h30.
- Deux pauses de 15 minutes le matin et l'après-midi.
- 1 heure de pause déjeuner

DEROULEMENT

- Les horaires de fin de journée sont adaptés en fonction des horaires des trains ou des avions des différents participants.

A retenir

Durée : **3 jours** soit 21h.
Réf. **GKDAPHS**

01 42 93 52 72

Dates des sessions

Cette formation est également proposée en formule **INTRA-ENTREPRISE.**



Inclus dans cette formation



Coaching Après-COURS

Pendant 30 jours, votre formateur sera disponible pour vous aider. CERTyou s'engage dans la réalisation de vos objectifs.

100%
SATISFACTION
GARANTIE

Votre garantie 100% SATISFACTION

Notre engagement 100% satisfaction vous garantit la plus grande qualité de formation.

- Une attestation de suivi de formation vous sera remise en fin de formation.
- Cette formation est organisée pour un maximum de 14 participants.

PROCHAINES FORMATIONS

[Réussir la Certification Gestion de Projet PMP du PMI](#)

[Réussir la Certification PRINCE2 Foundation](#)

[Réussir les Certifications PRINCE2 Foundation et PRINCE2 Practitioner](#)

[Réussir la Certification ITIL Foundation](#)

[Réussir la Certification Agile certifié SCRUM Master](#)

[Réussir les Certifications TOGAF Certified et TOGAF Foundation](#)

Retrouvez cette formation sur notre site :

[Data Analyse Pig, Hive et Spark](#)