



BIG DATA : ANALYSE DE DONNÉES EN ENVIRONNEMENT HADOOP

1950 € HT (tarif inter) | REF : BAS212
TARIF SPÉCIAL : particuliers et demandeurs d'emploi

Hadoop est un framework logiciel open source permettant de stocker des données, et de lancer des applications sur des grappes de ... et prend en charge Microsoft Windows et d'autres composants afin d'étendre la polyvalence du framework pour le traitement et l'analyse de données.

PROGRAMME

Rappel de l'architecture BI

- Rappel de l'architecture décisionnelle.
- Les limites de la BI.

Introduction au BIG Data et à Hadoop

- Définition du Big Data et ses cas d'usages.
- Les 3V du big Data.
- Big Data vs BI classique.
- Présentation de Hadoop.

Les sources de données

- Pourquoi capturer les données ?.
- Les données internes de l'entreprise.
- Les données externes : Web.
- Les données externes : Open Data.
- Les données externes : Les réseaux sociaux.

Data Lake

- Définition du Data Lake.
- Comparaison des deux approches (BI vs Big Data).
- Avantages et points de vigilance.
- Architecture du Data Lake.

NOSQL

- L'origine du NOSQL.
- Définition du NOSQL.
- ACID (SQL) vs CAP (Big Data).
- Les différentes approches NoSQL.

L'écosystème Hadoop

- L'origine et pourquoi Hadoop.
- Le paradigme MapReduce et l'utilisation à travers YARN.
- Le système de fichier HDFS.

Chargement de données sur Hadoop

- Chargement des données depuis une base de données relationnelle.
- Chargement des données depuis Hadoop.
- Utilisation et paramétrage avancée.

Manipulation des données dans un cluster Hadoop

- HUE : Comment fonctionne cette interface Web ?.
- HIVE : paramétrages et requêtes sur HIVE / PIG : fonctionnement de Pig.
- HABSE : qu'est-ce que HABSE ?.
- Sqoop : Utilité de Sqoop / ElasticSearch.
- Oozie : paramétrage de workflow avec Oozie.



3

JOURS

21

HEURES

OBJECTIFS

Connaître les fonctionnements d'Hadoop Distributed File System (HDFS) et YARN / MapReduce
Savoir explorer HDFS et suivre l'exécution d'une application YARN
Déterminer les fonctionnements et l'utilisation des différents outils de manipulation des données

PUBLIC | PRÉREQUIS

PUBLIC

Développeurs, chefs de projet, Data Miners, chargés d'études statistiques, consultants en informatique décisionnelle

PRÉREQUIS

Connaissances de base des modèles relationnels, des statistiques et des langages de programmation
Connaissances de base des concepts de la Business Intelligence

INFOS PRATIQUES

HORAIRE DE LA FORMATION

de 9 h 00 à 12 h 30 et de 13 h 30 à 17 h 00

MÉTHODOLOGIE

PÉDAGOGIQUE

Théorie | Cas pratiques | Synthèse

MODALITÉS D'ÉVALUATION

Évaluation qualitative des acquis tout au long de la formation et appréciation des résultats

DATES ET LIEUX

Aucune session ouverte