

Vue d'ensemble de technologies Big-Data

Programme *(Mis à jour le 27/01/2025)*

Introduction (2H)

- Les origines du Big Data : un monde de données numériques, l'e-santé, chronologie
- Une définition par les quatre V : la provenance des données
- Une rupture : changements de quantité, de qualité, d'habitudes
- La valeur de la donnée : un changement d'importance
- La donnée en tant que matière première
- Le quatrième paradigme de la découverte scientifique

Big Data : traitements depuis l'acquisition jusqu'au résultat (1H30)

- L'enchaînement des opérations
- L'acquisition
- Le recueil des données : crawling, scraping
- La gestion de flux événementiels (Complex Event Processing, CEP)
- L'indexation du flux entrant
- L'intégration avec les anciennes données
- La qualité des données : un cinquième V ?
- Les différents types de traitement : recherche, apprentissage (Machine Learning, transactionnel, data mining)
- D'autres modèles d'enchaînement : Amazon, e-Santé
- Un ou plusieurs gisements de données ? De Hadoop à l'in-memory
- De l'analyse de tonalité à la découverte de connaissances

Relations entre Cloud et Big Data (1H30)

- Le modèle d'architecture des Clouds publics et privés
- Les services XaaS
- Les objectifs et avantages des architectures Cloud
- Les infrastructures
- Les égalités et les différences entre Cloud et Big Data
- Les Clouds de stockage
- Classification, sécurité et confidentialité des données
- La structure comme critère de classification : non structurée, structurée, semi-structurée
- Classification selon le cycle de vie : données temporaires ou permanentes, archives actives
- Difficultés en matière de sécurité : augmentation des volumétries, la distribution

Introduction à l'Open Data (0H30)

- La philosophie des données ouvertes et les objectifs
- La libération des données publiques
- Les difficultés de la mise en œuvre
- Les caractéristiques essentielles des données ouvertes
- Les domaines d'application
- Les bénéfices escomptés

Matériel pour les architectures de stockage (0H30)

- Les serveurs, disques, réseaux et l'usage des disques SSD, l'importance de l'infrastructure réseau

Référence

THBI3495

Durée

1 jour / 7 heures

Prix HT / stagiaire

1250€

Objectifs pédagogiques

- Découvrir les principaux concepts du Big Data
- Identifier les enjeux économiques
- Évaluer les avantages et les inconvénients du Big Data
- Comprendre les principaux problèmes et les solutions potentielles
- Identifier les principales méthodes et champs d'application du Big Data

Niveau requis

- Connaissances de base des architectures techniques

Public concerné

- DSI, directeurs techniques, chefs de projet, architectes, responsables SI

Formateur

Les formateurs intervenants pour Themanis sont qualifiés par notre Responsable Technique Olivier Astre pour les formations informatiques et bureautiques et par Didier Payen pour les formations management.

Conditions d'accès à la formation

Délai : 3 mois à 1 semaine avant le démarrage de la formation dans la limite des effectifs indiqués

Moyens pédagogiques et techniques

Salles de formation (les personnes en situation de handicap peuvent avoir des besoins spécifiques pour suivre la formation. N'hésitez pas à nous contacter pour en discuter) équipée d'un ordinateur de dernière génération par stagiaire, réseau haut débit et vidéo-projection UHD

Documents supports de formation projetés
Apports théoriques, étude de cas concrets et exercices

Mise à disposition en ligne de documents supports à la suite de la formation

Dispositif de suivi de l'exécution de l'évaluation des résultats de la formation

Feuilles d'émargement (signature électronique privilégiée)

Evaluations formatives et des acquis sous forme de questions orales et/ou écrites (QCM) et/ou mises en situation

Questionnaires de satisfaction (enquête électronique privilégiée)

- Les architectures Cloud et les architectures plus traditionnelles
- Les avantages et les difficultés
- Le TCO
 - La consommation électrique : serveurs (IPNM), disques (MAID)
- Le stockage objet : principe et avantages
- Le stockage objet par rapport aux stockages traditionnels NAS et SAN
- L'architecture logicielle
- Niveaux d'implantation de la gestion du stockage
- Le « Software Defined Storage »
- Architecture centralisée (Hadoop File System)
- L'architecture Peer-to-Peer et l'architecture mixte
- Les interfaces et connecteurs : S3, CDMI, FUSE, et avenir des autres stockages (NAS, SAN) par rapport au stockage objet

Protection des données (1H)

- La conservation dans le temps face aux accroissements de volumétrie
- La sauvegarde, en ligne ou locale ?
- L'archive traditionnelle et l'archive active
- Les liens avec la gestion de hiérarchie de stockage : avenir des bandes magnétiques
- La réplication multisites
- La dégradation des supports de stockage

Méthodes de traitement et champs d'application (1H)

- Classification des méthodes d'analyse selon le volume des données et la puissance des traitements
- Hadoop : le modèle de traitement Map Reduce
- L'écosystème Hadoop : Hive, Pig. Les difficultés d'Hadoop
- OpenStack et le gestionnaire de données Ceph
- Le Complex Event Processing : un exemple ? Storm
- Le décisionnel et le transactionnel renouvelés : les bases de données NoSQL. Typologie
- L'ingestion de données et l'indexation
- Les crawlers Open Source
- Recherche et analyse : Elasticsearch
- L'apprentissage : Mahout. In-memory
- Visualisation : temps réel ou non, sur le Cloud (Bime), comparaison QlikView, Tibco Spotfire, Tableau
- Une architecture générale du data mining via le Big Data

Exemples et conclusions durant la journée