Big Data, état de l'art
découvrez les concepts, architectures techniques, les outils et les champs d'application du Big Data

Séminaire de 2 jours - 14h
Réf : BGA - Prix 2024 : 2 090CHF HT

L'accroissement continuel des données numériques dans les organisations a conduit à l'émergence du Big Data. Ce concept recouvre les questions de stockage et de conservation de vastes quantités de données mais aussi celles liées au gisement potentiel de valeur que représentent ces masses d'informations. Ce séminaire présente les problèmes spécifiques du Big Data et les solutions techniques potentielles, de la gestion des données aux différents types de traitement.

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

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

Appréhender les avantages et les contraintes du Big Data

LE PROGRAMME

dernière mise à jour : 11/2022

1) Introduction

- 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.

2) Big Data : traitements depuis l'acquisition jusqu'au résultat

- 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.

3) Relations entre Cloud et Big Data

- Le modèle d'architecture des Clouds publics et privés.

PARTICIPANTS

Directeurs SI, Responsables SI, Chefs de projets, Architectes, Consultants ou toute personne amenée à participer à un projet Big Data.

PRÉREQUIS

Connaissances de base des architectures techniques.

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques…
Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

• Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
• À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
• Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

- 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.
- Les solutions potentielles.

4) Introduction à l'Open Data

- 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.

5) Matériel pour les architectures de stockage

- Les serveurs, disques, réseaux et l'usage des disques SSD, l'importance de l'infrastructure réseau.
- 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, etc.
- Avenir des autres stockages (NAS, SAN) par rapport au stockage objet.

6) Protection des données

- 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.

7) Méthodes de traitement et champs d'application

- 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.
- Du BI au Big Data.
- Le décisionnel et le transactionnel renouvelés : les bases de données NoSQL.Typologie et exemples.
- L'ingestion de données et l'indexation. Deux exemples : Splunk et Logstash.
- 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.

8) Cas d'usage à travers des exemples et conclusion

- L'anticipation : besoins des utilisateurs dans les entreprises, maintenance des équipements.
- La sécurité des personnes, détection de fraudes (postale, taxes), le réseau.
- La recommandation. Analyses marketing et analyses d'impact.
- Analyses de parcours. Distribution de contenu vidéo.
- Le Big Data pour l'industrie automobile ? Pour l'industrie pétrolière ?
- Faut-il se lancer dans un projet Big Data ?
- Quel avenir pour les données ?
- Gouvernance du stockage des données : rôle et recommandations, le Data Scientist, les compétences d'un projet Big Data.

LES DATES

CLASSE À DISTANCE
2024 : 18 juin, 01 oct., 17 déc.