

Big Data - Architectures

Panorama des solutions

 Présentiel ou en classe à distance



2 jours (14 h)

Prix inter : 1.700,00 € HT

Réf.: BD551

La variété et la maturité des offres dédiées au traitement des données de masse posent le problème de la définition d'architectures propres aux spécificités des différents projets Big Data. En effet, en fonction des volumes de données à analyser, de leur fréquence d'analyse, de leur nature et des résultats attendus dépend le choix des briques technologiques à mettre en oeuvre. Cette introduction technique de 2 jours apportera aux participants une vision d'ensemble des différentes briques techniques pouvant entrer en jeu dans les projets Big Data. Il leur permettra de disposer d'une première approche pratique utile dans la définition d'une architecture cible.

A qui s'adresse cette formation ?



Pour qui

- Chefs de projets, architectes, développeurs, data scientists et toute personne souhaitant connaître les outils et solutions pour concevoir et mettre en oeuvre une architecture Big Data



Prérequis

- Posséder une bonne culture générale des systèmes d'information

Programme

1 - Introduction

- L'essentiel du Big Data : calcul distribué, données non structurées
- Besoins fonctionnels et caractéristiques techniques des projets
- La valorisation des données
- Le positionnement respectif des technologies de Cloud, Big Data et NoSQL et les liens, implications
- Concepts clés : ETL, Extract Transform Load, CAP, 3V, 4V, données non structurées, prédictif, Machine Learning
- Exemple d'application : Amazon Rekognition, Polly, EMR
- L'écosystème du Big Data : les acteurs, les produits, état de l'art
- Cycle de vie des projets Big Data
- Emergence de nouveaux métiers : Data scientists, Data labs, Hadoop scientists, CDO, ...
- Rôle de la DSI dans la démarche Big Data
- Gouvernance des données : importance de la qualité des données, fiabilité, durée de validité, sécurité des données
- Aspects législatifs : sur le stockage, la conservation de données, etc ... sur les traitements, la commercialisation des données, des résultats

2 - Stockage distribué

- Caractéristiques NoSQL
- Les différents modes et formats de stockage
- Les types de bases de données : clé/valeur, document, colonne, graphe
- Besoin de distribution
- Définition de la notion d'élasticité
- Principe du stockage réparti
- Définitions : réplication, sharding, gossip, hachage

- Systèmes de fichiers distribués : GFS, HDFS, Ceph
- Les bases de données : Redis, Cassandra, DynamoDB, Accumulo, HBase, MongoDB, BigTable, Neo4j, ...
- Données structurées et non structurées, documents, images, fichiers XML, JSON, CSV, ...
- Moteurs de recherche
- Principe de fonctionnement
- Méthodes d'indexation
- Recherche dans les bases de volumes importants
- Présentation d'Elasticsearch et SolR
- Principe du schemaless, schéma de stockage, clé de distribution, clé de hachage

3 - Calcul et restitution, intégration

- Différentes solutions : calculs en mode batch, ou en temps réel, sur des flux de données ou des données statiques
- Les produits : langage de calculs statistiques, R Statistics Language, sas, RStudio
- Outils de visualisation : Tableau, QlikView
- Ponts entre les outils statistiques et les bases Big Data
- Outils de calcul sur des volumes importants : Kafka/Spark Streaming/Storm en temps réel, Hadoop/Spark en mode batch
- Zoom sur Hadoop : complémentarité de HDFS et YARN
- Restitution et analyse : logstash, kibana, elk, zeppelin
- Principe de map/reduce et exemples d'implémentations, langages et sur-couches
- Présentation de Pig pour la conception de tâches map/reduce sur une grappe Hadoop



Les objectifs de la formation

- Comprendre les concepts essentiels du Big Data et les technologies implémentées
- Savoir analyser les difficultés propres à un projet Big Data, les freins et les apports, tant sur les aspects techniques que sur ceux liés à la gestion du projet



Evaluation

- Cette formation fait l'objet d'une évaluation formative.



Les points forts de la formation

- Une approche exhaustive des concepts et technologies mises en oeuvre dans les architectures Big Data : des outils aux solutions d'hébergement dans le Cloud, tous les aspects sont couverts par le programme.
- Les retours d'expérience et conseils de consultants experts.
- 100% des participants à cette formation se sont déclarés satisfaits ou très satisfaits au cours des 12 derniers mois.



Dates et villes 2026 - Référence BD551



Dernières places disponibles



Session garantie

A distance

du 12 mars au 13 mars

du 10 sept. au 11 sept.

du 26 mai au 27 mai

du 9 nov. au 10 nov.

Paris

du 12 mars au 13 mars

du 10 sept. au 11 sept.

du 26 mai au 27 mai

du 9 nov. au 10 nov.