BIG DATA

Architecture et technologies

Plan de cours N° : 1112

Durée : 2 jours (14h)

Participants

Chef de projets, architecte, développeur, data-scientist. Et toute personne souhaitant connaître les outils et solutions pour concevoir et mettre en oeuvre une architecture BigData.

Pré-Requis

Bonne culture générale sur les systèmes d'information.

Objectifs

Comprendre les concepts essentiels du BigData et les technologies implémentées. Savoir analyser les difficultés propres à un projet BigData, les freins, les apports, tant sur les aspects techniques que sur les points liés à la gestion du projet.

Méthode pédagogique

Alternance entre apports théoriques et exercices pratiques

Support de cours fourni lors de la formation

Moyens d'encadrement mis en oeuvre

1 à 8 personnes maximum par session
1 poste informatique par personne
Une assistance post-formation, d'une durée d'un an, sur le contenu de la formation

Moyens permettant de suivre son exécution et d'en apprécier les résultats

Emargement par demi-journée
Evaluation des acquis par mise en situation de travail
Evaluation qualitative de fin de stage
Remise d'une attestation individuelle de formation en fin de stage

Assistance

formateurs@atp-formation.com

Introduction

- L'essentiel du BigData
  - Calcul distribué, données non structurées
  - Besoins fonctionnels et caractéristiques techniques des projets
  - La valorisation des données
  - Le positionnement respectif des technologies de cloud, BigData et noSQL
  - Liens et 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 BigData
  - Les acteurs, produits, état de l'art
  - Cycle de vie des projets Big Data
- Emergence de nouveaux métiers
  - DataScientist, Data lab, Hadoop scientist, CDO…
- Rôle de la DSI dans la démarche BigData
  - Gouvernance des données
- Aspects législatifs
  - Sur le stockage, la conservation de données
  - Sur les traitements, la commercialisation des données, des résultats

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 répart
- 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 SoIR
- Principe du schemaless
  - Schéma de stockage
  - Clé de distribution
  - Clé de hachage

31, av. du Granier
38240 Meylan
Tél. 04 76 41 14 20
Fax 04 76 41 38 45
contact@atp-formation.com

SAS au Capital de 40.000€ - 390 032 050 RCS GRENOBLE
SIRET 390 032 050 00010 - APE 8559A - TVA Intracom. FR03390032050

BIG DATA

Architecture et technologies

Plan de cours N° : 1112

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 BigData
  - Outils de calcul sur des volumes importants Kafka, Spark, Hadoop …
- Zoom sur Hadoop
  - Complémentarité de HDFS et YARN
- Restitution et analyse
  - Logstash, Kibana, elk, zeppelin
- Principe de map/reduce
  - Exemples d'implémentations
  - Langage et sur-couches
- Présentation de pig pour la conception de tâches map/reduce sur une grappe Hadoop

31, av. du Granier
38240 Meylan
Tél. 04 76 41 14 20
Fax 04 76 41 38 45
contact@atp-formation.com

SAS au Capital de 40.000€ - 390 032 050 RCS GRENOBLE
SIRET 390 032 050 00010 - APE 8559A - TVA Intracom. FR03390032050