

BIGDATA, ARCHITECTURE ET TECHNOLOGIES

Durée

-1 jours

Référence Formation

4-PP-NEW

Objectifs

Comprendre les concepts du BigData et connaître les technologies associées.

Participants

Chefs de projets, architectes, et toute personne souhaitant connaître les outils et solutions pour mettre en place une architecture BigData.

Pré-requis

Bonne culture générale sur les systèmes d'information.

Moyens pédagogiques

Accueil des stagiaires dans une salle dédiée à la formation équipée d'un vidéo projecteur, tableau blanc et paperboard ainsi qu'un ordinateur par participant pour les formations informatiques.

Positionnement préalable oral ou écrit sous forme de tests d'évaluation, feuille de présence signée en demi-journée, évaluation des acquis tout au long de la formation.

En fin de stage : QCM, exercices pratiques ou mises en situation professionnelle, questionnaire de satisfaction, attestation de stage, support de cours remis à chaque participant.

Formateur expert dans son domaine d'intervention

Apports théoriques et exercices pratiques du formateur

Utilisation de cas concrets issus de l'expérience professionnelle des participants

Réflexion de groupe et travail d'échanges avec les participants

Pour les formations à distance : Classe virtuelle organisée principalement avec l'outil ZOOM.

Assistance technique et pédagogique : envoi des coordonnées du formateur par mail avant le début de la formation pour accompagner le bénéficiaire dans le déroulement de son parcours à distance.

PROGRAMME

- Introduction

Le besoin : volumes importants de données, traitements optimisés de flux de données au fil de l'eau (nouvelles technologies et nouveaux usages)

Domaines concernés : recherche scientifique, médical, e-commerce, sécurité

Développement des techniques sur différents aspects : stockage, indexation/recherche, calcul

Définition de ETL : Extract Transform Load

Les acteurs aujourd'hui

- Stockage

Caractéristiques NoSQL :

Structure de données proches des utilisateurs, développeurs

Données structurées et non structurées, documents, images

Fichiers XML, JSON, CSV

Les différents modes et formats de stockage

Stockage réparti : réplication, sharding, gossip protocol, hachage

CAP ÉLAN FORMATION

www.capelanformation.fr - Tél : 04.86.01.20.50

Mail : contact@capelanformation.fr

Organisme enregistré sous le N° 76 34 0908834

[version 2023]

Systèmes de fichiers distribués : GFS, HDFS, BigTable □

Les bases de données

Quelques exemples de produits et leurs caractéristiques : cassandra, MongoDB, CouchDB, DynamoDB...

- Indexation et recherche

Moteurs de recherche

Principe de fonctionnement

Méthodes d'indexation

Exemple de Lucene, et mise en oeuvre avec solr

Recherche dans les bases de volumes importants

Exemples de produits et comparaison : dremel, drill, elasticsearch, MapReduce □

- 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

Outils de calcul sur des volumes importants : storm en temps réel, hadoop en mode batch

Zoom sur Hadoop : complémentarité de HDMS et MapReduce

- Evolutions

Les offres Saas BigData, type Google BigQuery

Les limites actuelles. Les avancées annoncées