



Formation Building Batch Data Analytics Solutions on AWS

N° ACTIVITÉ : 11 92 18558 92

TÉLÉPHONE : 01 85 77 07 07

E-MAIL : inscription@hubformation.com

Dans ce cours, vous apprendrez à créer des solutions d'analyse de données par lots à l'aide d'Amazon EMR, un service géré Apache Spark et Apache Hadoop de niveau entreprise. Vous apprendrez comment Amazon EMR s'intègre à des projets open source tels qu'Apache Hive, Hue et HBase, et à des services AWS tels qu'AWS Glue et AWS Lake Formation.

Le cours aborde les composants de collecte, d'ingestion, de catalogage, de stockage et de traitement des données dans le contexte de Spark et Hadoop. Vous apprendrez à utiliser les blocs-notes EMR pour prendre en charge les charges de travail d'analyse et d'apprentissage automatique. Vous apprendrez également à appliquer les meilleures pratiques en matière de sécurité, de performances et de gestion des coûts au fonctionnement d'Amazon EMR.

Objectifs

- | Comparer les fonctionnalités et les avantages des entrepôts de données, des lacs de données et des architectures de données modernes
- | Concevoir et mettre en oeuvre une solution d'analyse de données par lots
- | Identifier et appliquer les techniques appropriées, y compris la compression, pour optimiser le stockage des données
- | Sélectionner et déployer les options appropriées pour ingérer, transformer et stocker des données
- | Choisir les types d'instance et de noeud, les clusters, la mise à l'échelle automatique et la topologie de réseau appropriés pour un cas d'utilisation métier particulier
- | Identifier comment le stockage et le traitement des données affectent les mécanismes d'analyse et de visualisation nécessaires pour obtenir des informations commerciales exploitables
- | Sécuriser les données au repos et en transit
- | Surveiller les charges de travail analytiques pour identifier et résoudre les problèmes
- | Appliquer les meilleures pratiques de gestion des coûts

Prérequis

- | Avoir suivi les cours :
- | AWS Technical Essentials ou Architecting sur AWS
- | Building Data Lakes on AWS

Public

- | Ingénieurs plateformes de données
- | Architectes et opérateurs qui construisent et gèrent des pipelines d'analyse de données

Programme de la formation

Présentation de l'analyse des données et du pipeline de données

- | Cas d'utilisation de l'analyse de données
- | Utilisation du pipeline de données pour l'analyse

Référence	AWSBBDAS
Durée	1 jour (7h)
Tarif	700 €HT

SESSIONS PROGRAMMÉES

A DISTANCE (FRA)

- le 6 mai 2024
- le 3 juin 2024
- le 8 juillet 2024
- le 12 août 2024
- le 23 septembre 2024
- le 14 octobre 2024
- le 18 novembre 2024
- le 16 décembre 2024

[VOIR TOUTES LES DATES](#)

Présentation d'Amazon EMR

- | Utilisation d'Amazon EMR dans les solutions d'analyse
- | Architecture de cluster Amazon EMR
- | Stratégies de gestion des coûts

Pipeline d'analyse de données à l'aide d'Amazon EMR : ingestion et stockage

- | Optimisation du stockage avec Amazon EMR
- | Techniques d'ingestion de données

Analyse de données par lots hautes performances à l'aide d'Apache Spark sur Amazon EMR

- | Cas d'utilisation d'Apache Spark sur Amazon EMR
- | Pourquoi Apache Spark sur Amazon EMR
- | Concepts de Spark
- | Transformation, traitement et analytique
- | Utilisation de blocs-notes avec Amazon EMR
- | Mise en pratique 1 : Analyse de données à faible latence à l'aide d'Apache Spark sur Amazon EMR

Traitement et analyse des données de lot avec Amazon EMR et Apache Hive

- | Utilisation d'Amazon EMR avec Hive pour traiter les données par lots
- | Transformation, traitement et analytique
- | Introduction à Apache HBase sur Amazon EMR
- | Mise en pratique 2 : traitement de données par lots à l'aide d'Amazon EMR avec Hive

Méthode pédagogique

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

Méthode d'évaluation

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

Suivre cette formation à distance

Voici les prérequis techniques pour pouvoir suivre le cours à distance :

- | Un ordinateur avec webcam, micro, haut-parleur et un navigateur (de préférence Chrome ou Firefox). Un casque n'est pas nécessaire suivant l'environnement.
- | Une connexion Internet de type ADSL ou supérieure. Attention, une connexion Internet ne permettant pas, par exemple, de recevoir la télévision par Internet, ne sera pas suffisante, cela engendrera des déconnexions intempestives du stagiaire et dérangera toute la classe.
- | Privilégier une connexion filaire plutôt que le Wifi.
- | Avoir accès au poste depuis lequel vous suivrez le cours à distance au moins 2 jours avant la formation pour effectuer les tests de connexion préalables.
- | Votre numéro de téléphone portable (pour l'envoi du mot de passe d'accès aux supports de cours et pour une messagerie instantanée autre que celle intégrée à la classe virtuelle).
- | Selon la formation, une configuration spécifique de votre machine peut être attendue, merci de nous contacter.
- | Pour les formations incluant le passage d'une certification la dernière journée, un voucher vous est fourni pour passer l'examen en ligne.
- | Pour les formations logiciel (Adobe, Microsoft Office...), il est nécessaire d'avoir le logiciel installé sur votre machine, nous ne fournissons pas de licence ou de version test.
- | Horaires identiques au présentiel.

Accessibilité

Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.

Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.