[sfeir] Institute

AWS | AWSBIGDATA

Big Data on AWS

Utiliser la plate-forme Big Data d'AWS pour traiter les données et créer des environnements de Big Data

3 jour(s) / 21h

Objectifs pédagogiques

- Utiliser Apache Hadoop avec Amazon EMR
- Lancer et configurer un cluster Amazon EMR
- Utilisez des frameworks de programmation communs pour Amazon EMR, notamment Hive, Pig et Streaming
- Utilisez Hue pour améliorer la facilité d'utilisation d'Amazon EMR
- Utilisez les analyses en mémoire avec Spark sur Amazon EMR
- Comprendre comment des services comme AWS Glue, Amazon Kinesis, Amazon Redshift, Amazon Athena et Amazon QuickSight peuvent être utilisés avec des charges de travail Big Data

Public cible

- Les personnes responsables de la conception et de la mise en œuvre de solutions Big Data, à savoir les architectes de solutions et les administrateurs SysOps
- Data Scientists et Data Analysts intéressés à en savoir plus sur les solutions Big Data sur AWS

Prérequis

- Connaissance de base des technologies Big Data, notamment Apache Hadoop, HDFS et les requêtes SQL/NoSQL
- Formation numérique gratuite Data Analytics Fundamentals ou expérience équivalente
- Connaissance pratique des services AWS de base et de la mise en œuvre du cloud public
- Avoir suivi la formation en classe AWS Technical Essentials ou posséder une expérience équivalente
- Compréhension de base de l'entreposage de données, des systèmes de bases de données relationnelles et de la conception de bases de données

Programme

Jour 1

Module 1 : Présentation du Big Data

- Qu'est-ce que le big data
- Le pipeline big data
- Principes architecturaux du Big Data

Module 2 : Ingestion et transfert Big Data

- Présentation : Ingestion de données
- Transfert de données

Module 3 : Streaming Big Data et Amazon Kinesis

- Traitement de flux de données volumineuses
- Amazon Kinesis
- Amazon Kinesis Data Firehose
- Flux vidéo Amazon Kinesis
- Analyse de données Amazon Kinesis
- Atelier pratique 1 : Diffusion et traitement des logs du serveur Apache à l'aide d'Amazon Kinesis

Module 4 : Solutions de stockage de Big Data

- Options de stockage de données AWS
- Concepts de solutions de stockage
- Facteurs dans le choix d'un magasin de données

Module 5 : Traitement et analyse Big Data

- Traitement et analyse de données volumineuses
- Amazon Athena
- Atelier pratique 2 : Utilisation d'Amazon Athena pour analyser les données de journal

Jour 2

Module 6 : Apache Hadoop et Amazon EMR

- Introduction à Amazon EMR et Apache Hadoop
- Bonnes pratiques pour l'ingestion de données
- Amazon EMR
- Architecture Amazon EMR
- Atelier pratique 3 : Stockage et interrogation de données sur Amazon DynamoDB

Module 7 : Utilisation d'Amazon EMR

- Développer et exécuter votre application
- Lancement de votre cluster
- Gestion de la sortie de vos travaux terminés

Module 8 : Frameworks de programmation Hadoop

- Frameworks Hadoop
- Autres frameworks à utiliser sur Amazon EMR
- Atelier pratique 4 : Traitement des journaux de serveur avec Hive sur Amazon EMR

Module 9 : Interfaces Web sur Amazon EMR

- Hue sur Amazon EMR
- Surveillance de votre cluster
- Atelier pratique 5 : Exécution de scripts Pig dans Hue sur Amazon EMR

Module 10 : Apache Spark sur Amazon EMR

- Apache Spark
- Utilisation de Spark
- Atelier pratique 6 : Traiter les données de NY Taxi à l'aide d'Apache Spark

Jour 3

Module 11 : Utilisation d'AWS Glue pour automatiser les charges de travail ETL

- Qu'est-ce qu'AWS Glue ?
- AWS Glue : Orchestration des tâches

Module 12 : Amazon Redshift et les mégadonnées

- Entrepôts de données vs bases de données traditionnelles
- Amazon Redshift
- Architecture Amazon Redshift

Module 13 : Sécuriser vos déploiements Amazon

- Sécuriser vos déploiements Amazon
- Présentation de la sécurité Amazon EMR
- Présentation d'AWS Identity and Access Management (IAM)
- Sécurisation des données
- Présentation de la sécurité Amazon Kinesis
- Présentation de la sécurité d'Amazon DynamoDB
- Présentation de la sécurité Amazon Redshift

Module 14 : Gérer les coûts du Big Data

- Considérations relatives au coût total pour Amazon EMR
- Modèles de tarification Amazon EC2
- Modèles de tarification Amazon Kinesis
- Considérations de coût pour Amazon DynamoDB
- Considérations sur les coûts et modèles de tarification pour Amazon Redshift
- Optimisation des coûts avec AWS

Module 15 : Visualiser et orchestrer le Big Data

- Visualisation du big data
- Amazon QuickSight
- Orchestrer un workflow big data
- Atelier pratique 7 : Utiliser TIBCO Spotfire pour visualiser les données

Module 16 : Modèles de conception Big Data

- Architectures communes

Module 17 : Conclusion du cours

- Et après?