

Building Batch Data Analytics Solutions on AWS

Cursusduur: 1 Dag Cursuscode: GK7378

Beschrijving:

In deze cursus leert u oplossingen voor batchgegevensanalyse te bouwen met behulp van Amazon EMR, een beheerde Apache Spark- en Apache Hadoop-service op bedrijfsniveau. Je leert hoe Amazon EMR integreert met open-sourceprojecten zoals Apache Hive, Hue en HBase, en met AWS-services zoals AWS Glue en AWS Lake Formation. De cursus behandelt het verzamelen, opnemen, catalogiseren, opslaan en verwerken van componenten voor gegevensverwerking in de context van Spark en Hadoop. U leert EMR Notebooks te gebruiken om zowel analyse- als machine learning-workloads te ondersteunen. Je leert ook om best practices op het gebied van beveiliging, prestaties en kostenbeheer toe te passen op de werking van Amazon EMR.

Doelgroep:

Deze cursus is bedoeld voor: Ingenieurs van dataplatforms Architecten en operators die pijplijnen voor gegevensanalyse bouwen en beheren

Doelstelling:

- In deze cursus leer je:
- Vergelijk de functies en voordelen van datawarehouses, data lakes en moderne data-architecturen
- Ontwerp en implementeer een oplossing voor batchgegevensanalyse
- Identificeer en pas de juiste technieken, waaronder compressie, toe om de gegevensopslag te optimaliseren
- Selecteer en implementeer de juiste opties om gegevens op te nemen, te transformeren en op te slaan
- Kies de juiste instantie- en knooppunttypen, clusters, automatisch schalen en netwerktopologie voor een bepaald zakelijk gebruiksscenario
- Begrijp hoe gegevensopslag en -verwerking van invloed zijn op de analyse- en visualisatiemechanismen die nodig zijn om bruikbare zakelijke inzichten te verkrijgen
- Gegevens beveiligen in rust en tijdens overdracht
- Bewaak analyseworkloads om problemen te identificeren en op te lossen
- Best practices voor kostenbeheer toepassen

Vereiste kennis en vaardigheden:

■

Cursusinhoud:

Module A: Overzicht van data-analyse en de datapijplijn

- Gebruiksscenario's voor gegevensanalyse
- De gegevenspijplijn gebruiken voor analyses

Module 1: Inleiding tot Amazon EMR

- Amazon EMR gebruiken in analyseoplossingen
- Architectuur van Amazon EMR-clusters
- Interactieve demo 1: Lancering van een Amazon EMR-cluster
- Strategieën voor kostenbeheersing

Module 2: Pijplijn voor gegevensanalyse met behulp van Amazon EMR: opname en opslag

- Opslagoptimalisatie met Amazon EMR
- Technieken voor gegevensopname

Module 3: Krachtige analyse van batchgegevens met behulp van Apache Spark op Amazon EMR

- Apache Spark op Amazon EMR-gebruiksscenario's
- Waarom Apache Spark op Amazon EMR
- Spark concepten
- Interactieve demo 2: Maak verbinding met een EMR-cluster en voer Scala-opdrachten uit met behulp van de
- De shell van de vonk
- Transformatie, verwerking en analyse
- Notebooks gebruiken met Amazon EMR
- Praktijklab 1: Gegevensanalyse met lage latentie met behulp van Apache Spark op Amazon EMR

Module 4: Batchgegevens verwerken en analyseren met Amazon EMR en Apache Hive

- Amazon EMR gebruiken met Hive om batchgegevens te verwerken
- Transformatie, verwerking en analyse
- Praktijklab 2: Batchgegevensverwerking met behulp van Amazon EMR met Hive
- Inleiding tot Apache HBase op Amazon EMR

Module 5: Serverloze gegevensverwerking

- Serverloze gegevensverwerking, -transformatie en -analyse
- AWS Glue gebruiken met Amazon EMR-workloads
- Practice Lab 3: Gegevensverwerking in Spark orkestreren met behulp van AWS Step Functions

Module 6: Beveiliging en monitoring van Amazon EMR-clusters

- EPD-clusters beveiligen
- Interactieve Demo 3: Client-side encryptie met EMRFS
- Amazon EMR-clusters bewaken en problemen oplossen
- Demo: De geschiedenis van Apache Spark-clusters bekijken

Module 7: Oplossingen voor batchgegevensanalyse ontwerpen

- Gebruiksscenario's voor batchgegevensanalyse
- Activiteit: Een workflow voor batchgegevensanalyse ontwerpen

Module B: Moderne data-architecturen ontwikkelen op AWS

- Moderne data-architecturen

Nadere informatie:

Neem voor nadere informatie of boekingen contact op met onze Customer Service Desk 030 - 60 89 444

info@globalknowledge.nl

www.globalknowledge.com/nl-nl/

Iepenhoeve 5, 3438 MR Nieuwegein