

Building Data Lakes on AWS

Duration: 1 Day **Course Code: GK7377** **Delivery Method: Virtuel deltagelse**

Overview:

In this course, you will learn how to build an operational data lake that supports analysis of both structured and unstructured data. You will learn the components and functionality of the services involved in creating a data lake. You will use AWS Lake Formation to build a data lake, AWS Glue to build a data catalog, and Amazon Athena to analyze data. The course lectures and labs further your learning with the exploration of several common data lake architectures.

Course level: Intermediate

Duration: 1 day

Virtuel deltagelse

Et V&C Select kursus indholder nøjagtig det samme som et almindeligt kursus. Før kursusstart modtager man kursusmaterialet. Dernæst logger man på kurset via internettet og ser via sin pc den selvsamme præsentation som de øvrige deltagere, man kommunikerer via chat med underviseren og de øvrige deltagere på kurset. Denne uddannelsesmodel er både tids-og omkostningsbesparende og kan være et oplagt alternativ til almindelig klasseundervisning, hvis man f.eks. har et begrænset rejsebudget.

Target Audience:

This course is intended for Data platform engineers, Solutions architects, & IT professionals

Objectives:

- In this course, you will learn to:
 - Secure a data lake with appropriate permission
 - Apply data lake methodologies in planning and designing a data lake
 - Ingest, store, and transform data in a data lake
 - Articulate the components and services required for building an AWS data lake
 - Query, analyze, and visualize data within a data lake
-

Prerequisites:

We recommend that attendees of this course have:

- One year of experience building data analytics pipelines or have completed the Data Analytics Fundamentals digital course
 - GK4501 - AWS Technical Essentials
-

Content:

Module 1: Introduction to data lakes

- Describe the value of data lakes
- Compare data lakes and data warehouses
- Describe the components of a data lake
- Recognize common architectures built on data lakes

Module 2: Data ingestion, cataloging, and preparation

- Describe the relationship between data lake storage and data ingestion
- Describe AWS Glue crawlers and how they are used to create a data catalog
- Identify data formatting, partitioning, and compression for efficient storage and query
- Lab 1: Set up a simple data lake

Module 3: Data processing and analytics

- Recognize how data processing applies to a data lake
- Use AWS Glue to process data within a data lake
- Describe how to use Amazon Athena to analyze data in a data lake

Module 4: Building a data lake with AWS Lake Formation

- Describe the features and benefits of AWS Lake Formation
- Use AWS Lake Formation to create a data lake
- Understand the AWS Lake Formation security model
- Lab 2: Build a data lake using AWS Lake Formation

Module 5: Additional Lake Formation configurations

- Automate AWS Lake Formation using blueprints and workflows
- Apply security and access controls to AWS Lake Formation
- Match records with AWS Lake Formation FindMatches
- Visualize data with Amazon QuickSight
- Lab 3: Automate data lake creation using AWS Lake Formation blueprints
- Lab 4: Data visualization using Amazon QuickSight

Module 6: Architecture and course review

- Post course knowledge check
- Architecture review
- Course review

Further Information:

For More information, or to book your course, please call us on tlf.nr.: 44 88 18 00

training@globalknowledge.dk

www.globalknowledge.com/da-dk/

Global Knowledge, Stamholmen 110, 2650 Hvidovre