

## Data Warehousing on AWS

**Duration: 3 Days**    **Course Code: GK4375**    **Delivery Method: Virtual Learning**

### Overview:

Data Warehousing on AWS introduces you to concepts, strategies, and best practices for designing a cloud-based data warehousing solution using Amazon Redshift. This course demonstrates how to ingest, store, and transform data in the data warehouse. Topics covered include: the purpose of Amazon Redshift, how Amazon Redshift addresses business and technical challenges, features and capabilities of Amazon Redshift, designing a Data Warehousing Solution on AWS by applying best practices based on the Well-Architected Framework, integration with AWS and non-AWS products and services, performance tuning, orchestration, and securing and monitoring Amazon Redshift.

Course level: Advanced

Duration: 3 days

#### Activities

This course includes presentations, hands-on labs, and demonstrations.

#### Virtual Learning

This interactive training can be taken from any location, your office or home and is delivered by a trainer. This training does not have any delegates in the class with the instructor, since all delegates are virtually connected. Virtual delegates do not travel to this course, Global Knowledge will send you all the information needed before the start of the course and you can test the logins.

### Target Audience:

This course is intended for:

- Data engineers
- Data architects
- Database architects
- Database administrators
- Database developers

### Objectives:

- In this course, you will learn to:
- Describe Amazon Redshift architecture and its roles in a modern data architecture
- Design and implement a data warehouse in the cloud using Amazon Redshift
- Identify and load data into an Amazon Redshift data warehouse from a variety of sources
- Analyze data using SQL QV2 notebooks
- Design and implement a disaster recovery strategy for an Amazon Redshift data warehouse
- Perform maintenance and performance tuning on an Amazon Redshift data warehouse
- Secure and manage access to an Amazon Redshift data warehouse
- Share data between multiple Redshift clusters in an organization
- Orchestrate workflows in the data warehouse using AWS Step Functions state machines
- Create an ML model and configure predictors using Amazon Redshift ML

### Prerequisites:

We recommend that attendees of this course have completed the following courses:

- Fundamentals of Analytics on AWS – Part 1 (Digital course)
- Fundamentals of Analytics on AWS – Part 2 (Digital course)
- Building Data Lakes on AWS (Instructor led Training)

## Content:

### Day 1

#### Module 1: Data Warehouse Concepts

- Modern data architecture
- Introduction to the course story
- Data warehousing with Amazon Redshift
- Amazon Redshift Serverless architecture
- Hands-On Lab: Launch and Configure an Amazon Redshift Serverless Data Warehouse

#### Module 2: Setting up Amazon Redshift

- Data models for Amazon Redshift
- Data management in Amazon Redshift
- Managing permissions in Amazon Redshift
- Hands-On Lab: Setting up a Data Warehouse using Amazon Redshift Serverless

#### Module 3: Loading Data

- Overview of data sources
- Loading data from Amazon Simple Storage Service (Amazon S3)
- Extract, transform, and load (ETL) and extract, load, and transform (ELT)
- Loading streaming data
- Loading data from relational databases
- Hands-On Lab: Populating the data warehouse

### Day 2

#### Module 4: Deep Dive into SQL Query Editor v2 and Notebooks

- Features of Amazon Redshift Query Editor v2
- Demonstration: Using Amazon Redshift Query Editor v2
- Advanced queries
- Hands-On Lab: Data Wrangling on AWS

#### Module 5: Backup and Recovery

- Disaster recovery
- Backing up and restoring Amazon Redshift provisioned
- Backing up and restoring Amazon Redshift Serverless

#### Module 6: Amazon Redshift Performance Tuning

- Factors that impact query performance
- Table maintenance and materialized views
- Query analysis
- Workload management
- Tuning guidance
- Amazon Redshift monitoring
- Hands-On Lab: Performance Tuning the Data Warehouse

#### Module 7: Securing Amazon Redshift

- Introduction to Amazon Redshift security and compliance
- Authentication with Amazon Redshift
- Access control with Amazon Redshift
- Data encryption with Amazon Redshift
- Auditing and compliance with Amazon Redshift
- Hands-On Lab: Securing Amazon Redshift

### Day 3

#### Module 8: Orchestration

- Overview of data orchestration
- Orchestration with AWS Step Functions
- Orchestration with Amazon Managed Workflows for Apache Airflow (MWAA)
- Hands-On Lab: Orchestrating the Data Warehouse Pipeline

#### Module 9: Amazon Redshift ML

- Machine Learning Overview
- Getting started with Amazon Redshift ML
- Amazon Redshift ML workflow scenarios
- Amazon Redshift ML Usage
- Hands-On Lab: Predicting customer churn with Amazon Redshift ML

#### Module 10: Amazon Redshift Data Sharing

- Overview of data sharing in Amazon Redshift
- Amazon DataZone for Data as a service

#### Module 11: Wrap-Up

- Hands-On Lab: End of course challenge lab

## Further Information:

For More information, or to book your course, please call us on 0800/84.009

[info@globalknowledge.be](mailto:info@globalknowledge.be)

[www.globalknowledge.com/en-be/](http://www.globalknowledge.com/en-be/)