BUILDING MODERN DATA ANALYTICS SOLUTIONS ON AWS

Course Code: 821673

Gain the skills to modernize data architecture from expert AWS instructors

The Building Modern Data Analytics Solutions on AWS collection of one-day, intermediate level instructor-led courses dives deep into Amazon Lake Formation, Amazon Glue, Amazon EMR, Amazon Kinesis, and Amazon Redshift and the current thinking in building and operating data analytics pipelines to turn data into insights.

What You'll Learn

In the four-course Building Modern Data Analytics Solutions on AWS collection, you'll receive comprehensive training for developing modern data skills, including:

- How to leverage AWS data Services to store, process, analyze, stream, and query data to make decisions with speed and agility at scale
- How to modernize data solutions end to end
- Skills to put your data to work to make better, more informed decisions, respond faster to the unexpected, and uncover new opportunities

Who Needs to Attend

- Data warehouse engineers
- Data platform engineers
- Solutions architects

Prerequisites

AWS Technical Essentials introduces you to AWS products, services, and common solutions. It provides you with fundamentals to become more proficient in identifying AWS services.

BUILDING MODERN DATA ANALYTICS SOLUTIONS ON AWS

Course Code: 821673

CLASSROOM LIVE	$2,700 USD	4 Day

Classroom Live Outline

Day 1: Building Data Lakes on AWS (GK 821336)

- **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

Day 2: Building Batch Data Analytics Solutions on AWS (GK 821564)

- **Module 1: Overview of Data Analytics and the Data Pipeline**
  - Data analytics use cases
  - Using the data pipeline for analytics
  - Module 1: Introduction to Amazon EMR
  - Using Amazon EMR in analytics solutions
  - Amazon EMR cluster architecture
  - Interactive Demo 1: Launching an Amazon EMR cluster
  - Cost management strategies
- **Module 2: Data Analytics Pipeline Using Amazon EMR: Ingestion and Storage**
  - Storage optimization with Amazon EMR
  - Data ingestion techniques
- **Module 3: High-Performance Batch Data Analytics Using Apache Spark on Amazon EMR**
  - Apache Spark on Amazon EMR use cases
  - Why Apache Spark on Amazon EMR
  - Spark concepts
  - Interactive Demo 2: Connect to an EMR cluster and perform Scala commands using the Spark shell
  - Transformation, processing, and analytics
  - Using notebooks with Amazon EMR
- **Module 4: Processing and Analyzing Batch Data with Amazon EMR and Apache Hive**
  - Using Amazon EMR with Hive to process batch data
  - Transformation, processing, and analytics
  - Introduction to Apache HBase on Amazon EMR
- **Module 5: Serverless Data Processing**
  - Serverless data processing, transformation, and analytics
  - Using AWS Glue with Amazon EMR workloads
  - Practice Lab 3: Orchestrate data processing in Spark using AWS Step Functions
- **Module 6: Security and Monitoring of Amazon EMR Clusters**
  - Securing EMR clusters
  - Interactive Demo 3: Client-side encryption with EMRFS
  - Monitoring and troubleshooting Amazon EMR clusters
  - Demo: Reviewing Apache Spark cluster history
- **Module 7: Designing Batch Data Analytics Solutions**
  - Batch data analytics use cases

- Activity: Designing a batch data analytics workflow
- Module B: Developing Modern Data Architectures on AWS
- Modern data architectures

Day 3: Building Streaming Data Analytics Solutions on AWS (GK 821672)

- **Module A: Overview of Data Analytics and the Data Pipeline**
  - Data analytics use cases
  - Using the data pipeline for analytics
- **Module 1: Using Amazon Redshift in the Data Analytics Pipeline**
  - Why Amazon Redshift for data warehousing?
  - Overview of Amazon Redshift
- **Module 2: Introduction to Amazon Redshift**
  - Amazon Redshift architecture
  - Interactive Demo 1: Touring the Amazon Redshift console
  - Amazon Redshift features
  - Practice Lab 1: Load and query data in an Amazon Redshift cluster
- **Module 3: Ingestion and Storage**
  - Ingestion
  - Interactive Demo 2: Connecting your Amazon Redshift cluster using a Jupyter notebook with Data API
  - Data distribution and storage
  - Interactive Demo 3: Analyzing semi-structured data using the SUPER data type
  - Querying data in Amazon Redshift
  - Practice Lab 2: Data analytics using Amazon Redshift Spectrum
- **Module 4: Processing and Optimizing Data**
  - Data transformation
  - Advanced querying
  - Practice Lab 3: Data transformation and querying in Amazon Redshift
  - Resource management
  - Interactive Demo 4: Applying mixed workload management on Amazon Redshift
  - Automation and optimization
  - Interactive demo 5: Amazon Redshift cluster resizing from the dc2.large to ra3.xlplus cluster
- **Module 5: Security and Monitoring of Amazon Redshift Clusters**
  - Securing the Amazon Redshift cluster
  - Monitoring and troubleshooting Amazon Redshift clusters
  - Module 6: Designing Data Warehouse Analytics Solutions
  - Data warehouse use case review
  - Activity: Designing a data warehouse analytics workflow
- **Module B: Developing Modern Data Architectures on AWS**
  - Modern data architectures

Day 4: Building Data Analytics Solutions Using Amazon Redshift (GK 821497)

- **Module A: Overview of Data Analytics and the Data Pipeline**

- Data analytics use cases
- Using the data pipeline for analytics

- **Module 1: Using Amazon Redshift in the Data Analytics Pipeline**
  - Why Amazon Redshift for data warehousing?
  - Overview of Amazon Redshift
- **Module 2: Introduction to Amazon Redshift**
  - Amazon Redshift architecture
  - Interactive Demo 1: Touring the Amazon Redshift console
  - Amazon Redshift features
  - Practice Lab 1: Load and query data in an Amazon Redshift cluster
- **Module 3: Ingestion and Storage**
  - Ingestion
  - Interactive Demo 2: Connecting your Amazon Redshift cluster using a Jupyter notebook with Data API
  - Data distribution and storage
  - Interactive Demo 3: Analyzing semi-structured data using the SUPER data type
  - Querying data in Amazon Redshift
  - Practice Lab 2: Data analytics using Amazon Redshift Spectrum
- **Module 4: Processing and Optimizing Data**
  - Data transformation
  - Advanced querying
  - Practice Lab 3: Data transformation and querying in Amazon Redshift
  - Resource management
  - Interactive Demo 4: Applying mixed workload management on Amazon Redshift
  - Automation and optimization
  - Interactive demo 5: Amazon Redshift cluster resizing from the dc2.large to ra3.xlplus cluster
- **Module 5: Security and Monitoring of Amazon Redshift Clusters**
  - Securing the Amazon Redshift cluster
  - Monitoring and troubleshooting Amazon Redshift clusters
- **Module 6: Designing Data Warehouse Analytics Solutions**
  - Data warehouse use case review
  - Activity: Designing a data warehouse analytics workflow
- **Module B: Developing Modern Data Architectures on AWS**
  - Modern data architectures

BUILDING MODERN DATA ANALYTICS SOLUTIONS ON AWS

Course Code: 821673

VIRTUAL CLASSROOM LIVE	$2,700 USD	4 Day

Virtual Classroom Live Outline

Day 1: Building Data Lakes on AWS (GK 821336)

- **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

Day 2: Building Batch Data Analytics Solutions on AWS (GK 821564)

- **Module 1: Overview of Data Analytics and the Data Pipeline**
  - Data analytics use cases
  - Using the data pipeline for analytics
  - Module 1: Introduction to Amazon EMR
  - Using Amazon EMR in analytics solutions
  - Amazon EMR cluster architecture
  - Interactive Demo 1: Launching an Amazon EMR cluster
  - Cost management strategies
- **Module 2: Data Analytics Pipeline Using Amazon EMR: Ingestion and Storage**
  - Storage optimization with Amazon EMR
  - Data ingestion techniques
- **Module 3: High-Performance Batch Data Analytics Using Apache Spark on Amazon EMR**
  - Apache Spark on Amazon EMR use cases
  - Why Apache Spark on Amazon EMR
  - Spark concepts
  - Interactive Demo 2: Connect to an EMR cluster and perform Scala commands using the Spark shell
  - Transformation, processing, and analytics
  - Using notebooks with Amazon EMR
- **Module 4: Processing and Analyzing Batch Data with Amazon EMR and Apache Hive**
  - Using Amazon EMR with Hive to process batch data
  - Transformation, processing, and analytics
  - Introduction to Apache HBase on Amazon EMR
- **Module 5: Serverless Data Processing**
  - Serverless data processing, transformation, and analytics
  - Using AWS Glue with Amazon EMR workloads
  - Practice Lab 3: Orchestrate data processing in Spark using AWS Step Functions
- **Module 6: Security and Monitoring of Amazon EMR Clusters**
  - Securing EMR clusters
  - Interactive Demo 3: Client-side encryption with EMRFS
  - Monitoring and troubleshooting Amazon EMR clusters
  - Demo: Reviewing Apache Spark cluster history
- **Module 7: Designing Batch Data Analytics Solutions**
  - Batch data analytics use cases

- Activity: Designing a batch data analytics workflow
- Module B: Developing Modern Data Architectures on AWS
- Modern data architectures

Day 3: Building Streaming Data Analytics Solutions on AWS (GK 821672)

- **Module A: Overview of Data Analytics and the Data Pipeline**
  - Data analytics use cases
  - Using the data pipeline for analytics
- **Module 1: Using Amazon Redshift in the Data Analytics Pipeline**
  - Why Amazon Redshift for data warehousing?
  - Overview of Amazon Redshift
- **Module 2: Introduction to Amazon Redshift**
  - Amazon Redshift architecture
  - Interactive Demo 1: Touring the Amazon Redshift console
  - Amazon Redshift features
  - Practice Lab 1: Load and query data in an Amazon Redshift cluster
- **Module 3: Ingestion and Storage**
  - Ingestion
  - Interactive Demo 2: Connecting your Amazon Redshift cluster using a Jupyter notebook with Data API
  - Data distribution and storage
  - Interactive Demo 3: Analyzing semi-structured data using the SUPER data type
  - Querying data in Amazon Redshift
  - Practice Lab 2: Data analytics using Amazon Redshift Spectrum
- **Module 4: Processing and Optimizing Data**
  - Data transformation
  - Advanced querying
  - Practice Lab 3: Data transformation and querying in Amazon Redshift
  - Resource management
  - Interactive Demo 4: Applying mixed workload management on Amazon Redshift
  - Automation and optimization
  - Interactive demo 5: Amazon Redshift cluster resizing from the dc2.large to ra3.xlplus cluster
- **Module 5: Security and Monitoring of Amazon Redshift Clusters**
  - Securing the Amazon Redshift cluster
  - Monitoring and troubleshooting Amazon Redshift clusters
  - Module 6: Designing Data Warehouse Analytics Solutions
  - Data warehouse use case review
  - Activity: Designing a data warehouse analytics workflow
- **Module B: Developing Modern Data Architectures on AWS**
  - Modern data architectures

Day 4: Building Data Analytics Solutions Using Amazon Redshift (GK 821497)

- **Module A: Overview of Data Analytics and the Data Pipeline**

- Data analytics use cases
- Using the data pipeline for analytics

- **Module 1: Using Amazon Redshift in the Data Analytics Pipeline**
  - Why Amazon Redshift for data warehousing?
  - Overview of Amazon Redshift
- **Module 2: Introduction to Amazon Redshift**
  - Amazon Redshift architecture
  - Interactive Demo 1: Touring the Amazon Redshift console
  - Amazon Redshift features
  - Practice Lab 1: Load and query data in an Amazon Redshift cluster
- **Module 3: Ingestion and Storage**
  - Ingestion
  - Interactive Demo 2: Connecting your Amazon Redshift cluster using a Jupyter notebook with Data API
  - Data distribution and storage
  - Interactive Demo 3: Analyzing semi-structured data using the SUPER data type
  - Querying data in Amazon Redshift
  - Practice Lab 2: Data analytics using Amazon Redshift Spectrum
- **Module 4: Processing and Optimizing Data**
  - Data transformation
  - Advanced querying
  - Practice Lab 3: Data transformation and querying in Amazon Redshift
  - Resource management
  - Interactive Demo 4: Applying mixed workload management on Amazon Redshift
  - Automation and optimization
  - Interactive demo 5: Amazon Redshift cluster resizing from the dc2.large to ra3.xlplus cluster
- **Module 5: Security and Monitoring of Amazon Redshift Clusters**
  - Securing the Amazon Redshift cluster
  - Monitoring and troubleshooting Amazon Redshift clusters
- **Module 6: Designing Data Warehouse Analytics Solutions**
  - Data warehouse use case review
  - Activity: Designing a data warehouse analytics workflow
- **Module B: Developing Modern Data Architectures on AWS**
  - Modern data architectures

Oct 13 - 16, 2025 | 8:30 AM - 5:00 PM EDT

Dec 8 - 11, 2025 | 8:30 AM - 5:00 PM EST

Jan 5 - 8, 2026 | 8:30 AM - 5:00 PM EST

Mar 16 - 19, 2026 | 8:30 AM - 5:00 PM EDT

BUILDING MODERN DATA ANALYTICS SOLUTIONS ON AWS

Course Code: 821673

PRIVATE GROUP TRAINING	4 Day

Visit us at www.globalknowledge.com or call us at 1-866-716-6688.

Date created: 8/30/2025 11:08:08 AM