Data Engineering on AWS

Category: Cloud Platforms | **Vendor:** Amazon Web Services

Duration: 24.00 hours (3 days) **19.5 CPD Hours** **Rating:** ⭐ 4.6 (5,878 reviews)

Course Information

Delivery Format: Instructor Led - Online

Course Overview

Data Engineering on AWS is a 3-day intermediate course, designed for professionals seeking a deep dive into data engineering practices and solutions on AWS. Through a balanced combination of theory, practical labs, and activities, participants learn to design, build, optimize, and secure data engineering solutions using AWS services. From foundational concepts to hands-on implementation of data lakes, data warehouses, and both batch and streaming data pipelines, this course equips data professionals with the skills needed to architect and manage modern data solutions at scale.

About This Course

Data Engineering on AWS is a 3-day intermediate course, designed for professionals seeking a deep dive into data engineering practices and solutions on AWS. Through a balanced combination of theory, practical labs, and activities, participants learn to design, build, optimize, and secure data engineering solutions using AWS services. From foundational concepts to hands-on implementation of data lakes, data warehouses, and both batch and streaming data pipelines, this course equips data professionals with the skills needed to architect and manage modern data solutions at scale.

Who Should Attend

This course is designed for professionals who are interested in designing, building, optimizing, and securing data engineering solutions using AWS services.

Prerequisites & Entry Requirements

General Prerequisites:

We recommend that attendees of this course have:

- Familiarity with basic machine learning concepts, such as supervised and unsupervised learning, regression, classification, and clustering algorithms.
- Working knowledge of Python programming language and common data science libraries like NumPy, Pandas, and Scikit-learn.
- Basic understanding of cloud computing concepts and familiarity with the AWS platform.
- Familiarity with SQL and relational databases is recommended but not mandatory.
- Experience with version control systems like Git is beneficial but not required.

Learning Outcomes

Upon successful completion of this course, participants will be able to:

In this course, you will learn to do the following:

- Understand the foundational roles and key concepts of data engineering, including data personas, data discovery, and relevant AWS services.
- Identify and explain the various AWS tools and services crucial for data engineering, encompassing orchestration, security, monitoring, CI/CD, IaC, networking, and cost optimization.
- Design and implement a data lake solution on AWS, including storage, data ingestion, transformation, and serving data for consumption.
- Optimize and secure a data lake solution by implementing open table formats, security measures, and troubleshooting common issues.
- Design and set up a data warehouse using Amazon Redshift Serverless, understanding its architecture, data ingestion, processing, and serving capabilities.
- Apply performance optimization techniques to data warehouses in Amazon Redshift, including monitoring, data optimization, query optimization, and orchestration.
- Manage security and access control for data warehouses in Amazon Redshift, understanding authentication, data security, auditing, and compliance.
- Design effective batch data pipelines using appropriate AWS services for processing and transforming data.
- Implement comprehensive strategies for batch data pipelines, covering data processing, transformation, integration, cataloging, and serving data for consumption.
- Optimize, orchestrate, and secure batch data pipelines, demonstrating advanced skills in data processing automation and security.
- Architect streaming data pipelines, understanding various use cases, ingestion, storage, processing, and analysis using AWS services.
- Optimize and secure streaming data solutions, including compliance considerations and access control.

Detailed Course Outline

Day 1

Module 1: Data Engineering Roles and Key Concepts

- Role of a Data Engineer
- Key functions of a Data Engineer
- Data Personas
- Data Discovery
- AWS Data Services

Module 2: AWS Data Engineering Tools and Services

- Orchestration and Automation
- Data Engineering Security
- Monitoring
- Continuous Integration and Continuous Delivery
- Infrastructure as Code
- AWS Serverless Application Model
- Networking Considerations
- Cost Optimization Tools

Module 3: Designing and Implementing Data Lakes

- Data lake introduction
- Data lake storage
- Ingest data into a data lake
- Catalog data

- Transform data
- Server data for consumption

Hands-on lab: Setting up a Data Lake on AWS

Module 4: Optimizing and Securing a Data Lake Solution

- Open Table Formats
- Security using AWS Lake Formation
- Setting permissions with Lake Formation
- Security and governance
- Troubleshooting

Hand-on lab: Automating Data Lake Creation using AWS Lake Formation Blueprints

Day 2

Module 5: Data Warehouse Architecture and Design Principles

- Introduction to data warehouses
- Amazon Redshift Overview
- Ingesting data into Redshift
- Processing data
- Serving data for consumption

Hands-on Lab: Setting up a Data Warehouse using Amazon Redshift Serverless

Module 6: Performance Optimization Techniques for Data Warehouses

- Monitoring and optimization options
- Data optimization in Amazon Redshift
- Query optimization in Amazon Redshift
- Orchestration options

Module 7: Security and Access Control for Data Warehouses

- Authentication and access control in Amazon Redshift
- Data security in Amazon Redshift
- Auditing and compliance in Amazon Redshift

Hands-on lab: Managing Access Control in Redshift

Module 8: Designing Batch Data Pipelines

- Introduction to batch data pipelines
- Designing a batch data pipeline
- AWS services for batch data processing

Module 9: Implementing Strategies for Batch Data Pipeline

- Elements of a batch data pipeline
- Processing and transforming data
- Integrating and cataloging your data

- Serving data for consumption

Hands-on lab: A Day in the Life of a Data Engineer

Day 3

Module 10: Optimizing, Orchestrating, and Securing Batch Data Pipelines

- Optimizing the batch data pipeline
- Orchestrating the batch data pipeline
- Securing the batch data pipeline

Hands-on lab: Orchestrating Data Processing in Spark using AWS Step Functions

Module 11: Streaming Data Architecture Patterns

- Introduction to streaming data pipelines
- Ingesting data from stream sources
- Streaming data ingestion services
- Storing streaming data
- Processing Streaming Data
- Analyzing Streaming Data with AWS Services

Hands-on lab: Streaming Analytics with Amazon Managed Service for Apache Flink

Module 12: Optimizing and Securing Streaming Solutions

- Optimizing a streaming data solution
- Securing a streaming data pipeline
- Compliance considerations

Hands-on lab: Access Control with Amazon Managed Streaming for Apache Kafka, Data Engineering Roles and Key Concepts

The role of a data engineer

Data discovery for a data analytics system

AWS services for data workflows

Continuous integration and continuous delivery

Networking considerations

Designing and Implementing Data Lakes

Data lake introduction

Data lake storage

Ingest data

Catalog data

Transform data

Serve data for consumption

Lab: Setting up a Data Lake on AWS

Optimizing and Securing Data Lake Solutions

Optimizing performance

Security using Lake Formation

Setting permissions with Lake Formation

Security and governance

Troubleshooting

Lab: Automating Data Lake Creation using AWS Lake Formation Blueprints

Data Warehouse Architecture and Design Principles

Introduction to data warehouses

Amazon Redshift overview

Ingesting data into Amazon Redshift

Processing data

Serving data for consumption

Lab: Setting up a Data Warehouse using Amazon Redshift Serverless

Performance Optimization Techniques for Data Warehouses

Monitoring and optimization options

Data optimization in Amazon Redshift

Query optimization in Amazon Redshift

Data orchestration

Security and Access Control for Data Warehouses

Authentication and access control in Amazon Redshift

Data security in Amazon Redshift

Lab: Working with Amazon Redshift

Designing Batch Data Pipelines

Introduction to batch data pipelines

Designing a batch data pipeline

Ingesting batch data

Implementing Strategies for Batch Data Pipelines

Processing and transforming data

Transforming data formats

Integrating your data

Cataloging data

Serving data for consumption

Lab: A Day in the Life of a Data Engineer

Optimizing, and Securing Batch Data Pipelines

Optimizing the batch data pipeline

Orchestrating the batch data pipeline

Securing the batch data pipeline

Lab: Orchestrating Data Processing in Spark using AWS Step Functions

Streaming Data Architecture Patterns

Introduction to streaming data pipelines

Ingesting data from stream sources

Storing streaming data

Processing streaming data

Analyzing streaming data

Lab: Streaming Analytics with Amazon Managed Service for Apache Flink

Optimizing and Securing Streaming Solutions

Optimizing a streaming data solution

Securing a streaming data pipeline

Lab: Access Control with Amazon Managed Streaming for Apache Kafka

Compliance and Cost Optimization

Compliance considerations

Cost optimization tools

Course Wrap-Up

Skills You Will Learn:

In this course, you will learn to do the following:

- Understand the foundational roles and key concepts of data engineering, including data personas, data discovery, and relevant AWS services.
- Identify and explain the various AWS tools and services crucial for data engineering, encompassing orchestration, security, monitoring, CI/CD, IaC, networking, and cost optimization.
- Design and implement a data lake solution on AWS, including storage, data ingestion, transformation, and serving data for consumption.
- Optimize and secure a data lake solution by implementing open table formats, security measures, and troubleshooting common issues.
- Design and set up a data warehouse using Amazon Redshift Serverless, understanding its architecture, data ingestion, processing, and serving capabilities.
- Apply performance optimization techniques to data warehouses in Amazon Redshift, including monitoring, data optimization, query optimization, and orchestration.
- Manage security and access control for data warehouses in Amazon Redshift, understanding authentication, data security, auditing, and compliance.
- Design effective batch data pipelines using appropriate AWS services for processing and transforming data.
- Implement comprehensive strategies for batch data pipelines, covering data processing, transformation, integration, cataloging, and serving data for consumption.
- Optimize, orchestrate, and secure batch data pipelines, demonstrating advanced skills in data processing automation and security.
- Architect streaming data pipelines, understanding various use cases, ingestion, storage, processing, and analysis using AWS services.
- Optimize and secure streaming data solutions, including compliance considerations and access control.

Frequently Asked Questions

Q: What delivery options are available for Data Engineering on AWS?

We offer multiple delivery formats:

- Live Instructor-Led Classroom Online (Virtual/Live Online)
- Traditional Instructor-Led Classroom Training (ILT)
- On-site delivery at your offices anywhere in United Kingdom
- Private dedicated courses customized for your team

Q: How many CPD hours does this course provide?

The 3-day Data Engineering on AWS course provides up to 19.5 CPD hours of structured learning. CPD certificates can be provided upon request.

Q: What is the duration of the Data Engineering on AWS training?

The training takes place over 3 day(s), with each day lasting approximately 24.00 hours including breaks for lunch and refreshments.

Q: Do you provide corporate training for Data Engineering on AWS?

Yes, we provide corporate training, dedicated training, and closed classes for Data Engineering on AWS. Training can take place anywhere in United Kingdom including London, Manchester, Birmingham, Edinburgh, or live online allowing teams from across United Kingdom or internationally to attend.

Q: What related terms do people search for?

Popular related searches include: Data Engineering on AWS, AWS Operations, DEAWS, Amazon Web Services, vILT

Q: Why choose Nexus Human for Data Engineering on AWS?

Nexus Human is recognized as one of the leading training providers. Our trainers have won multiple awards including:

- Small Firms Best Trainer Award
- National Training Partner of the Year (Ireland) - Multiple Years
- Global Top 30 Instructor Awards (2012, 2019, 2021)
- Tech Excellence Award Nominations
- Learning Performance Institute (LPI) External Training Provider Sponsor 2024

Q: Are there any discount codes available?

Yes! Use discount code **PENPAL5** when booking your Data Engineering on AWS training. Please note that only one discount code can be used per booking and cannot be combined with other special offers.

Nexus Human

Professional Training & Development

Email: info@nexushuman.com

Website: www.nexushuman.com

Phone: +353 1 889 8700 (Ireland) | +44 20 8142 8961 (UK)

© 2026 Nexus Human. All rights reserved. This brochure was generated on 19/10/2026.