

Table of Content

No. of Days: 3

No. of Hours : 24

Module 1: Data Ingestion and Transformation - Part 1

Perform Data Ingestion

- Reading Data from Various Sources: Kinesis, MSK, Redshift
- Implementing Configuration Options for Batch Ingestion
- Consuming Data APIs
- Setting Up Schedulers and Event Triggers
- Throttling and Overcoming Rate Limits
- Managing Fan-In and Fan-Out for Data Distribution

Module 2: Data Ingestion and Transformation - Part 2

Transform and Process Data

- Optimizing Container Usage for Performance
- Connecting to Different Data Sources: JDBC, ODBC
- Integrating Data from Multiple Sources
- Optimizing Data Processing Costs
- Implementing Data Transformation Services: EMR, Glue, Redshift, Lambda
- Data Transformation Between Formats: CSV to Parquet
- Debugging Common Transformation Failures
- Creating Data APIs Using AWS Services

Module 3: Data Store Management - Part 1

Choose a Data Store

- Implementing Appropriate Storage Services: Redshift, EMR, Lake Formation, RDS, DynamoDB, Kinesis Data Streams, MSK
- Configuring Storage Services for Access Patterns
- Applying Storage Services to Use Cases: S3
- Integrating Migration Tools: AWS Transfer Family
- Data Migration or Remote Access Methods: Redshift Federated Queries, Materialized Views, Spectrum

Module 4: Data Store Management - Part 2

1. Understand Data Cataloging Systems

- Using Data Catalogs to Consume Data
- Building and Referencing Data Catalogs: Glue Data Catalog, Hive Metastore
- Discovering Schemas and Using Glue Crawlers
- Synchronizing Partitions and Creating Connections: Glue

2. Manage the Lifecycle of Data

- Performing Load and Unload Operations Between S3 and Redshift
- Managing S3 Lifecycle Policies for Storage Tiers
- Expiring S3 Data and Managing Versioning
- Managing DynamoDB TTL

Module 5: Data Store Management - Part 3

1. Design Data Models and Schema Evolution

- Designing Schemas for Redshift, DynamoDB, and Lake Formation
- Addressing Changes to Data Characteristics
- Performing Schema Conversion: AWS SCT and DMS Schema Conversion
- Establishing Data Lineage with AWS Tools: SageMaker ML Lineage Tracking

Module 6: Data Operations and Support - Part 1

1. Automate Data Processing by Using AWS Services

- Orchestrating Data Pipelines: MWAA, Step Functions
- Troubleshooting AWS Managed Workflows
- Calling SDKs to Access AWS Features
- Using Features of Services for Data Processing: EMR, Redshift, Glue
- Consuming and Maintaining Data APIs
- Preparing Data Transformation: AWS Glue DataBrew
- Querying Data: Amazon Athena
- Using Lambda to Automate Data Processing
- Managing Events and Schedulers: EventBridge

Module 7: Data Operations and Support - Part 2

1. Analyze Data by Using AWS Services

- Visualizing Data by Using AWS Services and Tools: AWS Glue DataBrew, Amazon QuickSight
- Verifying and Cleaning Data: Lambda, Athena, QuickSight, Jupyter Notebooks, Amazon SageMaker Data Wrangler
- Using Athena to Query Data or Create Views
- Using Athena Notebooks that Use Apache Spark to Explore Data

2. Maintain and Monitor Data Pipelines

- Extracting Logs for Audits
- Deploying Logging and Monitoring Solutions
- Using Notifications During Monitoring to Send Alerts
- Troubleshooting Performance Issues
- Using CloudTrail to Track API Calls
- Troubleshooting and Maintaining Pipelines: AWS Glue, Amazon EMR
- Using Amazon CloudWatch Logs to Log Application Data
- Analyzing Logs with AWS Services: Athena, Amazon EMR, Amazon OpenSearch Service, CloudWatch Logs Insights, Big Data Application Logs

Module 8: Data Operations and Support - Part 3

1. Ensure Data Quality

- Running Data Quality Checks During Processing: Checking for Empty Fields
- Defining Data Quality Rules: AWS Glue DataBrew
- Investigating Data Consistency: AWS Glue DataBrew

Module 9: Data Security and Governance - Part 1

1. Apply Authentication Mechanisms

- Updating VPC Security Groups
- Creating and Updating IAM Groups, Roles, Endpoints, and Services
- Creating and Rotating Credentials for Password Management: using AWS Secrets Manager
- Setting Up IAM Roles for Access: for Lambda, Amazon API Gateway, AWS CLI, CloudFormation
- Applying IAM Policies to Roles, Endpoints, and Services: for S3 Access Points, AWS PrivateLink

2. Apply Authorization Mechanisms

- Creating Custom IAM Policies
- Storing Application and Database Credentials: using Secrets Manager, AWS Systems Manager Parameter Store
- Managing Database User Permissions: for Amazon Redshift
- Managing Permissions through Lake Formation: for Amazon Redshift, EMR, Athena, S3

Module 10: Data Security and Governance - Part 2

1. Ensure Data Encryption and Masking

- Applying Data Encryption in AWS Analytics Services: Redshift, EMR, Glue
- Understanding Client-Side vs. Server-Side Encryption
- Protecting Sensitive Data with Masking and Anonymization
- Using Encryption Keys for Data Security: AWS Key Management Service

2. Prepare Logs for Audit

- Logging Application Data
- Logging Access to AWS Services
- Using CloudTrail for API Call Tracking
- Using CloudWatch Logs for Application Log Storage
- Analyzing Logs with AWS Services: Athena, CloudWatch Logs Insights, OpenSearch Service

3. Understand Data Privacy and Governance

- Protecting Personally Identifiable Information (PII)
- Understanding Data Sovereignty
- Granting Permissions for Data Sharing: Redshift
- Implementing PII Identification: Macie with Lake Formation
- Implementing Data Privacy Strategies to Control Backups and Replication
- Managing Configuration Changes with AWS Config