Associate Exam Guide (DEA-C01)

AWS Certified Data Engineer

AWS Certified Data Engineer: Associate Exam Guide (DEA-C01)

Table of Contents

AWS Certified Data Engineer - Associate (DEA-C01)

The AWS Certified Data Engineer - Associate (DEA-C01) exam validates a candidate's ability to implement data pipelines and to monitor, troubleshoot, and optimize cost and performance issues in accordance with best practices.

Note: AWS exam guides are periodically reviewed and revised to ensure that each certification exam tests skills and AWS services and features that are current and relevant for the job role(s) that the certification is designed to target. Exam guide revisions will be published at least one month before changes are reflected on your exam. Check the Revisions section for a summary of changes.

Topics

Introduction

The [AWS Certified Data Engineer - Associate (DEA-C01)](#) exam validates a candidate's ability to implement data pipelines and to monitor, troubleshoot, and optimize cost and performance issues in accordance with best practices.

The exam also validates a candidate's ability to complete the following tasks:

- Ingest and transform data, and orchestrate data pipelines while applying programming concepts.

- Choose an optimal data store, design data models, catalog data schemas, and manage data lifecycles.

- Operationalize, maintain, and monitor data pipelines. Analyze data and ensure data quality.

- Implement appropriate authentication, authorization, data encryption, privacy, and governance. Enable logging.

Target Candidate Description

The target candidate should have the equivalent of 2–3 years of experience in data engineering. The target candidate should understand the effects of volume, variety, and velocity on data ingestion, transformation, modeling, security, governance, privacy, schema design, and optimal data store design. Additionally, the target candidate should have at least 1–2 years of hands-on experience with AWS services.

Recommended general IT knowledge

The target candidate should have the following general IT knowledge:

- Setup and maintenance of extract, transform, and load (ETL) pipelines from ingestion to destination

- Application of high-level but language-agnostic programming concepts as required by the pipeline

- How to use Git commands for source control

- How to use data lakes to store data

- General concepts for networking, storage, and compute

- General concepts of vectors

Recommended AWS knowledge

The target candidate should have the following AWS knowledge:

- How to use AWS services to accomplish the tasks listed in the Introduction section of this exam guide

- An understanding of the AWS services for encryption, governance, protection, and logging of all data that is part of data pipelines
- The ability to compare AWS services to understand the cost, performance, and functional differences between services
- How to structure SQL queries and how to run SQL queries on AWS services
- An understanding of how to analyze data, verify data quality, and ensure data consistency by using AWS services

Job tasks that are out of scope for the target candidate

The following list contains job tasks that the target candidate is not expected to be able to perform. This list is non-exhaustive. These tasks are out of scope for the exam:

- Perform ML training and inferences.
- Demonstrate knowledge of programming language-specific syntax.
- Draw business conclusions based on data.

Exam content

Response types

There are two types of questions on the exam:

- **Multiple choice:** Has one correct response and three incorrect responses (distractors)
- **Multiple response:** Has two or more correct responses out of five or more response options

Select one or more responses that best complete the statement or answer the question. Distractors, or incorrect answers, are response options that a candidate with incomplete knowledge or skill might choose. Distractors are generally plausible responses that match the content area.

Unanswered questions are scored as incorrect; there is no penalty for guessing. The exam includes 50 questions that affect your score.

The exam includes 15 unscored questions that do not affect your score. AWS collects information about performance on these unscored questions to evaluate these questions for future use as scored questions. These unscored questions are not identified on the exam.

The AWS Certified Data Engineer - Associate (DEA-C01) exam has a pass or fail designation. The exam is scored against a minimum standard established by AWS professionals who follow certification industry best practices and guidelines.

Your results for the exam are reported as a scaled score of 100–1,000. The minimum passing score is 720. Your score shows how you performed on the exam as a whole and whether you passed. Scaled scoring models help equate scores across multiple exam forms that might have slightly different difficulty levels.

Your score report could contain a table of classifications of your performance at each section level. The exam uses a compensatory scoring model, which means that you do not need to achieve a passing score in each section. You need to pass only the overall exam.

Each section of the exam has a specific weighting, so some sections have more questions than other sections have. The table of classifications contains general information that highlights your strengths and weaknesses. Use caution when you interpret section-level feedback.

Unscored content

The exam includes 15 unscored questions that do not affect your score. AWS collects information about performance on these unscored questions to evaluate these questions for future use as scored questions. These unscored questions are not identified on the exam.

Exam results

The AWS Certified Data Engineer - Associate (DEA-C01) exam has a pass or fail designation. The exam is scored against a minimum standard established by AWS professionals who follow certification industry best practices and guidelines.

Your results for the exam are reported as a scaled score of 100–1,000. The minimum passing score is 720. Your score shows how you performed on the exam as a whole and whether you passed. Scaled scoring models help equate scores across multiple exam forms that might have slightly different difficulty levels.

Your score report could contain a table of classifications of your performance at each section level. The exam uses a compensatory scoring model, which means that you do not need to achieve a passing score in each section. You need to pass only the overall exam.

Each section of the exam has a specific weighting, so some sections have more questions than other sections have. The table of classifications contains general information that highlights your strengths and weaknesses. Use caution when you interpret section-level feedback.

Content outline

This exam guide includes weightings, content domains, and task statements for the exam. This guide does not provide a comprehensive list of the content on the exam. However, additional context for each task statement is available to help you prepare for the exam.

The exam has the following content domains and weightings:

- Content Domain 1: Data Ingestion and Transformation (34% of scored content)
- Content Domain 2: Data Store Management (26% of scored content)
- Content Domain 3: Data Operations and Support (22% of scored content)
- Content Domain 4: Data Security and Governance (18% of scored content)

AWS Services for the Exam

The AWS Certified Data Engineer - Associate exam covers specific AWS services that are relevant to data engineers. Understanding which services are in scope can help you focus your preparation efforts.

For detailed information about the AWS services covered in the exam, see the following section:

- In-Scope AWS Services

Content Domain 1: Data Ingestion and Transformation

Tasks

- Task 1.1: Perform data ingestion
- Task 1.2: Transform and process data
- Task 1.3: Orchestrate data pipelines
- Task 1.4: Apply programming concepts

Task 1.1: Perform data ingestion

- Skill 1.1.1: Read data from streaming sources (for example, Amazon Kinesis, Amazon Managed Streaming for Apache Kafka [Amazon MSK], Amazon DynamoDB Streams, AWS Database Migration Service [AWS DMS], AWS Glue, Amazon Redshift).

- Skill 1.1.2: Read data from batch sources (for example, Amazon S3, AWS Glue, Amazon EMR, AWS DMS, Amazon Redshift, AWS Lambda, Amazon AppFlow).

- Skill 1.1.3: Implement appropriate configuration options for batch ingestion.

- Skill 1.1.4: Consume data APIs.

- Skill 1.1.5: Set up schedulers by using Amazon EventBridge, Apache Airflow, or time-based schedules for jobs and crawlers.

- Skill 1.1.6: Set up event triggers (for example, Amazon S3 Event Notifications, EventBridge).

- Skill 1.1.7: Call a Lambda function from Kinesis.

- Skill 1.1.8: Create allowlists for IP addresses to allow connections to data sources.

- Skill 1.1.9: Implement throttling and overcoming rate limits (for example, DynamoDB, Amazon RDS, Kinesis).

- Skill 1.1.10: Manage fan-in and fan-out for streaming data distribution.

- Skill 1.1.11: Describe replayability of data ingestion pipelines.

- Skill 1.1.12: Define stateful and stateless data transactions.

Task 1.2: Transform and process data

- Skill 1.2.1: Optimize container usage for performance needs (for example, Amazon Elastic Kubernetes Service [Amazon EKS], Amazon Elastic Container Service [Amazon ECS]).

- Skill 1.2.2: Connect to different data sources (for example, Java Database Connectivity [JDBC], Open Database Connectivity [ODBC]).

- Skill 1.2.3: Integrate data from multiple sources.

- Skill 1.2.4: Optimize costs while processing data.

- Skill 1.2.5: Implement data transformation services based on requirements (for example, Amazon EMR, AWS Glue, Lambda, Amazon Redshift).

- Skill 1.2.6: Transform data between formats (for example, from .csv to Apache Parquet).

- Skill 1.2.7: Troubleshoot and debug common transformation failures and performance issues.

- Skill 1.2.8: Create data APIs to make data available to other systems by using AWS services.

- Skill 1.2.9: Define volume, velocity, and variety of data (for example, structured data, unstructured data).

- Skill 1.2.10: Integrate large language models (LLMs) for data processing.

Task 1.3: Orchestrate data pipelines

- Skill 1.3.1: Use orchestration services to build workflows for data ETL pipelines (for example, Lambda, EventBridge, Amazon Managed Workflows for Apache Airflow [Amazon MWAA], AWS Step Functions, AWS Glue workflows).

- Skill 1.3.2: Build data pipelines for performance, availability, scalability, resiliency, and fault tolerance.

- Skill 1.3.3: Implement and maintain serverless workflows.

- Skill 1.3.4: Use notification services to send alerts (for example, Amazon Simple Notification Service [Amazon SNS], Amazon Simple Queue Service [Amazon SQS]).

Task 1.4: Apply programming concepts

- Skill 1.4.1: Optimize code to reduce runtime for data ingestion and transformation.

- Skill 1.4.2: Configure Lambda functions to meet concurrency and performance needs.

- Skill 1.4.3: Use programming languages and frameworks for data engineering (for example, Python, SQL, Scala, R, Java, Bash, PowerShell).

- Skill 1.4.4: Use software engineering best practices for data engineering (for example, version control, testing, logging, monitoring).

- Skill 1.4.5: Use Infrastructure as Code (IaC) to deploy data engineering solutions.

- Skill 1.4.6: Use the AWS Serverless Application Model (AWS SAM) to package and deploy serverless data pipelines (for example, Lambda functions, Step Functions, DynamoDB tables).

- Skill 1.4.7: Use and mount storage volumes from within Lambda functions.

- Skill 1.4.8: Use infrastructure as code (IaC) for repeatable resource deployment (for example, AWS CloudFormation and AWS Cloud Development Kit [AWS CDK]).

- Skill 1.4.9: Describe continuous integration and continuous delivery (CI/CD) (implementation, testing, and deployment of data pipelines).

- Skill 1.4.10: Define distributed computing.

- Skill 1.4.11: Describe data structures and algorithms (for example, graph data structures and tree data structures).

Content Domain 2: Data Store Management

Tasks

- [Task 2.1: Choose a data store](#)

- [Task 2.2: Understand data cataloging systems](#)

- [Task 2.3: Manage the lifecycle of data](#)

- [Task 2.4: Design data models and schema evolution](#)

Task 2.1: Choose a data store

- Skill 2.1.1: Implement the appropriate storage services for specific cost and performance requirements (for example, Amazon Redshift, Amazon EMR, AWS Lake Formation, Amazon RDS, Amazon DynamoDB, Amazon Kinesis Data Streams, Amazon Managed Streaming for Apache Kafka [Amazon MSK]).

- Skill 2.1.2: Configure the appropriate storage services for specific access patterns and requirements (for example, Amazon Redshift, Amazon EMR, Lake Formation, Amazon RDS, DynamoDB).

- Skill 2.1.3: Apply storage services to appropriate use cases (for example, using indexing algorithms like Hierarchical Navigable Small Worlds [HNSW] with Amazon Aurora PostgreSQL and using Amazon MemoryDB for fast key/value pair access).

- Skill 2.1.4: Integrate migration tools into data processing systems (for example, AWS Transfer Family).

- Skill 2.1.5: Implement data migration or remote access methods (for example, Amazon Redshift federated queries, Amazon Redshift materialized views, Amazon Redshift Spectrum).

- Skill 2.1.6: Manage locks to prevent access to data (for example, Amazon Redshift, Amazon RDS).

- Skill 2.1.7: Manage open table formats (for example Apache Iceberg).

- Skill 2.1.8: Describe vector index types (for example, HNSW, IVF).

Task 2.2: Understand data cataloging systems

- Skill 2.2.1: Use data catalogs to consume data from the data's source.

- Skill 2.2.2: Build and reference a technical data catalog (for example, AWS Glue Data Catalog, Apache Hive metastore).

- Skill 2.2.3: Discover schemas and use AWS Glue crawlers to populate data catalogs.

- Skill 2.2.4: Synchronize partitions with a data catalog.

- Skill 2.2.5: Create new source or target connections for cataloging (for example, AWS Glue).

- Skill 2.2.6: Create and manage business data catalogs (for example, Amazon SageMaker Catalog).

Task 2.3: Manage the lifecycle of data

- Skill 2.3.1: Perform load and unload operations to move data between Amazon S3 and Amazon Redshift.

- Skill 2.3.2: Manage S3 Lifecycle policies to change the storage tier of S3 data.

- Skill 2.3.3: Expire data when it reaches a specific age by using S3 Lifecycle policies.

- Skill 2.3.4: Manage S3 versioning and DynamoDB TTL.

- Skill 2.3.5: Delete data to meet business and legal requirements.

- Skill 2.3.6: Protect data with appropriate resiliency and availability.

Task 2.4: Design data models and schema evolution

- Skill 2.4.1: Design schemas for Amazon Redshift, DynamoDB, and Lake Formation.

- Skill 2.4.2: Address changes to the characteristics of data.

- Skill 2.4.3: Perform schema conversion (for example, by using the AWS Schema Conversion Tool [AWS SCT] and AWS Database Migration Service [AWS DMS] Schema Conversion).

- Skill 2.4.4: Establish data lineage by using AWS tools (for example, Amazon SageMaker ML Lineage Tracking and Amazon SageMaker Catalog).

- Skill 2.4.5: Describe best practices for indexing, partitioning strategies, compression, and other data optimization techniques.

- Skill 2.4.6: Describe vectorization concepts (for example, Amazon Bedrock knowledge base).

Content Domain 3: Data Operations and Support

Tasks

- [Task 3.1: Automate data processing by using AWS services](#)
- [Task 3.2: Analyze data by using AWS services](#)
- [Task 3.3: Maintain and monitor data pipelines](#)
- [Task 3.4: Ensure data quality](#)

Task 3.1: Automate data processing by using AWS services

- Skill 3.1.1: Orchestrate data pipelines (for example, Amazon Managed Workflows for Apache Airflow [Amazon MWAA], AWS Step Functions).
- Skill 3.1.2: Troubleshoot Amazon managed workflows.
- Skill 3.1.3: Call SDKs to access Amazon features from code.
- Skill 3.1.4: Use the features of AWS services to process data (for example, Amazon EMR, Amazon Redshift, AWS Glue).
- Skill 3.1.5: Consume and maintain data APIs.
- Skill 3.1.6: Prepare data for transformation (for example, AWS Glue DataBrew and Amazon SageMaker Unified Studio).
- Skill 3.1.7: Query data (for example, Amazon Athena).
- Skill 3.1.8: Use AWS Lambda to automate data processing.
- Skill 3.1.9: Manage events and schedulers (for example, Amazon EventBridge).

Task 3.2: Analyze data by using AWS services

- Skill 3.2.1: Visualize data by using AWS services and tools (for example, DataBrew, Amazon QuickSight).
- Skill 3.2.2: Verify and clean data (for example, Lambda, Athena, QuickSight, Jupyter Notebooks, Amazon SageMaker Data Wrangler).
- Skill 3.2.3: Use SQL in Amazon Redshift and Athena to query data or to create views.
- Skill 3.2.4: Use Athena notebooks that use Apache Spark to explore data.
- Skill 3.2.5: Describe tradeoffs between provisioned services and serverless services.

- Skill 3.2.6: Define data aggregation, rolling average, grouping, and pivoting.

Task 3.3: Maintain and monitor data pipelines

- Skill 3.3.1: Extract logs for audits.

- Skill 3.3.2: Deploy logging and monitoring solutions to facilitate auditing and traceability.

- Skill 3.3.3: Use notifications during monitoring to send alerts.

- Skill 3.3.4: Troubleshoot performance issues.

- Skill 3.3.5: Use AWS CloudTrail to track API calls.

- Skill 3.3.6: Troubleshoot and maintain pipelines (for example, AWS Glue, Amazon EMR).

- Skill 3.3.7: Use Amazon CloudWatch Logs to log application data (with a focus on configuration and automation).

- Skill 3.3.8: Analyze logs with AWS services (for example, Athena, Amazon EMR, Amazon OpenSearch Service, CloudWatch Logs Insights, big data application logs).

Task 3.4: Ensure data quality

- Skill 3.4.1: Run data quality checks while processing the data (for example, checking for empty fields).

- Skill 3.4.2: Define data quality rules (for example, DataBrew).

- Skill 3.4.3: Investigate data consistency (for example, DataBrew).

- Skill 3.4.4: Describe data sampling techniques.

- Skill 3.4.5: Implement data skew mechanisms.

Content Domain 4: Data Security and Governance

Tasks

- Task 4.1: Apply authentication mechanisms

- Task 4.2: Apply authorization mechanisms

- Task 4.3: Ensure data encryption and masking

- Task 4.4: Prepare logs for audit

- Task 4.5: Understand data privacy and governance

Task 4.1: Apply authentication mechanisms

- Skill 4.1.1: Update VPC security groups.

- Skill 4.1.2: Create and update AWS Identity and Access Management (IAM) groups, roles, endpoints, and services.

- Skill 4.1.3: Create and rotate credentials for password management (for example, AWS Secrets Manager).

- Skill 4.1.4: Set up IAM roles for access (for example, AWS Lambda, Amazon API Gateway, AWS CLI, AWS CloudFormation).

- Skill 4.1.5: Apply IAM policies to roles, endpoints, and services (for example, S3 Access Points, AWS PrivateLink).

- Skill 4.1.6: Describe the differences between managed services and unmanaged services.

- Skill 4.1.7: Use domain, domain units, and projects for SageMaker Unified Studio.

Task 4.2: Apply authorization mechanisms

- Skill 4.2.1: Create custom IAM policies when a managed policy does not meet the needs.

- Skill 4.2.2: Store application and database credentials (for example, Secrets Manager, AWS Systems Manager Parameter Store).

- Skill 4.2.3: Provide database users, groups, and roles access and authority in a database (for example, for Amazon Redshift).

- Skill 4.2.4: Manage permissions through AWS Lake Formation (for Amazon Redshift, Amazon EMR, Amazon Athena, and Amazon S3).

- Skill 4.2.5: Apply authorization methods that address business needs (role-based, tag-based, and attribute-based).

- Skill 4.2.6: Construct custom policies that meet the principle of least privilege.

Task 4.3: Ensure data encryption and masking

- Skill 4.3.1: Apply data masking and anonymization according to compliance laws or company policies.

- Skill 4.3.2: Use encryption keys to encrypt or decrypt data (for example, AWS Key Management Service [AWS KMS]).

- Skill 4.3.3: Configure encryption across AWS account boundaries.

- Skill 4.3.4: Enable encryption in transit or before transit for data.

Task 4.4: Prepare logs for audit

- Skill 4.4.1: Use AWS CloudTrail to track API calls.

- Skill 4.4.2: Use Amazon CloudWatch Logs to store application logs.

- Skill 4.4.3: Use AWS CloudTrail Lake for centralized logging queries.

- Skill 4.4.4: Analyze logs by using AWS services (for example, Athena, CloudWatch Logs Insights, Amazon OpenSearch Service).

- Skill 4.4.5: Integrate various AWS services to perform logging (for example, Amazon EMR in cases of large volumes of log data).

Task 4.5: Understand data privacy and governance

- Skill 4.5.1: Grant permissions for data sharing (for example, data sharing for Amazon Redshift).

- Skill 4.5.2: Implement PII identification (for example, Amazon Macie with Lake Formation).

- Skill 4.5.3: Implement data privacy strategies to prevent backups or replications of data to disallowed AWS Regions.

- Skill 4.5.4: Viewing configuration changes that have occurred in an account (for example, AWS Config).

- Skill 4.5.5: Maintain data sovereignty.

- Skill 4.5.6: Manage data access through Amazon SageMaker Catalog projects.

- Skill 4.5.7: Describe governance data framework and data sharing patterns.

In-Scope AWS Services

The following list contains AWS services and features that are in scope for the AWS Certified Data Engineer - Associate (DEA-C01) exam. This list is non-exhaustive and is subject to change. AWS offerings appear in categories that align with the offerings' primary functions.

Topics

- [Analytics](#)

- [Application Integration](#)

- [Cloud Financial Management](#)

- [Compute](#)

- [Containers](#)

- [Database](#)

- [Developer Tools](#)

- [Web and Mobile](#)

- [Machine Learning](#)

- [Management and Governance](#)

- [Migration and Transfer](#)

- [Networking and Content Delivery](#)

- [Security, Identity, and Compliance](#)

- [Storage](#)

Analytics

- Amazon Athena

- Amazon EMR

- AWS Glue

- AWS Glue DataBrew

- AWS Lake Formation

- Amazon Kinesis Data Firehose

- Amazon Kinesis Data Streams

- Amazon Managed Service for Apache Flink

- Amazon Managed Streaming for Apache Kafka (Amazon MSK)

- Amazon OpenSearch Service

- Amazon QuickSuite

- Amazon SageMaker AI

Application Integration

- Amazon AppFlow

- Amazon EventBridge

- Amazon Managed Workflows for Apache Airflow (Amazon MWAA)

- Amazon Simple Notification Service (Amazon SNS)

- Amazon Simple Queue Service (Amazon SQS)

- AWS Step Functions

Cloud Financial Management

- AWS Budgets

- AWS Cost Explorer

Compute

- AWS Batch

- Amazon EC2

- AWS Lambda

- AWS Serverless Application Model (AWS SAM)

Containers

- Amazon Elastic Container Registry (Amazon ECR)

- Amazon Elastic Container Service (Amazon ECS)

- Amazon Elastic Kubernetes Service (Amazon EKS)

Database

- Amazon DocumentDB (with MongoDB compatibility)

- Amazon DynamoDB

- Amazon Keyspaces (for Apache Cassandra)

- Amazon MemoryDB for Redis

- Amazon Neptune

- Amazon RDS

- Amazon Aurora

- Amazon Redshift

Developer Tools

- AWS CLI

- AWS CloudFormation

- AWS Cloud Development Kit (AWS CDK)

- AWS CodeBuild

- AWS CodeDeploy

- AWS CodePipeline

- Amazon Q

Web and Mobile

- Amazon API Gateway

Machine Learning

- Amazon SageMaker AI

- Amazon Bedrock

- Amazon Kendra

Management and Governance

- AWS CloudTrail

- Amazon CloudWatch

- Amazon CloudWatch Logs

- AWS Config

- Amazon Managed Grafana

- AWS Systems Manager

- AWS Well-Architected Tool

- AWS Data Exchange

Migration and Transfer

- AWS Application Discovery Service

- AWS Application Migration Service

- AWS Database Migration Service (AWS DMS)

- AWS DataSync

- AWS Snow Family

- AWS Transfer Family

Networking and Content Delivery

- Amazon CloudFront

- AWS PrivateLink

- Amazon Route 53

- Amazon VPC

Security, Identity, and Compliance

- AWS Identity and Access Management (IAM)

- AWS Key Management Service (AWS KMS)

- Amazon Macie

- AWS Secrets Manager

- AWS Shield

- AWS WAF

Storage

- AWS Backup

- Amazon Elastic Block Store (Amazon EBS)

- Amazon Elastic File System (Amazon EFS)

- Amazon S3

- Amazon S3 Tables

- Amazon S3 Glacier

- Amazon S3 Tables

Out-of-Scope AWS Services

The following list contains AWS services and features that are out of scope for the exam. This list is non-exhaustive and is subject to change. AWS offerings that are entirely unrelated to the target job roles for the exam are excluded from this list.

Topics

- Analytics
- Business Applications
- Compute
- Containers
- Developer Tools
- Frontend Web and Mobile
- Internet of Things (IoT)
- Machine Learning
- Management and Governance
- Media Services
- Migration and Transfer
- Storage

Analytics

- Amazon FinSpace

Business Applications

- Alexa for Business

- Amazon Chime

- Amazon Connect

- AWS IQ

- Amazon WorkMail

Compute

- AWS App Runner

- AWS Elastic Beanstalk

- Amazon Lightsail

- AWS Outposts

- AWS Serverless Application Repository

Containers

- Red Hat OpenShift Service on AWS (ROSA)

Developer Tools

- AWS Fault Injection Simulator (AWS FIS)

- AWS X-Ray

Frontend Web and Mobile

- AWS Amplify

- AWS AppSync

- AWS Device Farm

- Amazon Location Service

- Amazon Pinpoint

- Amazon Simple Email Service (Amazon SES)

Internet of Things (IoT)

- FreeRTOS
- AWS IoT 1-Click
- AWS IoT Device Defender
- AWS IoT Device Management
- AWS IoT Events
- AWS IoT FleetWise
- AWS IoT RoboRunner
- AWS IoT SiteWise
- AWS IoT TwinMaker

Machine Learning

- Amazon DevOps Guru

Management and Governance

- AWS Activate
- AWS Managed Services (AMS)

Media Services

- Amazon Elastic Transcoder
- AWS Elemental Appliances and Software
- AWS Elemental MediaConnect
- AWS Elemental MediaConvert
- AWS Elemental MediaLive
- AWS Elemental MediaPackage
- AWS Elemental MediaStore

- AWS Elemental MediaTailor
- Amazon Interactive Video Service (Amazon IVS)
- Amazon Nimble Studio

Migration and Transfer

- AWS Mainframe Modernization
- AWS Migration Hub

Storage

- EC2 Image Builder

Revisions

AWS exam guides are periodically reviewed and updated to ensure that our certification exams test skills and AWS services and features that are relevant for the job role(s) that a certification is designed to target. Exam guide updates will be published approximately one month before updates will be reflected on your exam.

Topics

- [Change History](#)
- [Changes to knowledge and skills](#)
- [New skills added](#)
- [Skills removed](#)
- [Changes to in- and out-of-scope services](#)

Change History

Version	Publication date
1.0	This is the first release.
1.1	December 12, 2025

The separate knowledge and skills in Version 1.0 of the exam guide were consolidated into one list of skills under each task. Knowledge items in Version 1.0 that overlapped with existing skills were removed in Version 1.1.

Changes to knowledge and skills

Version 1.1	Version 1.0
Skill 1.4.8: Use infrastructure as code (IaC) for repeatable resource deployment (for example AWS CloudFormation and AWS Cloud Development Kit (AWS CDK)).	Knowledge of: Infrastructure as code (IaC) for repeatable deployments (for example, AWS Cloud Development Kit [AWS CDK], AWS CloudFormation)
Skill 2.1.3: Apply storage services to appropriate use cases (for example, using indexing algorithms like HNSW with Amazon Aurora PostgreSQL and using Amazon MemoryDB for fast key/value pair access).	Skills in: Applying storage services to appropriate use cases (for example, Amazon S3)
Skill 2.2.2: Build and reference a technical data catalog (for example, AWS Glue Data Catalog, Apache Hive metastore).	Skills in: Building and referencing a data catalog (for example, AWS Glue Data Catalog, Apache Hive metastore)
Skill 2.4.4: Establish data lineage by using AWS tools (for example, Amazon SageMaker ML Lineage Tracking and Amazon SageMaker Catalog).	Skills in: Establishing data lineage by using AWS tools (for example, Amazon SageMaker ML Lineage Tracking)
Skill 3.1.6: Prepare data for transformation (for example, AWS Glue DataBrew and Amazon SageMaker Unified Studio).	Skills in: Preparing data transformation (for example, AWS Glue DataBrew)
Skill 3.2.3: Use SQL in Redshift and Athena to query data or to create views.	Skills in: Using Athena to query data or to create views
Skill 4.2.5: Apply authorization methods that address business needs (role-based, tag-based, and attribute-based).	Knowledge of: Authorization methods (role-based, policy-based, tag-based, and attribute based)

Version 1.1	Version 1.0
Skill 4.2.6: Construct custom policies that meet the principle of least privilege.	Knowledge of: Principle of least privilege as it applies to AWS security
Skill 4.3.4: Enable encryption in transit or before transit for data.	Skills in: Enabling encryption in transit for data
Skill 4.5.4: Viewing configuration changes that have occurred in an account (for example, AWS Config).	Skills in: Managing configuration changes that have occurred in an account (for example, AWS Config)
Skill 4.5.5: Maintain data sovereignty.	Knowledge of: Data sovereignty

New skills added

- Skill 1.2.10: Integrate Large Language Models (LLM) for data processing.
- Skill 2.1.7: Manage open table formats (for example Apache Iceberg).
- Skill 2.1.8: Describe vector index types (for example, HNSW, IVF).
- Skill 2.2.6: Create and manage business data catalogs (for example Amazon SageMaker Catalog).
- Skill 2.4.6: Describe vectorization concepts (for example, Amazon Bedrock knowledge base).
- Skill 4.1.7: Use domain, domain units, and projects for SageMaker Unified Studio.
- Skill 4.5.6: Manage data access through Amazon SageMaker Catalog projects.
- Skill 4.5.7: Describe governance data framework and data sharing patterns.

Skills removed

There are no knowledge and skills removed on Version 1.1 other than knowledge items that were already covered by existing skills.

Changes to in- and out-of-scope services

Services added to the in-scope list

- Amazon Aurora
- Amazon Q

- Amazon Bedrock
- Amazon Kendra
- AWS Data Exchange
- Amazon S3 Tables

Services removed from the in-scope list

- AWS Cloud9
- AWS CodeCommit
- AWS Schema Conversion Tool (AWS SCT)

Services added to the out-of-scope list

No services were added to the out-of-scope list

Services removed from the out-of-scope list (all due to service changes or deprecations)

- Amazon Honeycode
- Amazon WorkDocs
- Amazon Timestream
- Amazon CodeWhisperer

Survey

How useful was this exam guide? Let us know by taking our survey.