

## **AWS Certified Data Analytics – Specialty (DAS-C01) Exam Guide**

### **Introduction**

The AWS Certified Data Analytics – Specialty (DAS-C01) exam is intended for individuals who perform a data analytics role. The exam validates a candidate's ability to use AWS services to design, build, secure, and maintain analytics solutions that provide insight from data.

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

- Define AWS data analytics services and understand how they integrate with each other.
- Explain how AWS data analytics services fit in the data lifecycle of collection, storage, processing, and visualization.

### **Target candidate description**

The target candidate should have 5 years of experience with common data analytics technologies. The target candidate also should have at least 2 years of hands-on experience and expertise working with AWS services to design, build, secure, and maintain analytics solutions.

#### **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:

- Design and implement machine learning algorithms.
- Implement container-based solutions.
- Use high performance computing (HPC).
- Design online transactional processing (OLTP) database solutions.

Refer to the Appendix for a list of in-scope AWS services and features and a list of out-of-scope AWS services and features.

## 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.

### 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 Analytics – Specialty (DAS-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 750. 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:

- Domain 1: Collection (18% of scored content)
- Domain 2: Storage and Data Management (22% of scored content)
- Domain 3: Processing (24% of scored content)
- Domain 4: Analysis and Visualization (18% of scored content)
- Domain 5: Security (18% of scored content)

## **Domain 1: Collection**

Task Statement 1.1: Determine the operational characteristics of the collection system.

- Confirm that data loss is within tolerance limits in the event of failures.
- Evaluate costs associated with data acquisition, transfer, and provisioning from various sources into the collection system (for example, networking, bandwidth, ETL, data migration).
- Assess the failure scenarios that the collection system may experience, and take remediation actions based on impact.
- Determine data persistence at various points of data capture.
- Identify the latency characteristics of the collection system.

Task Statement 1.2: Select a collection system that handles the frequency, volume, and source of data.

- Describe and characterize the volume and flow characteristics of incoming data (for example, streaming, transactional, batch).
- Match the flow characteristics of data to potential solutions.
- Assess the tradeoffs between various ingestion services, and take into account scalability, cost, fault tolerance, and latency.
- Explain the throughput capability of a variety of types of data collection solutions, and identify bottlenecks.
- Choose a collection solution that satisfies connectivity constraints of the source data system.

Task Statement 1.3: Select a collection system that addresses the key properties of data, such as order, format, and compression.

- Describe how to capture data changes at the source.
- Discuss data structure and format, compression applied, and encryption requirements.
- Distinguish the impact of out-of-order delivery of data, duplicate delivery of data, and the tradeoffs between at-most-once, exactly-once, and at-least-once processing.
- Describe how to transform and filter data during the collection process.

## **Domain 2: Storage and Data Management**

Task Statement 2.1: Determine the operational characteristics of the storage solution for analytics.

- Determine the appropriate storage service or services on the basis of cost compared to performance.
- Understand the durability, reliability, and latency characteristics of the storage solution based on requirements.
- Determine the requirements of a system for strong or eventual consistency of the storage system.
- Determine the appropriate storage solution to address data freshness requirements.

Task Statement 2.2: Determine data access and retrieval patterns.

- Determine the appropriate storage solution based on update patterns (for example, bulk, transactional, micro batching).
- Determine the appropriate storage solution based on access patterns (for example, sequential or random access, continuous usage or one-time usage).
- Determine the appropriate storage solution to address change characteristics of data (append-only changes or updates).
- Determine the appropriate storage solution for long-term storage and transient storage.
- Determine the appropriate storage solution for structured data and semi-structured data.
- Determine the appropriate storage solution to address query latency requirements.

Task Statement 2.3: Select appropriate data layout, schema, structure, and format.

- Determine appropriate mechanisms to address schema evolution requirements.
- Select the appropriate storage format for a specific task.
- Select the appropriate compression and encoding strategies for a chosen storage format.
- Select the appropriate data sorting and distribution strategies and the storage layout for efficient data access.
- Explain the cost and performance implications of different data distributions, layouts, and formats (for example, size and number of files).
- Implement data formatting and partitioning schemes for data-optimized analysis.

Task Statement 2.4: Define data lifecycles based on usage patterns and business requirements.

- Determine the appropriate strategy to address data lifecycle requirements.
- Apply appropriate lifecycle and data retention policies to different storage solutions.

Task Statement 2.5: Determine the appropriate system to catalog data and to manage metadata.

- Evaluate mechanisms to discover new and updated data sources.
- Evaluate mechanisms to create and update data catalogs and metadata.
- Explain mechanisms to search and retrieve data catalogs and metadata.
- Explain mechanisms to tag and classify data.

### **Domain 3: Processing**

Task Statement 3.1: Determine appropriate data processing solution requirements.

- Understand data preparation and usage requirements.
- Understand different types of data sources and targets.
- Evaluate performance and orchestration needs.
- Evaluate appropriate services for cost, scalability, and availability.

Task Statement 3.2: Design a solution to transform and prepare data for analysis.

- Apply appropriate ETL and ELT techniques for batch workloads and real-time workloads.
- Implement failover, scaling, and replication mechanisms.
- Implement techniques to address concurrency needs.
- Implement techniques to improve cost-optimization efficiencies.
- Orchestrate workflows.
- Aggregate and enrich data for downstream consumption.

Task Statement 3.3: Automate and operationalize data processing solutions.

- Implement automated techniques for repeatable workflows.
- Apply methods to identify and recover from processing failures.
- Deploy logging and monitoring solutions to enable auditing and traceability.

## Domain 4: Analysis and Visualization

Task Statement 4.1: Determine the operational characteristics of an analysis and visualization solution.

- Determine costs associated with analysis and visualization.
- Determine scalability associated with analysis.
- Determine failover recovery and fault tolerance within the RPO and RTO.
- Determine the availability characteristics of an analysis tool.
- Evaluate dynamic, interactive, and static presentations of data.
- Translate performance requirements to an appropriate visualization approach (for example, pre-compute and consume static data, consume dynamic data).

Task Statement 4.2: Select the appropriate data analysis solution for a given scenario.

- Evaluate and compare analysis solutions.
- Select the right type of analysis based on the customer use case (for example, streaming, interactive, collaborative, operational).

Task Statement 4.3: Select the appropriate data visualization solution for a given scenario.

- Evaluate output capabilities for a given analysis solution (for example, metrics, KPIs, tabular, API).
- Choose the appropriate method for data delivery (for example, web, mobile, email, collaborative notebooks).
- Choose and define the appropriate data refresh schedule.
- Choose appropriate tools for different data freshness requirements (for example, Amazon OpenSearch Service, Amazon QuickSight, Amazon EMR notebooks).
- Understand the capabilities of visualization tools for interactive use cases (for example, drill down, drill through, pivot).
- Implement the appropriate data access mechanism (for example, in memory, direct access).
- Implement an integrated solution from multiple heterogeneous data sources.

## **Domain 5: Security**

Task Statement 5.1: Select appropriate authentication and authorization mechanisms.

- Implement appropriate authentication methods (for example, federated access, SSO, AWS Identity and Access Management [IAM]).
- Implement appropriate authorization methods (for example, policies, ACLs, table and column level permissions).
- Implement appropriate access control mechanisms (for example, security groups, role-based controls).

Task Statement 5.2: Apply data protection and encryption techniques.

- Determine data encryption and masking needs.
- Apply different encryption approaches (for example, server-side encryption, client-side encryption, AWS Key Management Service [AWS KMS], AWS CloudHSM).
- Implement at-rest and in-transit encryption mechanisms.
- Implement data obfuscation and masking techniques.
- Apply basic principles of key rotation and secrets management.

Task Statement 5.3: Apply data governance and compliance controls.

- Determine data governance and compliance requirements.
- Understand and configure access, and audit logging across data analytics services.
- Implement appropriate controls to meet compliance requirements.



## Appendix

### In-scope AWS services and features

The following list contains AWS services and features that are in scope for the exam. This list is non-exhaustive and is subject to change. AWS offerings appear in categories that align with the offerings' primary functions:

#### Analytics:

- Amazon Athena
- Amazon CloudSearch
- Amazon EMR
- AWS Glue
- Amazon Kinesis
- AWS Lake Formation
- Amazon Managed Streaming for Apache Kafka (Amazon MSK)
- Amazon OpenSearch Service
- Amazon QuickSight

#### Application Integration:

- Amazon MQ
- Amazon Simple Notification Service (Amazon SNS)
- Amazon Simple Queue Service (Amazon SQS)
- AWS Step Functions

#### Compute:

- AWS Auto Scaling
- Amazon EC2
- AWS Lambda

#### Database:

- Amazon DocumentDB (with MongoDB compatibility)
- Amazon DynamoDB
- Amazon ElastiCache
- Amazon Neptune
- Amazon RDS
- Amazon Redshift
- Amazon Timestream

#### Frontend Web and Mobile:

- Amazon API Gateway
- AWS AppSync
- Amazon Simple Email Service (Amazon SES)

#### Management and Governance:

- AWS CloudFormation
- AWS CloudTrail
- Amazon CloudWatch
- AWS Trusted Advisor

#### Machine Learning:

- Amazon SageMaker

#### Migration and Transfer:

- AWS Database Migration Service (AWS DMS)
- AWS DataSync
- AWS Snowball
- AWS Transfer Family

#### Networking and Content Delivery:

- AWS Direct Connect
- Elastic Load Balancing (ELB)
- Amazon VPC

### Security, Identity, and Compliance:

- AWS Artifact
- AWS Certificate Manager (ACM)
- AWS CloudHSM
- Amazon Cognito
- AWS IAM Identity Center (successor to AWS Single Sign-On)
- AWS Identity and Access Management (IAM)
- AWS Key Management Service (AWS KMS)
- Amazon Macie
- AWS Secrets Manager

### Storage:

- Amazon Elastic Block Store (Amazon EBS)
- Amazon S3
- Amazon S3 Glacier

### **Out-of-scope AWS services and features**

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:

### Internet of Things:

- AWS IoT Core

### Media Services:

- Amazon Kinesis Video Streams

### **Survey**

How useful was this exam guide? Let us know by [taking our survey](#).