

## 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 comprehensive understanding of how 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 a minimum of 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.

#### What is considered out of scope for the target candidate?

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

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

For a detailed list of specific tools and technologies that might be covered on the exam, as well as lists of in-scope and out-of-scope AWS services, refer to the Appendix.

### 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 will affect your score.

## Unscored content

The exam includes 15 unscored questions that do not affect your score. AWS collects information about candidate 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 is a pass or fail exam. 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 or not 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. This information is intended to provide general feedback about your exam performance. 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 contains general information that highlights your strengths and weaknesses. Use caution when interpreting section-level feedback.

## Content outline

This exam guide includes weightings, test domains, and objectives for the exam. It is not a comprehensive listing of the content on the exam. However, additional context for each of the objectives is available to help guide your preparation for the exam. The following table lists the main content domains and their weightings. The table precedes the complete exam content outline, which includes the additional context. The percentage in each domain represents only scored content.

Domain	% of Exam
Domain 1: Collection	18%
Domain 2: Storage and Data Management	22%
Domain 3: Processing	24%
Domain 4: Analysis and Visualization	18%
Domain 5: Security	18%
<b>TOTAL</b>	<b>100%</b>

## Domain 1: Collection

- 1.1 Determine the operational characteristics of the collection system
  - Evaluate that the 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 (e.g., networking, bandwidth, ETL/data migration costs)
  - Assess the failure scenarios that the collection system may undergo, and take remediation actions based on impact
  - Determine data persistence at various points of data capture
  - Identify the latency characteristics of the collection system
- 1.2 Select a collection system that handles the frequency, volume, and the source of data
  - Describe and characterize the volume and flow characteristics of incoming data (streaming, transactional, batch)
  - Match flow characteristics of data to potential solutions
  - Assess the tradeoffs between various ingestion services taking into account scalability, cost, fault tolerance, latency, etc.
  - Explain the throughput capability of a variety of different types of data collection and identify bottlenecks
  - Choose a collection solution that satisfies connectivity constraints of the source data system
- 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

- 2.1 Determine the operational characteristics of the storage solution for analytics
  - Determine the appropriate storage service(s) on the basis of cost vs. performance
  - Understand the durability, reliability, and latency characteristics of the storage solution based on requirements
  - Determine the requirements of a system for strong vs. eventual consistency of the storage system
  - Determine the appropriate storage solution to address data freshness requirements
- 2.2 Determine data access and retrieval patterns
  - Determine the appropriate storage solution based on update patterns (e.g., bulk, transactional, micro batching)
  - Determine the appropriate storage solution based on access patterns (e.g., sequential vs. random access, continuous usage vs. ad hoc)
  - Determine the appropriate storage solution to address change characteristics of data (append-only changes vs. updates)
  - Determine the appropriate storage solution for long-term storage vs. transient storage
  - Determine the appropriate storage solution for structured vs. semi-structured data
  - Determine the appropriate storage solution to address query latency requirements

- 2.3 Select appropriate data layout, schema, structure, and format
  - Determine appropriate mechanisms to address schema evolution requirements
  - Select the storage format for the task
  - Select the compression/encoding strategies for the chosen storage format
  - Select the 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 (e.g., size and number of files)
  - Implement data formatting and partitioning schemes for data-optimized analysis
- 2.4 Define data lifecycle based on usage patterns and business requirements
  - Determine the strategy to address data lifecycle requirements
  - Apply the lifecycle and data retention policies to different storage solutions
- 2.5 Determine the appropriate system for cataloging data and managing metadata
  - Evaluate mechanisms for discovery of new and updated data sources
  - Evaluate mechanisms for creating and updating data catalogs and metadata
  - Explain mechanisms for searching and retrieving data catalogs and metadata
  - Explain mechanisms for tagging and classifying data

### **Domain 3: Processing**

- 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
- 3.2 Design a solution for transforming and preparing data for analysis
  - Apply appropriate ETL/ELT techniques for batch and real-time workloads
  - Implement failover, scaling, and replication mechanisms
  - Implement techniques to address concurrency needs
  - Implement techniques to improve cost-optimization efficiencies
  - Apply orchestration workflows
  - Aggregate and enrich data for downstream consumption
- 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**

- 4.1 Determine the operational characteristics of the 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/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 (pre-compute and consume static data vs. consume dynamic data)

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

## Domain 5: Security

- 5.1 Select appropriate authentication and authorization mechanisms
  - Implement appropriate authentication methods (e.g., federated access, SSO, IAM)
  - Implement appropriate authorization methods (e.g., policies, ACL, table/column level permissions)
  - Implement appropriate access control mechanisms (e.g., security groups, role-based control)
- 5.2 Apply data protection and encryption techniques
  - Determine data encryption and masking needs
  - Apply different encryption approaches (server-side encryption, client-side encryption, 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
- 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

### Which key tools, technologies, and concepts might be covered on the exam?

The following is a non-exhaustive list of the tools and technologies that could appear on the exam. This list is subject to change and is provided to help you understand the general scope of services, features, or technologies on the exam. AWS services are grouped according to their primary functions. While some of these technologies will likely be covered more than others on the exam, the order and placement of them in this list is no indication of relative weight or importance:

#### AWS services and features

Analytics:

- Amazon Athena
- Amazon CloudSearch
- Amazon Elasticsearch Service (Amazon ES)
- Amazon EMR
- AWS Glue
- Amazon Kinesis (excluding Kinesis Video Streams)
- AWS Lake Formation
- Amazon Managed Streaming for Apache Kafka
- Amazon QuickSight
- Amazon Redshift

Application Integration:

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

Compute:

- Amazon EC2
- Elastic Load Balancing
- AWS Lambda

Customer Engagement:

- Amazon Simple Email Service (Amazon SES)

Database:

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

#### Management and Governance:

- AWS Auto Scaling
- 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 for SFTP

#### Networking and Content Delivery:

- Amazon API Gateway
- AWS Direct Connect
- Amazon VPC (and associated features)

#### Security, Identity, and Compliance:

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

#### Storage:

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

## Out-of-scope AWS services and features

The following is a non-exhaustive list of AWS services and features that are not covered on the exam. These services and features do not represent every AWS offering that is excluded from the exam content. Services or features that are entirely unrelated to the target job roles for the exam are excluded from this list because they are assumed to be irrelevant.

Out-of-scope AWS services and features include the following:

- AWS IoT Core