

Introduction

The AWS Certified Big Data - Specialty examination validates technical skills and experience in designing and implementing AWS services to derive value from data. The examination is for individuals who perform complex Big Data analyses, and validates an individual's ability to:

- Implement core AWS Big Data services according to basic architectural best practices
- Design and maintain Big Data
- Leverage tools to automate Data Analysis

Required Prerequisite Certification

- Valid AWS Associate Certification (*AWS Certified Solutions Architect — Associate, or AWS Certified Developer — Associate, or AWS Certified SysOps Administrator — Associate*)

Recommended AWS Knowledge

- A minimum of 2 years' experience using AWS technology
- AWS Security best practices
- Independently define AWS architecture and services and understand how they integrate with each other.
- Define and architect AWS big data services and explain how they fit in the data lifecycle of collection, ingestion, storage, processing, and visualization.

Recommended General IT Knowledge

- At least 5 years' experience in a data analytics field
- Understand how to control access to secure data
- Understand the frameworks that underpin large scale distributed systems like Hadoop/Spark and MPP data warehouses
- Understand the tools and design platforms that allow processing of data from multiple heterogeneous sources with different frequencies (batch/real-time)
- Capable of designing a scalable and cost-effective architecture to process data

Examination Preparation

These training courses and materials may be helpful for examination preparation:

AWS Training (aws.amazon.com/training)

- Big Data Technology Fundamentals
<https://aws.amazon.com/training/course-descriptions/bigdata-fundamentals/>
- Big Data on AWS
<https://aws.amazon.com/training/course-descriptions/bigdata/>

AWS Whitepapers (aws.amazon.com/whitepapers)

- AWS Cloud Computing Whitepapers (aws.amazon.com/whitepapers), specifically [Database](#) and [Analytics](#)
- AWS Documentation (aws.amazon.com/documentation)

Exam Content

Response Types

There are two types of multiple-choice questions on the examination. One type has one correct response and 3-4 incorrect responses (distractors). The other type has two or more correct responses out of a total of 5-6 responses. Always choose the best response(s). Incorrect responses are plausible, and are designed to be attractive to candidates who do not know the correct response. Unanswered questions are scored as incorrect; there is no penalty for guessing.

Your examination may include non-scored questions that are placed on the test to gather statistical information. These questions will not be identified on the examination, and will not affect your score.

Content Outline

This examination blueprint includes weighting, test objectives, and example content. Example topics and concepts are included to clarify the test objectives. They should not be construed as a comprehensive listing of all of the content of this examination.

The table below lists the main content domains and their weighting on the examination.

Domain	% of Examination
1.0 Collection	17%
2.0 Storage	17%
3.0 Processing	17%
4.0 Analysis	17%
5.0 Visualization	12%
6.0 Data Security	20%
TOTAL	100%

Domain 1.0: Collection

- 1.1 Determine the operational characteristics of the collection system
- 1.2 Select a collection system that handles the frequency of data change and type of data being ingested
- 1.3 Identify the properties that need to be enforced by the collection system: order, data structure, metadata, etc.
- 1.4 Explain the durability and availability characteristics for the collection approach

Domain 2.0: Storage

- 2.1 Determine and optimize the operational characteristics of the storage solution
- 2.2 Determine data access and retrieval patterns
- 2.3 Evaluate mechanisms for capture, update, and retrieval of catalog entries
- 2.4 Determine appropriate data structure and storage format

Domain 3.0: Processing

- 3.1 Identify the appropriate data processing technology needed for a given scenario
- 3.2 Determine how to design and architect the data processing solution
- 3.3 Determine the operational characteristics of the solution implemented

Domain 4.0: Analysis

- 4.1 Determine the tools and techniques required for analysis
- 4.2 Determine how to design and architect the analytical solution
- 4.3 Determine and optimize the operational characteristics of the Analysis System

Domain 5.0: Visualization

- 5.1 Determine the appropriate techniques for delivering the results/output
- 5.2 Determine how to design and create the Visualization platform
- 5.3 Determine and optimize the operational characteristics of the Visualization system

Domain 6.0: Data Security

- 6.1 Determine encryption requirements and/or implementation technologies
- 6.2 Choose the appropriate technology to enforce data governance
- 6.3 Identify how to ensure data integrity
- 6.4 Evaluate regulatory requirements