AWS DAS-C01

AWS DATA ANALYTICS SPECIALTY CERTIFICATION QUESTIONS & ANSWERS

Exam Summary – Syllabus – Questions

DAS-C01

AWS Certified Data Analytics - Specialty

65 Questions Exam – 750 / 1000 Cut Score – Duration of 180 minutes

www.VMExam.com

Table of Contents

Know Your DAS-C01 Certification Well:

The DAS-C01 is best suitable for candidates who want to gain knowledge in the AWS Specialty. Before you start your DAS-C01 preparation you may struggle to get all the crucial Data Analytics Specialty materials like DAS-C01 syllabus, sample questions, study guide.

But don't worry the DAS-C01 PDF is here to help you prepare in a stress free manner.

The PDF is a combination of all your queries like-

- What is in the DAS-C01 syllabus?
- How many questions are there in the DAS-C01 exam?
- Which Practice test would help me to pass the DAS-C01 exam at the first attempt?

Passing the DAS-C01 exam makes you AWS Certified Data Analytics - Specialty. Having the Data Analytics Specialty certification opens multiple opportunities for you. You can grab a new job, get a higher salary or simply get recognition within your current organization.

AWS DAS-C01 Data Analytics Specialty Certification Details:

Exam Name	AWS Certified Data Analytics - Specialty (Data Analytics Specialty)
Exam Code	DAS-C01
Exam Price	$300 USD
Duration	180 minutes
Number of Questions	65
Passing Score	750 / 1000
Recommended Training / Books	**Data Analytics Fundamentals** **Big Data on AWS**
Schedule Exam	**PEARSON VUE**
Sample Questions	**AWS DAS-C01 Sample Questions**
Recommended Practice	**AWS Certified Data Analytics - Specialty Practice Test**

DAS-C01 Syllabus:

Section	Objectives
Collection - 18%	
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
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
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
Storage and Data Management - 22%	
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
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

Section	Objectives
	patterns (e.g., sequential vs. random access, continuous usage vs.ad hoc) - Determine the appropriate storage solution to address change characteristics of data (appendonly 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
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
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
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
	Processing - 24%
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
Design a solution for transforming	- Apply appropriate ETL/ELT techniques for batch and real-time workloads - Implement failover, scaling, and replication mechanisms

Section	Objectives
and preparing data for analysis	- Implement techniques to address concurrency needs - Implement techniques to improve cost-optimization efficiencies - Apply orchestration workflows - Aggregate and enrich data for downstream consumption
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

Analysis and Visualization - 18%

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)
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)
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

Security - 18%

Select appropriate authentication and	- Implement appropriate authentication methods (e.g., federated access, SSO, IAM)

Section	Objectives
authorization mechanisms	- Implement appropriate authorization methods (e.g., policies, ACL, table/column level permissions) - Implement appropriate access control mechanisms (e.g., security groups, role-based control)
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
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

AWS DAS-C01 Sample Questions:

Question: 1

A company needs to implement a near-real-time fraud prevention feature for its ecommerce site.

User and order details need to be delivered to an Amazon SageMaker endpoint to flag suspected fraud. The amount of input data needed for the inference could be as much as 1.5 MB.

Which solution meets the requirements with the LOWEST overall latency?

a) Create an Amazon Managed Streaming for Kafka cluster and ingest the data for each order into a topic. Use a Kafka consumer running on Amazon EC2 instances to read these messages and invoke the Amazon SageMaker endpoint.
b) Create an Amazon Kinesis Data Streams stream and ingest the data for each order into the stream. Create an AWS Lambda function to read these messages and invoke the Amazon SageMaker endpoint.
c) Create an Amazon Kinesis Data Firehose delivery stream and ingest the data for each order into the stream. Configure Kinesis Data Firehose to deliver the data to an Amazon S3 bucket. Trigger an AWS Lambda function with an S3 event notification to read the data and invoke the Amazon SageMaker endpoint.
d) Create an Amazon SNS topic and publish the data for each order to the topic. Subscribe the Amazon SageMaker endpoint to the SNS topic.

Answer: a

Question: 2

A financial company uses Amazon EMR for its analytics workloads. During the company's annual security audit, the security team determined that none of the EMR clusters' root volumes are encrypted. The security team recommends the company encrypt its EMR clusters' root volume as soon as possible.

Which solution would meet these requirements?

a) Enable at-rest encryption for EMR File System (EMRFS) data in Amazon S3 in a security configuration. Re-create the cluster using the newly created security configuration.
b) Specify local disk encryption in a security configuration. Re-create the cluster using the newly created security configuration.
c) Detach the Amazon EBS volumes from the master node. Encrypt the EBS volume and attach it back to the master node.
d) Re-create the EMR cluster with LZO encryption enabled on all volumes.

Answer: b

Question: 3

A company is providing analytics services to its marketing and human resources (HR) departments. The departments can only access the data through their business intelligence (BI) tools, which run Presto queries on an Amazon EMR cluster that uses the EMR File System (EMRFS).

The marketing data analyst must be granted access to the advertising table only. The HR data analyst must be granted access to the personnel table only.

Which approach will satisfy these requirements?

a) Create separate IAM roles for the marketing and HR users. Assign the roles with AWS Glue resourcebased policies to access their corresponding tables in the AWS Glue Data Catalog. Configure Presto to use the AWS Glue Data Catalog as the Apache Hive metastore.
b) Create the marketing and HR users in Apache Ranger. Create separate policies that allow access to the user's corresponding table only. Configure Presto to use Apache Ranger and an external Apache Hive metastore running in Amazon RDS.
c) Create separate IAM roles for the marketing and HR users. Configure EMR to use IAM roles for EMRFS access. Create a separate bucket for the HR and marketing data. Assign appropriate permissions so the users will only see their corresponding datasets.
d) Create the marketing and HR users in Apache Ranger. Create separate policies that allows access to the user's corresponding table only. Configure Presto to use Apache Ranger and the AWS Glue Data Catalog as the Apache Hive metastore.

Answer: a

Question: 4

A data engineer needs to create a dashboard to display social media trends during the last hour of a large company event. The dashboard needs to display the associated metrics with a consistent latency of less than 2 minutes.

Which solution meets these requirements?

a) Publish the raw social media data to an Amazon Kinesis Data Firehose delivery stream. Use Kinesis Data Analytics for SQL Applications to perform a sliding window analysis to compute the metrics and output the results to a Kinesis Data Streams data stream. Configure an AWS Lambda function to save the stream data to an Amazon DynamoDB table. Deploy a real-time dashboard hosted in an Amazon S3 bucket to read and display the metrics data stored in the DynamoDB table.

b) Publish the raw social media data to an Amazon Kinesis Data Firehose delivery stream. Configure the stream to deliver the data to an Amazon Elasticsearch Service cluster with a buffer interval of 0 seconds. Use Kibana to perform the analysis and display the results.

c) Publish the raw social media data to an Amazon Kinesis Data Streams data stream. Configure an AWS Lambda function to compute the metrics on the stream data and save the results in an Amazon S3 bucket. Configure a dashboard in Amazon QuickSight to query the data using Amazon Athena and display the results.

d) Publish the raw social media data to an Amazon SNS topic. Subscribe an Amazon SQS queue to the topic. Configure Amazon EC2 instances as workers to poll the queue, compute the metrics, and save the results to an Amazon Aurora MySQL database. Configure a dashboard in Amazon QuickSight to query the data in Aurora and display the results.

Answer: a

Question: 5

A company ingests a large set of clickstream data in nested JSON format from different sources and stores it in Amazon S3.

Data analysts need to analyze this data in combination with data stored in an Amazon Redshift cluster. Data analysts want to build a cost-effective and automated solution for this need.

Which solution meets these requirements?

a) Use Apache Spark SQL on Amazon EMR to convert the clickstream data to a tabular format. Use the Amazon Redshift COPY command to load the data into the Amazon Redshift cluster.

b) Use AWS Lambda to convert the data to a tabular format and write it to Amazon S3. Use the Amazon Redshift COPY command to load the data into the Amazon Redshift cluster.

c) Use the Relationalize class in an AWS Glue ETL job to transform the data and write the data back to Amazon S3. Use Amazon Redshift Spectrum to create external tables and join with the internal tables.

d) Use the Amazon Redshift COPY command to move the clickstream data directly into new tables in the Amazon Redshift cluster.

Answer: c

Question: 6

A media company is migrating its on-premises legacy Hadoop cluster with its associated data processing scripts and workflow to an Amazon EMR environment running the latest Hadoop release. The developers want to reuse the Java code that was written for data processing jobs for the on-premises cluster.

Which approach meets these requirements?

a) Deploy the existing Oracle Java Archive as a custom bootstrap action and run the job on the EMR cluster.
b) Compile the Java program for the desired Hadoop version and run it using a CUSTOM_JAR step on the EMR cluster.
c) Submit the Java program as an Apache Hive or Apache Spark step for the EMR cluster.
d) Use SSH to connect the master node of the EMR cluster and submit the Java program using the AWS CLI.

Answer: b

Question: 7

A publisher website captures user activity and sends clickstream data to Amazon Kinesis Data Streams.

The publisher wants to design a cost-effective solution to process the data to create a timeline of user activity within a session. The solution must be able to scale depending on the number of active sessions.

Which solution meets these requirements?

a) Include a variable in the clickstream data from the publisher website to maintain a counter for the number of active user sessions. Use a timestamp for the partition key for the stream. Configure the consumer application to read the data from the stream and change the number of processor threads based upon the counter. Deploy the consumer application on Amazon EC2 instances in an EC2 Auto Scaling group.
b) Include a variable in the clickstream to maintain a counter for each user action during their session. Use the action type as the partition key for the stream. Use the Kinesis Client Library (KCL) in the consumer application to retrieve the data from the stream and perform the processing. Configure the consumer application to read the data from the stream and change the number of processor threads based upon the counter. Deploy the consumer application on AWS Lambda.
c) Include a session identifier in the clickstream data from the publisher website and use as the partition key for the stream. Use the Kinesis Client Library (KCL) in the consumer application to retrieve the data from the stream and perform the processing. Deploy the consumer application on Amazon EC2 instances in an EC2 Auto Scaling group. Use an AWS Lambda function to reshard the stream based upon Amazon CloudWatch alarms.
d) Include a variable in the clickstream data from the publisher website to maintain a counter for the number of active user sessions. Use a timestamp for the partition key for the stream. Configure the consumer application to read the data from the stream and change the number of processor threads based upon the counter. Deploy the consumer application on AWS Lambda.

Answer: c

Question: 8

A company is currently using Amazon DynamoDB as the database for a user support application.

The company is developing a new version of the application that will store a PDF file for each support case ranging in size from 1–10 MB. The file should be retrievable whenever the case is accessed in the application.

How can the company store the file in the MOST cost-effective manner?

a) Store the file in Amazon Document DB and the document ID as an attribute in the DynamoDB table.
b) Store the file in Amazon S3 and the object key as an attribute in the DynamoDB table.
c) Split the file into smaller parts and store the parts as multiple items in a separate DynamoDB table.
d) Store the file as an attribute in the DynamoDB table using Base64 encoding.

Answer: b

Question: 9

An online retail company wants to perform analytics on data in large Amazon S3 objects using Amazon EMR.

An Apache Spark job repeatedly queries the same data to populate an analytics dashboard. The analytics team wants to minimize the time to load the data and create the dashboard.

Which approaches could improve the performance?

(Select TWO.)

a) Copy the source data into Amazon Redshift and rewrite the Apache Spark code to create analytical reports by querying Amazon Redshift.
b) Copy the source data from Amazon S3 into Hadoop Distributed File System (HDFS) using s3distcp.
c) Load the data into Spark Data Frames.
d) Stream the data into Amazon Kinesis and use the Kinesis Connector Library (KCL) in multiple Spark jobs to perform analytical jobs.
e) Use Amazon S3 Select to retrieve the data necessary for the dashboards from the S3 objects.

Answer: c, e

A real estate company is receiving new property listing data from its agents through .csv files every day and storing these files in Amazon S3.

The data analytics team created an Amazon Quick Sight visualization report that uses a dataset imported from the S3 files. The data analytics team wants the visualization report to reflect the current data up to the previous day.

How can a data analyst meet these requirements?

a) Schedule an AWS Lambda function to drop and re-create the dataset daily.
b) Configure the visualization to query the data in Amazon S3 directly without loading the data into SPICE.
c) Schedule the dataset to refresh daily.
d) Close and open the Amazon Quick Sight visualization.

Answer: c

Study Guide to Crack AWS Data Analytics Specialty DAS-C01 Exam:

- Getting details of the DAS-C01 syllabus, is the first step of a study plan. This pdf is going to be of ultimate help. Completion of the syllabus is must to pass the DAS-C01 exam.

- Making a schedule is vital. A structured method of preparation leads to success. A candidate must plan his schedule and follow it rigorously to attain success.

- Joining the AWS provided training for DAS-C01 exam could be of much help. If there is specific training for the exam, you can discover it from the link above.

- Read from the DAS-C01 sample questions to gain your idea about the actual exam questions. In this PDF useful sample questions are provided to make your exam preparation easy.

- Practicing on DAS-C01 practice tests is must. Continuous practice will make you an expert in all syllabus areas.

Reliable Online Practice Test for DAS-C01 Certification

Make VMExam.com your best friend during your AWS Certified Data Analytics - Specialty exam preparation. We provide authentic practice tests for the DAS-C01 exam. Experts design these online practice tests, so we can offer you an exclusive experience of taking the actual DAS-C01 exam. We guarantee you 100% success in your first exam attempt if you continue practicing regularly. Don't bother if you don't get 100% marks in initial practice exam attempts. Just utilize the result section to know your strengths and weaknesses and prepare according to that until you get 100% with our practice tests. Our evaluation makes you confident, and you can score high in the DAS-C01 exam.

Start Online practice of DAS-C01 Exam by visiting URL

https://www.vmexam.com/aws/das-c01-aws-certified-data-analytics-specialty