DAS-C01^Q&As

AWS Certified Data Analytics - Specialty (DAS-C01)

Pass Amazon DAS-C01 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.pass2lead.com/das-c01.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Amazon Official Exam Center

⚙ **Instant Download** After Purchase

⚙ **100% Money Back** Guarantee

⚙ **365 Days** Free Update

⚙ **800,000+** Satisfied Customers

QUESTION 1

A data analyst notices the following error message while loading data to an Amazon Redshift cluster:

"The bucket you are attempting to access must be addressed using the specified endpoint."

What should the data analyst do to resolve this issue?

A. Specify the correct AWS Region for the Amazon S3 bucket by using the REGION option with the COPY command.

B. Change the Amazon S3 object\\'s ACL to grant the S3 bucket owner full control of the object.

C. Launch the Redshift cluster in a VPC.

D. Configure the timeout settings according to the operating system used to connect to the Redshift cluster.

Correct Answer: D

QUESTION 2

A real estate company has a mission-critical application using Apache HBase in Amazon EMR. Amazon EMR is configured with a single master node. The company has over 5 TB of data stored on an Hadoop Distributed File System (HDFS). The company wants a cost-effective solution to make its HBase data highly available.

Which architectural pattern meets company\\'s requirements?

A. Use Spot Instances for core and task nodes and a Reserved Instance for the EMR master node. Configure the EMR cluster with multiple master nodes. Schedule automated snapshots using Amazon EventBridge.

B. Store the data on an EMR File System (EMRFS) instead of HDFS. Enable EMRFS consistent view. Create an EMR HBase cluster with multiple master nodes. Point the HBase root directory to an Amazon S3 bucket.

C. Store the data on an EMR File System (EMRFS) instead of HDFS and enable EMRFS consistent view. Run two separate EMR clusters in two different Availability Zones. Point both clusters to the same HBase root directory in the same Amazon S3 bucket.

D. Store the data on an EMR File System (EMRFS) instead of HDFS and enable EMRFS consistent view. Create a primary EMR HBase cluster with multiple master nodes. Create a secondary EMR HBase read-replica cluster in a separate Availability Zone. Point both clusters to the same HBase root directory in the same Amazon S3 bucket.

Correct Answer: C

Reference: https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hbase-s3.html

QUESTION 3

A company\\'s system operators and security engineers need to analyze activities within specific date ranges of AWS CloudTrail logs. All log files are stored in an Amazon S3 bucket, and the size of the logs is more than 5 TB. The solution must be cost-effective and maximize query performance.

![Pass2Lead](https://Pass2Lead.com)
Which solution meets these requirements?

A. Copy the logs to a new S3 bucket with a prefix structure of . Use the date column as a partition key. Create a table on Amazon Athena based on the objects in the new bucket. Automatically add metadata partitions by using the MSCK REPAIR TABLE command in Athena. Use Athena to query the table and partitions.

B. Create a table on Amazon Athena. Manually add metadata partitions by using the ALTER TABLE ADD PARTITION statement, and use multiple columns for the partition key. Use Athena to query the table and partitions.

C. Launch an Amazon EMR cluster and use Amazon S3 as a data store for Apache HBase. Load the logs from the S3 bucket to an HBase table on Amazon EMR. Use Amazon Athena to query the table and partitions.

D. Create an AWS Glue job to copy the logs from the S3 source bucket to a new S3 bucket and create a table using Apache Parquet file format, Snappy as compression codec, and partition by date. Use Amazon Athena to query the table and partitions.

Correct Answer: D

QUESTION 4

An online retail company is using Amazon Redshift to run queries and perform analytics on customer shopping behavior. When multiple queries are running on the cluster, runtime for small queries increases significantly. The company\\'s data analytics team to decrease the runtime of these small queries by prioritizing them ahead of large queries.

Which solution will meet these requirements?

A. Use Amazon Redshift Spectrum for small queries

B. Increase the concurrency limit in workload management (WLM)

C. Configure short query acceleration in workload management (WLM)

D. Add a dedicated compute node for small queries

Correct Answer: C

Short query acceleration (SQA) prioritizes selected short-running queries ahead of longer-running queries. SQA executes short-running queries in a dedicated space, so that SQA queries aren\\'t forced to wait in queues behind longer queries.

Reference: https://docs.aws.amazon.com/redshift/latest/mgmt/workload-mgmt-config.html

---

QUESTION 5

A data analyst at a fast-growing retail company needs to store data coming in from several dozen marketing campaigns. Each source will write its output to a CSV file that is stored in Amazon S3. The data will later be analyzed by individual campaign managers using Amazon Athena to roughly track the number of daily unique visits to their specific campaign websites over time. The company wants to minimize the cost of data analysis.

Which combination of actions would lead to the MOST efficient one-time analysis of the data? (Choose two.)

A. Use an AWS Glue job to convert all files to Apache ORC format. Use the COUNT(DISTINCT column) function to

![Pass2Lead](https://Pass2Lead.com)
obtain a count of unique visitors.

B. Create one S3 bucket for all the data. Partition the data by date.

C. Create a separate S3 bucket for each campaign. Partition the data by date.

D. Create a separate S3 bucket for each month. Partition the data by campaign.

E. Use an AWS Glue job to convert all files to Apache Parquet format. Use the approx_distinct() function to obtain a count of unique visitors.

Correct Answer: AE

QUESTION 6

A reseller that has thousands of AWS accounts receives AWS Cost and Usage Reports in an Amazon S3 bucket. The reports are delivered to the S3 bucket in the following format: //yyyymmdd-yyyymmdd/.parquet An AWS Glue crawler crawls the S3 bucket and populates an AWS Glue Data Catalog with a table. Business analysts use Amazon Athena to query the table and create monthly summary reports for the AWS accounts. The business analysts

are experiencing slow queries because of the accumulation of reports from the last 5 years. The business analysts want the operations team to make changes to improve query performance. Which action should the operations team take to meet these requirements?

A. Change the file format to .csv.zip

B. Partition the data by date and account ID

C. Partition the data by month and account ID

D. Partition the data by account ID, year, and month

Correct Answer: A

Reference: https://docs.aws.amazon.com/cur/latest/userguide/access-cur-s3.html

QUESTION 7

A data architect at a large financial institution is building a data platform on AWS with the intent of implementing fraud detection by identifying duplicate customer accounts. The fraud detection algorithm will run in a batch mode to identify when a newly created account matches one for a user that was previously fraudulent.

Which approach MOST cost-effectively meets these requirements?

A. Build a custom deduplication script by using Apache Spark on an Amazon EMR cluster. Use PySpark to compare the data frames that represent the new customers and the fraudulent customer set to identify matches.

B. Load the data to an Amazon Redshift cluster. Use custom SQL to build deduplication logic.

C. Load the data to Amazon S3 to form the basis of a data lake. Use Amazon Athena to build a deduplication script.

D. Load the data to Amazon S3. Use the AWS Glue FindMatches transform to implement deduplication logic.

Correct Answer: D

QUESTION 8

A company wants to research user turnover by analyzing the past 3 months of user activities. With millions of users, 1.5 TB of uncompressed data is generated each day. A 30-node Amazon Redshift cluster with 2.56 TB of solid state drive (SSD) storage for each node is required to meet the query performance goals.

The company wants to run an additional analysis on a year\\'s worth of historical data to examine trends indicating which features are most popular. This analysis will be done once a week.

What is the MOST cost-effective solution?

A. Increase the size of the Amazon Redshift cluster to 120 nodes so it has enough storage capacity to hold 1 year of data. Then use Amazon Redshift for the additional analysis.

B. Keep the data from the last 90 days in Amazon Redshift. Move data older than 90 days to Amazon S3 and store it in Apache Parquet format partitioned by date. Then use Amazon Redshift Spectrum for the additional analysis.

C. Keep the data from the last 90 days in Amazon Redshift. Move data older than 90 days to Amazon S3 and store it in Apache Parquet format partitioned by date. Then provision a persistent Amazon EMR cluster and use Apache Presto for the additional analysis.

D. Resize the cluster node type to the dense storage node type (DS2) for an additional 16 TB storage capacity on each individual node in the Amazon Redshift cluster. Then use Amazon Redshift for the additional analysis.

Correct Answer: B

QUESTION 9

A company has a marketing department and a finance department. The departments are storing data in Amazon S3 in their own AWS accounts in AWS Organizations. Both departments use AWS Lake Formation to catalog and secure their

data. The departments have some databases and tables that share common names.

The marketing department needs to securely access some tables from the finance department.

Which two steps are required for this process? (Choose two.)

A. The finance department grants Lake Formation permissions for the tables to the external account for the marketing department.

B. The finance department creates cross-account IAM permissions to the table for the marketing department role.

C. The marketing department creates an IAM role that has permissions to the Lake Formation tables.

Correct Answer: AB

Granting Lake Formation Permissions Creating an IAM role (AWS CLI)

Reference: https://docs.aws.amazon.com/lake-formation/latest/dg/lake-formation-permissions.html

---

https://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_create_for-user.html

QUESTION 10

A media company is using Amazon QuickSight dashboards to visualize its national sales data. The dashboard is using a dataset with these fields: ID, date, time_zone, city, state, country, longitude, latitude, sales_volume, and number_of_items.

To modify ongoing campaigns, the company wants an interactive and intuitive visualization of which states across the country recorded a significantly lower sales volume compared to the national average.

Which addition to the company\\'s QuickSight dashboard will meet this requirement?

A. A geospatial color-coded chart of sales volume data across the country.

B. A pivot table of sales volume data summed up at the state level.

C. A drill-down layer for state-level sales volume data.

D. A drill through to other dashboards containing state-level sales volume data.

Correct Answer: B

Reference: https://docs.aws.amazon.com/quicksight/latest/user/pivot-table.html

QUESTION 11

A data analytics specialist is building a solution to securely collect and store data from multiple applications for analytics. The solution must store the data across multiple Availability Zones less than 2 minutes after collecting the data. The solution also must provide a secure way to authenticate and authorize the source application users.

Which solution will meet these requirements with the LEAST operational overhead?

A. Authenticate by using an Amazon Cognito user pool that is authorized to write to an Amazon API Gateway REST API. Configure the API as a proxy for an Amazon Kinesis Data Firehose delivery stream that has an Amazon S3 destination.

B. Write data to an Amazon S3 bucket by using the AWS SDK. Configure a bucket policy that limits writes to specific IAM roles.

C. Create an IAM access key that is authorized to write to an Amazon Kinesis Data Firehose delivery stream that has an Amazon S3 destination. Embed the access key into the applications.

D. Call a REST-based service such as Amazon API Gateway that uses a custom authentication service to store data on a Kubernetes cluster.

Correct Answer: A

QUESTION 12

A company uses an Amazon QuickSight business intelligence (BI) dashboard to make decisions. A data warehouse that is hosted on an Amazon Redshift cluster is the primary data source for the dashboard.

The user experience has degraded because some BI queries are complex and include multiple table joins. A data analytics specialist needs to reduce query response times to improve the user experience and meet SLAs.

Which solution will meet these requirements with the LEAST operational overhead?

A. Create automated materialized views. Turn on the query rewrite feature in Amazon Redshift.

B. Create views in Amazon Redshift. Change the QuickSight code to select from the views.

C. Create a new consolidated table by using CREATE TABLE AS (CTAS). Change the QuickSight code to select from the consolidated table.

D. Create an external table. Turn on the query rewrite feature in Amazon Redshift.

Correct Answer: A

QUESTION 13

A mortgage company has a microservice for accepting payments. This microservice uses the Amazon DynamoDB encryption client with AWS KMS managed keys to encrypt the sensitive data before writing the data to DynamoDB. The finance team should be able to load this data into Amazon Redshift and aggregate the values within the sensitive fields. The Amazon Redshift cluster is shared with other data analysts from different business units.

Which steps should a data analyst take to accomplish this task efficiently and securely?

A. Create an AWS Lambda function to process the DynamoDB stream. Decrypt the sensitive data using the same KMS key. Save the output to a restricted S3 bucket for the finance team. Create a finance table in Amazon Redshift that is accessible to the finance team only. Use the COPY command to load the data from Amazon S3 to the finance table.

B. Create an AWS Lambda function to process the DynamoDB stream. Save the output to a restricted S3 bucket for the finance team. Create a finance table in Amazon Redshift that is accessible to the finance team only. Use the COPY command with the IAM role that has access to the KMS key to load the data from S3 to the finance table.

C. Create an Amazon EMR cluster with an EMR_EC2_DefaultRole role that has access to the KMS key. Create Apache Hive tables that reference the data stored in DynamoDB and the finance table in Amazon Redshift. In Hive, select the data from DynamoDB and then insert the output to the finance table in Amazon Redshift.

D. Create an Amazon EMR cluster. Create Apache Hive tables that reference the data stored in DynamoDB. Insert the output to the restricted Amazon S3 bucket for the finance team. Use the COPY command with the IAM role that has access to the KMS key to load the data from Amazon S3 to the finance table in Amazon Redshift.

Correct Answer: B

QUESTION 14

A data analytics specialist has a 50 GB data file in .csv format and wants to perform a data transformation task. The data analytics specialist is using the Amazon Athena CREATE TABLE AS SELECT (CTAS) statement to perform the transformation. The resulting output will be used to query the data from Amazon Redshift Spectrum.

Which CTAS statement should the data analytics specialist use to provide the MOST efficient performance?

![Pass2Lead logo](https://Pass2Lead.com)
A. CREATE TABLE new_Table

WITH (

 format = \\'TEXTFILE\\',

 orc_compression = \\'SNAPPY\\')

AS SELECT *

FROM old_table;

B. CREATE TABLE new_Table

WITH (

 format = \\'TEXTFILE\\',

)

AS SELECT *

FROM old_table;

C. CREATE TABLE new_Table

WITH (

 format = \\'PARQUET\\',

 parquet_compression = \\'SNAPPY\\')

AS SELECT *

FROM old_table;

D. CREATE TABLE new_Table

WITH (

 format = JSON,

)

AS SELECT *

FROM old_table;

Correct Answer: C

Reference: https://docs.aws.amazon.com/athena/latest/ug/performance-tuning.html#performance-tuning-file-formats

---

QUESTION 15

An online gaming company is using an Amazon Kinesis Data Analytics SQL application with a Kinesis data stream as its

![Pass2Lead](https://Pass2Lead.com)
source. The source sends three non-null fields to the application: player_id, score, and us_5_digit_zip_code.

A data analyst has a .csv mapping file that maps a small number of us_5_digit_zip_code values to a territory code. The data analyst needs to include the territory code, if one exists, as an additional output of the Kinesis Data Analytics application.

How should the data analyst meet this requirement while minimizing costs?

A. Store the contents of the mapping file in an Amazon DynamoDB table. Preprocess the records as they arrive in the Kinesis Data Analytics application with an AWS Lambda function that fetches the mapping and supplements each record to include the territory code, if one exists. Change the SQL query in the application to include the new field in the SELECT statement.

B. Store the mapping file in an Amazon S3 bucket and configure the reference data column headers for the .csv file in the Kinesis Data Analytics application. Change the SQL query in the application to include a join to the file\\'s S3 Amazon Resource Name (ARN), and add the territory code field to the SELECT columns.

C. Store the mapping file in an Amazon S3 bucket and configure it as a reference data source for the Kinesis Data Analytics application. Change the SQL query in the application to include a join to the reference table and add the territory code field to the SELECT columns.

D. Store the contents of the mapping file in an Amazon DynamoDB table. Change the Kinesis Data Analytics application to send its output to an AWS Lambda function that fetches the mapping and supplements each record to include the territory code, if one exists. Forward the record from the Lambda function to the original application destination.

Correct Answer: C

[DAS-C01 PDF Dumps](https://www.pass2lead.com/das-c01.html)   [DAS-C01 VCE Dumps](https://www.pass2lead.com/das-c01.html)   [DAS-C01 Study Guide](https://www.pass2lead.com/das-c01.html)