

AWS-Certified-Data-Analytics-Specialty Dumps

AWS Certified Data Analytics - Specialty

<https://www.certleader.com/AWS-Certified-Data-Analytics-Specialty-dumps.html>



NEW QUESTION 1

A company is sending historical datasets to Amazon S3 for storage. A data engineer at the company wants to make these datasets available for analysis using Amazon Athena. The engineer also wants to encrypt the Athena query results in an S3 results location by using AWS solutions for encryption. The requirements for encrypting the query results are as follows:

Use custom keys for encryption of the primary dataset query results. Use generic encryption for all other query results.

Provide an audit trail for the primary dataset queries that shows when the keys were used and by whom.

Which solution meets these requirements?

- A. Use server-side encryption with S3 managed encryption keys (SSE-S3) for the primary dataset
- B. Use SSE-S3 for the other datasets.
- C. Use server-side encryption with customer-provided encryption keys (SSE-C) for the primary dataset. Use server-side encryption with S3 managed encryption keys (SSE-S3) for the other datasets.
- D. Use server-side encryption with AWS KMS managed customer master keys (SSE-KMS CMKs) for the primary dataset
- E. Use server-side encryption with S3 managed encryption keys (SSE-S3) for the other datasets.
- F. Use client-side encryption with AWS Key Management Service (AWS KMS) customer managed keys for the primary dataset
- G. Use S3 client-side encryption with client-side keys for the other datasets.

Answer: A

NEW QUESTION 2

A manufacturing company uses Amazon S3 to store its data. The company wants to use AWS Lake Formation to provide granular-level security on those data assets. The data is in Apache Parquet format. The company has set a deadline for a consultant to build a data lake.

How should the consultant create the MOST cost-effective solution that meets these requirements?

- A. Run Lake Formation blueprints to move the data to Lake Formation
- B. Once Lake Formation has the data, apply permissions on Lake Formation.
- C. To create the data catalog, run an AWS Glue crawler on the existing Parquet data
- D. Register the Amazon S3 path and then apply permissions through Lake Formation to provide granular-level security.
- E. Install Apache Ranger on an Amazon EC2 instance and integrate with Amazon EM
- F. Using Ranger policies, create role-based access control for the existing data assets in Amazon S3.
- G. Create multiple IAM roles for different users and group
- H. Assign IAM roles to different data assets in Amazon S3 to create table-based and column-based access controls.

Answer: A

Explanation:

<https://aws.amazon.com/blogs/big-data/building-securing-and-managing-data-lakes-with-aws-lake-formation/>

NEW QUESTION 3

A company's data analyst needs to ensure that queries executed in Amazon Athena cannot scan more than a prescribed amount of data for cost control purposes. Queries that exceed the prescribed threshold must be canceled immediately.

What should the data analyst do to achieve this?

- A. Configure Athena to invoke an AWS Lambda function that terminates queries when the prescribed threshold is crossed.
- B. For each workgroup, set the control limit for each query to the prescribed threshold.
- C. Enforce the prescribed threshold on all Amazon S3 bucket policies
- D. For each workgroup, set the workgroup-wide data usage control limit to the prescribed threshold.

Answer: B

Explanation:

<https://docs.aws.amazon.com/athena/latest/ug/manage-queries-control-costs-with-workgroups.html>

NEW QUESTION 4

Once a month, a company receives a 100 MB .csv file compressed with gzip. The file contains 50,000 property listing records and is stored in Amazon S3 Glacier. The company needs its data analyst to query a subset of the data for a specific vendor.

What is the most cost-effective solution?

- A. Load the data into Amazon S3 and query it with Amazon S3 Select.
- B. Query the data from Amazon S3 Glacier directly with Amazon Glacier Select.
- C. Load the data to Amazon S3 and query it with Amazon Athena.
- D. Load the data to Amazon S3 and query it with Amazon Redshift Spectrum.

Answer: A

NEW QUESTION 5

A regional energy company collects voltage data from sensors attached to buildings. To address any known dangerous conditions, the company wants to be alerted when a sequence of two voltage drops is detected within 10 minutes of a voltage spike at the same building. It is important to ensure that all messages are delivered as quickly as possible. The system must be fully managed and highly available. The company also needs a solution that will automatically scale up as it covers additional cities with this monitoring feature. The alerting system is subscribed to an Amazon SNS topic for remediation.

Which solution meets these requirements?

- A. Create an Amazon Managed Streaming for Kafka cluster to ingest the data, and use an Apache Spark Streaming with Apache Kafka consumer API in an automatically scaled Amazon EMR cluster to process the incoming data
- B. Use the Spark Streaming application to detect the known event sequence and send the SNS message.
- C. Create a REST-based web service using Amazon API Gateway in front of an AWS Lambda function. Create an Amazon RDS for PostgreSQL database with sufficient Provisioned IOPS (PIOPS). In the Lambda function, store incoming events in the RDS database and query the latest data to detect the known event

sequence and send the SNS message.

- D. Create an Amazon Kinesis Data Firehose delivery stream to capture the incoming sensor data
- E. Use an AWS Lambda transformation function to detect the known event sequence and send the SNS message.
- F. Create an Amazon Kinesis data stream to capture the incoming sensor data and create another stream for alert message
- G. Set up AWS Application Auto Scaling on bot
- H. Create a Kinesis Data Analytics for Java application to detect the known event sequence, and add a message to the message stream
- I. Configure an AWS Lambda function to poll the message stream and publish to the SNS topic.

Answer: D

NEW QUESTION 6

An advertising company has a data lake that is built on Amazon S3. The company uses AWS Glue Data Catalog to maintain the metadata. The data lake is several years old and its overall size has increased exponentially as additional data sources and metadata are stored in the data lake. The data lake administrator wants to implement a mechanism to simplify permissions management between Amazon S3 and the Data Catalog to keep them in sync. Which solution will simplify permissions management with minimal development effort?

- A. Set AWS Identity and Access Management (IAM) permissions for AWS Glue
- B. Use AWS Lake Formation permissions
- C. Manage AWS Glue and S3 permissions by using bucket policies
- D. Use Amazon Cognito user pools.

Answer: B

NEW QUESTION 7

An online retail company is migrating its reporting system to AWS. The company's legacy system runs data processing on online transactions using a complex series of nested Apache Hive queries. Transactional data is exported from the online system to the reporting system several times a day. Schemas in the files are stable between updates.

A data analyst wants to quickly migrate the data processing to AWS, so any code changes should be minimized. To keep storage costs low, the data analyst decides to store the data in Amazon S3. It is vital that the data from the reports and associated analytics is completely up to date based on the data in Amazon S3. Which solution meets these requirements?

- A. Create an AWS Glue Data Catalog to manage the Hive metadata
- B. Create an AWS Glue crawler over Amazon S3 that runs when data is refreshed to ensure that data changes are updated
- C. Create an Amazon EMR cluster and use the metadata in the AWS Glue Data Catalog to run Hive processing queries in Amazon EMR.
- D. Create an AWS Glue Data Catalog to manage the Hive metadata
- E. Create an Amazon EMR cluster with consistent view enabled
- F. Run emrfs sync before each analytics step to ensure data changes are updated
- G. Create an EMR cluster and use the metadata in the AWS Glue Data Catalog to run Hive processing queries in Amazon EMR.
- H. Create an Amazon Athena table with CREATE TABLE AS SELECT (CTAS) to ensure data is refreshed from underlying queries against the raw dataset
- I. Create an AWS Glue Data Catalog to manage the Hive metadata over the CTAS table
- J. Create an Amazon EMR cluster and use the metadata in the AWS Glue Data Catalog to run Hive processing queries in Amazon EMR.
- K. Use an S3 Select query to ensure that the data is properly updated
- L. Create an AWS Glue Data Catalog to manage the Hive metadata over the S3 Select table
- M. Create an Amazon EMR cluster and use the metadata in the AWS Glue Data Catalog to run Hive processing queries in Amazon EMR.

Answer: A

NEW QUESTION 8

A large university has adopted a strategic goal of increasing diversity among enrolled students. The data analytics team is creating a dashboard with data visualizations to enable stakeholders to view historical trends. All access must be authenticated using Microsoft Active Directory. All data in transit and at rest must be encrypted.

Which solution meets these requirements?

- A. Amazon QuickSight Standard edition configured to perform identity federation using SAML 2.0 and the default encryption settings.
- B. Amazon QuickSight Enterprise edition configured to perform identity federation using SAML 2.0 and the default encryption settings.
- C. Amazon QuickSight Standard edition using AD Connector to authenticate using Active Directory. Configure Amazon QuickSight to use customer-provided keys imported into AWS KMS.
- D. Amazon QuickSight Enterprise edition using AD Connector to authenticate using Active Directory. Configure Amazon QuickSight to use customer-provided keys imported into AWS KMS.

Answer: D

NEW QUESTION 9

An analytics software as a service (SaaS) provider wants to offer its customers business intelligence (BI) reporting capabilities that are self-service. The provider is using Amazon QuickSight to build these reports. The data for the reports resides in a multi-tenant database, but each customer should only be able to access their own data.

The provider wants to give customers two user role options:

- Read-only users for individuals who only need to view dashboards
 - Power users for individuals who are allowed to create and share new dashboards with other users
- Which QuickSight feature allows the provider to meet these requirements?

- A. Embedded dashboards
- B. Table calculations
- C. Isolated namespaces
- D. SPICE

Answer: A

NEW QUESTION 10

A marketing company is using Amazon EMR clusters for its workloads. The company manually installs third-party libraries on the clusters by logging in to the master nodes. A data analyst needs to create an automated solution to replace the manual process. Which options can fulfill these requirements? (Choose two.)

- A. Place the required installation scripts in Amazon S3 and execute them using custom bootstrap actions.
- B. Place the required installation scripts in Amazon S3 and execute them through Apache Spark in Amazon EMR.
- C. Install the required third-party libraries in the existing EMR master node.
- D. Create an AMI out of that master node and use that custom AMI to re-create the EMR cluster.
- E. Use an Amazon DynamoDB table to store the list of required application.
- F. Trigger an AWS Lambda function with DynamoDB Streams to install the software.
- G. Launch an Amazon EC2 instance with Amazon Linux and install the required third-party libraries on the instance.
- H. Create an AMI and use that AMI to create the EMR cluster.

Answer: AE

Explanation:

[https://aws.amazon.com/about-aws/whats-new/2017/07/amazon-emr-now-supports-launching-clusters-with-cust](https://aws.amazon.com/about-aws/whats-new/2017/07/amazon-emr-now-supports-launching-clusters-with-custom-bootstrap-actions/)
https://docs.aws.amazon.com/de_de/emr/latest/ManagementGuide/emr-plan-bootstrap.html

NEW QUESTION 10

A company uses Amazon Redshift as its data warehouse. A new table includes some columns that contain sensitive data and some columns that contain non-sensitive data. The data in the table eventually will be referenced by several existing queries that run many times each day. A data analytics specialist must ensure that only members of the company's auditing team can read the columns that contain sensitive data. All other users must have read-only access to the columns that contain non-sensitive data. Which solution will meet these requirements with the LEAST operational overhead?

- A. Grant the auditing team permission to read from the table.
- B. Load the columns that contain non-sensitive data into a second table.
- C. Grant the appropriate users read-only permissions to the second table.
- D. Grant all users read-only permissions to the columns that contain non-sensitive data. Use the GRANT SELECT command to allow the auditing team to access the columns that contain sensitive data.
- E. Grant all users read-only permissions to the columns that contain non-sensitive data. Attach an IAM policy to the auditing team with an explicit Allow action that grants access to the columns that contain sensitive data.
- F. Grant the auditing team permission to read from the table. Create a view of the table that includes the columns that contain non-sensitive data. Grant the appropriate users read-only permissions to that view.

Answer: B

Explanation:

[https://aws.amazon.com/jp/about-aws/whats-new/2020/03/announcing-column-level-access-control-for-amazon](https://aws.amazon.com/jp/about-aws/whats-new/2020/03/announcing-column-level-access-control-for-amazon-redshift/)

NEW QUESTION 15

A company uses the Amazon Kinesis SDK to write data to Kinesis Data Streams. Compliance requirements state that the data must be encrypted at rest using a key that can be rotated. The company wants to meet this encryption requirement with minimal coding effort. How can these requirements be met?

- A. Create a customer master key (CMK) in AWS KMS.
- B. Assign the CMK an alias.
- C. Use the AWS Encryption SDK, providing it with the key alias to encrypt and decrypt the data.
- D. Create a customer master key (CMK) in AWS KMS.
- E. Assign the CMK an alias.
- F. Enable server-side encryption on the Kinesis data stream using the CMK alias as the KMS master key.
- G. Create a customer master key (CMK) in AWS KMS.
- H. Create an AWS Lambda function to encrypt and decrypt the data.
- I. Set the KMS key ID in the function's environment variables.
- J. Enable server-side encryption on the Kinesis data stream using the default KMS key for Kinesis Data Streams.

Answer: B

NEW QUESTION 20

A marketing company is storing its campaign response data in Amazon S3. A consistent set of sources has generated the data for each campaign. The data is saved into Amazon S3 as .csv files. A business analyst will use Amazon Athena to analyze each campaign's data. The company needs the cost of ongoing data analysis with Athena to be minimized. Which combination of actions should a data analytics specialist take to meet these requirements? (Choose two.)

- A. Convert the .csv files to Apache Parquet.
- B. Convert the .csv files to Apache Avro.
- C. Partition the data by campaign.
- D. Partition the data by source.
- E. Compress the .csv files.

Answer: AC

Explanation:

<https://aws.amazon.com/blogs/big-data/top-10-performance-tuning-tips-for-amazon-athena/>

NEW QUESTION 23

A smart home automation company must efficiently ingest and process messages from various connected devices and sensors. The majority of these messages

are comprised of a large number of small files. These messages are ingested using Amazon Kinesis Data Streams and sent to Amazon S3 using a Kinesis data stream consumer application. The Amazon S3 message data is then passed through a processing pipeline built on Amazon EMR running scheduled PySpark jobs. The data platform team manages data processing and is concerned about the efficiency and cost of downstream data processing. They want to continue to use PySpark.

Which solution improves the efficiency of the data processing jobs and is well architected?

- A. Send the sensor and devices data directly to a Kinesis Data Firehose delivery stream to send the data to Amazon S3 with Apache Parquet record format conversion enable
- B. Use Amazon EMR running PySpark to process the data in Amazon S3.
- C. Set up an AWS Lambda function with a Python runtime environmen
- D. Process individual Kinesis data stream messages from the connected devices and sensors using Lambda.
- E. Launch an Amazon Redshift cluste
- F. Copy the collected data from Amazon S3 to Amazon Redshift and move the data processing jobs from Amazon EMR to Amazon Redshift.
- G. Set up AWS Glue Python jobs to merge the small data files in Amazon S3 into larger files and transform them to Apache Parquet forma
- H. Migrate the downstream PySpark jobs from Amazon EMR to AWS Glue.

Answer: D

Explanation:

<https://aws.amazon.com/it/about-aws/whats-new/2020/04/aws-glue-now-supports-serverless-streaming-etl/>

NEW QUESTION 28

A company is streaming its high-volume billing data (100 MBps) to Amazon Kinesis Data Streams. A data analyst partitioned the data on account_id to ensure that all records belonging to an account go to the same Kinesis shard and order is maintained. While building a custom consumer using the Kinesis Java SDK, the data analyst notices that, sometimes, the messages arrive out of order for account_id. Upon further investigation, the data analyst discovers the messages that are out of order seem to be arriving from different shards for the same account_id and are seen when a stream resize runs.

What is an explanation for this behavior and what is the solution?

- A. There are multiple shards in a stream and order needs to be maintained in the shar
- B. The data analyst needs to make sure there is only a single shard in the stream and no stream resize runs.
- C. The hash key generation process for the records is not working correctl
- D. The data analyst should generate an explicit hash key on the producer side so the records are directed to the appropriate shard accurately.
- E. The records are not being received by Kinesis Data Streams in orde
- F. The producer should use the PutRecords API call instead of the PutRecord API call with the SequenceNumberForOrdering parameter.
- G. The consumer is not processing the parent shard completely before processing the child shards after a stream resiz
- H. The data analyst should process the parent shard completely first before processing the child shards.

Answer: D

Explanation:

<https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html> the parent shards that remain after the reshard could still contain data that you haven't read yet that was added to the stream before the reshard. If you read data from the child shards before having read all data from the parent shards, you could read data for a particular hash key out of the order given by the data records' sequence numbers. Therefore, assuming that the order of the data is important, you should, after a reshard, always continue to read data from the parent shards until it is exhausted. Only then should you begin reading data from the child shards.

NEW QUESTION 33

A company uses Amazon Redshift as its data warehouse. A new table has columns that contain sensitive data. The data in the table will eventually be referenced by several existing queries that run many times a day.

A data analyst needs to load 100 billion rows of data into the new table. Before doing so, the data analyst must ensure that only members of the auditing group can read the columns containing sensitive data.

How can the data analyst meet these requirements with the lowest maintenance overhead?

- A. Load all the data into the new table and grant the auditing group permission to read from the tabl
- B. Load all the data except for the columns containing sensitive data into a second tabl
- C. Grant the appropriate users read-only permissions to the second table.
- D. Load all the data into the new table and grant the auditing group permission to read from the tabl
- E. Use the GRANT SQL command to allow read-only access to a subset of columns to the appropriate users.
- F. Load all the data into the new table and grant all users read-only permissions to non-sensitive columns.Attach an IAM policy to the auditing group with explicit ALLOW access to the sensitive data columns.
- G. Load all the data into the new table and grant the auditing group permission to read from the table.Create a view of the new table that contains all the columns, except for those considered sensitive, and grant the appropriate users read-only permissions to the table.

Answer: B

Explanation:

<https://aws.amazon.com/blogs/big-data/achieve-finer-grained-data-security-with-column-level-access-control-in>

NEW QUESTION 38

A company hosts an on-premises PostgreSQL database that contains historical data. An internal legacy application uses the database for read-only activities. The company's business team wants to move the data to a data lake in Amazon S3 as soon as possible and enrich the data for analytics.

The company has set up an AWS Direct Connect connection between its VPC and its on-premises network. A data analytics specialist must design a solution that achieves the business team's goals with the least operational overhead.

Which solution meets these requirements?

- A. Upload the data from the on-premises PostgreSQL database to Amazon S3 by using a customized batch upload proces
- B. Use the AWS Glue crawler to catalog the data in Amazon S3. Use an AWS Glue job to enrich and store the result in a separate S3 bucket in Apache Parquet forma
- C. Use Amazon Athena to query the data.
- D. Create an Amazon RDS for PostgreSQL database and use AWS Database Migration Service (AWS DMS) to migrate the data into Amazon RD

- E. Use AWS Data Pipeline to copy and enrich the data from the Amazon RDS for PostgreSQL table and move the data to Amazon S3. Use Amazon Athena to query the data.
- F. Configure an AWS Glue crawler to use a JDBC connection to catalog the data in the on-premises databases
- G. Use an AWS Glue job to enrich the data and save the result to Amazon S3 in Apache Parquet format
- H. Create an Amazon Redshift cluster and use Amazon Redshift Spectrum to query the data.
- I. Configure an AWS Glue crawler to use a JDBC connection to catalog the data in the on-premises databases
- J. Use an AWS Glue job to enrich the data and save the result to Amazon S3 in Apache Parquet format
- K. Use Amazon Athena to query the data.

Answer: B

NEW QUESTION 39

A US-based sneaker retail company launched its global website. All the transaction data is stored in Amazon RDS and curated historic transaction data is stored in Amazon Redshift in the us-east-1 Region. The business intelligence (BI) team wants to enhance the user experience by providing a dashboard for sneaker trends. The BI team decides to use Amazon QuickSight to render the website dashboards. During development, a team in Japan provisioned Amazon QuickSight in ap-northeast-1. The team is having difficulty connecting Amazon QuickSight from ap-northeast-1 to Amazon Redshift in us-east-1. Which solution will solve this issue and meet the requirements?

- A. In the Amazon Redshift console, choose to configure cross-Region snapshots and set the destination Region as ap-northeast-1. Restore the Amazon Redshift Cluster from the snapshot and connect to Amazon QuickSight launched in ap-northeast-1.
- B. Create a VPC endpoint from the Amazon QuickSight VPC to the Amazon Redshift VPC so Amazon QuickSight can access data from Amazon Redshift.
- C. Create an Amazon Redshift endpoint connection string with Region information in the string and use this connection string in Amazon QuickSight to connect to Amazon Redshift.
- D. Create a new security group for Amazon Redshift in us-east-1 with an inbound rule authorizing access from the appropriate IP address range for the Amazon QuickSight servers in ap-northeast-1.

Answer: B

NEW QUESTION 41

A large company has a central data lake to run analytics across different departments. Each department uses a separate AWS account and stores its data in an Amazon S3 bucket in that account. Each AWS account uses the AWS Glue Data Catalog as its data catalog. There are different data lake access requirements based on roles. Associate analysts should only have read access to their departmental data. Senior data analysts can have access in multiple departments including theirs, but for a subset of columns only. Which solution achieves these required access patterns to minimize costs and administrative tasks?

- A. Consolidate all AWS accounts into one account
- B. Create different S3 buckets for each department and move all the data from every account to the central data lake account
- C. Migrate the individual data catalogs into a central data catalog and apply fine-grained permissions to give to each user the required access to tables and databases in AWS Glue and Amazon S3.
- D. Keep the account structure and the individual AWS Glue catalogs on each account
- E. Add a central data lake account and use AWS Glue to catalog data from various accounts
- F. Configure cross-account access for AWS Glue crawlers to scan the data in each departmental S3 bucket to identify the schema and populate the catalog
- G. Add the senior data analysts into the central account and apply highly detailed access controls in the Data Catalog and Amazon S3.
- H. Set up an individual AWS account for the central data lake
- I. Use AWS Lake Formation to catalog the cross-account location
- J. On each individual S3 bucket, modify the bucket policy to grant S3 permissions to the Lake Formation service-linked role
- K. Use Lake Formation permissions to add fine-grained access controls to allow senior analysts to view specific tables and columns.
- L. Set up an individual AWS account for the central data lake and configure a central S3 bucket
- M. Use an AWS Lake Formation blueprint to move the data from the various buckets into the central S3 bucket
- N. On each individual bucket, modify the bucket policy to grant S3 permissions to the Lake Formation service-linked role
- O. Use Lake Formation permissions to add fine-grained access controls for both associate and senior analysts to view specific tables and columns.

Answer: C

Explanation:

Lake Formation provides secure and granular access to data through a new grant/revoke permissions model that augments AWS Identity and Access Management (IAM) policies. Analysts and data scientists can use the full portfolio of AWS analytics and machine learning services, such as Amazon Athena, to access the data. The configured Lake Formation security policies help ensure that users can access only the data that they are authorized to access. Source : <https://docs.aws.amazon.com/lake-formation/latest/dg/how-it-works.html>

NEW QUESTION 42

A central government organization is collecting events from various internal applications using Amazon Managed Streaming for Apache Kafka (Amazon MSK). The organization has configured a separate Kafka topic for each application to separate the data. For security reasons, the Kafka cluster has been configured to only allow TLS encrypted data and it encrypts the data at rest.

A recent application update showed that one of the applications was configured incorrectly, resulting in writing data to a Kafka topic that belongs to another application. This resulted in multiple errors in the analytics pipeline as data from different applications appeared on the same topic. After this incident, the organization wants to prevent applications from writing to a topic different than the one they should write to. Which solution meets these requirements with the least amount of effort?

- A. Create a different Amazon EC2 security group for each application
- B. Configure each security group to have access to a specific topic in the Amazon MSK cluster
- C. Attach the security group to each application based on the topic that the applications should read and write to.
- D. Install Kafka Connect on each application instance and configure each Kafka Connect instance to write to a specific topic only.
- E. Use Kafka ACLs and configure read and write permissions for each topic
- F. Use the distinguished name of the clients' TLS certificates as the principal of the ACL.
- G. Create a different Amazon EC2 security group for each application
- H. Create an Amazon MSK cluster and Kafka topic for each application
- I. Configure each security group to have access to the specific cluster.

Answer: B

NEW QUESTION 45

A company uses Amazon Elasticsearch Service (Amazon ES) to store and analyze its website clickstream data. The company ingests 1 TB of data daily using Amazon Kinesis Data Firehose and stores one day's worth of data in an Amazon ES cluster.

The company has very slow query performance on the Amazon ES index and occasionally sees errors from Kinesis Data Firehose when attempting to write to the index. The Amazon ES cluster has 10 nodes running a single index and 3 dedicated master nodes. Each data node has 1.5 TB of Amazon EBS storage attached and the cluster is configured with 1,000 shards. Occasionally, JVMMemoryPressure errors are found in the cluster logs.

Which solution will improve the performance of Amazon ES?

- A. Increase the memory of the Amazon ES master nodes.
- B. Decrease the number of Amazon ES data nodes.
- C. Decrease the number of Amazon ES shards for the index.
- D. Increase the number of Amazon ES shards for the index.

Answer: C

Explanation:

<https://aws.amazon.com/premiumsupport/knowledge-center/high-jvm-memory-pressure-elasticsearch/>

NEW QUESTION 48

An Amazon Redshift database contains sensitive user data. Logging is necessary to meet compliance requirements. The logs must contain database authentication attempts, connections, and disconnections. The logs must also contain each query run against the database and record which database user ran each query.

Which steps will create the required logs?

- A. Enable Amazon Redshift Enhanced VPC Routin
- B. Enable VPC Flow Logs to monitor traffic.
- C. Allow access to the Amazon Redshift database using AWS IAM onl
- D. Log access using AWS CloudTrail.
- E. Enable audit logging for Amazon Redshift using the AWS Management Console or the AWS CLI.
- F. Enable and download audit reports from AWS Artifact.

Answer: C

NEW QUESTION 49

An education provider's learning management system (LMS) is hosted in a 100 TB data lake that is built on Amazon S3. The provider's LMS supports hundreds of schools. The provider wants to build an advanced analytics reporting platform using Amazon Redshift to handle complex queries with optimal performance. System users will query the most recent 4 months of data 95% of the time while 5% of the queries will leverage data from the previous 12 months.

Which solution meets these requirements in the MOST cost-effective way?

- A. Store the most recent 4 months of data in the Amazon Redshift cluste
- B. Use Amazon Redshift Spectrum to query data in the data lak
- C. Use S3 lifecycle management rules to store data from the previous 12 months in Amazon S3 Glacier storage.
- D. Leverage DS2 nodes for the Amazon Redshift cluste
- E. Migrate all data from Amazon S3 to Amazon Redshif
- F. Decommission the data lake.
- G. Store the most recent 4 months of data in the Amazon Redshift cluste
- H. Use Amazon Redshift Spectrum to query data in the data lak
- I. Ensure the S3 Standard storage class is in use with objects in the data lake.
- J. Store the most recent 4 months of data in the Amazon Redshift cluste
- K. Use Amazon Redshift federated queries to join cluster data with the data lake to reduce cost
- L. Ensure the S3 Standard storage class is in use with objects in the data lake.

Answer: C

NEW QUESTION 50

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 dat
- B. Then use Amazon Redshift for the additional analysis.
- C. Keep the data from the last 90 days in Amazon Redshif
- D. Move data older than 90 days to Amazon S3 and store it in Apache Parquet format partitioned by dat
- E. Then use Amazon Redshift Spectrum for the additional analysis.
- F. Keep the data from the last 90 days in Amazon Redshif
- G. Move data older than 90 days to Amazon S3 and store it in Apache Parquet format partitioned by dat
- H. Then provision a persistent Amazon EMR cluster and use Apache Presto for the additional analysis.
- I. Resize the cluster node type to the dense storage node type (DS2) for an additional 16 TB storagecapacity on each individual node in the Amazon Redshift cluste
- J. Then use Amazon Redshift for the additional analysis.

Answer: B

NEW QUESTION 55

A company wants to improve user satisfaction for its smart home system by adding more features to its recommendation engine. Each sensor asynchronously pushes its nested JSON data into Amazon Kinesis Data Streams using the Kinesis Producer Library (KPL) in Java. Statistics from a set of failed sensors showed

that, when a sensor is malfunctioning, its recorded data is not always sent to the cloud.

The company needs a solution that offers near-real-time analytics on the data from the most updated sensors. Which solution enables the company to meet these requirements?

- A. Set the RecordMaxBufferedTime property of the KPL to "1" to disable the buffering on the sensor side. Use Kinesis Data Analytics to enrich the data based on a company-developed anomaly detection SQL scrip
- B. Push the enriched data to a fleet of Kinesis data streams and enable the data transformation feature to flatten the JSON fil
- C. Instantiate a dense storage Amazon Redshift cluster and use it as the destination for the Kinesis Data Firehose delivery stream.
- D. Update the sensors code to use the PutRecord/PutRecords call from the Kinesis Data Streams API with the AWS SDK for Jav
- E. Use Kinesis Data Analytics to enrich the data based on a company-developed anomaly detection SQL scrip
- F. Direct the output of KDA application to a Kinesis Data Firehose delivery stream, enable the data transformation feature to flatten the JSON file, and set the Kinesis Data Firehose destination to an Amazon Elasticsearch Service cluster.
- G. Set the RecordMaxBufferedTime property of the KPL to "0" to disable the buffering on the sensor side. Connect for each stream a dedicated Kinesis Data Firehose delivery stream and enable the data transformation feature to flatten the JSON file before sending it to an Amazon S3 bucke
- H. Load the S3 data into an Amazon Redshift cluster.
- I. Update the sensors code to use the PutRecord/PutRecords call from the Kinesis Data Streams API with the AWS SDK for Jav
- J. Use AWS Glue to fetch and process data from the stream using the Kinesis Client Library (KCL). Instantiate an Amazon Elasticsearch Service cluster and use AWS Lambda to directly push data into it.

Answer: B

Explanation:

<https://docs.aws.amazon.com/streams/latest/dev/developing-producers-with-kpl.html>

The KPL can incur an additional processing delay of up to RecordMaxBufferedTime within the library (user-configurable). Larger values of RecordMaxBufferedTime results in higher packing efficiencies and better performance. Applications that cannot tolerate this additional delay may need to use the AWS SDK directly.

NEW QUESTION 58

A company wants to use an automatic machine learning (ML) Random Cut Forest (RCF) algorithm to visualize complex real-world scenarios, such as detecting seasonality and trends, excluding outliers, and imputing missing values.

The team working on this project is non-technical and is looking for an out-of-the-box solution that will require the LEAST amount of management overhead.

Which solution will meet these requirements?

- A. Use an AWS Glue ML transform to create a forecast and then use Amazon QuickSight to visualize the data.
- B. Use Amazon QuickSight to visualize the data and then use ML-powered forecasting to forecast the key business metrics.
- C. Use a pre-build ML AMI from the AWS Marketplace to create forecasts and then use Amazon QuickSight to visualize the data.
- D. Use calculated fields to create a new forecast and then use Amazon QuickSight to visualize the data.

Answer: A

NEW QUESTION 62

A company is migrating from an on-premises Apache Hadoop cluster to an Amazon EMR cluster. The cluster runs only during business hours. Due to a company requirement to avoid intraday cluster failures, the EMR cluster must be highly available. When the cluster is terminated at the end of each business day, the data must persist.

Which configurations would enable the EMR cluster to meet these requirements? (Choose three.)

- A. EMR File System (EMRFS) for storage
- B. Hadoop Distributed File System (HDFS) for storage
- C. AWS Glue Data Catalog as the metastore for Apache Hive
- D. MySQL database on the master node as the metastore for Apache Hive
- E. Multiple master nodes in a single Availability Zone
- F. Multiple master nodes in multiple Availability Zones

Answer: ACE

Explanation:

<https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-ha.html> "Note : The cluster can reside only in one Availability Zone or subnet."

NEW QUESTION 65

A large financial company is running its ETL process. Part of this process is to move data from Amazon S3 into an Amazon Redshift cluster. The company wants to use the most cost-efficient method to load the dataset into Amazon Redshift.

Which combination of steps would meet these requirements? (Choose two.)

- A. Use the COPY command with the manifest file to load data into Amazon Redshift.
- B. Use S3DistCp to load files into Amazon Redshift.
- C. Use temporary staging tables during the loading process.
- D. Use the UNLOAD command to upload data into Amazon Redshift.
- E. Use Amazon Redshift Spectrum to query files from Amazon S3.

Answer: AC

NEW QUESTION 67

A retail company has 15 stores across 6 cities in the United States. Once a month, the sales team requests a visualization in Amazon QuickSight that provides the ability to easily identify revenue trends across cities and stores. The visualization also helps identify outliers that need to be examined with further analysis.

Which visual type in QuickSight meets the sales team's requirements?

- A. Geospatial chart
- B. Line chart
- C. Heat map

D. Tree map

Answer: A

NEW QUESTION 71

A large ride-sharing company has thousands of drivers globally serving millions of unique customers every day. The company has decided to migrate an existing data mart to Amazon Redshift. The existing schema includes the following tables.

A trips fact table for information on completed rides. A drivers dimension table for driver profiles. A customers fact table holding customer profile information. The company analyzes trip details by date and destination to examine profitability by region. The drivers data rarely changes. The customers data frequently changes.

What table design provides optimal query performance?

- A. Use DISTSTYLE KEY (destination) for the trips table and sort by date
- B. Use DISTSTYLE ALL for the drivers and customers tables.
- C. Use DISTSTYLE EVEN for the trips table and sort by date
- D. Use DISTSTYLE ALL for the drivers table. Use DISTSTYLE EVEN for the customers table.
- E. Use DISTSTYLE KEY (destination) for the trips table and sort by date
- F. Use DISTSTYLE ALL for the drivers table
- G. Use DISTSTYLE EVEN for the customers table.
- H. Use DISTSTYLE EVEN for the drivers table and sort by date
- I. Use DISTSTYLE ALL for both fact tables.

Answer: C

Explanation:

<https://www.matillion.com/resources/blog/aws-redshift-performance-choosing-the-right-distribution-styles/#:~:t>

https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-best-dist-key.html

NEW QUESTION 72

A data analyst is using Amazon QuickSight for data visualization across multiple datasets generated by applications. Each application stores files within a separate Amazon S3 bucket. AWS Glue Data Catalog is used as a central catalog across all application data in Amazon S3. A new application stores its data within a separate S3 bucket. After updating the catalog to include the new application data source, the data analyst created a new Amazon QuickSight data source from an Amazon Athena table, but the import into SPICE failed.

How should the data analyst resolve the issue?

- A. Edit the permissions for the AWS Glue Data Catalog from within the Amazon QuickSight console.
- B. Edit the permissions for the new S3 bucket from within the Amazon QuickSight console.
- C. Edit the permissions for the AWS Glue Data Catalog from within the AWS Glue console.
- D. Edit the permissions for the new S3 bucket from within the S3 console.

Answer: B

NEW QUESTION 74

A company uses Amazon Kinesis Data Streams to ingest and process customer behavior information from application users each day. A data analytics specialist notices that its data stream is throttling. The specialist has turned on enhanced monitoring for the Kinesis data stream and has verified that the data stream did not exceed the data limits. The specialist discovers that there are hot shards

Which solution will resolve this issue?

- A. Use a random partition key to ingest the records.
- B. Increase the number of shards Split the size of the log records.
- C. Limit the number of records that are sent each second by the producer to match the capacity of the stream.
- D. Decrease the size of the records that are sent from the producer to match the capacity of the stream.

Answer: A

NEW QUESTION 79

A company is reading data from various customer databases that run on Amazon RDS. The databases contain many inconsistent fields For example, a customer record field that is place_id in one database is location_id in another database. The company wants to link customer records across different databases, even when many customer record fields do not match exactly

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

- A. Create an Amazon EMR cluster to process and analyze data in the databases Connect to the Apache Zeppelin notebook, and use the FindMatches transform to find duplicate records in the data.
- B. Create an AWS Glue crawler to crawl the database
- C. Use the FindMatches transform to find duplicate records in the data Evaluate and tune the transform by evaluating performance and results of finding matches
- D. Create an AWS Glue crawler to crawl the data in the databases Use Amazon SageMaker to construct Apache Spark ML pipelines to find duplicate records in the data
- E. Create an Amazon EMR cluster to process and analyze data in the database
- F. Connect to the Apache Zeppelin notebook, and use Apache Spark ML to find duplicate records in the data
- G. Evaluate and tune the model by evaluating performance and results of finding duplicates

Answer: B

NEW QUESTION 80

A company has an application that ingests streaming data. The company needs to analyze this stream over a 5-minute timeframe to evaluate the stream for anomalies with Random Cut Forest (RCF) and summarize the current count of status codes. The source and summarized data should be persisted for future use.

Which approach would enable the desired outcome while keeping data persistence costs low?

- A. Ingest the data stream with Amazon Kinesis Data Stream
- B. Have an AWS Lambda consumer evaluate the stream, collect the number status codes, and evaluate the data against a previously trained RCF mode
- C. Persist the source and results as a time series to Amazon DynamoDB.
- D. Ingest the data stream with Amazon Kinesis Data Stream
- E. Have a Kinesis Data Analytics application evaluate the stream over a 5-minute window using the RCF function and summarize the count of status code
- F. Persist the source and results to Amazon S3 through output delivery to Kinesis Data Firehouse.
- G. Ingest the data stream with Amazon Kinesis Data Firehose with a delivery frequency of 1 minute or 1 MB in Amazon S3. Ensure Amazon S3 triggers an event to invoke an AWS Lambda consumer that evaluates the batch data, collects the number status codes, and evaluates the data against a previously trained RCF mode
- H. Persist the source and results as a time series to Amazon DynamoDB.
- I. Ingest the data stream with Amazon Kinesis Data Firehose with a delivery frequency of 5 minutes or 1 MB into Amazon S3. Have a Kinesis Data Analytics application evaluate the stream over a 1-minute window using the RCF function and summarize the count of status code
- J. Persist the results to Amazon S3 through a Kinesis Data Analytics output to an AWS Lambda integration.

Answer: B

NEW QUESTION 82

A software company wants to use instrumentation data to detect and resolve errors to improve application recovery time. The company requires API usage anomalies, like error rate and response time spikes, to be detected in near-real time (NRT) The company also requires that data analysts have access to dashboards for log analysis in NRT
Which solution meets these requirements'?

- A. Use Amazon Kinesis Data Firehose as the data transport layer for logging data Use Amazon Kinesis Data Analytics to uncover the NRT API usage anomalies Use Kinesis Data Firehose to deliver log data to Amazon OpenSearch Service (Amazon Elasticsearch Service) for search, log analytics, and application monitoring Use OpenSearch Dashboards (Kibana) in Amazon OpenSearch Service (Amazon Elasticsearch Service) for the dashboards.
- B. Use Amazon Kinesis Data Analytics as the data transport layer for logging dat
- C. Use Amazon Kinesis Data Streams to uncover NRT monitoring metric
- D. Use Amazon Kinesis Data Firehose to deliver log data to Amazon OpenSearch Service (Amazon Elasticsearch Service) for search, log analytics, and application monitoring Use Amazon QuickSight for the dashboards
- E. Use Amazon Kinesis Data Analytics as the data transport layer for logging data and to uncover NRT monitoring metrics Use Amazon Kinesis Data Firehose to deliver log data to Amazon OpenSearch Service (Amazon Elasticsearch Service) for search, log analytics, and application monitoring Use OpenSearch Dashboards (Kibana) in Amazon OpenSearch Service (Amazon Elasticsearch Service) for the dashboards
- F. Use Amazon Kinesis Data Firehose as the data transport layer for logging data Use Amazon Kinesis Data Analytics to uncover NRT monitoring metrics Use Amazon Kinesis Data Streams to deliver logdata to Amazon OpenSearch Service (Amazon Elasticsearch Service) for search, log analytics, and application monitoring Use Amazon QuickSight for the dashboards.

Answer: C

NEW QUESTION 84

A company hosts an Apache Flink application on premises. The application processes data from several Apache Kafka clusters. The data originates from a variety of sources, such as web applications mobile apps and operational databases The company has migrated some of these sources to AWS and now wants to migrate the Flink application. The company must ensure that data that resides in databases within the VPC does not traverse the internet The application must be able to process all the data that comes from the company's AWS solution, on-premises resources and the public internet
Which solution will meet these requirements with the LEAST operational overhead?

- A. Implement Flink on Amazon EC2 within the company's VPC Create Amazon Managed Streaming for Apache Kafka (Amazon MSK) clusters in the VPC to collect data that comes from applications and databases within the VPC Use Amazon Kinesis Data Streams to collect data that comes from the public internet Configure Flink to have sources from Kinesis Data Streams Amazon MSK and any on-premises Kafka clusters by using AWS Client VPN or AWS Direct Connect
- B. Implement Flink on Amazon EC2 within the company's VPC Use Amazon Kinesis Data Streams to collect data that comes from applications and databases within the VPC and the public internet Configure Flink to have sources from Kinesis Data Streams and any on-premises Kafka clusters by using AWS Client VPN or AWS Direct Connect
- C. Create an Amazon Kinesis Data Analytics application by uploading the compiled Flink jar file Use Amazon Kinesis Data Streams to collect data that comes from applications and databases within the VPC and the public internet Configure the Kinesis Data Analytics application to have sources from Kinesis Data Streams and any on-premises Kafka clusters by using AWS Client VPN or AWS Direct Connect
- D. Create an Amazon Kinesis Data Analytics application by uploading the compiled Flink jar file Create Amazon Managed Streaming for Apache Kafka (Amazon MSK) clusters in the company's VPC to collect data that comes from applications and databases within the VPC Use Amazon Kinesis Data Streams to collect data that comes from the public internet Configure the Kinesis Data Analytics application to have sources from Kinesis Data Stream
- E. Amazon MSK and any on-premises Kafka clusters by using AWS Client VPN or AWS Direct Connect

Answer: D

NEW QUESTION 89

.....

Thank You for Trying Our Product

* **100% Pass or Money Back**

All our products come with a 90-day Money Back Guarantee.

* **One year free update**

You can enjoy free update one year. 24x7 online support.

* **Trusted by Millions**

We currently serve more than 30,000,000 customers.

* **Shop Securely**

All transactions are protected by VeriSign!

100% Pass Your AWS-Certified-Data-Analytics-Specialty Exam with Our Prep Materials Via below:

<https://www.certleader.com/AWS-Certified-Data-Analytics-Specialty-dumps.html>