Amazon-Web-Services

Exam Questions DAS-C01

AWS Certified Data Analytics - Specialty

About Exambible

Your Partner of IT Exam

Found in 1998

Exambible is a company specialized on providing high quality IT exam practice study materials, especially Cisco CCNA, CCDA, CCNP, CCIE, Checkpoint CCSE, CompTIA A+, Network+ certification practice exams and so on. We guarantee that the candidates will not only pass any IT exam at the first attempt but also get profound understanding about the certificates they have got. There are so many alike companies in this industry, however, Exambible has its unique advantages that other companies could not achieve.

Our Advances

* 99.9% Uptime

    All examinations will be up to date.

* 24/7 Quality Support

    We will provide service round the clock.

* 100% Pass Rate

    Our guarantee that you will pass the exam.

* Unique Gurantee

    If you do not pass the exam at the first time, we will not only arrange FULL REFUND for you, but also provide you another exam of your claim, ABSOLUTELY FREE!

NEW QUESTION 1
A financial company uses Amazon S3 as its data lake and has set up a data warehouse using a multi-node Amazon Redshift cluster. The data files in the data lake are organized in folders based on the data source of each data file. All the data files are loaded to one table in the Amazon Redshift cluster using a separate COPY command for each data file location. With this approach, loading all the data files into Amazon Redshift takes a long time to complete. Users want a faster solution with little or no increase in cost while maintaining the segregation of the data files in the S3 data lake.
Which solution meets these requirements?

A. Use Amazon EMR to copy all the data files into one folder and issue a COPY command to load the data into Amazon Redshift.
B. Load all the data files in parallel to Amazon Aurora, and run an AWS Glue job to load the data into Amazon Redshift.
C. Use an AWS Glue job to copy all the data files into one folder and issue a COPY command to load the data into Amazon Redshift.
D. Create a manifest file that contains the data file locations and issue a COPY command to load the data into Amazon Redshift.

Answer: D

Explanation:
https://docs.aws.amazon.com/redshift/latest/dg/loading-data-files-using-manifest.html "You can use a manifest to ensure that the COPY command loads all of the required files, and only the required files, for a data load"

NEW QUESTION 2
Three teams of data analysts use Apache Hive on an Amazon EMR cluster with the EMR File System (EMRFS) to query data stored within each teams Amazon S3 bucket. The EMR cluster has Kerberos enabled and is configured to authenticate users from the corporate Active Directory. The data is highly sensitive, so access must be limited to the members of each team.
Which steps will satisfy the security requirements?

A. For the EMR cluster Amazon EC2 instances, create a service role that grants no access to Amazon S3.Create three additional IAM roles, each granting access to each team's specific bucke
B. Add the additional IAM roles to the cluster's EMR role for the EC2 trust polic
C. Create a security configuration mapping for the additional IAM roles to Active Directory user groups for each team.
D. For the EMR cluster Amazon EC2 instances, create a service role that grants no access to Amazon S3.Create three additional IAM roles, each granting access to each team's specific bucke
E. Add the service role for the EMR cluster EC2 instances to the trust policies for the additional IAM role
F. Create a security configuration mapping for the additional IAM roles to Active Directory user groups for each team.
G. For the EMR cluster Amazon EC2 instances, create a service role that grants full access to Amazon S3.Create three additional IAM roles, each granting access to each team's specific bucke
H. Add the service role for the EMR cluster EC2 instances to the trust polices for the additional IAM role
I. Create a security configuration mapping for the additional IAM roles to Active Directory user groups for each team.
J. For the EMR cluster Amazon EC2 instances, create a service role that grants full access to Amazon S3.Create three additional IAM roles, each granting access to each team's specific bucke
K. Add the service role for the EMR cluster EC2 instances to the trust polices for the base IAM role
L. Create a security configuration mapping for the additional IAM roles to Active Directory user groups for each team.

Answer: C

NEW QUESTION 3
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 4
A company wants to optimize the cost of its data and analytics platform. The company is ingesting a number of .c sv and JSON files in Amazon S3 from various data sources. Incoming data is expected to be 50 GB each day. The company is using Amazon Athena to query the raw data in Amazon S3 directly. Most queries aggregate data from the past 12 months, and data that is older than 5 years is infrequently queried. The typical query scans about 500 MB of data and is expected to return results in less than 1 minute. The raw data must be retained indefinitely for compliance requirements.
Which solution meets the company's requirements?

A. Use an AWS Glue ETL job to compress, partition, and convert the data into a columnar data forma
B. Use Athena to query the processed datase
C. Configure a lifecycle policy to move the processed data into the Amazon S3 Standard-Infrequent Access (S3 Standard-IA) storage class 5 years after object creation.Configure a second lifecycle policy to move the raw data into Amazon S3 Glacier for long-term archival 7 days after object creation.
D. Use an AWS Glue ETL job to partition and convert the data into a row-based data forma
E. Use Athena to query the processed datase
F. Configure a lifecycle policy to move the data into the Amazon S3 Standard- Infrequent Access (S3 Standard-IA) storage class 5 years after object creatio
G. Configure a second lifecycle policy to move the raw data into Amazon S3 Glacier for long-term archival 7 days after object creation.
H. Use an AWS Glue ETL job to compress, partition, and convert the data into a columnar data forma
I. Use Athena to query the processed datase
J. Configure a lifecycle policy to move the processed data into the Amazon S3 Standard-Infrequent Access (S3 Standard-IA) storage class 5 years after the object was last accesse
K. Configure a second lifecycle policy to move the raw data into Amazon S3 Glacier forlong-term archival 7 days after the last date the object was accessed.
L. Use an AWS Glue ETL job to partition and convert the data into a row-based data forma

M. Use Athena to query the processed datase

N. Configure a lifecycle policy to move the data into the Amazon S3 Standard- Infrequent Access (S3 Standard-IA) storage class 5 years after the object was last accesse

O. Configure a second lifecycle policy to move the raw data into Amazon S3 Glacier for long-term archival 7 days after the last date the object was accessed.

Answer: A

NEW QUESTION 5

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 dat

B. Use DISTSTYLE ALL for the drivers and customers tables.

C. Use DISTSTYLE EVEN for the trips table and sort by dat

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 dat

F. Use DISTSTYLE ALL for the drivers tabl

G. Use DISTSTYLE EVEN for the customers table.

H. Use DISTSTYLE EVEN for the drivers table and sort by dat

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 6

A company's marketing team has asked for help in identifying a high performing long-term storage service for their data based on the following requirements:

≫ The data size is approximately 32 TB uncompressed.

≫ There is a low volume of single-row inserts each day.

≫ There is a high volume of aggregation queries each day.

≫ Multiple complex joins are performed.

≫ The queries typically involve a small subset of the columns in a table.

Which storage service will provide the MOST performant solution?

A. Amazon Aurora MySQL

B. Amazon Redshift

C. Amazon Neptune

D. Amazon Elasticsearch

Answer: B

NEW QUESTION 7

A healthcare company uses AWS data and analytics tools to collect, ingest, and store electronic health record (EHR) data about its patients. The raw EHR data is stored in Amazon S3 in JSON format partitioned by hour, day, and year and is updated every hour. The company wants to maintain the data catalog and metadata in an AWS Glue Data Catalog to be able to access the data using Amazon Athena or Amazon Redshift Spectrum for analytics.

When defining tables in the Data Catalog, the company has the following requirements:

Choose the catalog table name and do not rely on the catalog table naming algorithm. Keep the table updated with new partitions loaded in the respective S3 bucket prefixes.

Which solution meets these requirements with minimal effort?

A. Run an AWS Glue crawler that connects to one or more data stores, determines the data structures, and writes tables in the Data Catalog.

B. Use the AWS Glue console to manually create a table in the Data Catalog and schedule an AWS Lambda function to update the table partitions hourly.

C. Use the AWS Glue API CreateTable operation to create a table in the Data Catalo

D. Create an AWS Glue crawler and specify the table as the source.

E. Create an Apache Hive catalog in Amazon EMR with the table schema definition in Amazon S3, and update the table partition with a scheduled jo

F. Migrate the Hive catalog to the Data Catalog.

Answer: C

Explanation:
Updating Manually Created Data Catalog Tables Using Crawlers: To do this, when you define a crawler, instead of specifying one or more data stores as the source of a crawl, you specify one or more existing Data Catalog tables. The crawler then crawls the data stores specified by the catalog tables. In this case, no new tables are created; instead, your manually created tables are updated.

NEW QUESTION 8

A company stores its sales and marketing data that includes personally identifiable information (PII) in Amazon S3. The company allows its analysts to launch their own Amazon EMR cluster and run analytics reports with the data. To meet compliance requirements, the company must ensure the data is not publicly accessible throughout this process. A data engineer has secured Amazon S3 but must ensure the individual EMR clusters created by the analysts are not exposed to the public internet.

Which solution should the data engineer to meet this compliance requirement with LEAST amount of effort?

A. Create an EMR security configuration and ensure the security configuration is associated with the EMR clusters when they are created.
B. Check the security group of the EMR clusters regularly to ensure it does not allow inbound traffic from IPv4 0.0.0.0/0 or IPv6 ::/0.
C. Enable the block public access setting for Amazon EMR at the account level before any EMR cluster is created.
D. Use AWS WAF to block public internet access to the EMR clusters across the board.

Answer: C

Explanation:
https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-block-public-access.html

NEW QUESTION 9
A software company hosts an application on AWS, and new features are released weekly. As part of the application testing process, a solution must be developed that analyzes logs from each Amazon EC2 instance to ensure that the application is working as expected after each deployment. The collection and analysis solution should be highly available with the ability to display new information with minimal delays.
Which method should the company use to collect and analyze the logs?

A. Enable detailed monitoring on Amazon EC2, use Amazon CloudWatch agent to store logs in Amazon S3, and use Amazon Athena for fast, interactive log analytics.
B. Use the Amazon Kinesis Producer Library (KPL) agent on Amazon EC2 to collect and send data to Kinesis Data Streams to further push the data to Amazon Elasticsearch Service and visualize using Amazon QuickSight.
C. Use the Amazon Kinesis Producer Library (KPL) agent on Amazon EC2 to collect and send data to Kinesis Data Firehose to further push the data to Amazon Elasticsearch Service and Kibana.
D. Use Amazon CloudWatch subscriptions to get access to a real-time feed of logs and have the logs delivered to Amazon Kinesis Data Streams to further push the data to Amazon Elasticsearch Service and Kibana.

Answer: D

NEW QUESTION 10
A manufacturing company wants to create an operational analytics dashboard to visualize metrics from equipment in near-real time. The company uses Amazon Kinesis Data Streams to stream the data to other applications. The dashboard must automatically refresh every 5 seconds. A data analytics specialist must design a solution that requires the least possible implementation effort.
Which solution meets these requirements?

A. Use Amazon Kinesis Data Firehose to store the data in Amazon S3. Use Amazon QuickSight to build the dashboard.
B. Use Apache Spark Streaming on Amazon EMR to read the data in near-real tim
C. Develop a custom application for the dashboard by using D3.js.
D. Use Amazon Kinesis Data Firehose to push the data into an Amazon Elasticsearch Service (Amazon ES) cluste
E. Visualize the data by using a Kibana dashboard.
F. Use AWS Glue streaming ETL to store the data in Amazon S3. Use Amazon QuickSight to build the dashboard.

Answer: B

NEW QUESTION 10
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 14
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.

Answer: B

NEW QUESTION 17
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 18
A company wants to collect and process events data from different departments in near-real time. Before storing the data in Amazon S3, the company needs to clean the data by standardizing the format of the address and timestamp columns. The data varies in size based on the overall load at each particular point in time. A single data record can be 100 KB-10 MB.
How should a data analytics specialist design the solution for data ingestion?

A. Use Amazon Kinesis Data Stream
B. Configure a stream for the raw dat
C. Use a Kinesis Agent to write data to the strea
D. Create an Amazon Kinesis Data Analytics application that reads data from the raw stream, cleanses it, and stores the output to Amazon S3.
E. Use Amazon Kinesis Data Firehos
F. Configure a Firehose delivery stream with a preprocessing AWS Lambda function for data cleansin
G. Use a Kinesis Agent to write data to the delivery strea
H. Configure Kinesis Data Firehose to deliver the data to Amazon S3.
I. Use Amazon Managed Streaming for Apache Kafk
J. Configure a topic for the raw dat
K. Use a Kafka producer to write data to the topi
L. Create an application on Amazon EC2 that reads data from the topic by using the Apache Kafka consumer API, cleanses the data, and writes to Amazon S3.
M. Use Amazon Simple Queue Service (Amazon SQS). Configure an AWS Lambda function to read events from the SQS queue and upload the events to Amazon S3.

Answer: B

NEW QUESTION 23
A hospital is building a research data lake to ingest data from electronic health records (EHR) systems from multiple hospitals and clinics. The EHR systems are independent of each other and do not have a common patient identifier. The data engineering team is not experienced in machine learning (ML) and has been asked to generate a unique patient identifier for the ingested records.
Which solution will accomplish this task?

A. An AWS Glue ETL job with the FindMatches transform
B. Amazon Kendra
C. Amazon SageMaker Ground Truth
D. An AWS Glue ETL job with the ResolveChoice transform

Answer: A

Explanation:
Matching Records with AWS Lake Formation FindMatches

NEW QUESTION 28
A manufacturing company has been collecting IoT sensor data from devices on its factory floor for a year and is storing the data in Amazon Redshift for daily analysis. A data analyst has determined that, at an expected ingestion rate of about 2 TB per day, the cluster will be undersized in less than 4 months. A long-term solution is needed. The data analyst has indicated that most queries only reference the most recent 13 months of data, yet there are also quarterly reports that need to query all the data generated from the past 7 years. The chief technology officer (CTO) is concerned about the costs, administrative effort, and performance of a long-term solution.
Which solution should the data analyst use to meet these requirements?

A. Create a daily job in AWS Glue to UNLOAD records older than 13 months to Amazon S3 and delete those records from Amazon Redshif
B. Create an external table in Amazon Redshift to point to the S3 locatio
C. Use Amazon Redshift Spectrum to join to data that is older than 13 months.
D. Take a snapshot of the Amazon Redshift cluste
E. Restore the cluster to a new cluster using dense storage nodes with additional storage capacity.
F. Execute a CREATE TABLE AS SELECT (CTAS) statement to move records that are older than 13 months to quarterly partitioned data in Amazon Redshift Spectrum backed by Amazon S3.
G. Unload all the tables in Amazon Redshift to an Amazon S3 bucket using S3 Intelligent-Tierin
H. Use AWS Glue to crawl the S3 bucket location to create external tables in an AWS Glue Data Catalog.Create an Amazon EMR cluster using Auto Scaling for any daily analytics needs, and use Amazon Athena for the quarterly reports, with both using the same AWS Glue Data Catalog.

Answer: A

NEW QUESTION 29
A company using Amazon QuickSight Enterprise edition has thousands of dashboards analyses and datasets. The company struggles to manage and assign permissions for granting users access to various items within QuickSight. The company wants to make it easier to implement sharing and permissions management.
Which solution should the company implement to simplify permissions management?

A. Use QuickSight folders to organize dashboards, analyses, and datasets Assign individual users permissions to these folders
B. Use QuickSight folders to organize dashboards analyses, and datasets Assign group permissions by using these folders.
C. Use AWS 1AM resource-based policies to assign group permissions to QuickSight items
D. Use QuickSight user management APIs to provision group permissions based on dashboard naming conventions

Answer: C

NEW QUESTION 34
A company with a video streaming website wants to analyze user behavior to make recommendations to users in real time Clickstream data is being sent to Amazon Kinesis Data Streams and reference data is stored in Amazon S3 The company wants a solution that can use standard SQL quenes The solution must

also provide a way to look up pre-calculated reference data while making recommendations
Which solution meets these requirements?

A. Use an AWS Glue Python shell job to process incoming data from Kinesis Data Streams Use the Boto3 library to write data to Amazon Redshift
B. Use AWS Glue streaming and Scale to process incoming data from Kinesis Data Streams Use the AWS Glue connector to write data to Amazon Redshift
C. Use Amazon Kinesis Data Analytics to create an in-application table based upon the reference data Process incoming data from Kinesis Data Streams Use a data stream to write results to Amazon Redshift
D. Use Amazon Kinesis Data Analytics to create an in-application table based upon the reference dataProcess incoming data from Kinesis Data Streams Use an Amazon Kinesis Data Firehose delivery stream to write results to Amazon Redshift

Answer: D

NEW QUESTION 35
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 37
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 42
A company stores Apache Parquet-formatted files in Amazon S3 The company uses an AWS Glue Data Catalog to store the table metadata and Amazon Athena to query and analyze the data The tables have a large number of partitions The queries are only run on small subsets of data in the table A data analyst adds new time partitions into the table as new data arrives The data analyst has been asked to reduce the query runtime
Which solution will provide the MOST reduction in the query runtime?

A. Convert the Parquet files to the csv file format..Then attempt to query the data again
B. Convert the Parquet files to the Apache ORC file forma
C. Then attempt to query the data again
D. Use partition projection to speed up the processing of the partitioned table
E. Add more partitions to be used over the tabl
F. Then filter over two partitions and put all columns in the WHERE clause

Answer: C

NEW QUESTION 47
A company wants to improve the data load time of a sales data dashboard. Data has been collected as .csv files and stored within an Amazon S3 bucket that is partitioned by date. The data is then loaded to an Amazon Redshift data warehouse for frequent analysis. The data volume is up to 500 GB per day.
Which solution will improve the data loading performance?

A. Compress .csv files and use an INSERT statement to ingest data into Amazon Redshift.
B. Split large .csv files, then use a COPY command to load data into Amazon Redshift.
C. Use Amazon Kinesis Data Firehose to ingest data into Amazon Redshift.
D. Load the .csv files in an unsorted key order and vacuum the table in Amazon Redshift.

Answer: B

Explanation:
https://docs.aws.amazon.com/redshift/latest/dg/c_loading-data-best-practices.html

NEW QUESTION 52
A company is migrating its existing on-premises ETL jobs to Amazon EMR. The code consists of a series of jobs written in Java. The company needs to reduce overhead for the system administrators without changing the underlying code. Due to the sensitivity of the data, compliance requires that the company use root device volume encryption on all nodes in the cluster. Corporate standards require that environments be provisioned though AWS CloudFormation when possible.
Which solution satisfies these requirements?

A. Install open-source Hadoop on Amazon EC2 instances with encrypted root device volume
B. Configure the cluster in the CloudFormation template.
C. Use a CloudFormation template to launch an EMR cluste
D. In the configuration section of the cluster, define a bootstrap action to enable TLS.
E. Create a custom AMI with encrypted root device volume
F. Configure Amazon EMR to use the custom AMI using the CustomAmild property in the CloudFormation template.
G. Use a CloudFormation template to launch an EMR cluste
H. In the configuration section of the cluster, define a bootstrap action to encrypt the root device volume of every node.

Answer: C

NEW QUESTION 54
A media content company has a streaming playback application. The company wants to collect and analyze the data to provide near-real-time feedback on playback issues. The company needs to consume this data and return results within 30 seconds according to the service-level agreement (SLA). The company needs the consumer to identify playback issues, such as quality during a specified timeframe. The data will be emitted as JSON and may change schemas over time.
Which solution will allow the company to collect data for processing while meeting these requirements?

A. Send the data to Amazon Kinesis Data Firehose with delivery to Amazon S3. Configure an S3 event trigger an AWS Lambda function to process the dat
B. The Lambda function will consume the data and process it to identify potential playback issue
C. Persist the raw data to Amazon S3.
D. Send the data to Amazon Managed Streaming for Kafka and configure an Amazon Kinesis Analytics for Java application as the consume
E. The application will consume the data and process it to identify potential playback issue
F. Persist the raw data to Amazon DynamoDB.
G. Send the data to Amazon Kinesis Data Firehose with delivery to Amazon S3. Configure Amazon S3 to trigger an event for AWS Lambda to proces
H. The Lambda function will consume the data and process it to identify potential playback issue
I. Persist the raw data to Amazon DynamoDB.
J. Send the data to Amazon Kinesis Data Streams and configure an Amazon Kinesis Analytics for Java application as the consume
K. The application will consume the data and process it to identify potential playback issue
L. Persist the raw data to Amazon S3.

Answer: D

Explanation:
https://aws.amazon.com/blogs/aws/new-amazon-kinesis-data-analytics-for-java/

NEW QUESTION 59
A company launched a service that produces millions of messages every day and uses Amazon Kinesis Data Streams as the streaming service.
The company uses the Kinesis SDK to write data to Kinesis Data Streams. A few months after launch, a data analyst found that write performance is significantly reduced. The data analyst investigated the metrics and determined that Kinesis is throttling the write requests. The data analyst wants to address this issue without significant changes to the architecture.
Which actions should the data analyst take to resolve this issue? (Choose two.)

A. Increase the Kinesis Data Streams retention period to reduce throttling.
B. Replace the Kinesis API-based data ingestion mechanism with Kinesis Agent.
C. Increase the number of shards in the stream using the UpdateShardCount API.
D. Choose partition keys in a way that results in a uniform record distribution across shards.
E. Customize the application code to include retry logic to improve performance.

Answer: CD

Explanation:
https://aws.amazon.com/blogs/big-data/under-the-hood-scaling-your-kinesis-data-streams/

NEW QUESTION 63
A company leverages Amazon Athena for ad-hoc queries against data stored in Amazon S3. The company wants to implement additional controls to separate query execution and query history among users, teams, or applications running in the same AWS account to comply with internal security policies.
Which solution meets these requirements?

A. Create an S3 bucket for each given use case, create an S3 bucket policy that grants permissions to appropriate individual IAM user
B. and apply the S3 bucket policy to the S3 bucket.
C. Create an Athena workgroup for each given use case, apply tags to the workgroup, and create an IAM policy using the tags to apply appropriate permissions to the workgroup.
D. Create an IAM role for each given use case, assign appropriate permissions to the role for the given usecase, and add the role to associate the role with Athena.
E. Create an AWS Glue Data Catalog resource policy for each given use case that grants permissions to appropriate individual IAM users, and apply the resource policy to the specific tables used by Athena.

Answer: B

Explanation:
https://docs.aws.amazon.com/athena/latest/ug/user-created-workgroups.html
Amazon Athena Workgroups - A new resource type that can be used to separate query execution and query history between Users, Teams, or Applications running under the same AWS account https://aws.amazon.com/about-aws/whats-new/2019/02/athena_workgroups/

NEW QUESTION 65
A marketing company has data in Salesforce, MySQL, and Amazon S3. The company wants to use data from these three locations and create mobile dashboards for its users. The company is unsure how it should create the dashboards and needs a solution with the least possible customization and coding.
Which solution meets these requirements?

A. Use Amazon Athena federated queries to join the data source
B. Use Amazon QuickSight to generate the mobile dashboards.
C. Use AWS Lake Formation to migrate the data sources into Amazon S3. Use Amazon QuickSight to generate the mobile dashboards.
D. Use Amazon Redshift federated queries to join the data source
E. Use Amazon QuickSight to generate the mobile dashboards.
F. Use Amazon QuickSight to connect to the data sources and generate the mobile dashboards.

Answer: C

NEW QUESTION 67
A data engineering team within a shared workspace company wants to build a centralized logging system for all weblogs generated by the space reservation system. The company has a fleet of Amazon EC2 instances that process requests for shared space reservations on its website. The data engineering team wants to ingest all weblogs into a service that will provide a near-real-time search engine. The team does not want to manage the maintenance and operation of the logging system.
Which solution allows the data engineering team to efficiently set up the web logging system within AWS?

A. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis data stream to CloudWatc
B. Choose Amazon Elasticsearch Service as the end destination of the weblogs.
C. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis Data Firehose delivery stream to CloudWatc
D. Choose Amazon Elasticsearch Service as the end destination of the weblogs.
E. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis data stream to CloudWatc
F. Configure Splunk as the end destination of the weblogs.
G. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis Firehose delivery stream to CloudWatc
H. Configure Amazon DynamoDB as the end destination of the weblogs.

Answer: B

Explanation:
https://docs.aws.amazon.com/AmazonCloudWatch/latest/logs/CWL_ES_Stream.html

NEW QUESTION 70
A data analyst is designing an Amazon QuickSight dashboard using centralized sales data that resides in Amazon Redshift. The dashboard must be restricted so that a salesperson in Sydney, Australia, can see only the Australia view and that a salesperson in New York can see only United States (US) data.
What should the data analyst do to ensure the appropriate data security is in place?

A. Place the data sources for Australia and the US into separate SPICE capacity pools.
B. Set up an Amazon Redshift VPC security group for Australia and the US.
C. Deploy QuickSight Enterprise edition to implement row-level security (RLS) to the sales table.
D. Deploy QuickSight Enterprise edition and set up different VPC security groups for Australia and the US.

Answer: D

NEW QUESTION 75
A company has an application that uses the Amazon Kinesis Client Library (KCL) to read records from a Kinesis data stream.
After a successful marketing campaign, the application experienced a significant increase in usage. As a result, a data analyst had to split some shards in the data stream. When the shards were split, the application started
throwing an ExpiredIteratorExceptions error sporadically. What should the data analyst do to resolve this?

A. Increase the number of threads that process the stream records.
B. Increase the provisioned read capacity units assigned to the stream's Amazon DynamoDB table.
C. Increase the provisioned write capacity units assigned to the stream's Amazon DynamoDB table.
D. Decrease the provisioned write capacity units assigned to the stream's Amazon DynamoDB table.

Answer: C

NEW QUESTION 78
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 databas
G. Use an AWS Glue job to enrich the data and save the result to Amazon S3 in Apache Parquet forma
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 databas

J. Use an AWS Glue job to enrich the data and save the result to Amazon S3 in Apache Parquet forma
K. Use Amazon Athena to query the data.

Answer: B

NEW QUESTION 79
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 81
A company needs to collect streaming data from several sources and store the data in the AWS Cloud. The dataset is heavily structured, but analysts need to
perform several complex SQL queries and need consistent performance. Some of the data is queried more frequently than the rest. The company wants a solution
that meets its performance requirements in a cost-effective manner.
Which solution meets these requirements?

A. Use Amazon Managed Streaming for Apache Kafka to ingest the data to save it to Amazon S3. Use Amazon Athena to perform SQL queries over the ingested
data.
B. Use Amazon Managed Streaming for Apache Kafka to ingest the data to save it to Amazon Redshift.Enable Amazon Redshift workload management (WLM) to
prioritize workloads.
C. Use Amazon Kinesis Data Firehose to ingest the data to save it to Amazon Redshif
D. Enable Amazon Redshift workload management (WLM) to prioritize workloads.
E. Use Amazon Kinesis Data Firehose to ingest the data to save it to Amazon S3. Load frequently queried data to Amazon Redshift using the COPY comman
F. Use Amazon Redshift Spectrum for less frequently queried data.

Answer: B

NEW QUESTION 82
A marketing company collects clickstream data The company sends the data to Amazon Kinesis Data Firehose and stores the data in Amazon S3 The company
wants to build a series of dashboards that will be used by hundreds of users across different departments The company will use Amazon QuickSight to develop
these dashboards The company has limited resources and wants a solution that could scale and provide daily updates about clickstream activity
Which combination of options will provide the MOST cost-effective solution? (Select TWO)

A. Use Amazon Redshift to store and query the clickstream data
B. Use QuickSight with a direct SQL query
C. Use Amazon Athena to query the clickstream data in Amazon S3
D. Use S3 analytics to query the clickstream data
E. Use the QuickSight SPICE engine with a daily refresh

Answer: BD

NEW QUESTION 86
A company analyzes historical data and needs to query data that is stored in Amazon S3. New data is generated daily as .csv files that are stored in Amazon S3.
The company's analysts are using Amazon Athena to perform SQL queries against a recent subset of the overall data. The amount of data that is ingested into
Amazon S3 has increased substantially over time, and the query latency also has increased.
Which solutions could the company implement to improve query performance? (Choose two.)

A. Use MySQL Workbench on an Amazon EC2 instance, and connect to Athena by using a JDBC or ODBC connecto
B. Run the query from MySQL Workbench instead of Athena directly.
C. Use Athena to extract the data and store it in Apache Parquet format on a daily basi
D. Query the extracted data.
E. Run a daily AWS Glue ETL job to convert the data files to Apache Parquet and to partition the converted file
F. Create a periodic AWS Glue crawler to automatically crawl the partitioned data on a daily basis.
G. Run a daily AWS Glue ETL job to compress the data files by using the .gzip forma
H. Query the compressed data.
I. Run a daily AWS Glue ETL job to compress the data files by using the .lzo forma
J. Query the compressed data.

Answer: BC

NEW QUESTION 90

A company operates toll services for highways across the country and collects data that is used to understand usage patterns. Analysts have requested the ability to run traffic reports in near-real time. The company is interested in building an ingestion pipeline that loads all the data into an Amazon Redshift cluster and alerts operations personnel when toll traffic for a particular toll station does not meet a specified threshold. Station data and the corresponding threshold values are stored in Amazon S3.
Which approach is the MOST efficient way to meet these requirements?

A. Use Amazon Kinesis Data Firehose to collect data and deliver it to Amazon Redshift and Amazon Kinesis Data Analytics simultaneousl
B. Create a reference data source in Kinesis Data Analytics to temporarily store the threshold values from Amazon S3 and compare the count of vehicles for a particular toll station against its corresponding threshold valu
C. Use AWS Lambda to publish an Amazon Simple Notification Service (Amazon SNS) notification if the threshold is not met.
D. Use Amazon Kinesis Data Streams to collect all the data from toll station
E. Create a stream in Kinesis Data Streams to temporarily store the threshold values from Amazon S3. Send both streams to Amazon Kinesis Data Analytics to compare the count of vehicles for a particular toll station against its corresponding threshold valu
F. Use AWS Lambda to publish an Amazon Simple Notification Service (Amazon SNS) notification if the threshold is not me
G. Connect Amazon Kinesis Data Firehose to Kinesis Data Streams to deliver the data to Amazon Redshift.
H. Use Amazon Kinesis Data Firehose to collect data and deliver it to Amazon Redshif
I. Then, automatically trigger an AWS Lambda function that queries the data in Amazon Redshift, compares the count of vehicles for a particular toll station against its corresponding threshold values read from Amazon S3, and publishes an Amazon Simple Notification Service (Amazon SNS) notification if the threshold is not met.
J. Use Amazon Kinesis Data Firehose to collect data and deliver it to Amazon Redshift and Amazon Kinesis Data Analytics simultaneousl
K. Use Kinesis Data Analytics to compare the count of vehicles against the threshold value for the station stored in a table as an in-application stream based on information stored in Amazon S3. Configure an AWS Lambda function as an output for the application that will publish an Amazon Simple Queue Service (Amazon SQS) notification to alert operations personnel if the threshold is not met.

Answer: D

NEW QUESTION 93
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 strea
B. Decrypt the sensitive data using the same KMS ke
C. Save the output to a restricted S3 bucket for the finance tea
D. Create a finance table in Amazon Redshift that is accessible to the finance team onl
E. Use the COPY command to load the data from Amazon S3 to the finance table.
F. Create an AWS Lambda function to process the DynamoDB strea
G. Save the output to a restricted S3 bucket for the finance tea
H. Create a finance table in Amazon Redshift that is accessible to the finance team onl
I. 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.
J. 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 Redshif
K. In Hive, select the data from DynamoDB and then insert the output to the finance table in Amazon Redshift.
L. Create an Amazon EMR cluste
M. Create Apache Hive tables that reference the data stored in DynamoD
N. Insert the output to the restricted Amazon S3 bucket for the finance tea
O. 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.

Answer: B

NEW QUESTION 95
A manufacturing company uses Amazon Connect to manage its contact center and Salesforce to manage its customer relationship management (CRM) data. The data engineering team must build a pipeline to ingest data from the contact center and CRM system into a data lake that is built on Amazon S3.
What is the MOST efficient way to collect data in the data lake with the LEAST operational overhead?

A. Use Amazon Kinesis Data Streams to ingest Amazon Connect data and Amazon AppFlow to ingest Salesforce data.
B. Use Amazon Kinesis Data Firehose to ingest Amazon Connect data and Amazon Kinesis Data Streams to ingest Salesforce data.
C. Use Amazon Kinesis Data Firehose to ingest Amazon Connect data and Amazon AppFlow to ingest Salesforce data.
D. Use Amazon AppFlow to ingest Amazon Connect data and Amazon Kinesis Data Firehose to ingest Salesforce data.

Answer: B

NEW QUESTION 97
A banking company wants to collect large volumes of transactional data using Amazon Kinesis Data Streams for real-time analytics. The company uses PutRecord to send data to Amazon Kinesis, and has observed network outages during certain times of the day. The company wants to obtain exactly once semantics for the entire processing pipeline.
What should the company do to obtain these characteristics?

A. Design the application so it can remove duplicates during processing be embedding a unique ID in each record.
B. Rely on the processing semantics of Amazon Kinesis Data Analytics to avoid duplicate processing of events.
C. Design the data producer so events are not ingested into Kinesis Data Streams multiple times.
D. Rely on the exactly one processing semantics of Apache Flink and Apache Spark Streaming included in Amazon EMR.

Answer: A

NEW QUESTION 98
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 nod
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 instanc
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://docs.aws.amazon.com/de_de/emr/latest/ManagementGuide/emr-plan-bootstrap.html

NEW QUESTION 102
A large retailer has successfully migrated to an Amazon S3 data lake architecture. The company's marketing team is using Amazon Redshift and Amazon
QuickSight to analyze data, and derive and visualize insights. To ensure the marketing team has the most up-to-date actionable information, a data analyst
implements nightly refreshes of Amazon Redshift using terabytes of updates from the previous day.
After the first nightly refresh, users report that half of the most popular dashboards that had been running correctly before the refresh are now running much
slower. Amazon CloudWatch does not show any alerts.
What is the MOST likely cause for the performance degradation?

A. The dashboards are suffering from inefficient SQL queries.
B. The cluster is undersized for the queries being run by the dashboards.
C. The nightly data refreshes are causing a lingering transaction that cannot be automatically closed by Amazon Redshift due to ongoing user workloads.
D. The nightly data refreshes left the dashboard tables in need of a vacuum operation that could not be automatically performed by Amazon Redshift due to
ongoing user workloads.

Answer: D

Explanation:
https://github.com/awsdocs/amazon-redshift-developer-guide/issues/21

NEW QUESTION 106
An ecommerce company stores customer purchase data in Amazon RDS. The company wants a solution to store and analyze historical data. The most recent 6
months of data will be queried frequently for analytics workloads. This data is several terabytes large. Once a month, historical data for the last 5 years must be
accessible and will be joined with the more recent data. The company wants to optimize performance and cost.
Which storage solution will meet these requirements?

A. Create a read replica of the RDS database to store the most recent 6 months of dat
B. Copy the historical data into Amazon S3. Create an AWS Glue Data Catalog of the data in Amazon S3 and Amazon RD
C. Run historical queries using Amazon Athena.
D. Use an ETL tool to incrementally load the most recent 6 months of data into an Amazon Redshift cluste
E. Run more frequent queries against this cluste
F. Create a read replica of the RDS database to run queries on the historical data.
G. Incrementally copy data from Amazon RDS to Amazon S3. Create an AWS Glue Data Catalog of the data in Amazon S3. Use Amazon Athena to query the
data.
H. Incrementally copy data from Amazon RDS to Amazon S3. Load and store the most recent 6 months of data in Amazon Redshif
I. Configure an Amazon Redshift Spectrum table to connect to all historical data.

Answer: D

NEW QUESTION 108
A company is planning to create a data lake in Amazon S3. The company wants to create tiered storage based on access patterns and cost objectives. The
solution must include support for JDBC connections from legacy clients, metadata management that allows federation for access control, and batch-based ETL
using PySpark and Scala. Operational management should be limited.
Which combination of components can meet these requirements? (Choose three.)

A. AWS Glue Data Catalog for metadata management
B. Amazon EMR with Apache Spark for ETL
C. AWS Glue for Scala-based ETL
D. Amazon EMR with Apache Hive for JDBC clients
E. Amazon Athena for querying data in Amazon S3 using JDBC drivers
F. Amazon EMR with Apache Hive, using an Amazon RDS with MySQL-compatible backed metastore

Answer: BEF

NEW QUESTION 112
A company wants to run analytics on its Elastic Load Balancing logs stored in Amazon S3. A data analyst needs to be able to query all data from a desired year,
month, or day. The data analyst should also be able to query a subset of the columns. The company requires minimal operational overhead and the most cost-
effective solution.
Which approach meets these requirements for optimizing and querying the log data?

A. Use an AWS Glue job nightly to transform new log files into .csv format and partition by year, month, and da
B. Use AWS Glue crawlers to detect new partition

C. Use Amazon Athena to query data.
D. Launch a long-running Amazon EMR cluster that continuously transforms new log files from Amazon S3 into its Hadoop Distributed File System (HDFS) storage and partitions by year, month, and da
E. Use Apache Presto to query the optimized format.
F. Launch a transient Amazon EMR cluster nightly to transform new log files into Apache ORC format and partition by year, month, and da
G. Use Amazon Redshift Spectrum to query the data.
H. Use an AWS Glue job nightly to transform new log files into Apache Parquet format and partition by year, month, and da
I. Use AWS Glue crawlers to detect new partition
J. Use Amazon Athena to querydata.

Answer: C

NEW QUESTION 117
An online retail company uses Amazon Redshift to store historical sales transactions. The company is required to encrypt data at rest in the clusters to comply with the Payment Card Industry Data Security Standard (PCI DSS). A corporate governance policy mandates management of encryption keys using an on-premises hardware security module (HSM).
Which solution meets these requirements?

A. Create and manage encryption keys using AWS CloudHSM Classi
B. Launch an Amazon Redshift cluster in a VPC with the option to use CloudHSM Classic for key management.
C. Create a VPC and establish a VPN connection between the VPC and the on-premises networt
D. Create an HSM connection and client certificate for the on-premises HS
E. Launch a cluster in the VPC with theoption to use the on-premises HSM to store keys.
F. Create an HSM connection and client certificate for the on-premises HS
G. Enable HSM encryption on the existing unencrypted cluster by modifying the cluste
H. Connect to the VPC where the Amazon Redshift cluster resides from the on-premises network using a VPN.
I. Create a replica of the on-premises HSM in AWS CloudHS
J. Launch a cluster in a VPC with the option to use CloudHSM to store keys.

Answer: B

NEW QUESTION 120
A data analyst is designing a solution to interactively query datasets with SQL using a JDBC connection. Users will join data stored in Amazon S3 in Apache ORC format with data stored in Amazon Elasticsearch Service (Amazon ES) and Amazon Aurora MySQL.
Which solution will provide the MOST up-to-date results?

A. Use AWS Glue jobs to ETL data from Amazon ES and Aurora MySQL to Amazon S3. Query the data with Amazon Athena.
B. Use Amazon DMS to stream data from Amazon ES and Aurora MySQL to Amazon Redshif
C. Query the data with Amazon Redshift.
D. Query all the datasets in place with Apache Spark SQL running on an AWS Glue developer endpoint.
E. Query all the datasets in place with Apache Presto running on Amazon EMR.

Answer: C

NEW QUESTION 123
A transport company wants to track vehicular movements by capturing geolocation records. The records are 10 B in size and up to 10,000 records are captured each second. Data transmission delays of a few minutes are acceptable, considering unreliable network conditions. The transport company decided to use Amazon Kinesis Data Streams to ingest the data. The company is looking for a reliable mechanism to send data to Kinesis Data Streams while maximizing the throughput efficiency of the Kinesis shards.
Which solution will meet the company's requirements?

A. Kinesis Agent
B. Kinesis Producer Library (KPL)
C. Kinesis Data Firehose
D. Kinesis SDK

Answer: B

NEW QUESTION 124
A large energy company is using Amazon QuickSight to build dashboards and report the historical usage data of its customers This data is hosted in Amazon Redshift The reports need access to all the fact tables' billions ot records to create aggregation in real time grouping by multiple dimensions
A data analyst created the dataset in QuickSight by using a SQL query and not SPICE Business users have noted that the response time is not fast enough to meet their needs
Which action would speed up the response time for the reports with the LEAST implementation effort?

A. Use QuickSight to modify the current dataset to use SPICE
B. Use AWS Glue to create an Apache Spark job that joins the fact table with the dimension
C. Load the data into a new table
D. Use Amazon Redshift to create a materialized view that joins the fact table with the dimensions
E. Use Amazon Redshift to create a stored procedure that joins the fact table with the dimensions Load the data into a new table

Answer: A

NEW QUESTION 127
A financial company uses Apache Hive on Amazon EMR for ad-hoc queries. Users are complaining of sluggish performance.
A data analyst notes the following:

⟩ Approximately 90% of queries are submitted 1 hour after the market opens.

Hadoop Distributed File System (HDFS) utilization never exceeds 10%.
Which solution would help address the performance issues?

A. Create instance fleet configurations for core and task node
B. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch CapacityRemainingGB metri
C. Create an automatic scaling policy to scale in the instance fleet based on the CloudWatch CapacityRemainingGB metric.
D. Create instance fleet configurations for core and task node
E. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch YARNMemoryAvailablePercentage metri
F. Create an automatic scaling policy to scale in the instance fleet based on the CloudWatch YARNMemoryAvailablePercentage metric.
G. Create instance group configurations for core and task node
H. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch CapacityRemainingGB metri
I. Create anautomatic scaling policy to scale in the instance groups based on the CloudWatch CapacityRemainingGB metric.
J. Create instance group configurations for core and task node
K. Create an automatic scaling policy to scale out the instance groups based on the Amazon CloudWatch YARNMemoryAvailablePercentage metri
L. Create an automatic scaling policy to scale in the instance groups based on the CloudWatch YARNMemoryAvailablePercentage metric.

Answer: D

Explanation:
https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html

NEW QUESTION 132
A global company has different sub-organizations, and each sub-organization sells its products and services in various countries. The company's senior leadership wants to quickly identify which sub-organization is the strongest performer in each country. All sales data is stored in Amazon S3 in Parquet format.
Which approach can provide the visuals that senior leadership requested with the least amount of effort?

A. Use Amazon QuickSight with Amazon Athena as the data sourc
B. Use heat maps as the visual type.
C. Use Amazon QuickSight with Amazon S3 as the data sourc
D. Use heat maps as the visual type.
E. Use Amazon QuickSight with Amazon Athena as the data sourc
F. Use pivot tables as the visual type.
G. Use Amazon QuickSight with Amazon S3 as the data sourc
H. Use pivot tables as the visual type.

Answer: A

NEW QUESTION 134
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 139
A company needs to store objects containing log data in JSON format. The objects are generated by eight applications running in AWS. Six of the applications generate a total of 500 KiB of data per second, and two of the applications can generate up to 2 MiB of data per second.
A data engineer wants to implement a scalable solution to capture and store usage data in an Amazon S3
bucket. The usage data objects need to be reformatted, converted to .csv format, and then compressed before they are stored in Amazon S3. The company requires the solution to include the least custom code possible and has authorized the data engineer to request a service quota increase if needed.
Which solution meets these requirements?

A. Configure an Amazon Kinesis Data Firehose delivery stream for each applicatio
B. Write AWS Lambda functions to read log data objects from the stream for each applicatio
C. Have the function perform reformatting and .csv conversio
D. Enable compression on all the delivery streams.
E. Configure an Amazon Kinesis data stream with one shard per applicatio
F. Write an AWS Lambda function to read usage data objects from the shard
G. Have the function perform .csv conversion, reformatting, and compression of the dat
H. Have the function store the output in Amazon S3.
I. Configure an Amazon Kinesis data stream for each applicatio
J. Write an AWS Lambda function to read usage data objects from the stream for each applicatio
K. Have the function perform .csv conversion, reformatting, and compression of the dat
L. Have the function store the output in Amazon S3.
M. Store usage data objects in an Amazon DynamoDB tabl
N. Configure a DynamoDB stream to copy the objects to an S3 bucke
O. Configure an AWS Lambda function to be triggered when objects are written to the S3 bucke
P. Have the function convert the objects into .csv format.

Answer: A

NEW QUESTION 142

A company analyzes its data in an Amazon Redshift data warehouse, which currently has a cluster of three dense storage nodes. Due to a recent business acquisition, the company needs to load an additional 4 TB of user data into Amazon Redshift. The engineering team will combine all the user data and apply complex calculations that require I/O intensive resources. The company needs to adjust the cluster's capacity to support the change in analytical and storage requirements.

Which solution meets these requirements?

A. Resize the cluster using elastic resize with dense compute nodes.
B. Resize the cluster using classic resize with dense compute nodes.
C. Resize the cluster using elastic resize with dense storage nodes.
D. Resize the cluster using classic resize with dense storage nodes.

Answer: C

NEW QUESTION 143
......

Relate Links

100% Pass Your DAS-C01 Exam with Exambible Prep Materials

https://www.exambible.com/DAS-C01-exam/

Contact us

We are proud of our high-quality customer service, which serves you around the clock 24/7.

Viste - https://www.exambible.com/