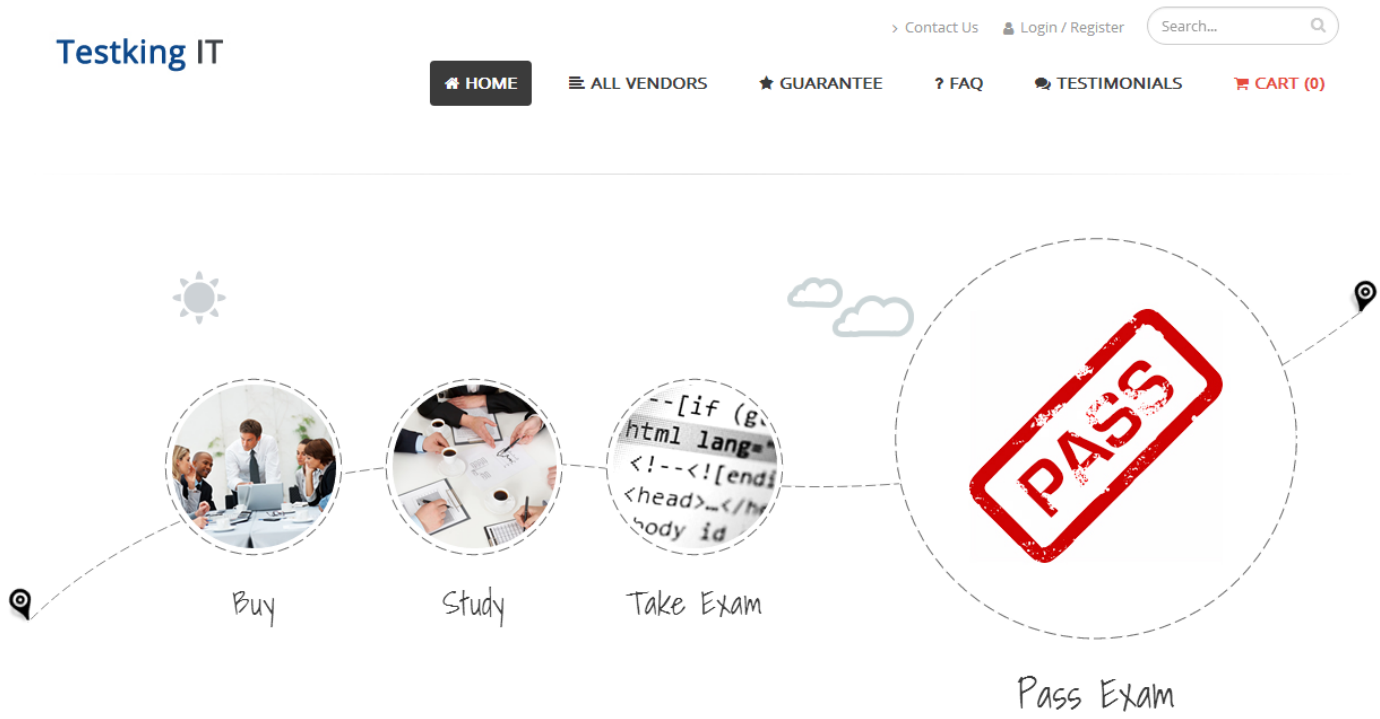
TestkingIT



Try **Desktop Test Engine** before you buy

We're not the only ones **happy** about TestKingsIT Practice Material ...

48236+ customers in 100+ countries use TestKingsIT Test Engine. Meet our customers.

iMessenger VOREED GetCustom JET ORANGE iCompany Paradoxx

http://www.testkingit.com/

Latest practice material - Exam Cram - TestKingIT

Exam : **DAS-C01**

Title : AWS Certified Data Analytics - Specialty (DAS-C01) Exam

Vendor : Amazon

Version : DEMO

NO.1 A financial services company is building a data lake solution on Amazon S3. The company plans to use analytics offerings from AWS to meet user needs for one-time querying and business intelligence reports. A portion of the columns will contain personally identifiable information (PII). Only authorized users should be able to see plaintext PII data.
What is the MOST operationally efficient solution that meets these requirements?

A. Define a bucket policy for each S3 bucket of the data lake to allow access to users who have authorization to see PII data. Catalog the data by using AWS Glue. Create two IAM roles. Attach a permissions policy with access to PII columns to one role. Attach a policy without these permissions to the other role.

B. Register the S3 locations with AWS Lake Formation. Create two IAM roles. Use Lake Formation data permissions to grant Select permissions to all of the columns for one role. Grant Select permissions to only columns that contain non-PII data for the other role.

C. Register the S3 locations with AWS Lake Formation. Create an AWS Glue job to create an E TL workflow that removes the PII columns from the data and creates a separate copy of the data in another data lake S3 bucket. Register the new S3 locations with Lake Formation. Grant users the permissions to each data lake data based on whether the users are authorized to see PII data.

D. Register the S3 locations with AWS Lake Formation. Create two IAM roles. Attach a permissions policy with access to PII columns to one role. Attach a policy without these permissions to the other role. For each downstream analytics service, use its native security functionality and the IAM roles to secure the PII data.

Answer: B

Explanation:
This solution meets the requirements because:
AWS Lake Formation is a fully managed service that allows you to build, secure, and manage data lakes on AWS1. You can use Lake Formation to register your S3 locations as data sources and catalog your data using AWS Glue1.
AWS Lake Formation provides fine-grained data permissions that enable you to control access to your data at the column or row level1. You can use Lake Formation to create two IAM roles and grant them different Select permissions based on the PII status of the columns1.
AWS Lake Formation integrates with various analytics services from AWS, such as Amazon Athena, Amazon Redshift, Amazon EMR, and Amazon QuickSight1. You can use these services to query and visualize your data in S3 using the IAM roles and permissions defined by Lake Formation1.

NO.2 A power utility company is deploying thousands of smart meters to obtain real-time updates about power consumption. The company is using Amazon Kinesis Data Streams to collect the data streams from smart meters. The consumer application uses the Kinesis Client Library (KCL) to retrieve the stream dat a. The company has only one consumer application.
The company observes an average of 1 second of latency from the moment that a record is written to the stream until the record is read by a consumer application. The company must reduce this latency to 500 milliseconds.
Which solution meets these requirements?

A. Use enhanced fan-out in Kinesis Data Streams.

B. Increase the number of shards for the Kinesis data stream.

C. Reduce the propagation delay by overriding the KCL default settings.

D. Develop consumers by using Amazon Kinesis Data Firehose.

Answer: C
Explanation:
The KCL defaults are set to follow the best practice of polling every 1 second. This default results in average propagation delays that are typically below 1 second.

NO.3 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 data. Use Amazon Kinesis Data Streams 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 Amazon QuickSight for the dashboards

C. 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

D. 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 log data to Amazon OpenSearch Service (Amazon Elasticsearch Service) for search, log analytics, and application monitoring Use Amazon QuickSight for the dashboards.

Answer: C

NO.4 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: B,D

NO.5 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

NO.6 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 Classic. Launch an Amazon Redshift cluster in a VPC with the option to use CloudHSM Classic for key management.

B. Create a VPC and establish a VPN connection between the VPC and the on-premises network. Create an HSM connection and client certificate for the on-premises HSM. Launch a cluster in the VPC with the option to use the on-premises HSM to store keys.

C. Create an HSM connection and client certificate for the on-premises HSM. Enable HSM encryption on the existing unencrypted cluster by modifying the cluster. Connect to the VPC where the Amazon Redshift cluster resides from the on-premises network using a VPN.

D. Create a replica of the on-premises HSM in AWS CloudHSM. Launch a cluster in a VPC with the option to use CloudHSM to store keys.

Answer: B

NO.7 A machinery company wants to collect data from sensors. A data analytics specialist needs to implement a solution that aggregates the data in near-real time and saves the data to a persistent data store. The data must be stored in nested JSON format and must be queried from the data store with a latency of single-digit milliseconds.
Which solution will meet these requirements?

A. Use Amazon Kinesis Data Streams to receive the data from the sensors. Use Amazon Kinesis Data Analytics to read the stream, aggregate the data, and send the data to an AWS Lambda function. Configure the Lambda function to store the data in Amazon DynamoDB.

B. Use Amazon Kinesis Data Firehose to receive the data from the sensors. Use Amazon Kinesis Data

Analytics to aggregate the data. Use an AWS Lambda function to read the data from Kinesis Data Analytics and store the data in Amazon S3.

C. Use Amazon Kinesis Data Firehose to receive the data from the sensors. Use an AWS Lambda function to aggregate the data during capture. Store the data from Kinesis Data Firehose in Amazon DynamoDB.

D. Use Amazon Kinesis Data Firehose to receive the data from the sensors. Use an AWS Lambda function to aggregate the data during capture. Store the data in Amazon S3.

Answer: C

Explanation:

This solution meets the requirements because:

Amazon Kinesis Data Firehose is a fully managed service that can capture, transform, and load streaming data into AWS data stores, such as Amazon S3, Amazon Redshift, Amazon Elasticsearch Service, and Amazon DynamoDB1. It can receive data from sensors and other sources and deliver it to a destination with near-real time latency.

AWS Lambda is a serverless compute service that can run code in response to events and automatically manage the underlying compute resources2. It can be used to perform custom transformations on the data during capture by Kinesis Data Firehose3. It can aggregate the data according to the desired logic and output format.

Amazon DynamoDB is a fully managed NoSQL database service that supports key-value and document data models4. It can store nested JSON data as document attributes and provide single-digit millisecond latency for queries. It can be used as a persistent data store for the aggregated sensor data.

NO.8 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

NO.9 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

NO.10 A media company has been performing analytics on log data generated by its applications. There has been a recent increase in the number of concurrent analytics jobs running, and the overall performance of existing jobs is decreasing as the number of new jobs is increasing. The partitioned data is stored in Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA) and the analytic processing is performed on Amazon EMR clusters using the EMR File System (EMRFS) with consistent view enabled. A data analyst has determined that it is taking longer for the EMR task nodes to list objects in Amazon S3.
Which action would MOST likely increase the performance of accessing log data in Amazon S3?
A. Use a hash function to create a random string and add that to the beginning of the object prefixes when storing the log data in Amazon S3.
B. Use a lifecycle policy to change the S3 storage class to S3 Standard for the log data.
C. Increase the read capacity units (RCUs) for the shared Amazon DynamoDB table.
D. Redeploy the EMR clusters that are running slowly to a different Availability Zone.
Answer: C
Explanation:
https://docs.aws.amazon.com/emr/latest/ManagementGuide/emrfs-metadata.html

NO.11 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

NO.12 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

NO.13 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 data Process incoming data from Kinesis Data Streams Use an Amazon Kinesis Data Firehose delivery stream to write results to Amazon Redshift

Answer: D

NO.14 A company wants to optimize the cost of its data and analytics platform. The company is ingesting a number of
.csv 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 format. Use Athena to query the processed dataset. 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.

B. Use an AWS Glue ETL job to partition and convert the data into a row-based data format. Use Athena to query the processed dataset. Configure a lifecycle policy to move the 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.

C. Use an AWS Glue ETL job to compress, partition, and convert the data into a columnar data

format. Use Athena to query the processed dataset. 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 accessed. 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.

D. Use an AWS Glue ETL job to partition and convert the data into a row-based data format. Use Athena to query the processed dataset. 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 accessed. 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

NO.15 An ecommerce company uses Amazon Aurora PostgreSQL to process and store live transactional data and uses Amazon Redshift for its data warehouse solution. A nightly ET L job has been implemented to update the Redshift cluster with new data from the PostgreSQL database. The business has grown rapidly and so has the size and cost of the Redshift cluster. The company's data analytics team needs to create a solution to archive historical data and only keep the most recent 12 months of data in Amazon Redshift to reduce costs. Data analysts should also be able to run analytics queries that effectively combine data from live transactional data in PostgreSQL, current data in Redshift, and archived historical data.

Which combination of tasks will meet these requirements? (Select THREE.)

A. Configure the Amazon Redshift Federated Query feature to query live transactional data in the PostgreSQL database.

B. Configure Amazon Redshift Spectrum to query live transactional data in the PostgreSQL database.

C. Schedule a monthly job to copy data older than 12 months to Amazon S3 by using the UNLOAD command, and then delete that data from the Redshift cluster. Configure Amazon Redshift Spectrum to access historical data in Amazon S3.

D. Schedule a monthly job to copy data older than 12 months to Amazon S3 Glacier Flexible Retrieval by using the UNLOAD command, and then delete that data from the Redshift cluster. Configure Redshift Spectrum to access historical data with S3 Glacier Flexible Retrieval.

E. Create a late-binding view in Amazon Redshift that combines live, current, and historical data from different sources.

F. Create a materialized view in Amazon Redshift that combines live, current, and historical data from different sources.

Answer: A,C,E

NO.16 A company uses Amazon Elasticsearch Service (Amazon ES) to store and analyze its website clickstream dat a. 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/

NO.17 An ecommerce company stores customer purchase data in Amazon RDS. The company wants a solution to store and analyze historical dat a. 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 data. Copy the historical data into Amazon S3. Create an AWS Glue Data Catalog of the data in Amazon S3 and Amazon RDS. Run historical queries using Amazon Athena.
B. Use an ETL tool to incrementally load the most recent 6 months of data into an Amazon Redshift cluster. Run more frequent queries against this cluster. Create a read replica of the RDS database to run queries on the historical data.
C. 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.
D. Incrementally copy data from Amazon RDS to Amazon S3. Load and store the most recent 6 months of data in Amazon Redshift. Configure an Amazon Redshift Spectrum table to connect to all historical data.
Answer: D
Explanation:
Section: (none)

NO.18 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. Create an AMI out of that master node and use that custom AMI to re-create the EMR cluster.
D. Use an Amazon DynamoDB table to store the list of required applications. Trigger an AWS Lambda function with DynamoDB Streams to install the software.
E. Launch an Amazon EC2 instance with Amazon Linux and install the required third-party libraries on the instance. Create an AMI and use that AMI to create the EMR cluster.
Answer: A,E

Explanation:
https://aws.amazon.com/about-aws/whats-new/2017/07/amazon-emr-now-supports-launching-clusters-with-custom-amazon-linux-amis/
https://docs.aws.amazon.com/de_de/emr/latest/ManagementGuide/emr-plan-bootstrap.html

NO.19 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. Assign the CMK an alias. Use the AWS Encryption SDK, providing it with the key alias to encrypt and decrypt the data.
B. Create a customer master key (CMK) in AWS KMS. Assign the CMK an alias. Enable server-side encryption on the Kinesis data stream using the CMK alias as the KMS master key.
C. Create a customer master key (CMK) in AWS KMS. Create an AWS Lambda function to encrypt and decrypt the data. Set the KMS key ID in the function's environment variables.
D. Enable server-side encryption on the Kinesis data stream using the default KMS key for Kinesis Data
Answer: B
Streams.

NO.20 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 CloudWatch. Choose Amazon Elasticsearch Service as the end destination of the weblogs.
B. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis Data Firehose delivery stream to CloudWatch. 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 stream to CloudWatch. Configure Splunk as the end destination of the weblogs.
D. Set up the Amazon CloudWatch agent to stream weblogs to CloudWatch logs and subscribe the Amazon Kinesis Firehose delivery stream to CloudWatch. 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

NO.21 A manufacturing company uses Amazon S3 to store its dat
a. 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. Once Lake Formation has the data, apply permissions on Lake Formation.

B. To create the data catalog, run an AWS Glue crawler on the existing Parquet data. Register the Amazon S3 path and then apply permissions through Lake Formation to provide granular-level security.

C. Install Apache Ranger on an Amazon EC2 instance and integrate with Amazon EMR. Using Ranger policies, create role-based access control for the existing data assets in Amazon S3.

D. Create multiple IAM roles for different users and groups. 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/