

Q1)

Let us suppose Star Bank Limited is a leading Japanese monetary institution that provides a full range of financial products and services to both institutional and individual customers with headquartered in Tokyo. Star Bank is hosting their existing infrastructure on AWS with many internal segments.

The firm is planning to launch a self-data discovery platform running out of AWS on QuickSight. With the help of QuickSight, multiple datasets are created and multiple analyses are generated respectively. The Team has been designated to work on visuals.

Which of the following options must the team choose for a Customized table view of data?

- Pie Chart
- Tabular Reports

Explanation:-Use tabular reports to see a customized table view of your data.

To create a table visual, choose at least one field of any data type. You can add as many columns as you need. Plus, you can add calculated columns.

- Tree Map
 - Heat Maps
-

Q2)

Let us suppose WMS Media is a social media monitoring firm with headquarters in London, England. Such that WMS Media sells three different products - Analytics, Audiences, and Insights. WMS Media Analytics is a "self-serve application" or software as a service, which archives social media data in order to provide companies with information and the means to track specific segments to analyze their brands' online presence. The tool's coverage for WMS Media includes blogs, news sites, forums, videos, reviews, images and social networks like Twitter and Facebook.

Users can search data by using Text and Image Search, and use charting, categorization, sentiment analysis and other features to provide further information and analysis. WMS Media has access to over 80 million sources. WMS Media offers Image and text analysis capabilities to the applications which includes identify objects, people, text, scenes, and activities and also provides highly accurate facial analysis and facial recognition.

Which of the following service can be used to provide this capability?

- Amazon Polly
- Amazon Recognition

Explanation:-Amazon Recognition makes it easy to add image and video analysis to your applications. You just provide an image or video to the Recognition API, and the service can identify objects, people, text, scenes, and activities. It can detect any inappropriate content as well. Amazon Recognition also provides highly accurate facial analysis and facial recognition. You can detect, analyze, and compare faces for a wide variety of use cases, including user verification, cataloging, people counting etc.

- Amazon Comprehend
 - Amazon SageMaker
-

Q3)

Let us suppose AMex is a multimedia company running a messaging app. Such that one of the principal features of QuickDialog is that pictures and messages are usually only available for a short time before they become inaccessible to users. The app has evolved from originally centering on person-to-person photo sharing to present users' where "Stories" of 24 hours of sequential content, along with "Discover", allowing brands show ad-supported short-form media.

AMex use DynamoDB to support the mobile application and S3 to host the images and other documents shared between users. Given that 'PNX' has a large customer base spread across multiple geographic areas where customers needs to update their profile information while using the application. You are required to propose a solution from the following given options that can be easily implemented and provides full consistency.

Choose the most suitable option.

- Use global tables, a fully managed solution across multiple regions, multi-master databases

Explanation:-Amazon DynamoDB global tables provide a fully managed solution for deploying a multi- region, multi-master database, without having to build and maintain your own replication solution. Replication is performed by DynamoDB streams

- Create CustomerProfile table in a region, create replication copies in different AWS regions and enable replication through AWS Kinesis Data Streams
 - Create CustomerProfile table in a region, create replication copies in different AWS regions and enable replication through AWS Data Pipeline
 - Create CustomerProfile table in a region, create replication copies in different AWS regions and enable replication through AWS Kinesis Data Firehose
-

Q4) "GBX-Bank is a privately held Internet retailer of both physical and digital products founded in 2008. The company has more than six-million clients worldwide. GBX-Bank aims to serve as a connection between digital content makers and affiliate dealers, who then promote them to clients. GBX-Bank's technology aids in payments, tax calculations and a variety of customer service tasks. GBX-Bank assists in building perceptibility and revenue making opportunities for entrepreneurs.

GBX-Bank runs multiple java based web applications running on windows based EC2 machines in AWS managed by internal IT Java team, to serve various business functions. GBX-Bank is looking to enable web-site traffic analytics there by understanding user navigational behavior, preferences and other click related info. The amount of data captured per click is in tens of bytes. GBX-Bank has the following objectives in mind for the solution.

GBX-Bank uses KPL to process the data and KCL library to consume the records. Thousands of events are being generated every second and every event is sensitive and equally important and Gluebush.com wants to treat every record as a separate stream."

Which of the following details the implementation guidelines (Choose the most appropriate answer)?

- Each record in a separate Kinesis Data Streams record and make one HTTP request to send it to Kinesis Data Streams
- Each HTTP request carries multiple Kinesis Stream records which is sent to kinesis Data streams
- Batching is implemented as the target implementation
- Batching is not implemented as the target implementation

Explanation:-Batching refers to performing a single action on multiple items instead of repeatedly performing the action on each individual item. In this context, the "item" is a record, and the action is sending it to Kinesis Data Streams. In a non-batchingsituation, you would place each record in a separate Kinesis Data Streams record and make one HTTP request to send it to Kinesis Data Streams. With batching, each HTTP request can carry multiple records instead of just one

Q5)

Let us suppose PNX Ltd. (PNX) is a global environmental solutions company running its operations in Asia Pacific, the Middle East, Africa and America. It maintains more than 10 exploration labs around the world, including a knowledge centre, an "innovative process development centre" in Bangkok, a materials and membrane products development centre as well as advanced machining, Prototyping and industrial design functions. PNX hosts their existing enterprise infrastructure on AWS and runs multiple applications to address the product life cycle management. The datasets are available in Aurora, RDS and S3 in file format. 'PNX' Management team is interested in building analytics around product life cycle and advanced machining, prototyping and other functions.

The IT team proposed Redshift to fulfill the EDW and analytics requirements. They adapt modeling approaches laid by Bill Inmon and Kimball to efficiently design the solution. The team understands that the data loaded into Redshift would be in terabytes and identified multiple massive dimensions, facts, summaries of millions of records and are working on establishing the best practices to address the design concerns.

We are specified that 6 tables they are currently working on -

ORDER_FCT is a Fact Table with billions of rows related to orders

SALES_FCT is a Fact Table with billions of rows related to sales transactions. This table is specifically used to generate reports EOD (End of Day),

EOW(End of Week), and EOM (End of Month) and also sales queries

CUST_DIM is a Dimension table with billions of rows related to customers. It is a TYPE 2 Dimension table

PART_DIM is a part dimension table with billions of records that defines the materials that were ordered

DATE_DIM is a dimension table

SUPPLIER_DIM holds the information about suppliers the 'PNX' work with

SALES_FCT and **DATE_DIM** are joined together frequently since EOD sales reports are generated every day.

Given the tables suggest suitable distribution style from the given options.

- Distribute the **SALES_FCT** with KEY DISTRIBUTION on its own Primary KEY (one of the columns) while **DATE_DIM** is distributed with KEY DISTRIBUTION on Its PRIMARY KEY
- Distribute the **SALES_FCT** with EVEN DISTRIBUTION on its own Primary KEY (one of the columns) while **DATE_DIM** is distributed with EVEN distribution on Its PRIMARY KEY
- Distribute the **SALES_FCT** with KEY DISTRIBUTION on its own Primary KEY (one of the columns) while **DATE_DIM** is distributed with ALL DISTRIBUTION on Its PRIMARY KEY

Explanation:-All distribution makes a copy of the entire table in every compute node. Being billion record tables, this is not a right approach to design. This is the perfect design for **DATE_DIM** table which has very low number and can be distributed to all tables

- Distribute the **SALES_FCT** with ALL DISTRIBUTION on its own Primary KEY (one of the columns) while **DATE_DIM** is distributed with EVEN distribution on Its PRIMARY KEY

Q6)

"Variant.Com (VT) is an online specialty retailer that sells clothing and outdoor refreshment gear for trekking, go camping, boulevard biking, mountain biking, rock hiking, ice mountaineering, skiing, avalanche protection, snowboarding, fly fishing, kayaking, rafting, road and trace running, and many more.

VT runs their entire online infrastructure on java based web applications running on AWS. The VT is capturing click stream data and use custom-build recommendation engine to recommend products which eventually improve sales, understand customer preferences and already using AWS kinesis KPL to collect events and transaction logs and process the stream. The event/log size is around 12 bytes.

The log stream generated into the stream is used for multiple purposes. VT proposes Kinesis Firehose to process the stream and capture information. Without writing applications or consumer code, what purposes can be fulfilled OOTB."

- 1.) Deliver real-time streaming data to Amazon Simple Storage Service (Amazon S3)**
- 2.) Deliver real-time streaming data to DynamoDB to support processing of digital documents**
- 3.) Deliver real-time streaming data to Redshift to support data warehousing and real-time analytics**
- 4.) Ingest data into ES domains to support Enterprise search built on Elasticsearch**

- Only 2, 3 and 4

Explanation:-Amazon Kinesis Data Firehose is a fully managed service for delivering real-time streaming data to destinations such as Amazon Simple Storage Service (Amazon S3), Amazon Redshift, Amazon Elasticsearch Service (Amazon ES), and Splunk. With Kinesis Data Firehose, you don't need to write applications or manage resources. Configure data producers to send data to Kinesis Data Firehose, and it automatically delivers the data to the destination that you specified.

- Only 1 and 4
- Only 1, 3 and 4

Explanation:-Amazon Kinesis Data Firehose is a fully managed service for delivering real-time streaming data to destinations such as Amazon Simple Storage Service (Amazon S3), Amazon Redshift, Amazon Elasticsearch Service (Amazon ES), and Splunk. With Kinesis Data Firehose, you don't need to write applications or manage resources. Configure data producers to send data to Kinesis Data Firehose, and it automatically delivers the data to the destination that you specified.

- Only 3 and 4

Explanation:-Amazon Kinesis Data Firehose is a fully managed service for delivering real-time streaming data to destinations such as Amazon Simple Storage Service (Amazon S3), Amazon Redshift, Amazon Elasticsearch Service (Amazon ES), and Splunk. With Kinesis Data Firehose, you don't need to write applications or manage resources. Configure data producers to send data to Kinesis Data Firehose, and it automatically delivers the data to the destination that you specified.

Q7)

ERB Ltd. is a global environmental solutions company running its operations in Asia Pacific, the Middle East, Africa and the Americas. It's maintaining more than 10 exploration labs around the world, including a knowledge centre, an "innovative process development centre" in Bangkok, a materials and membrane products development centre as well as advanced machining, prototyping and industrial design functions.

ERB Ltd. hosts their existing enterprise infrastructure on AWS and to address the product life cycle management runs multiple applications.

The datasets are available in Aurora, RDS and S3 in file format. ERB Ltd.'s Management team is interested in building analytics around product life cycle and advanced machining, prototyping and other functions.

The IT team proposed Redshift to fulfill the EDW and analytics requirements. They adapt modeling approaches laid by Bill Inmon and Kimball to efficiently design the solution. The team understands that the data loaded into Redshift would be in terabytes and identified multiple massive dimensions, facts, summaries of millions of records and are working on establishing the best practices to address the design concerns.

There are 6 tables that they are currently working on:

ORDER_FCT is a Fact Table with billions of rows related to orders

SALES_FCT is a Fact Table with billions of rows related to sales transactions. This table is specifically used to generate reports EOD (End of Day),

EOW(End of Week), and EOM (End of Month) and also sales queries

?CUST_DIM is a Dimension table with billions of rows related to customers. It is a TYPE 2 Dimension table

PART_DIM is a part dimension table with billions of records that defines the materials that were ordered

DATE_DIM is a dimension table

SUPPLIER_DIM holds the information about suppliers the PNLtd work with

One of the key requirements includes ORDER_FCT and PART_DIM are joined together in most of order related queries. ORDER_FCT has many other dimensions to support analysis.

How would Riley design this distribution?

- Distribute the ORDER_FCT with KEY distribution on its primary KEY (any one of the columns) and PART_DIM with KEY distribution on its PRIMARY KEY

Explanation:-This option is correct - KEY DISTRIBUTION distributes the rows are according to the values in one column. With distribution of data on same key in both the tables, there is no change of redistribution. This is the best approach to design.

- Distribute the ORDER_FCT with ALL distribution on its primary KEY (any one of the columns) and PART_DIM with ALL distribution on its PRIMARY KEY

Explanation:-This option is incorrect - EVEN DISTRIBUTION evenly distributes the rows across the slices in a round-robin fashion, regardless of the values in any particular column. EVEN distribution is appropriate when a table does not participate in joins. Definitely not a right approach.

- Distribute the ORDER_FCT and PART_DIM on same key with KEY distribution

Explanation:-This option is incorrect - KEY DISTRIBUTION distributes the rows are according to the values in one column. Queries initiate lot of redistribution of data of both ORDER_FCT and PART_DIM are not built on same key.

- Distribute the ORDER_FCT with EVEN distribution on its primary KEY (any one of the columns) and PART_DIM with EVEN distribution on its PRIMARY KEY

Explanation:-This option is incorrect - ALL distribution makes a copy of the entire table in every compute node. Being billion record tables, this is not a right approach to design.

Q8)

First Customer Bank, Limited is a leading varied American monetary institution Providing a full range of financial products and services to both institutional and individual customers. It is headquartered in New York. First Customer Bank is hosting their existing infrastructure on AWS.

First Customer bank has many segments internally and they are Planning to launch a self-data discovery platform running out of AWS on QuickSight.

Using QuickSight, multiple datasets are created and multiple analyses are generated respectively.

To track the records of actions taken by a user, role, or an AWS service in Amazon QuickSight the Team is working on enabling auditing.

Also the team need to capture the logs and storage it for long term archival to address compliance. Select the relevant solution of the given -

1.) Amazon QuickSight is integrated with AWS CloudTrail which provides a record of actions taken by a user, role, or an AWS service in Amazon QuickSight

2.) Amazon QuickSight is integrated with AWS CloudWatch which provides a record of actions taken by a user, role, or an AWS service in Amazon QuickSight

3.) when CloudTrail is enabled, you can enable continuous delivery of CloudTrail events to an Amazon S3 bucket, including events for Amazon QuickSight

4.) when CloudWatch is enabled, you can enable continuous delivery of CloudWatch events to an Amazon S3 bucket, including

- Only 1 and 4
- Only 2 and 4
- Only 3 and 4
- Only 1 and 3

Explanation:-Amazon QuickSight is integrated with AWS CloudTrail. This service provides a record of actions taken by a user, role, or an AWS service in Amazon QuickSight. The calls captured include calls from the Amazon QuickSight console. If you create a trail, you can enable continuous delivery of CloudTrail events to an Amazon S3 bucket, including events for Amazon QuickSight. If you don't configure a trail, you can still view the most recent events in the CloudTrail console in Event history.

Q9)

Canada Hotline (CH) is an Canadian department collection chain. There are 63 branches across 32 location in Canada, with clothing, accessories, bags, shoes, jewellery, scents, faces, health and exquisiteness products, home furnishing and decor products.

CH runs their existing operations and analytics infrastructure out of AWS which includes S3, EC2, Auto Scaling, CDN and also Redshift. The Redshift platform is being used for advanced analytics, real time analytics and being actively used for past 2 years. Suddenly performance issues are occurring in the application and administrator being a superuser needs to provide a list of reports in terms of current and historical performance of the cluster. for diagnosis

You are require to select the types of tables/views can help access the performance related info -

- 1.) STL system tables are generated from Amazon Redshift log files to provide a history of the system. They serve logging.**
- 2.) STL tables are actually virtual system tables that contain snapshots of the current system data. They serve snapshots.**
- 3.) STV system tables are generated from Amazon Redshift log files to provide a history of the system. They serve logging.**
- 4.) STV tables are actually virtual system tables that contain snapshots of the current system data. They serve snapshots.**

- Only 2 and 4

Explanation:-This option is correct.

Option 1 is correct - STL system tables are generated from Amazon Redshift log files to provide a history of the system.

Option 4 is correct - STV tables are actually virtual system tables that contain snapshots of the current system data.

- Only 3 and 4

Explanation:-Option 2 is incorrect - STL system tables are generated from Amazon Redshift log files to provide a history of the system.

Option 3 is incorrect - STV tables are actually virtual system tables that contain snapshots of the current system data.

- Only 1 and 3

Explanation:-Option 2 is incorrect - STL system tables are generated from Amazon Redshift log files to provide a history of the system.

Option 3 is incorrect - STV tables are actually virtual system tables that contain snapshots of the current system data.

- Only 1 and 4

Explanation:-This option is correct.

Option 1 is correct - STL system tables are generated from Amazon Redshift log files to provide a history of the system.

Option 4 is correct - STV tables are actually virtual system tables that contain snapshots of the current system data.

Q10) Virtual Trek (VT) is an online specialty retailer that sells clothing and outdoor refreshment gear for trekking, go camping, boulevard biking, mountain biking, rock hiking, ice mountaineering, skiing, avalanche protection, snowboarding, fly fishing, kayaking, rafting, road and trace running, and many more.

VT runs their entire online infrastructure on multiple java based web applications and other web framework applications running on AWS. The VT is capturing clickstream data and use custom-build recommendation engine to recommend products which eventually improve sales, understand customer preferences and already using AWS Kinesis Streams (KDS) to collect events and transaction logs and process the stream. Multiple departments from VT use different streams to address real-time integration and induce analytics into their applications and uses Kinesis as the backbone of real-time data integration across the enterprise.

VT uses a VPC to host all their applications and is looking at integration of kinesis into their web application. To understand the network flow behavior based on every 15 minutes, VT is looking at aggregating data based on the VPC logs for analytics. VPC Flow Logs have a capture window of approximately 10 minutes.

Which of the following queries can be used to capture aggregates based on each client for every 15 mins using Amazon Kinesis Data Analytics?

- Stagger Windows queries

Explanation:-This option is correct - Stagger windows query, A query that aggregates data using keyed time-based windows that open as data arrives. The keys allow for multiple overlapping windows. This is the recommended way to aggregate data using time-based windows.VPC Flow Logs have a capture window of approximately 10 minutes. But they can have a capture window of up to 15 minutes if you're aggregating data on the client. Stagger windows are ideal for aggregating these logs for analysis.

- Sliding windows queries

Explanation:-This option is incorrect - Sliding windows query, A query that aggregates data continuously, using a fixed time or rowcount interval.

- Continuous queries

Explanation:-This option is incorrect - Continuous Query is a query over a stream executes continuously over streaming data. This continuous execution enables scenarios, such as the ability for applications to continuously query a stream and generate alerts.

- Tumbling Windows queries

Explanation:-This option is incorrect - Tumbling Windows query, A query that aggregates data using distinct time-based windows that open and

Q11)

Virtual Trek (VT) is an online specialty retailer that sells clothing and outdoor refreshment gear for trekking, go camping, boulevard biking, mountain biking, rock hiking, ice mountaineering, skiing, avalanche protection, snowboarding, fly fishing, kayaking, rafting, road and trace running, and many more.

VT runs their entire online infrastructure on java based web applications running on AWS. The VT is capturing click stream data and use custom-build recommendation engine to recommend products which eventually improve sales, understand customer preferences and already using AWS kinesis KPL to collect events and transaction logs and process the stream. The event/log size is around 12 bytes.

The log stream generated into the stream is used for multiple purposes. VT proposes Kinesis Firehose to process the stream and capture information. Without writing applications or consumer code, what purposes can be fulfilled OOTB."

- 1.) Deliver real-time streaming data to Amazon Simple Storage Service (Amazon S3)
- 2.) Deliver real-time streaming data to DynamoDB to support processing of digital documents
- 3.) Deliver real-time streaming data to Redshift to support data warehousing and real-time analytics
- 4.) Ingest data into ES domains to support Enterprise search built on Elasticsearch

Only 1 and 4

Explanation:-This option is incorrect.

Only 1, 3 and 4

Explanation:-Amazon Kinesis Data Firehose is a fully managed service for delivering real-time streaming data to destinations such as Amazon Simple Storage Service (Amazon S3), Amazon Redshift, Amazon Elasticsearch Service (Amazon ES), and Splunk. With Kinesis Data Firehose, you don't need to write applications or manage resources. Configure data producers to send data to Kinesis Data Firehose, and it automatically delivers the data to the destination that you specified.

Only 2, 3 and 4

Explanation:-This option is incorrect.

Only 3 and 4

Explanation:-This option is incorrect.

Q12)

ERB is a global environmental solutions company running its operations in Asia Pacific, the Middle East, Africa and the Americas. It maintains more than 10 exploration labs around the world, including a knowledge centre, an "innovative process development centre" in Bangkok, a materials and membrane products development centre as well as advanced machining, Prototyping and industrial design functions. ERB hosts their existing enterprise infrastructure on AWS and runs multiple applications to address the product life cycle management. The datasets are available in Aurora, RDS and S3 in file format. 'ERB' Management team is interested in building analytics around product life cycle and advanced machining, prototyping and other functions.

The IT team proposed Redshift to fulfill the EDW and analytics requirements. They adapt modeling approaches laid by Bill Inmon and Kimball to efficiently design the solution. The team understands that the data loaded into Redshift would be in terabytes and identified multiple massive dimensions, facts, summaries of millions of records and are working on establishing the best practices to address the design concerns.

There are 6 tables given below that they are currently working on -

1. ORDER_FCT is a Fact Table with billions of rows related to orders
2. SALES_FCT is a Fact Table with billions of rows related to sales transactions. This table is specifically used to generate reports EOD (End of Day), EOW(End of Week), and EOM (End of Month) and also sales queries
3. CUST_DIM is a Dimension table with billions of rows related to customers. It is a TYPE 2 Dimension table
4. PART_DIM is a part dimension table with billions of records that defines the materials that were ordered
5. DATE_DIM is a dimension table
6. SUPPLIER_DIM holds the information about suppliers the 'ERB' work with
SALES_FCT and DATE_DIM are joined together frequently since EOD sales reports are generated every day.

For the given tables you are require to suggest the distribution style from the following options.

Distribute the SALES_FCT with KEY DISTRIBUTION on its own Primary KEY (one of the columns) while DATE_DIM is distributed with KEY DISTRIBUTION on Its PRIMARY KEY

Explanation:-This option is incorrect.

Distribute the SALES_FCT with EVEN DISTRIBUTION on its own Primary KEY (one of the columns) while DATE_DIM is distributed with EVEN distribution on Its PRIMARY KEY

Explanation:-This option is incorrect.

Distribute the SALES_FCT with KEY DISTRIBUTION on its own Primary KEY (one of the columns) while DATE_DIM is distributed with ALL DISTRIBUTION on Its PRIMARY KEY

Explanation:-This option is correct - ALL distribution makes a copy of the entire table in every compute node. Being billion record tables, this is not a right approach to design. This is the perfect design for DATE_DIM table which has very low number and can be distributed to all tables

Distribute the SALES_FCT with ALL DISTRIBUTION on its own Primary KEY (one of the columns) while DATE_DIM is distributed with EVEN distribution on Its PRIMARY KEY

Explanation:-This option is incorrect.

Q13)

"First Customer-Bank is a privately held Internet retailer of both physical and digital products founded in 2008. The company has more than six-million clients worldwide. First Customer-Bank aims to serve as a connection between digital content makers and affiliate dealers, who then promote them to clients. GBX-Bank's technology aids in payments, tax calculations and

a variety of customer service tasks. -Bank assists in building perceptibility and revenue making opportunities for entrepreneurs.

First Customer-Bank runs multiple java based web applications running on windows based EC2 machines in AWS managed by internal IT Java team, to serve various business functions. First Customer-Bank is looking to enable web-site traffic analytics there by understanding user navigational behavior, preferences and other click related info. The amount of data captured per click is in tens of bytes. First Customer-Bank has the following objectives in mind for the solution.

First Customer-Bank uses KPL to process the data and KCL library to consume the records. Thousands of events are being generated every second and every event is sensitive and equally important and Gluebush.com wants to treat every record as a separate stream."

Amongst the following given options which of the following detail the implementation guidelines?

1. Each record in a separate Kinesis Data Streams record and make one HTTP request to send it to Kinesis Data Streams.
2. Each HTTP request carries multiple Kinesis Stream records which is sent to kinesis Data streams.
3. Batching is implemented as the target implementation.
4. Batching is not implemented as the target implementation.



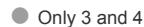
Explanation:-This option is incorrect.



Explanation:-This option is incorrect.



Explanation:-This option is correct - Batching refers to performing a single action on multiple items instead of repeatedly performing the action on each individual item. In this context, the "item" is a record, and the action is sending it to Kinesis Data Streams. In a non-batching situation, you would place each record in a separate Kinesis Data Streams record and make one HTTP request to send it to Kinesis Data Streams. With batching, each HTTP request can carry multiple records instead of just one. Also When batching is not implemented, each record is treated as a separate data stream record. In this context batching is not implemented.



Explanation:-This option is incorrect.

Q14)

Pear Fortunes Ltd is an American-based department store operator with an extensive network of 131 stores, spanning approximately 4.1 million m² of retail space across cities in Chicago, Texas, Philippines, Mexico and Brazil.

Pear has large assets of data around 20 TB's of structured data and 45 TB of unstructured data and is planning to host their data warehouse on AWS and unstructured data storage on S3. Parson IT team is well aware of the scalability, performance of AWS services capabilities.

Pear is currently using running their DWH, on-premises on Teradata and is concerned on the overall costs of the DWH on AWS. They want to initially migrate the platform onto AWS use it for basic analytics, and don't have any performance intensive workloads in place for time being. They have business needs around real-time data integration, data driven analytics as a roadmap of 5 years. Currently the number of users accessing the application would be around 100.

From the above given scenario choose the most suitable option available.



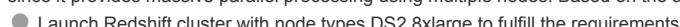
Explanation:-This option is incorrect.



Explanation:-This option is incorrect.



Explanation:-DS2 node types are optimized for large data workloads and use hard disk drive (HDD) storage. DS2.xlarge fulfills the requirements since it provides massive parallel processing using multiple nodes. Based on the amount of data loaded, this is the right option



Explanation:-This option is incorrect.

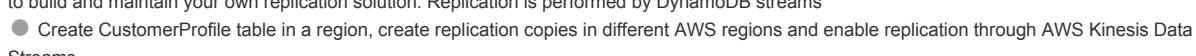
Q15) TWR is a multimedia company running a messaging app. One of the principal features of QuickDialog is that pictures and messages are usually only available for a short time before they become inaccessible to users. The app has evolved from originally centering on person-to-person photo sharing to present users' "Stories" of 24 hours of sequential content, along with "Discover", allowing brands show ad-supported short-form media.

They use DynamoDB to support the mobile application and S3 to host the images and other documents shared between users. 'TWR' has a large customer base spread across multiple geographic areas. Customers need to update their profile information while using the application.

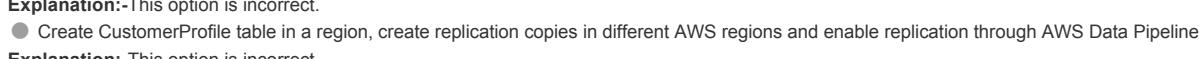
For the above given scenario propose a solution that can be easily implemented and provides full consistency.



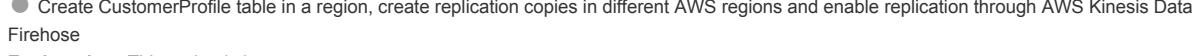
Explanation:-Amazon DynamoDB global tables provide a fully managed solution for deploying a multi-region, multi-master database, without having to build and maintain your own replication solution. Replication is performed by DynamoDB streams



Explanation:-This option is incorrect.



Explanation:-This option is incorrect.



Explanation:-This option is incorrect.

Q16)

As a part of the smart city , Bangalore (Bangalore Authority) , one of the largest cities in southern India is working on capturing massive volumes of video streams 24/7 captured from the large numbers of "Vivotek IB9371 – HT" cameras installed at traffic lights, parking lots, shopping malls, and just about every public venue to help solve traffic problems, help prevent crime, dispatch emergency responders, and much more. Bangalore Authority uses AWS to host their entire infrastructure.

The camera's write stream into Kinesis Video Stream securely and eventually consumed by applications for custom video processing, on-demand video playback and also consumed by AWS Recognition for video analytics. Along with the stream, different modes of streaming metadata are sent along with the stream.

There are 2 scenarios that need to be fulfilled.

·Requirement 1 - Affix metadata on a specific Adhoc basis to fragments in a stream, aka when smart camera detects motion in restricted areas, adds metadata [Motion = true] to the corresponding fragments that contain the motion before sending the fragments to its Kinesis Video Stream

·Requirement 2 - affix metadata to successive, consecutive fragments in a stream based on a continuing need, aka all smart cameras in the city sends the current latitude and longitude coordinates associated with all fragments it sends to its Kinesis Video Stream.

Choose from the following given options that should be used to achieve the above requirement.

1. Requirement 1 can be fulfilled by sending Non-persistent data
2. Requirement 2 can be fulfilled by sending Non-persistent data
3. Requirement 1 can be fulfilled by sending Persistent data
4. Requirement 2 can be fulfilled by sending Persistent data

Only 1 and 3

Explanation:-This option is incorrect.

Only 1 and 4

Explanation:-Option 1 is correct as Requirement 1 can be fulfilled by sending Nonpersistent data.

Also Option D is correct. Requirement 2 can be fulfilled by sending Persistent data.

Only 2 and 4

Explanation:-This option is incorrect.

Only 1 and 2

Explanation:-This option is incorrect.

Q17) Sunshine Media is a social media monitoring company headquartered in London, England. Sunshine Media sells three different products: Analytics, Audiences, and Insights. Sunshine Media Analytics is a "self-serve application" or software as a service, which archives social media data in order to provide companies with information and the means to track specific segments to analyze their brands' online presence.

The tool's coverage includes blogs, news sites, forums, videos, reviews, images and social networks such as Twitter and Facebook. Users can search data by using Text and Image Search, and use charting, categorization, sentiment analysis and other features to provide further information and analysis. Sunshine Media has access to over 80 million sources. Sunshine Media wants to provide Image and text analysis capabilities to the applications which includes identify objects, people, text, scenes, and activities and also provides highly accurate facial analysis and facial recognition.

Which of the given service can provide the given capability?

Amazon Comprehend

Explanation:-This option is incorrect.

Amazon Recognition

Explanation:-Amazon Recognition makes it easy to add image and video analysis to your applications. You just provide an image or video to the Recognition API, and the service can identify objects, people, text, scenes, and activities. It can detect any inappropriate content as well. Amazon Recognition also provides highly accurate facial analysis and facial recognition. You can detect, analyze, and compare faces for a wide variety of use cases, including user verification, cataloging, people counting, and pu

Amazon Polly

Explanation:-This option is incorrect.

Amazon SageMaker

Explanation:-This option is incorrect.

Q18)

First Customer Bank, Limited is a leading Japanese monetary institution that provides a full range of financial products and services to both institutional and individual customers. It is headquartered in Tokyo. First Customer Bank is hosting their existing infrastructure on AWS.

First Customer Bank has many segments internally and they are planning to launch a self-data discovery platform running out of AWS on QuickSight.

Using QuickSight, multiple datasets are created and multiple analyses are generated respectively. The Team is working on visuals.

Which of the following options will you choose in order to have a Customized table view of data?

Pie Chart

Explanation:-This option is incorrect.

Tabular Reports

Explanation:-Use tabular reports to see a customized table view of your data. To create a table visual, choose at least one field of any data type. You can add as many columns as you need. Plus, you can add calculated columns.

Heat Maps

Explanation:-This option is incorrect.

Tree Map

Explanation:-This option is incorrect.

Q19)

A company needs to deploy a data lake solution for their data scientists in which all company data is accessible and stored in a central S3 bucket. The company segregates the data by business unit, using specific prefixes.

Scientists can only access the data from their own business unit. The company needs a single sign-on identity and management solution based on Microsoft Active Directory (AD) to manage access to the data in Amazon S3.

Which of the given methods meets these requirements?

Use Amazon S3 API integration with AD to impersonate the users on access in a transparent manner.

Explanation:-This option is incorrect because there is no feature to integrate Amazon S3 directly with external identity providers

Deploy the AD Synchronization service to create AWS IAM users and groups based on AD information.

Explanation:-This option is incorrect because AD Synchronization will not sync directly with AWS IAM, and custom synchronization would not result in Amazon S3 being able to see group information.

Create bucket policies that only allow access to the authorized prefixes based on the users' group name in Active Directory.

Explanation:-This option is incorrect as bucket policies are linked to IAM principles and cannot recognize AD attributes.

Use AWS IAM Federation functions and specify the associated role based on the users' groups in AD.

Explanation:-This option is correct. Identity Federation allows organizations to associate temporary credentials to users authenticated through an external identity provider such as Microsoft Active Directory (AD). These temporary credentials are linked to AWS IAM roles that grant access to the S3 bucket.

Q20)

An administrator has a 500-GB file in Amazon S3. The administrator runs a nightly COPY command into a 10-node Amazon Redshift cluster.

The administrator wants to prepare the data to optimize performance of the COPY command.

How should the administrator prepare the data using the following given options?

Compress the file using gz compression.

Explanation:-This option is incorrect. As Compressing the files is a recommended practice and will also increase performance, but not to the same extent as loading multiple files in parallel.

Split the file into 500 smaller files

Explanation:-This option is correct. The critical aspect of this question is running the COPY command with the maximum amount of parallelism. This option will have a greater effect because it will allow Amazon Redshift to load multiple files per instance in parallel (COPY can process one file per slice on each node).

Convert the file format to AVRO.

Explanation:-This option is incorrect. This option will load one file per node in parallel, which will increase performance,

Split the file into 10 files of equal size

Explanation:-This option is incorrect.

Q21)

A customer needs to load a 550-GB data file into an Amazon Redshift cluster from Amazon S3, using the COPY command. The input file has both known and unknown issues that will probably cause the load process to fail. The customer needs the most efficient way to detect load errors without performing any cleanup if the load process fails.

Which of the following technique should the customer use in the above given scenario?

Split the input file into 50-GB blocks and load them separately.

Explanation:-This option is incorrect.

Write a script to delete the data from the tables in case of errors.

Explanation:-This option is incorrect.

Use COPY with NOLOAD parameter.

Explanation:-This option is correct. From the AWS Documentation for NOLOAD: NOLOAD checks the integrity of all of the data without loading it into the database. The NOLOAD option displays any errors that would occur if you had attempted to load the data. All other options will require subsequent processing on the cluster which will consume resources

Compress the input file before running COPY.

Explanation:-This option is incorrect.

Q22)

An organization needs a data store to handle the following data types and access patterns :

- Key-value access pattern

- Complex SQL queries and transactions

- Consistent reads

- Fixed schema

Which amongst the following data store must the organization choose?

Amazon S3

Explanation:-This option is incorrect.

Amazon Kinesis

Explanation:-This option is incorrect.

Amazon DynamoDB

Explanation:-This option is incorrect.

Amazon RDS

Explanation:-This option is correct. Amazon RDS handles all these requirements, and although Amazon RDS is not typically thought of as optimized for key-value based access, a schema with a good primary key selection can provide this functionality.

Amazon S3 provides no fixed schema and does not have consistent read after PUT support. Amazon Kinesis supports streaming data that is consistent as of a given sequence number but doesn't provide key/value access.

Finally, although Amazon DynamoDB provides key/value access and consist

Q23)

A web application emits multiple types of events to Amazon Kinesis Streams for operational reporting.

Critical events must be captured immediately before processing can continue, but informational events do not need to delay processing.

What is the most appropriate solution from the following given options to record these different types of events?

Log all events using the Kinesis Producer Library.

Log critical events using the PutRecords API method, and log informational events using the Kinesis Producer Library.

Explanation:-This option is correct. The core of this question is how to send event messages to Kinesis synchronously vs. asynchronously. The critical events must be sent synchronously, and the informational events can be sent asynchronously. The Kinesis Producer Library (KPL) implements an asynchronous send function, so it can be used for the informational messages. PutRecords is a synchronous send function, so it must be used for the critical events.

Log critical events using the Kinesis Producer Library, and log informational events using the PutRecords API method.

Log all events using the PutRecords API method.

Q24)

An administrator decides to use the Amazon Machine Learning service to classify social media posts that mention your company into two categories: posts that require a response and posts that do not.

The training dataset of 10,000 posts contains the details of each post, including the timestamp, author, and full text of the post. You are missing the target labels that are required for training.

Which of the given option will create valid target label data?

Use the sentiment analysis NLP library to determine whether a post requires a response.

Explanation:-This option is incorrect. As this would end up training an ML model using the output from a different machine learning model and therefore would significantly increase the possible error rate.

Ask the social media handling team to review each post and provide the label.

Explanation:-This option is incorrect.

Ask the social media handling team to review each post and provide the label & Use the Amazon Mechanical Turk web service to publish Human Intelligence Tasks that ask Turk workers to label the posts.

Explanation:-This option is correct. You need accurate data to train the service and get accurate results from future data. It is extremely important to have a very low error rate (if any!) in your training set, and therefore human-validated or assured labels are essential.

Using the a priori probability distribution of the two classes, use Monte-Carlo simulation to generate the labels.

Explanation:-This option is incorrect. As this would end up training an ML model using the output from a different machine learning model and therefore would significantly increase the possible error rate.

Q25) A mobile application collects data that must be stored in multiple Availability Zones within five minutes of being captured in the app.

Which of the following architecture securely meets these requirements?

The mobile app should authenticate with an Amazon Cognito identity that is authorized to write to an Amazon Kinesis Firehose with an Amazon S3 destination.

Explanation:-This option is correct because using Amazon Cognito gives you the ability to securely authenticate pools of users on any type of device at scale.

The mobile app should write to an S3 bucket that allows anonymous PutObject calls.

Explanation:-This option is incorrect.

The mobile app should authenticate with an embedded IAM access key that is authorized to write to an Amazon Kinesis Firehose with an Amazon S3 destination.

Explanation:-This option is incorrect.

The mobile app should call a REST-based service that stores data on Amazon EBS. Deploy the service on multiple EC2 instances across two Availability Zones

Explanation:-This option is incorrect. This option uses an anonymous Put, which may allow other apps to write counterfeit data;

Q26)

A company's social media manager requests more staff on the weekends to handle an increase in customer contacts from a particular region.

The company needs a report to visualize the trends on weekends over the past 6 months using QuickSight.

How should the data be represented?

- A map of regions with a heatmap overlay to show the volume of customer contacts
 - A pie chart per region plotting customer contacts per day of week
 - A line graph plotting customer contacts vs. time, with a line for each region
 - A bar graph plotting region vs. volume of social media contacts
-

Q27)

An online photo album app has a key design feature to support multiple screens (e.g., desktop, mobile phone, and tablet) with high-quality displays. Multiple versions of the image must be saved in different resolutions and layouts. The image-processing Java program takes an average of five seconds per upload, depending on the image size and format.

Each image upload captures the following image metadata - user, album, photo label, upload timestamp.

The app should support the following requirements-

Hundreds of user image uploads per second

Maximum image upload size of 10 MB

Maximum image metadata size of 1 KB

Image displayed in optimized resolution in all supported screens no later than one minute after image upload

Which strategy should be used to meet these requirements?

- Write image and metadata to Amazon Kinesis. Use Amazon Elastic MapReduce (EMR) with Spark Streaming to run image processing and save the images output to Amazon S3 and metadata to app repository DB.
 - Upload image with metadata to Amazon S3, use Lambda function to run the image processing and save the images output to Amazon S3 and metadata to the app repository DB.
 - Write image and metadata RDS with BLOB data type. Use AWS Data Pipeline to run the image processing and save the image output to Amazon S3 and metadata to the app repository DB.
 - Write images and metadata to Amazon Kinesis. Use a Kinesis Client Library (KCL) application to run the image processing and save the image output to Amazon S3 and metadata to the app repository DB.
-

Q28)

A customer is collecting clickstream data using Amazon Kinesis and is grouping the events by IP address into 5-minute chunks stored in Amazon S3.

Many analysts in the company use Hive on Amazon EMR to analyze this data. Their queries always reference a single IP address.

Data must be optimized for querying based on IP address using Hive running on Amazon EMR.

What is the most efficient method to query the data with Hive?

- Store the events for an IP address as a single file in Amazon S3 and add metadata with keys: Hive_Partitioned_IPAddress.
 - Store the data in an HBase table with the IP address as the row key
 - Store the Amazon S3 objects with the following naming scheme: bucket_name/source=ip_address/year=yy/month=mm/day=dd/hour=hh/filename.
 - Store an index of the files by IP address in the Amazon DynamoDB metadata store for EMRFS.
-

Q29)

A travel website needs to present a graphical quantitative summary of its daily bookings to website visitors for marketing purposes.

The website has millions of visitors per day, but wants to control costs by implementing the least-expensive solution for this visualization.

What is the most cost-effective solution?

1. Generate a static graph with a transient EMR cluster daily, and store it in Amazon S3.
2. Generate a graph using MicroStrategy backed by a transient EMR cluster
3. Implement a Jupiter front-end provided by a continuously running EMR cluster leveraging spot instances for task nodes.
4. Implement a Zeppelin application that runs on a long-running EMR cluster

- Only 2 and 4
 - Only 1 and 4
 - Only 2 and 3
 - Only 1 and 2
-

Q30)

An online gaming company uses DynamoDB to store user activity logs and is experiencing throttled writes on the company's DynamoDB table.

The company is NOT consuming close to the provisioned capacity. The table contains a large number of items and is partitioned on user and sorted by date. The table is 200GB and is currently provisioned at 10K WCU and 20K RCU.

Which two additional pieces of information are required to determine the cause of the throttling? (Choose two options)

- The structure of any LSIs that have been defined on the table
 - Application-level metrics showing the average item size and peak update rates for each attribute
 - CloudWatch data showing consumed and provisioned write capacity when writes are being throttled
 - The structure of any GSIs that have been defined on the table
-

Q31) A data engineer needs to collect data from multiple Amazon Redshift clusters within a business and consolidate the data into a single central data warehouse. Data must be encrypted at all times while at rest or in flight.

What is the most scalable way to build this data collection process?

- Run an ETL process that connects to the source clusters using SSL to issue a SELECT query for new data, and then write to the target data warehouse using an INSERT command over another SSL secured connection.

Explanation:-This option is incorrect.

- Use AWS KMS data key to run an UNLOAD ENCRYPTED command that stores the data in an unencrypted S3 bucket; run a COPY command to move the data into the target cluster

Explanation:-The most scalable solutions are the UNLOAD/COPY solutions because they will work in parallel, which eliminates two answers. The UNLOAD/COPY solutions meets the encryption requirements, the UNLOAD ENCRYPTED command automatically stores the data encrypted using client side encryption and uses HTTPS to encrypt the data during the transfer to S3

- Connect to the source cluster over an SSL client connection, and write data records to Amazon Kinesis Firehose to load into your target data warehouse.

Explanation:-This option is incorrect.

- Run an UNLOAD command that stores the data in an S3 bucket encrypted with an AWS KMS data key; run a COPY command to move the data into the target cluster

Explanation:-This option is incorrect.

Q32)

A company logs data from its application in large files and runs regular analytics of these logs to support internal reporting for three months after the logs are generated.

After three months, the logs are infrequently accessed for up to a year. The company also has a regulatory control requirement to store application logs for seven years.

Which of the following course of action should the company take to achieve these requirements in the most cost-efficient way?

- Store the files in S3 Glacier with a Deny Delete vault lock policy for archives less than seven years old and a vault access policy that restricts read access to the analytics IAM group and write access to the log writer service role.

Explanation:-This option is incorrect.

- Store the files in S3 Standard with a lifecycle policy to transition the storage class to Standard - IA after three months. After a year, transition the files to Glacier and add a Deny Delete vault lock policy for archives less than seven years old.

Explanation:-This option is incorrect.

- Store the files in S3 Standard with lifecycle policies to transition the storage class to Standard – IA after three months and delete them after a year. Simultaneously store the files in Amazon Glacier with a Deny Delete vault lock policy for archives less than seven years old.

Explanation:-There are two aspects to this question: setting up a lifecycle policy to ensure that objects are stored in the most cost-effective storage, and ensuring that the regulatory control is met. The lifecycle policy will store the objects on S3 Standard during the three months of active use, and then move the objects to S3 Standard – IA when access will be infrequent.

- Store the files in S3 Standard with a lifecycle policy to remove them after a year. Simultaneously store the files in Amazon S3 Glacier with a Deny Delete vault lock policy for archives less than seven years old.

Explanation:-This option is incorrect.

Q33)

A data engineer needs to architect a data warehouse for an online retail company to store historic purchases. The data engineer needs to use Amazon Redshift. To comply with PCI:DSS and meet corporate data protection standards, the data engineer must ensure that data is encrypted at rest and that the keys are managed by a corporate on-premises HSM.

Which of the following approach meets these requirements in the most cost-effective manner?

- Create a VPC, and then establish a VPN connection between the VPC and the on-premises network. Launch the Amazon Redshift cluster in the VPC, and configure it to use your corporate HSM.

Explanation:-This option is correct. Amazon Redshift can use an on-premises HSM for key management over the VPN, which ensures that the encryption keys are locally managed.

- Use the AWS CloudHSM service to establish a trust relationship between the CloudHSM and the corporate HSM over a Direct Connect connection. Configure Amazon Redshift to use the CloudHSM device.

Explanation:-This option is incorrect, as CloudHSM can cluster to an on-premises HSM.

But then key management could be performed on either the on-premises HSM or CloudHSM, and that doesn't meet the design goal.

- Use AWS Import/Export to import the corporate HSM device into the AWS Region where the Amazon Redshift cluster will launch, and configure Redshift to use the imported HSM.

Explanation:-This option is incorrect because you cannot put hardware into an AWS Region

- Configure the AWS Key Management Service to point to the corporate HSM device, and then launch the Amazon Redshift cluster with the KMS managing the encryption keys.

Explanation:-This option is incorrect as it describe a valid feature of KMS and violates the requirement for the corporate HSM to manage the keys requirement, even if it were possible.

Q34) A new algorithm has been written in Python to identify SPAM emails. The algorithm analyzes the free text contained within a sample set of 1 million emails stored on Amazon S3. The algorithm must be scaled across a production dataset of 5PB, which also resides in Amazon S3 storage.

Which of the following AWS service strategy is best for this use case?

- Copy the data into Amazon ElastiCache to perform text analysis on the in-memory data and export the results of the model into Amazon Machine Learning

Explanation:-This option is incorrect.

- Use Amazon Elasticsearch Service to store the text and then use the Python Elasticsearch Client to run analysis against the text index

Explanation:-This option is correct.

- Initial a Python job from AWS Data Pipeline to run directly against the Amazon S3 text files

Explanation:-This option is incorrect.

- Use Amazon EMR to parallelize the text analysis tasks across the cluster using a streaming program step

Explanation:-This option is incorrect.

Q35)

A Redshift data warehouse has different user teams that need to query the same table with very different query types.

These user teams are experiencing poor performance.

Which of the following action improves performance for the user teams in this situation ?

- Add interleaved sort keys per team

Explanation:-This option is incorrect.

- Create custom table views

Explanation:-This option is incorrect.

- Add Support for workload management queue hopping

Explanation:-This option is correct.

- Maintain team-specific copies of the table

Explanation:-This option is incorrect.

Q36)

A data engineer is about to perform a major upgrade to the DDL contained within an Amazon Redshift cluster to support a new data warehouse application. The upgrade scripts will include user permission updates, view and table structure changes as well as additional loading and data manipulation tasks. The data engineer must be able to restore the database to its existing state in the event of issues.

Which action should be taken prior to performing the upgrade task ?

- Make a copy of the automated snapshot on the Amazon Redshift cluster

Explanation:-This option is incorrect.

- Create a manual snapshot of the Amazon Redshift cluster

Explanation:-This option is incorrect.

- Run an UNLOAD command for all data in the warehouse and save it to S3

Explanation:-This option is incorrect.

- Call the waitForSnapshotAvailable command from either the AWS CLI or an AWS SDK

Explanation:-This option is correct.

Q37) What steps are required to grant S3 access to a Redshift cluster?

- Create an S3 bucket with anonymous access allowed and configure Redshift to use this bucket.

- Create an S3 bucket with permission to allow the Redshift service access.

- Create an IAM Service account for Redshift, set a secure password for the account and grant API access to it, and configure Redshift to use this account.

- Create an IAM role with an S3 access policy attached and configure Redshift to assume this role.

Explanation:-IAM roles allow you to create an entity that can have access policy attached and that can be assumed by users or services. Redshift can be configured to assume a role that you assign. In this case, you would create a role that's assumable by Redshift and that allows access to the S3 bucket containing the data.

Q38) What are the two states of data encryption? (Choose two.)

- Encryption at sea

- Encryption in transit

Explanation:-Data may be encrypted either at rest (stored) or in transit.

- Encryption in flight

- Encryption at rest

Explanation:-Data may be encrypted either at rest (stored) or in transit.

Q39)

You are developing a smart speaker that listens to conversations in people's homes at all times. All conversations must be transcribed in real time so that items can be added to users' shopping carts.

Bandwidth must be kept to a minimum to ensure that no data is lost. Which protocol should you use?

- WOPR

- HTTPS

- MQTT

Explanation:-Only MQTT is designed to minimize network bandwidth and provide reliable transport in unreliable networks.

Q40)

A developer on your team is working on a Lambda function that reads data from an RDS database and creates graphs that are stored in S3. The function works in the developer's environment, but when it's moved into production, it times out while trying to contact the database.

What is the most likely cause?

- The Lambda function is not associated with the VPC that contains the RDS instance.

Explanation:-Lambda functions must be associated with a VPC to access resources within the VPC, such as RDS instances.

- The Lambda function is written in Ruby, which is not supported.
 - The Lambda function is not configured to use TLS encryption, which is required by RDS.
 - The RDS instance is in read-only mode.
-

Q41)

The principle of least privilege states that individuals should have access to only a minimal set of resources needed to perform their legitimate duties.

How can this be implemented with AWS IAM?

- Create an IAM user for anyone who has legitimate duties that require access to AWS resources. Assign IAM roles to allow access only to a minimum set of resources each user requires to perform his or her duties.

Explanation:-Each user should have his or her own AWS IAM account, and the account should have access to the minimum set of privileges required for the user to perform his or her duties. In addition, audits should be performed regularly to ensure that privileges granted are still required.

- Create one IAM user for each department supervisor and have that person share the password with his or her trustworthy employees.
 - Give everyone the root password but make them swear an oath not to do anything other than their legitimate duties.
 - Create an IAM user for each employee but disable API access so that users are forced to use the console for all modifications.
-

Q42) What type of encryption is typically used to secure data at rest, such as in S3?

- Symmetric key encryption

Explanation:-Only symmetric key and asymmetric key are valid encryption types. Asymmetric key encryption is typically used to pass data between parties (such as between a website and a web browser), and symmetric key encryption is typically used to secure data against unwanted disclosure. S3 uses symmetric key encryption to secure data at rest.

- Multikey encryption
 - Asymmetric key encryption
 - Ultimate encryption
-

Q43)

You are collecting logs from 1,000 workstations across 18 international locations using fluentd. You need to retain these logs unaltered for 12 months to meet regulatory requirements, but you would also like to make these logs available for rapid search.

Which destination should set in fluentd?

- Redshift
- Elasticsearch
- DynamoDB
- S3

Explanation:-Only S3 meets the durability requirement to keep the log files unaltered for the specified time. In addition, S3 is the most cost-effective solution. From S3, the files could be loaded into Elasticsearch or another search engine to meet the search requirement.

Q44)

Let us suppose an Amazon Kinesis stream needs to be encrypted.

Which approach should be used to accomplish this task from the following options?

- Perform a client side encryption of the data before it enters the Amazon Kinesis stream on the producer

Explanation:-This option is correct.

- Use a partition key to segment the data by MD4 hash function, which makes it undecipherable while in transit

Explanation:-This option is incorrect.

- Use a shard to segment the data, which has built in functionality to make it indecipherable while in transit

Explanation:-This option is incorrect.

- Perform a client side encryption of the data before it enters the Amazon Kinesis stream on the consumer

Explanation:-This option is incorrect.

Q45)

Let us suppose a customer needs to determine the optimal distribution strategy for the ORDERS fact table in Redshift schema.

The ORDERS table has foreign key relationships with multiple dimension tables in this schema.

How should the company determine the most appropriate distribution key for the ORDERS table?

- Identify the largest dimension table and designate the key of this dimension table as the distribution key of the ORDERS table

Explanation:-This option is incorrect.

- Identify the largest and most frequently joined dimension table and ensure that it and the ORDERS table both have EVEN distribution

Explanation:-This option is incorrect.

- Identify the smallest dimension table and designate the key of this dimension table as the distribution key of the ORDERS table

Explanation:-This option is incorrect.

- Identify the largest and the most frequently joined dimension table and designate the key of this dimension table as the distribution key of the ORDERS table

Explanation:-This option is correct.

Q46) A company has several teams of analytics. Each team of analysts has their own cluster. The teams need to run SQL queries using Hive, Spark-SQL and Presto with Amazon EMR. The company needs to enable a centralized metadata layer to expose the Amazon S3 objects as tables to the analysts.

Which of the following approach meets the requirement for a centralized metadata layer?

- EMRFS consistent view with a common Amazon DynamoDB table

Explanation:-This option is correct.

- Bootstrap action to change the Hive Metastore to an Amazon RDS database

Explanation:-This option is incorrect.

- naming scheme support with automatic partition discovery from Amazon S3

Explanation:-This option is incorrect.

- s3distcp with the outputManifest option to generate RDS DDL

Explanation:-This option is incorrect.

Q47) Is there a limit to the number of groups you can have?

- Yes unless special permission granted

Explanation:-This option is incorrect.

- Yes for all users

Explanation:-This option is correct.

- Yes for all users except root

Explanation:-This option is incorrect.

- No

Explanation:-This option is incorrect.

Q48) Amazon S3 doesn't automatically give a user who creates _____ permission to perform other actions on that bucket or object.

- a file

Explanation:-This option is incorrect.

- a bucket or file

Explanation:-This option is incorrect.

- a bucket or object

Explanation:-This option is correct.

- a object or file

Explanation:-This option is incorrect.

Q49) What's an ECU?

- Elastic Compute Unit.

Explanation:-This option is correct.

- Extended Cluster User.

Explanation:-This option is incorrect.

- Elastic Computer Usage.

Explanation:-This option is incorrect.

- None of these.

Explanation:-This option is incorrect.

Q50)

A company needs to monitor the read and write IOPs metrics for their AWS MySQL RDS instances and send real-time alerts to their operations team.

Which AWS services can accomplish this?

1. Amazon Simple Queue Service

2. Amazon Simple Notification Service

3. Amazon CloudWatch

4. Amazon Simple Email Service

- Only 1 and 3

Explanation:-This option is incorrect.

- Only 2 and 3

Explanation:-This option is correct.

- Only 1 and 2

Explanation:-This option is incorrect.

- Only 2 and 4

Explanation:-This option is incorrect.

Q51)

An AWS customer is deploying a web application that is composed of a front-end running on Amazon EC2 and of confidential data that is stored on Amazon S3. The customer security policy that all access operations to this sensitive data must be authenticated and authorized by a centralized access management system that is operated by a separate security team.

In addition, the web application team that owns and administers the EC2 web front-end instances is prohibited from having any ability to access the data that circumvents this centralized access management system.

Which of the following configurations will support these requirements?

- Encrypt the data on Amazon S3 using a CloudHSM that is operated by the separate security team. Configure the web application to integrate with the CloudHSM for decrypting approved data access operations for trusted end-users.

Explanation:-This option is incorrect.

- Configure the web application to authenticate end-users against the centralized access management system. Have the web application provision trusted users STS tokens entitling the download of approved data directly from Amazon S3

Explanation:-This option is correct.

- Have the separate security team create and IAM role that is entitled to access the data on Amazon S3. Have the web application team provision their instances with this role while denying their IAM users access to the data on Amazon S3

Explanation:-This option is incorrect.

- Configure the web application to authenticate end-users against the centralized access management system using SAML. Have the end-users authenticate to IAM using their SAML token and download the approved data directly from S3.

Explanation:-This option is incorrect.

Q52)

Typically, you want your application to check whether a request generated an error before you spend any time processing results.

The easiest way to find out if an error occurred is to look for an _____ node in the response from the Amazon RDS API.

- Warning

Explanation:-This option is incorrect.

- Incorrect

Explanation:-This option is incorrect.

- Error

Explanation:-This option is correct.

- None of these

Explanation:-This option is incorrect.

Q53) 8 HTTP Query-based requests are HTTP requests that use the HTTP verb GET or POST and a Query parameter named _____.

- Retrieve

- Action

- Reset

- Value

Q54) In the 'Detailed' monitoring data available for your Amazon EBS volumes, Provisioned IOPS volumes automatically send _____ minute metrics to Amazon CloudWatch.

- 5

- 3

- 2

- 1

Q55)

You are working with customer who has 10 TB of archival data that they want to migrate to Amazon Glacier.

The customer has a 1Mbps connection to the Internet.

Which service or feature provide the fastest method of getting the data into Amazon Glacier?

- Amazon Glacier multipart upload

- AWS Storage Gateway

- AWS Import/Export

- VM Import/Export

Q56)

You need to configure an Amazon S3 bucket to serve static assets for your public-facing web application.

Which of the following methods ensure that all objects uploaded to the bucket are set to public read?

1. Configure the bucket ACL to sell all objects to public read

2. Amazon S3 objects default to public read, so no action is needed

3. Use AWS identity and access Management roles to set the bucket to public read

4. Configure the bucket policy to set all objects to public read

- Only 2 and 4
 - Only 1 and 4
 - Only 2 and 3
 - Only 1 and 2
-

Q57) What does RDS stands for in Amazon RDS stand for?

- Regional Database Service.
 - Regional Data Server.
 - Relational Database Service.
 - None of these
-

Q58) Without _____, you must either create multiple AWS accounts-each with its own billing and subscriptions to AWS products-or your employees must share the security credentials of a single AWS account.

- Amazon IAM
 - Amazon RDS
 - Amazon Glacier
 - Amazon EMR
-

Q59) If I modify a DB Instance or the DB parameter group associated with the instance, should I reboot the instance for the changes to take effect?

- Correct
 - Incorrect
-