



[Home](https://www.whizlabs.com/learn) (<https://www.whizlabs.com/learn>) > [My Courses](https://www.whizlabs.com/learn/my-courses) (<https://www.whizlabs.com/learn/my-courses>)
> [AWS Certified Big Data Specialty](https://www.whizlabs.com/learn/course/aws-bds-practice-tests#section-1) (<https://www.whizlabs.com/learn/course/aws-bds-practice-tests#section-1>)
> [Practice Test 2](https://www.whizlabs.com/learn/course/aws-bds-practice-tests/quiz/14898) (<https://www.whizlabs.com/learn/course/aws-bds-practice-tests/quiz/14898>) > **Report**

PRACTICE TEST 2

Attempt	1	Completed on	Sunday , 03 February 2019 , 11:39 PM
Marks Obtained	0 / 70	Time Taken	00 H 00 M 14 S
Your score is	0.0%	Result	Fail

Domains / Topics wise Quiz Performance Report

S.No.	Topic	Total Questions	Correct	Incorrect	Unattempted
1	Analysis	15	0	0	15
2	Storage	14	0	0	14
3	Collection	11	0	0	11
4	Processing	13	0	0	13
5	Visualization	10	0	1	9
6	Data Security	7	0	0	7

70 Questions	0 Correct	1 Incorrect	69 Unattempted	Show Answers	All	▼
------------------------	---------------------	-----------------------	--------------------------	---------------------	------------	---

QUESTION 1

UNATTEMPTED

ANALYSIS

Allianz Financial Services (AFS) is a banking group offering end-to-end banking and financial solutions in South East Asia through its consumer banking, business banking, Islamic banking, investment finance and stock broking businesses as well as unit trust and asset administration, having served the financial community over the past five decades. AFS being one the largest banks in the region is planning to improve its segment business by launching a campaign to identify potential customers for various new products launched based on their past behavior? AFS is looking for both batch and real-time predictive analytics.

Select 2 options.

- ☐ A. AWS recommends using Machine learning web service for batch and SPARK on EMR to address real-time to address predictive analytics both using same libraries
- ☐ B. AWS offers binary classification insights to address this problem with accuracy measured using Area Under the (Receiver Operating Characteristic) Curve (AUC) ✓
- ☐ C. AWS offers multiclass classification insights to address this problem with predictive accuracy being measured using macro-average F1 score
- ☐ D. AWS offers Regression model insights to address this problem with predictive accuracy being measured using Baseline RMSE
- ☐ E. AWS recommends using Machine learning web service to address both batch and real-time predictive analytics ✓

Explanation :

Answer : B and E

A. No. Machine Learning web service offers 2 types of predictions

- Batch Predictions asynchronously generate predictions for multiple input data observations
- Real-time Predictions synchronously generate predictions for individual data observations

<https://docs.aws.amazon.com/machine-learning/latest/dg/amazon-machine-learning-key-concepts.html> (<https://docs.aws.amazon.com/machine-learning/latest/dg/amazon-machine-learning-key-concepts.html>)

B. Yes. Problem being a binary classification whether the customer will subscribe to new product or not ML models for binary classification predict a binary outcome. To train binary classification models, Amazon

ML uses the industry-standard learning algorithm known as logistic regression. Amazon ML provides an industry-standard accuracy metric for binary classification models called Area Under the (Receiver Operating Characteristic) Curve (AUC)

<https://docs.aws.amazon.com/machine-learning/latest/dg/binary-model-insights.html> (<https://docs.aws.amazon.com/machine-learning/latest/dg/binary-model-insights.html>)

C. No. ML models for multiclass classification problems allow you to generate predictions for multiple classes.

In Amazon ML, the macro-average F1 score is used to evaluate the predictive accuracy of a multiclass metric.

<https://docs.aws.amazon.com/machine-learning/latest/dg/multiclass-model-insights.html> (<https://docs.aws.amazon.com/machine-learning/latest/dg/multiclass-model-insights.html>)

D. No. ML models for regression problems predict a numeric value. Amazon ML uses the industry standard root

mean square error (RMSE) metric.

<https://docs.aws.amazon.com/machine-learning/latest/dg/regression-model-insights.html> (<https://docs.aws.amazon.com/machine-learning/latest/dg/regression-model-insights.html>)

E. Yes. Machine Learning offers 2 types of predictions

- Batch Predictions asynchronously generate predictions for multiple input data observations

- Real-time Predictions synchronously generate predictions for individual data observations

<https://docs.aws.amazon.com/machine-learning/latest/dg/amazon-machine-learning-key-concepts.html> (<https://docs.aws.amazon.com/machine-learning/latest/dg/amazon-machine-learning-key-concepts.html>)

Ask our Experts



QUESTION 2

UNATTEMPTED

STORAGE

KindleYou is a location-based social search mobile app that allows users to like or dislike other users, and allows users to chat if both parties liked each other in the app. It has more than 1 billion customers across the world.

They use DynamoDB to support the mobile application and S3 to host the images and other documents shared between users.

KindleYou has collection named Interactions bulk-loaded with data infrequently; it doesn't incur very much write activity but always experience a high degree of read activity, which varies over time. KindleYou is looking at implementing Auto Scaling in DynamoDB. please advise on the critical steps! Select 3 options.

- ☐ A. Publishes consumed capacity metrics to AWS CloudTrail
- ☐ B. Publishes consumed capacity metrics to AWS CloudWatch ✓
- ☐ C. If the table's consumed capacity exceeds your target utilization (or falls below the target) for a specific length of time, Amazon CloudTrail triggers an alarm
- ☐ D. If the table's consumed capacity exceeds your target utilization (or falls below the target) for a specific length of time, AWS CloudWatch triggers an alarm using SNS ✓
- ☐ E. If the table's consumed capacity exceeds your target utilization (or falls below the target) for a specific length of time, AWS CloudWatch triggers an alarm using SQS
- ☐ F. The alarm invokes Application Auto Scaling to evaluate your scaling policy and process request using UpdateTable ✓

Explanation :

Answer: B, D,F

A. No. the metrics are published to CloudWatch

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWorks> (<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWorks>)

B. Yes. the metrics are published to CloudWatch

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWorks> (<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWorks>)

C. No cloudwatch triggers the alarm

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWorks>

(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWorks>)
D. Yes. the alarm on the AWS Management Console and receive notifications using Amazon Simple Notification Service
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWorks>)
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWorks>)
E. No. the user is notified using SNS
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWorks>)
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWorks>)

Ask our Experts



QUESTION 3

UNATTEMPTED

COLLECTION

S5Mart Corporation is an big box department store chain headquartered in United States. S5Mart operated 1,273 stores including 105 Super S5mart Center locations around the world.

S5Mart is adopting IoT solutions across a number of applications that are improving store operations, reducing theft, increasing purchases through cross selling, enabling precise inventory management, and most importantly enhancing the consumer's shopping experience. S5Mart is hosting their IOT solution on AWS IOT Core.

S5Mart is working defining standard approach to address provisioning of the devices which include creation and registration of devices. Please advise on the key steps in provisioning. Select 3 options.

- ☐ A. Enable a device with an existing certificate or have AWS IoT create and register one for you ✓
- ☐ B. Define device shadow service to get and set the state of a device over MQTT or HTTP
- ☐ C. Define AWS IoT rule actions to specify what to do when a rule is triggered and write data to other AWS services
- ☐ D. Assign a policy attached to the certificate and Define unique identifier for the thing (device). ✓
- ☐ E. Define set of attributes for the thing, including existing thing types and groups ✓

Explanation :

Answer : A,D,E

A. Yes. Even though the solution can be implemented by the combo, IOT rules provide the solution OOTB.

(<https://docs.aws.amazon.com/iot/latest/developerguide/iot-provision.html>)

(<https://docs.aws.amazon.com/iot/latest/developerguide/iot-provision.html>)

B. No. A device's shadow is a JSON document that is used to store and retrieve current state information for a device. The Device Shadow service maintains a shadow for each device you connect to AWS IoT

<https://docs.aws.amazon.com/iot/latest/developerguide/iot-device-shadows.html>
(<https://docs.aws.amazon.com/iot/latest/developerguide/iot-device-shadows.html>)

C. No. AWS IoT rule actions are used to specify what to do when a rule is triggered. You can define actions to write data to a DynamoDB database or a Kinesis Stream or to invoke a Lambda function, and more. This is not required for provisioning.

<https://docs.aws.amazon.com/iot/latest/developerguide/iot-rule-actions.html>
(<https://docs.aws.amazon.com/iot/latest/developerguide/iot-rule-actions.html>)

D. Yes. A policy attached to the certificate and a unique identifier for the thing (device).

<https://docs.aws.amazon.com/iot/latest/developerguide/iot-provision.html>
(<https://docs.aws.amazon.com/iot/latest/developerguide/iot-provision.html>)

E. Yes. A set of attributes for the thing, including existing thing types and groups.

<https://docs.aws.amazon.com/iot/latest/developerguide/iot-provision.html>
(<https://docs.aws.amazon.com/iot/latest/developerguide/iot-provision.html>)

Ask our Experts



QUESTION 4

UNATTEMPTED

COLLECTION

EduProvy is an Indian higher learning enrollment platform based on e-learning. Indian and international course providers offer programs such as MBA, Executive MBA, MCA, MSc IT, MA, MCom, BA, BCom BBA, BCA and BSc IT on the company platform. The courses are available as distance learning and online courses. It also features certificate courses in areas of retail, economics, telecom, digital marketing, programming languages, project management, and IT.

EduProvy already hosts their videos on S3 on AWS collected from various professors and lecturers and enables the content using content management application running out of EC2. The videos are segregated into different topics into different units and chapters and can be accessed by students. There is a requirement to convert large, high-quality digital media files into formats that users can play back on mobile devices, tablets, web browsers, and connected televisions.

EduProvy needs an end to end solution to process the master videos safely and secured and converted formats at a secured store easily accessible across the world and monitor end to end video conversion workflows. Master videos are seldom accessed

Build a cost-effective solution to fulfill the requirement. Select 5 options.

- ☐ A. use Amazon Glacier to store master content ✓
- ☐ B. use Amazon S3 to store master content

- ☐ C. Use Elastic Transcoder to perform video conversion to different formats ✓
- ☐ D. Use S3 to store videos of converted formats of master content ✓
- ☐ E. Use Amazon Glacier to store videos of converted formats of master content
- ☐ F. Use CloudFront to stream the renditions across the world ✓
- ☐ G. Use Kinesis Video Streams to perform video conversion and stream the renditions across the world
- ☐ H. Monitor health of the transcoding workflows using CloudWatch ✓
- ☐ I. Monitor health of the transcoding workflows using CloudTrail

Explanation :

Answer: A, C, D, F, H

Amazon Elastic Transcoder provides an important media building block for creating end-to-end media solutions on AWS. For example, you can use Amazon Glacier to store master content, Amazon Elastic Transcoder to transcode masters to renditions for distribution stored in Amazon S3, stream these renditions at scale over the Internet using Amazon CloudFront, and monitor the health of your transcoding workflow using CloudWatch.

<https://aws.amazon.com/elastictranscoder/> (<https://aws.amazon.com/elastictranscoder/>)

Ask our Experts



QUESTION 5

UNATTEMPTED

COLLECTION

HikeHills.com (HH) 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.

HH runs their entire online infrastructure on java based web applications running on AWS. The HH 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.

HH is planning to consider Kinesis Streams to address integration and the architecture team is planning to define naming standards for each component in the streams. Please advice. Select 3 options.

- ☐ A. Stream is a set of shards, A shard has a sequence of data records ✓
- ☐ B. Shard is a set of streams, A stream has a set of data records

- ☐ C. Stream is a set of records, A record has a sequence of shards
- ☐ D. Data records are composed of a sequence number, a partition key, and a data blob, which is an immutable sequence of bytes ✓
- ☐ E. Data streams are composed of a sequence number, a partition key, and a data blob, which is an immutable sequence of bytes
- ☐ F. Data shards are composed of a sequence number, a partition key, and a data blob, which is an immutable sequence of bytes
- ☐ G. A data stream's retention period is set to a default of 24 hours after creation ✓
- ☐ H. A data record's retention period is set to a default of 24 hours after creation

Explanation :

Answer: A, D, G

A. Yes. A Kinesis data stream is a set of shards. Each data record has a sequence number that is assigned by Kinesis Data Streams.

<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>

(<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>)

B. No. Each shard has a sequence of data records. A shard is a uniquely identified sequence of data records in a stream. A stream is composed of one or more shards, each of which provides a fixed unit of capacity. Each shard can support up to 5 transactions per second for reads, up to a maximum total data read rate of 2 MB per second and up to 1,000 records per second for writes, up to a maximum total data write rate of 1 MB per second (including partition keys). The data capacity of your stream is a function of the number of shards that you specify for the stream. The total capacity of the stream is the sum of the capacities of its shards.

<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>

(<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>)

C. No. A Kinesis data stream is a set of shards. Each shard has a sequence of data records. Each data record has a sequence number that is assigned by Kinesis Data Streams

<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>

(<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>)

D. Yes. A data record is the unit of data stored in a Kinesis data stream. Data records are composed of a sequence number, a partition key, and a data blob, which is an immutable sequence of bytes. Kinesis Data Streams does not inspect, interpret, or change the data in the blob in any way. A data blob can be up to 1 MB.

<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>

(<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>)

E. No. A Kinesis data stream is a set of shards. Each shard has a sequence of data records. Each data record has a sequence number that is assigned by Kinesis Data Streams

<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>

(<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>)

F. No. Each shard has a sequence of data records. A shard is a uniquely identified sequence of data records in a stream. A stream is composed of one or more shards, each of which provides a fixed unit of capacity. Each shard can support up to 5 transactions per second for reads, up to a maximum total data read rate of 2 MB per second and up to 1,000 records per second for writes, up to a maximum

total data write rate of 1 MB per second (including partition keys). The data capacity of your stream is a function of the number of shards that you specify for the stream. The total capacity of the stream is the sum of the capacities of its shards.

<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>

(<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>)

G. Yes. The retention period is the length of time that data records are accessible after they are added to the stream. A stream's retention period is set to a default of 24 hours after creation. You can increase the retention period up to 168 hours (7 days).

H. No. the retention period is set at stream level. The retention period is the length of time that data records are accessible after they are added to the stream. A stream's retention period is set to a default of 24 hours after creation. You can increase the retention period up to 168 hours (7 days).

<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>

(<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>)

H. No. the retention period is set at stream level. The retention period is the length of time that data records are accessible after they are added to the stream. A stream's retention period is set to a default of 24 hours after creation. You can increase the retention period up to 168 hours (7 days).

<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>

(<https://docs.aws.amazon.com/streams/latest/dev/key-concepts.html>)

Ask our Experts



QUESTION 6

UNATTEMPTED

STORAGE

Hymutabs Ltd (Hymutabs) 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 Singapore, a materials and membrane products development centre as well as advanced machining, prototyping and industrial design functions.

Hymutabs 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. Hymutabs 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 Hymutabs work with

SALES_FCT and DATE_DIM are joined together frequently since EOD sales reports are generated every day. please suggest your distribution style for both tables. Select 1 option.

- ☐ A. 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
- ☐ B. 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
- ☒ C. 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 ✓
- ☐ D. 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
- ☐ E. Distribute the SALES_FCT with EVEN DISTRIBUTION on its own Primary KEY (one of the columns) while DATE_DIM is distributed with ALL distribution on Its PRIMARY KEY

Explanation :

Answer : C

A. Incorrect. KEY DISTRIBUTION distributes the rows according to the values in one column. This is a right approach to design the table, but DATE_DIM with KEY DISTRIBUTION with number of records being very low, lot of data is copied between nodes. This approach is ok but not a perfect design to build the solution

<https://docs.aws.amazon.com/redshift/latest/dg/tutorial-tuning-tables-distribution.html>

(<https://docs.aws.amazon.com/redshift/latest/dg/tutorial-tuning-tables-distribution.html>)

B. 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. For a fact table like SALES_FCT, all the nodes participate in all queries even though the EOD reports is only for that particular day

<https://docs.aws.amazon.com/redshift/latest/dg/tutorial-tuning-tables-distribution.html>
(<https://docs.aws.amazon.com/redshift/latest/dg/tutorial-tuning-tables-distribution.html>)

C. 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

<https://docs.aws.amazon.com/redshift/latest/dg/tutorial-tuning-tables-distribution.html>
(<https://docs.aws.amazon.com/redshift/latest/dg/tutorial-tuning-tables-distribution.html>)

D. 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. Cannot be used for massive table like SALES_FCT.

<https://docs.aws.amazon.com/redshift/latest/dg/tutorial-tuning-tables-distribution.html>
(<https://docs.aws.amazon.com/redshift/latest/dg/tutorial-tuning-tables-distribution.html>)

E. 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. For a fact table like SALES_FCT, all the nodes participate in all queries even though the EOD reports is only for that particular day. SALES_FCT TABLE need to be designed on a table with a perfect distribution key in mind

<https://docs.aws.amazon.com/redshift/latest/dg/tutorial-tuning-tables-distribution.html>
(<https://docs.aws.amazon.com/redshift/latest/dg/tutorial-tuning-tables-distribution.html>)

Ask our Experts



QUESTION 7

UNATTEMPTED

PROCESSING

Tick-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. Tick-Bank aims to serve as a connection between digital content makers and affiliate dealers, who then promote them to clients. Tick-Bank's technology aids in payments, tax calculations and a variety of customer service tasks. Tick-Bank assists in building perceptibility and revenue making opportunities for entrepreneurs.

Tick-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. Tick-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. Tick-Bank has the following objectives in mind for the solution.

Tick-Bank has multiple Kinesis Streams supporting various business processes and uses enhanced fan out consumers to fulfill processing of data. What kind of activities are performed by the consumers using KCL library. Select 4 options.

- ☐ A. Connects to the stream ✓
- ☐ B. Enumerates the shards ✓

- ☐ C. Checkpoints processed records ✓
- ☐ D. Balances shard-worker associations when shards are split or merged ✓
- ☐ E. Connects to shards
- ☐ F. Enumerates multiple streams

Explanation :

Answer: A, B, C, D

The KCL acts as an intermediary between your record processing logic and Kinesis Data Streams. When you start a KCL application, it calls the KCL to instantiate a worker. This call provides the KCL with configuration information for the application, such as the stream name and AWS credentials. The KCL performs the following tasks:

- Connects to the stream
- Enumerates the shards
- Coordinates shard associations with other workers (if any)
- Instantiates a record processor for every shard it manages
- Pulls data records from the stream
- Pushes the records to the corresponding record processor
- Checkpoints processed records
- Balances shard-worker associations when the worker instance count changes
- Balances shard-worker associations when shards are split or merged

<https://docs.aws.amazon.com/streams/latest/dev/developing-consumers-with-kcl.html>
(<https://docs.aws.amazon.com/streams/latest/dev/developing-consumers-with-kcl.html>)

Ask our Experts



QUESTION 8

UNATTEMPTED

STORAGE

KindleYou is a location-based social search mobile app that allows users to like or dislike other users, and allows users to chat if both parties liked each other in the app. It has more than 1 billion customers across the world.

They use DynamoDB to support the mobile application and S3 to host the images and other documents shared between users. KindleYou has a large customer base spread across multiple geographic areas. Customers need to update their profile information while using the application. Team understands the Global tables are the right solution to implement replication. Suggest some best practices for designing replica table of the global table. Select 5 options.

- ☐ A. The table must have the same partition key as all of the other replicas. ✓
- ☐ B. The table must have the same write capacity but not read capacity ✓
- ☐ C. The table must have the same read capacity but not write capacity
- ☐ D. the table must have same read and write capacity.
- ☐ E. The table must have DynamoDB Streams enabled, with the stream containing both the new and the old images of the item. ✓
- ☐ F. The global secondary indexes must have the same partition key and sort key ✓
- ☐ G. The table and the global secondary names across all replicas must be same ✓
- ☐ H. The table and the local secondary names across all replicas must be same

Explanation :

Answer: A,B,E,F,G

A. Yes. The table must have the same partition key as all of the other replicas.

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html)

B. Yes. The table must have the same write capacity but not read capacity

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html)

C. No. The table must have the same write capacity but not read capacity

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html)

D. No. The table must have the same write capacity but not read capacity

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html)

E. Yes. The table must have DynamoDB Streams enabled, with the stream containing both the new and the old images of the item

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html)

F. Yes. The global secondary indexes must have the same partition key and sort key

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html)

G. Yes. The table and the global secondary names across all replicas must be same

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html)

H. No. The table and the global secondary names across all replicas must be same

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html)

Ask our Experts



QuickDialog 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. KindleYou has a large customer base spread across multiple geographic areas. Customers need to update their profile information while using the application. Team understands the Global tables are the right solution to implement replication. If you perform 10 writes to a local table that is replicated in two additional Regions, how many WCU's will be consumed? Select 1 option.

- ☐ A. 30 WCU's
- ☐ B. 40 WCU's
- ☒ C. 60 WCU's ✓
- ☐ D. 80 WCU's

Explanation :

Answer: C

1 local + 2 replica = 10+10+10=30, 30X2 = 60

Using DynamoDB auto scaling is the recommended way to manage throughput capacity settings for replica tables that use the provisioned mode. DynamoDB auto scaling automatically adjusts read capacity units (RCUs) and write capacity units (WCUs) for each replica table based upon your actual application workload. For more information, see [Managing Throughput Capacity Automatically with DynamoDB Auto Scaling](#).

If you create your replica tables using the AWS Management Console, auto scaling is enabled by default for each replica table, with default auto scaling settings for managing RCUs and WCUs.

Changes to auto scaling settings for a replica table or secondary index made through the DynamoDB console or using the `UpdateGlobalTableSettings` call are applied to all of the replica tables and matching secondary indexes in the global table automatically. These changes will overwrite any existing auto scaling settings. This ensures that provisioned write capacity settings are consistent across the replica tables and secondary indexes in your global table. If you use the `UpdateTable`, `RegisterScalableTarget`, or `PutScalingPolicy` calls, you should apply the change to each replica table and matching secondary index individually.

Note:

If auto scaling doesn't satisfy your application's capacity changes (unpredictable workload) or if you don't want to configure its settings (target settings for minimum, maximum, or utilization threshold),

you can use on-demand mode to manage capacity for your global tables. For more information, see On-Demand Mode.

If you enable on-demand mode on a global table, your consumption of replicated write request units (rWCUs) will be consistent with how rWCUs are provisioned. For example, if you perform 10 writes to a local table that is replicated in two additional Regions, you will consume 60 write request units ($10 + 10 + 10 = 30$; $30 \times 2 = 60$).

- https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html (https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_reqs_bestpractices.html)

Ask our Experts



QUESTION 10

UNATTEMPTED

VISUALIZATION

MSP 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. MSP Bank is hosting their existing infrastructure on AWS. MSP 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.

The team is looking for a chart to show a measure for the intersection of two dimensions, with color-coding to easily differentiate where values fall in the range. please help. Select 1 option.

- ☐ A. Tabular Reports
- ☒ B. Heat Maps ✓
- ☐ C. Line Chart
- ☐ D. Tree Maps

Explanation :

Answer: B

A. No. 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.

<https://docs.aws.amazon.com/quicksight/latest/user/tabular.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/tabular.html>)

B. Yes. Use heat maps to show a measure for the intersection of two dimensions, with color-coding to easily differentiate where values fall in the range. Heat maps can also be used to show the count of values for the intersection of the two dimensions.

<https://docs.aws.amazon.com/quicksight/latest/user/heat-map.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/heat-map.html>)

C. No. Use line charts to compare changes in measure values over period of time, for the following scenarios:

- One measure over a period of time, for example gross sales by month.
- Multiple measures over a period of time, for example gross sales and net sales by month.
- One measure for a dimension over a period of time, for example number of flight delays per day by airline.

<https://docs.aws.amazon.com/quicksight/latest/user/line-charts.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/line-charts.html>)

D. No. Use tree maps to visualize one or two measures for a dimension.

Each rectangle on the tree map represents one item in the dimension. Rectangle size represents the proportion of the value for the selected measure that the item represents compared to the whole for the dimension. You can optionally use rectangle color to represent another measure for the item. Rectangle color represents where the value for the item falls in the range for the measure, with darker colors indicating higher values and lighter colors indicating lower ones.

<https://docs.aws.amazon.com/quicksight/latest/user/tree-map.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/tree-map.html>)

Ask our Experts



QUESTION 11

UNATTEMPTED

PROCESSING

Tick-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. Tick-Bank aims to serve as a connection between digital content makers and affiliate dealers, who then promote them to clients. Tick-Bank's technology aids in payments, tax calculations and a variety of customer service tasks. Tick-Bank assists in building perceptibility and revenue making opportunities for entrepreneurs.

Tick-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. Tick-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. Tick-Bank has the following objectives in mind for the solution.

Optimize the overall platform costs by accumulating user records thereby improve throughput to the stream

Minimal changes to the web application by embedding simple code

Integrate seamlessly to de-aggregate batched records on the consumer
Long term storage into S3 storage bucket for future integration with the BI
application
Select 2 options.

- ☐ A. Use Kinesis Producer Library (KPL) micro services to make necessary code changes, aggregation of user records through batching thereby increasing throughput before stream record is processed and use Kinesis Client Library (KCL) to de-aggregate batched records from the stream ✓
- ☐ B. Use Kinesis Data Stream API with AWS SDK for java to aggregate using PutRecords and de-aggregate batched records using GetRecords
- ☐ C. Use Kinesis agent to pre-process, aggregate, batch the clickstream data and use Kinesis Data Streams to de-aggregate batched records
- ☐ D. Use KCL uses Kinesis connector library to de-aggregate and write data to S3 ✓
- ☐ E. Use Data Streams API to de-aggregate and write data to S3

Explanation :

Answer : A, D

A. Kinesis Producer Library (KPL) acts as an intermediary between your application and the Kinesis Data Streams API. KPL simplifies producer application development and also building batch of aggregation of user records by increasing payload size and improve throughput and optimize costs.

<https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html>)

KCL seamlessly integrate with Data streams to de-aggregate batched records on the consumer.

<https://docs.aws.amazon.com/streams/latest/dev/developing-consumers-with-kcl.html>
(<https://docs.aws.amazon.com/streams/latest/dev/developing-consumers-with-kcl.html>)

B. Kinesis Data Streams does not inspect, interpret, or change the data in any way. Each record also has an associated sequence number and partition key. Aggregation is not allowed. Besides each user record is a stream record and that improves throughput of the HTTP request but the overall platform costs cannot be optimized

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

C. The purpose of Kinesis agent is different. Kinesis Agent is a stand-alone Java application that can easily collect and send data to Kinesis Data Streams. The agent can continuously monitor set of files (more for log files) and Aggregation of data is not possible.

<https://docs.aws.amazon.com/streams/latest/dev/writing-with-agents.html>
(<https://docs.aws.amazon.com/streams/latest/dev/writing-with-agents.html>)

D. Kinesis connector library helps java developers integrate Kinesis Streams with other AWS services. The library provides connectors to various AWS services including S3. Each Amazon Kinesis connector application is a pipeline that understands how records from Kinesis Stream will be handled.

<https://github.com/aws-labs/amazon-kinesis-connectors> (<https://github.com/aws-labs/amazon-kinesis-connectors>)

E. Data Streams API can be consumer to the Data Stream but cannot de-aggregate the data. Also Streams API again needs kinesis connectors to connect to different sources

<https://docs.aws.amazon.com/streams/latest/dev/developing-consumers-with-sdk.html>
(<https://docs.aws.amazon.com/streams/latest/dev/developing-consumers-with-sdk.html>)

Ask our Experts



QUESTION 12

UNATTEMPTED

ANALYSIS

FlexiToner uses AWS to query 10 years' worth of historical data and get results in moments, with the flexibility to explore data for deeper insights. Movable Ink provides real-time personalization of marketing emails based on a wide range of user, device, and contextual data, driving higher response rates and better customer experiences. Also FlexiToner hosts log files captured from web servers running out of different EC2 machines

FlexiToner has lot of data assets available in structured, semi-structured and unstructured data forms containing emails, logs, structured data from databases in csv files with formats in CSV, LOG, JSON and binary formats like Parquet and ORC. FlexiToner is interested to build a data lake out of all the files stored on S3 and provide Data Lake as a service to users from different departments based on pay per queries run. What could be an effective solution to manage the environments? Select 3 options.

- ☐ A. Athena being serverless, allows to run ad-hoc queries on S3 using ANSI SQL interactively, without the need to aggregate or load the data ✓
- ☐ B. Athena running on scalable EC2 machines, allows to run ad-hoc queries on S3 using ANSI SQL interactively, without the need to aggregate or load the data
- ☐ C. Athena integrates with Amazon QuickSight for easy data visualization ✓
- ☐ D. Athena can act as a component of Amazon QuickSight for easy data visualization
- ☐ E. Athena integrates with the AWS Glue Data Catalog, which offers a persistent metadata store for your data in Amazon S3 ✓
- ☐ F. Athena integrates with the AWS RDS, which offers a persistent metadata store for your data in Amazon S3

Explanation :

Answer: A, C, E

A. Yes. Athena is serverless, so there is no infrastructure to set up or manage, and you pay only for the queries you run. Athena scales automatically—executing queries in parallel—so results are fast, even with large datasets and complex queries. Athena helps you analyze unstructured, semi-structured,

and structured data stored in Amazon S3. Examples include CSV, JSON, or columnar data formats such as Apache Parquet and Apache ORC. You can use Athena to run ad-hoc queries using ANSI SQL, without the need to aggregate or load the data into Athena

<https://docs.aws.amazon.com/athena/latest/ug/what-is.html>

(<https://docs.aws.amazon.com/athena/latest/ug/what-is.html>)

B. No. Athena is serverless, so there is no infrastructure to set up or manage, and you pay only for the queries you run. Athena scales automatically—executing queries in parallel—so results are fast, even with large datasets and complex queries. Athena helps you analyze unstructured, semi-structured, and structured data stored in Amazon S3. Examples include CSV, JSON, or columnar data formats such as Apache Parquet and Apache ORC. You can use Athena to run ad-hoc queries using ANSI SQL, without the need to aggregate or load the data into Athena

<https://docs.aws.amazon.com/athena/latest/ug/what-is.html>

(<https://docs.aws.amazon.com/athena/latest/ug/what-is.html>)

C. Yes. Athena integrates with Amazon QuickSight for easy data visualization. You can use Athena to generate reports or to explore data with business intelligence tools or SQL clients connected with a JDBC or an ODBC driver.

D. No. Athena integrates with Amazon QuickSight for easy data visualization. You can use Athena to generate reports or to explore data with business intelligence tools or SQL clients connected with a JDBC or an ODBC driver.

<https://docs.aws.amazon.com/athena/latest/ug/when-should-i-use-ate.html>

(<https://docs.aws.amazon.com/athena/latest/ug/when-should-i-use-ate.html>)

E. Yes. Athena integrates with the AWS Glue Data Catalog, which offers a persistent metadata store for your data in Amazon S3. This allows you to create tables and query data in Athena based on a central metadata store available throughout your AWS account and integrated with the ETL and data discovery features of AWS Glue.

<https://docs.aws.amazon.com/athena/latest/ug/when-should-i-use-ate.html>

(<https://docs.aws.amazon.com/athena/latest/ug/when-should-i-use-ate.html>)

F. No. Athena does not integrate with Amazon RDS. Instead Athena integrates with the AWS Glue Data Catalog, which offers a persistent metadata store for your data in Amazon S3. This allows you to create tables and query data in Athena based on a central metadata store available throughout your AWS account and integrated with the ETL and data discovery features of AWS Glue.

<https://docs.aws.amazon.com/athena/latest/ug/when-should-i-use-ate.html>

(<https://docs.aws.amazon.com/athena/latest/ug/when-should-i-use-ate.html>)

Ask our Experts



QUESTION 13

UNATTEMPTED

COLLECTION

HikeHills.com (HH) 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.

HH runs their entire online infrastructure on java based web applications running on AWS. The HH 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.

HH IT team identified lot of performance issues with the Kinesis Stream and based on the metrics captured, identified hot and cold shards. IT team wants to effectively improve the performance of the shards. How can they achieve that? Select 2 options.

- ☐ A. Split the hot shards to increase capacity for the hash keys that target those shards ✓
- ☐ B. Split the shards that receive more data to increase capacity for the hash keys that target those shards ✓
- ☐ C. Split the cold shards to increase capacity for the hash keys that target those shards
- ☐ D. Split the shards that receive less data to increase capacity for the hash keys that target those shards

Explanation :

Answer: A, B

A. Yes. Performance of a shard (hot shard) can be improve by splitting the shard based on the hash keys that target those shards

<https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html>)

B. Yes. Performance of a shard (more data) can be improve by splitting the shard based on the hash keys that target those shards

<https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html>)

C. No. Hot shards are split, cold shards are always merged to reduce unnecessary used capacity

<https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html>)

D. No. Hot shards are split, cold shards are always merged to reduce unnecessary used capacity

<https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html>)

Ask our Experts



MSP 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. MSP Bank is hosting their existing infrastructure on AWS. MSP bank has many organizations internally and they are planning to launch a self-data discovery platform running out of AWS on QuickSight.

Please advise what kind of capabilities does QuickSight Offer? Select 5 options.

- ☐ A. Data Preparation is the process of connecting sources for data analysis, including files, AWS services, and on-premises databases
- ☐ B. Data Preparation is the process of transforming raw data for use in an analysis which includes filtering, renaming fields, changing datatypes, adding calculated fields and creating custom SQL queries. ✓
- ☐ C. Super-fast, Parallel, In-memory Calculation Engine, SPICE with storage and processing capacity is engineered to rapidly perform advanced calculations and serve data and speeds up the analytical queries that you run against your imported data ✓
- ☐ D. Speed, Parallel, Intensive Calculation Engine, SPICE with storage and processing capacity is engineered to rapidly perform advanced calculations and serve data and speeds up the analytical queries that you run against your imported data
- ☐ E. Data Analysis is the basic workspace for creating and interacting with visuals, creating Sheets and Stories that you can play as a slideshow ✓
- ☐ F. Data Exploration is the basic workspace for creating and interacting with visuals, creating Sheets and Stories that you can play as a slideshow
- ☐ G. Dashboards is a read-only snapshot of an analysis that can be shared with other Amazon QuickSight users for reporting purposes ✓
- ☐ H. Dashboard allows to view and filter the dashboard visuals without changing the underlying data ✓

Explanation :

Answer: B,C,E,G,H

A. Data preparation is the process of transforming raw data for use in an analysis. This includes making changes like the following:

- Filtering out data so you can focus on what's important to you
- Renaming fields to make them easier to read
- Changing data types so they are more useful
- Adding calculated fields to enhance analysis
- Creating SQL queries to refine data

<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>)

B. Data preparation is the process of transforming raw data for use in an analysis. This includes making changes like the following:

- Filtering out data so you can focus on what's important to you
- Renaming fields to make them easier to read
- Changing data types so they are more useful
- Adding calculated fields to enhance analysis

Creating SQL queries to refine data

<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>)

C. SPICE is Amazon QuickSight's Super-fast, Parallel, In-memory Calculation Engine. SPICE is engineered to quickly perform advanced calculations and serve data. The storage and processing capacity available in SPICE speeds up the critical queries that you run against your imported data.

<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>)

D. SPICE is Amazon QuickSight's Super-fast, Parallel, In-memory Calculation Engine. SPICE is engineered to quickly perform advanced calculations and serve data. The storage and processing capacity available in SPICE speeds up the critical queries that you run against your imported data.

<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>)

E. A data analysis is the basic workspace for creating and interacting with visuals, which are graphical representations of your data. Each analysis contains a collection of visuals that you assemble and arrange for your purposes, such as a sales analysis, cost analysis, or tracking key performance indicators. Each analysis can contain stories, which you can use to save a sequential slide show of different iterations of the analysis.

<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>)

F. A data analysis is the basic workspace for creating and interacting with visuals, which are graphical representations of your data. Each analysis contains a collection of visuals that you assemble and arrange for your purposes, such as a sales analysis, cost analysis, or tracking key performance indicators. Each analysis can contain stories, which you can use to save a sequential slide show of different iterations of the analysis.

<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>)

G. A dashboard is a read-only snapshot of an analysis that you can share with other Amazon QuickSight users for reporting purposes.

<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>)

H. Dashboards let users view and filter the dashboard visuals without changing the underlying data.

<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>)

Ask our Experts



MSP 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. MSP Bank is hosting their existing infrastructure on on premise DC and AWS and maintains a hybrid environment.

MSP Bank hosts multiple web applications, CRM and ERP running on premise while moving storage, compute, DWH and AI running out of AWS. Also MSP is launching new applications running on AWS environment. MSP Banks hosts their Development, Testing and Production VPC to maintain different environments and maintains VPN connectivity between on premise DC and AWS.

MSP Bank is planning to build a data lake on all the log files stored in S3, captured from different applications running out of on premise and AWS and also identified data sets captured out of CRM, ERP and other Business applications . MSP Bank is looking at AWS Glue to acts as a fully managed ETL service that makes it simple and cost-effective to categorize your data, clean it, enrich it, and move it reliably between various data stores. What are the key components and functionalities of AWS Glue? Select 3 options.

- ☒ A. Crawlers scan data in all kinds of repositories, classify it, extract schema information from it, and store the metadata automatically in the AWS Glue Data Catalog ✓
- ☒ B. AWS Glue autogenerate Scala or PySpark (the Python API for Apache Spark) scripts with AWS Glue extensions that use and modify to perform ETL operations through AWS Glue ETL operations ✓
- ☒ C. AWS Glue Data Catalog, persistent metadata store that lets store, annotate, and share metadata in the AWS Cloud ✓
- ☐ D. Crawlers scan data in all kinds of repositories, classify it, extract schema information from it, and store the metadata automatically in the AWS Glue Meta Store
- ☐ E. AWS Glue Meta Store, persistent metadata store that lets store, annotate, and share metadata in the AWS Cloud
- ☐ F. AWS Glue autogenerate Scala or PySpark (the Python API for Apache Spark) scripts with AWS Glue extensions that use and modify to perform ETL operations through AWS Glue Jobs system

Explanation :

Answer: A, B, C

A. Yes. AWS Glue also lets you set up crawlers that can scan data in all kinds of repositories, classify it, extract schema information from it, and store the metadata automatically in the AWS Glue Data Catalog. From there it can be used to guide ETL operations.

<https://docs.aws.amazon.com/glue/latest/dg/components-overview.html>

(<https://docs.aws.amazon.com/glue/latest/dg/components-overview.html>)

B. Yes. Using the metadata in the Data Catalog, AWS Glue can autogenerate Scala or PySpark (the Python API for Apache Spark) scripts with AWS Glue extensions that you can use and modify to perform various ETL operations. For example, you can extract, clean, and transform raw data, and then store the result in a different repository, where it can be queried and analyzed.

<https://docs.aws.amazon.com/glue/latest/dg/components-overview.html>

(<https://docs.aws.amazon.com/glue/latest/dg/components-overview.html>)

C. Yes. The AWS Glue Data Catalog is your persistent metadata store. It is a managed service that lets you store, annotate, and share metadata in the AWS Cloud in the same way you would in an Apache Hive metastore.

<https://docs.aws.amazon.com/glue/latest/dg/components-overview.html>

(<https://docs.aws.amazon.com/glue/latest/dg/components-overview.html>)

D. No. AWS Glue also lets you set up crawlers that can scan data in all kinds of repositories, classify it, extract schema information from it, and store the metadata automatically in the AWS Glue Data Catalog. From there it can be used to guide ETL operations

<https://docs.aws.amazon.com/glue/latest/dg/components-overview.html>

(<https://docs.aws.amazon.com/glue/latest/dg/components-overview.html>)

E. No. The AWS Glue Data Catalog is your persistent metadata store. It is a managed service that lets you store, annotate, and share metadata in the AWS Cloud in the same way you would in an Apache Hive metastore.

<https://docs.aws.amazon.com/glue/latest/dg/components-overview.html>

(<https://docs.aws.amazon.com/glue/latest/dg/components-overview.html>)

F. No. Using the metadata in the Data Catalog, AWS Glue can autogenerate Scala or PySpark (the Python API for Apache Spark) scripts with AWS Glue extensions that you can use and modify to perform various ETL operations. For example, you can extract, clean, and transform raw data, and then store the result in a different repository, where it can be queried and analyzed.

<https://docs.aws.amazon.com/glue/latest/dg/components-overview.html>

(<https://docs.aws.amazon.com/glue/latest/dg/components-overview.html>)

Ask our Experts



QUESTION 16

UNATTEMPTED

PROCESSING

HikeHills.com (HH) 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.

HH runs their entire online infrastructure on java based web applications running on AWS. The HH 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. HH proposes Kinesis Firehose to process the stream and capture information. What purposes can be fulfilled OOTB without writing applications or consumer code? Select 4 options.

- ☒ A. Deliver real-time streaming data to Amazon Simple Storage Service (Amazon S3) ✓
- ☐ B. Deliver real-time streaming data to DynamoDB to support processing of digital documents
- ☐ C. Deliver real-time streaming data to Redshift to support data warehousing and real-time analytics ✓
- ☐ D. Ingest data into ES domains to support Enterprise search built on Elasticsearch ✓
- ☐ E. Allow Splunk to read and process data stream directly from Kinesis Firehose ✓
- ☐ F. Ingest data into Amazon EMR to support big data analytics

Explanation :

Answer: A, C, D, E

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. You can also configure Kinesis Data Firehose to transform your data before delivering it.

<https://docs.aws.amazon.com/firehose/latest/dev/what-is-this-service.html>
(<https://docs.aws.amazon.com/firehose/latest/dev/what-is-this-service.html>)

Ask our Experts



QUESTION 17

UNATTEMPTED

STORAGE

QuickDialog 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.

QuickDialog is using RDS to support their existing mobile application and are facing lot of issues in terms of scalability, performance and high availability. They are considering DynamoDB to migrate the platform. QuickDialog is evaluating whether the DynamoDB database fulfills the enterprise-grade platform capabilities. please advise why do you think is an enterprise-grade platform. Select 4 options.

- ☒ A. DynamoDB can support mission-critical workloads by enabling ACID transactions ✓
- ☒ B. Encryption of data at rest ✓
- ☒ C. Point-in-time replication (PITR) can be enabled for DynamoDB ✓
- ☐ D. Query data lake built on S3
- ☒ E. On-demand backup and restore ✓

Explanation :

Answer: A, B, C, E

A. Yes. DynamoDB provides native, server-side support for transactions, simplifying the developer experience of making coordinated, all-or-nothing changes to multiple items both within and across tables. With support for transactions, platform can be extended to scale, performance, and enterprise benefits of DynamoDB to a broader set of mission-critical workloads

<https://aws.amazon.com/dynamodb/features/> (<https://aws.amazon.com/dynamodb/features/>)

B. Yes. DynamoDB encrypts all customer data at rest by default. Encryption at rest enhances the security of your data by using encryption keys stored in AWS Key Management Service.

<https://aws.amazon.com/dynamodb/features/> (<https://aws.amazon.com/dynamodb/features/>)

C. Yes. Point-in-time recovery is achieved by protecting DynamoDB tables from accidental write or delete operations. PITR provides continuous backups of your DynamoDB table data, and you can restore that table to any point in time up to the second during the preceding 35 days.

<https://aws.amazon.com/dynamodb/features/> (<https://aws.amazon.com/dynamodb/features/>)

D. No. cannot query the data lake built on S3. Besides Redshift extend queries to your Amazon S3 data lake without loading data. You can query open file formats you already use, such as Avro, CSV, Grok, JSON, ORC, Parquet, and more, directly in S3

<https://aws.amazon.com/redshift/features/> (<https://aws.amazon.com/redshift/features/>)

E. Yes. On-demand backup and restore allows you to create full backups of your DynamoDB tables' data for data archiving, which can help you meet your corporate and governmental regulatory requirements.

<https://aws.amazon.com/dynamodb/features/> (<https://aws.amazon.com/dynamodb/features/>)

Ask our Experts



Gluebush.com is a British online confidential advertisement and public website. Classified ads are either free or paid for depending on the product category and the geographical market.

While the largest category of advertisements on Gluebush.com is "goods for sale", the site also supports around 100,000 motors listings across the UK at any one time, with an extensive social media presence on Twitter and Facebook, with 22,000 and 471,000 followers, respectively. Gumtree uses social media for communications and information about the brand as well as competitions and campaigns.

Gluebush.com runs multiple business applications both web and mobile based on AWS. Gluebush.com wants to collect log and event data from web servers, mobile devices, pre-process the data and process the data to feed live dashboards, and load data into data warehouse build on Redshift and on S3 for long term storage. Kinesis Data Streams are used to process the data. There is requirement to generate operational reports generated at the end of the day. Also looking for a solution to generate weekly and monthly reports. To what extent can Kinesis Stream support before moving to persistent data storage.

Select 3 options.

- ☐ A. Daily Reports can be generated from Kinesis Stream end-of-day since Kinesis data streams hosts data in stream since retention period for data stream is 24 hours by default ✓
- ☐ B. Weekly Reports can be generated from Kinesis Stream end-of-week since Kinesis data streams hosts data in stream since retention period for data stream is for 168 hours by default
- ☐ C. Daily Reports can be generated from Kinesis Stream end-of-month since Kinesis data streams hosts data in stream since retention period for data stream is for 744 hours by default
- ☐ D. Daily Reports can be generated from Kinesis Stream end-of-day if retention period for data stream is increased to 24 hours, with 4 hours by default
- ☐ E. weekly Reports can be generated from Kinesis Stream end-of-week if retention period for data stream is increased to 168 hours, with 24 hours by default ✓
- ☐ F. Monthly Reports can be generated from Kinesis Stream end-of-month if retention period for data stream is increased to 744 hours, with 24 hours by default
- ☐ G. Monthly reports can be generated from S3 or Redshift ✓

Explanation :

Answer: A, E, G

A. Yes. Daily Reports can be generated from Kinesis Stream end-of-day since Kinesis data streams

hosts data in stream since retention period for data stream is 24 hours by default
<https://docs.aws.amazon.com/streams/latest/dev/kinesis-extended-retention.html>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-extended-retention.html>)

B. No. Weekly Reports can be generated from Kinesis Stream end-of-day since Kinesis data streams hosts data in stream since retention period for data stream is 24 hours by default
<https://docs.aws.amazon.com/streams/latest/dev/kinesis-extended-retention.htm>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-extended-retention.htm>)

C. No. Monthly Reports can be generated from Kinesis Stream end-of-day since Kinesis data streams hosts data in stream since retention period for data stream is 24 hours by default
<https://docs.aws.amazon.com/streams/latest/dev/kinesis-extended-retention.html>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-extended-retention.html>)

D. No. Daily Reports can be generated from Kinesis Stream end-of-day since Kinesis data streams hosts data in stream since retention period for data stream is 24 hours by default
<https://docs.aws.amazon.com/streams/latest/dev/kinesis-extended-retention.html>
(<http://https://docs.aws.amazon.com/streams/latest/dev/kinesis-extended-retention.html>)

E. Yes. Weekly Reports can be generated from Kinesis Stream end-of-day since Kinesis data streams hosts data in stream since retention period for data stream is 24 hours by default but can be extended to 168 hours
<https://docs.aws.amazon.com/streams/latest/dev/kinesis-extended-retention.html>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-extended-retention.html>)

F. No. Monthly Reports cannot be generated from Kinesis Stream end-of-day since Kinesis data streams hosts data in stream since retention period for data stream is 24 hours by default but can be extended beyond 168 hours
<https://docs.aws.amazon.com/streams/latest/dev/kinesis-extended-retention.html>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-extended-retention.html>)

G. Yes. Data in S3 or Redshift need to be used to fulfill requirement for monthly reports. Kinesis data streams don't support that requirement.
<https://docs.aws.amazon.com/streams/latest/dev/kinesis-extended-retention.html>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-extended-retention.html>)

Ask our Experts



QUESTION 19

UNATTEMPTED

PROCESSING

HikeHills.com (HH) 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.

HH runs their entire online infrastructure on java based web applications running on AWS. The HH 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.

HH IT team identified lot of performance issues with the Kinesis Stream and based on the metrics captured, identified hot and cold shards. IT team wants to effectively remove the unused capacity. There are 2 shards SHARD 1 with a hash key range of 276...381 and SHARD 2 with a hash key range of 382...454. What Resharding strategy needs to be applied and how can it be applied? Select 1 option.

- ☐ A. SHARD 1 need to be split as SHARD 1A with hash keys 276...332 and SHARD 1B as 333...381, SHARD2 into SHARD2A 382..410 and SHARD @B as 410..454
- ☐ B. SHARD 1 need to be split as SHARD 1A with hash keys 276...332 and SHARD 1B as 333...381, SHARD2 into SHARD2A 382..410 and SHARD @B as 411..454
- ☐ C. MERGE AND SPLIT SHARD 1 and SHARD2 into 3 SHARDS, SHARD12A with hash keys 276...370, SHARD12B as 371...420, SHARD12C as 421..454
- ☐ D. MERGE 2 SHARDS into 1 SHARD with hash keys between 276..454 which remove un-usability ✓

Explanation :

Answer: D

Merging the shards removes unnecessary shards and improves usage thereby reducing total costs.

<https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-merge.html>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-merge.html>)

Ask our Experts



QUESTION 20

UNATTEMPTED

ANALYSIS

HikeHills.com (HH) 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.

HH runs their entire online infrastructure on java based web applications running on AWS. The HH 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 Streaming capabilities to collect events and transaction logs and process the stream.

HH wants to build analytics capability on the streams. What all data sources can be supported for ingestion of the stream data by kinesis analytics tool authoring using SQL Code using in-built editor? Select 2 options.

- ☐ A. AWS IoT
- ☐ B. AWS Kinesis data Streams ✓
- ☐ C. AWS Kinesis Video Streams
- ☐ D. AWS Kinesis Firehose ✓
- ☐ E. Spark Streaming on AWS EMR
- ☐ F. Storm on AWS EMR

Explanation :

Answer: B,D

Kinesis Data Analytics allows continuously reading and processing streaming data. The service supports ingesting data from Amazon Kinesis Data Streams and Amazon Kinesis Data Firehose streaming sources. Then, you author your SQL code using the interactive editor and test it with live streaming data.

<https://docs.aws.amazon.com/kinesisanalytics/latest/dev/what-is.html>
(<https://docs.aws.amazon.com/kinesisanalytics/latest/dev/what-is.html>)

Ask our Experts



QUESTION 21

UNATTEMPTED

ANALYSIS

MindPyramid Limited is a multinational information technology and outsourcing company headquartered in Vizag, India and New Jersey, USA. Founded in 2003, the company employs approximately 2000 employees. The company offers consulting services in cloud computing, big data and analytics. They offer services to major cloud providers including AWS. The team is working with one of the major clients having their infrastructure build on AWS. Currently the client is having lot of performance issues and wants to understand the design best practices from MindPyramid team. Please suggest the best practices in terms of loading data into tables. Select 4 options.

- ☐ A. Use single copy to load multiple files from Amazon S3, Amazon EMR, Amazon DynamoDB, or multiple data sources on remote hosts ✓
- ☐ B. for optimal parallelism, split your load data files so that the files are about equal size, between 1 MB and 1 GB after compression
- ☐ C. The number of files should be a multiple of the (number of slices+ X) in your cluster. X indicates the number of leader nodes
- ☐ D. Manage data consistency using a manifest file to load data and address eventual consistency issues ✓

- ☐ E. In order to reduce the need for VACCUUM, load data in the sort key order of the table ✓
- ☐ F. Load the data in sequential blocks according to sort order to eliminate the need to vacuum. ✓
- ☐ G. Staging tables benefits inserts and updates but create performance issues when upserts are performed

Explanation :

Answer : A, D, E, F

A. Yes. Single copy to load multiple files is a best practice to load the data from S3, EMR and Others

https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-copy.html

(https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-copy.html)

B. No. the optimal size of the files should be of equal size between 1 MB and 125 MB

https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-multiple-files.html

(https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-multiple-files.html)

C. No. the number of files should be a multiple of the number of slices in your cluster

https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-multiple-files.html

(https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-multiple-files.html)

D. Yes. Manifest files address eventual consistency issues during loads

<https://docs.aws.amazon.com/redshift/latest/dg/best-practices-preventing-load-data-errors.html>

(<https://docs.aws.amazon.com/redshift/latest/dg/best-practices-preventing-load-data-errors.html>)

E. Yes. Load your data in sort key order to avoid needing to vacuum.

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

(https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-sort-key-order.html)

F. Yes. load the data in sequential blocks according to sort order to eliminate the need to vacuum.

https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-load-data-in-sequential-blocks.html

(https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-load-data-in-sequential-blocks.html)

(https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-load-data-in-sequential-blocks.html)

G. No. the Use a staging table to perform a merge operations. load your data into a staging table and then join the staging table with your target table for an UPDATE statement and an INSERT statement.

https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-upsert.html

(https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-upsert.html)

Ask our Experts



QUESTION 22

UNATTEMPTED

PROCESSING

Allianz Financial Services (AFS) is a banking group offering end-to-end banking and financial solutions in South East Asia through its consumer banking, business banking, Islamic banking, investment finance and stock broking businesses as well as unit trust and asset administration, having served the financial community over the past five decades.

AFS launched EMR cluster to support their big data analytics requirements. AFS has a

large team of Hadoop developers who work on both Hive and Pig applications.

AFS is looking for a graphical user interface for use with Amazon EMR and Apache Hadoop and also groups together several different Hadoop ecosystem projects into a configurable interface Which EMR Hadoop ecosystem fulfills the requirements?

Select 1 option.

- ☒ A. Apache Hue ✓
- ☐ B. Apache Flink
- ☐ C. Apache Phoenix
- ☐ D. Apache Tez

Explanation :

Answer : A

A. Yes. Hue (Hadoop User Experience) is an open-source, web-based, graphical user interface for use with Amazon EMR and Apache Hadoop. Hue groups together several different Hadoop ecosystem projects into a configurable interface. Amazon EMR has also added customizations specific to Hue in Amazon EMR. Hue acts as a front-end for applications that run on your cluster, allowing you to interact with applications using an interface that may be more familiar or user-friendly. The applications in Hue, such as the Hive and Pig editors, replace the need to log in to the cluster to run scripts interactively using each application's respective shell

<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hue.html>

(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hue.html>)

B. No. Apache Flink is a streaming dataflow engine that you can use to run real-time stream processing on high-throughput data sources. Flink supports event time semantics for out-of-order events, exactly-once semantics, backpressure control, and APIs optimized for writing both streaming and batch applications. <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-flink.html>

(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-flink.html>)

C. No. Apache Phoenix is used for OLTP and operational analytics, allowing you to use standard SQL queries and JDBC APIs to work with an Apache HBase backing store.

<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-phoenix.html>

(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-phoenix.html>)

D. No. Apache Tez is a framework for creating a complex directed acyclic graph (DAG) of tasks for processing data. In some cases, it is used as an alternative to Hadoop MapReduce.

<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-tez.html>

(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-tez.html>)

Ask our Experts



MSP 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. MSP Bank is hosting their existing infrastructure on AWS. MSP 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. The team identified multiple charts to address the business requirements. Sort is one the key function the team wants to use in designing the charts. The team understands that some of the charts don't have the sort functionality. Select 3 options.

- ☐ A. Bar Charts
- ☐ B. Line Charts
- ☐ C. Geospatial Charts ✓
- ☐ D. KPI's ✓
- ☐ E. Heat Maps
- ☐ F. Scatter Plot ✓

Explanation :

Answer: C, D, F

Geospatial charts, KPI's, Scatter Plot does not support sort function

<https://docs.aws.amazon.com/quicksight/latest/user/sorting-visual-data.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/sorting-visual-data.html>)

Ask our Experts



QUESTION 24

UNATTEMPTED

ANALYSIS

Parson Fortunes Ltd is an Asian-based department store operator with an extensive network of 131 stores, spanning approximately 4.1 million square meters of retail space across cities in

India, China, Vietnam, Indonesia and Myanmar. Parson built a VPC to host their entire enterprise infrastructure on cloud. Parson 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. The files sent from their on premise data center are also hosted into S3 buckets. Parson IT team is well aware of the

scalability, performance of AWS services capabilities. Parson hosts their web applications, databases and the data warehouse built on Redshift in VPC The structured, semi-structured and unstructured formats are stored in S3 in various buckets.

This data be joined and queried along with data in Redshift using Redshift Spectrum.

What are the below best practices that can be enabled to improve query performance, overall costs and security? Select 5 options.

- ☐ A. Break large files into many smaller files, Typically of file size of 64 MB or larger ✓
- ☐ B. Group smaller files into a single large file. Typically of file size of 512 MB or larger
- ☐ C. Store files for a table in the same folder ✓
- ☐ D. Store files for a table into different sub-folders
- ☐ E. Keep all the files about the same size. If some files are much larger than others, Redshift Spectrum can't distribute the workload evenly ✓
- ☐ F. To reduce storage space, improve performance, and minimize costs, compress data files into gzip, snappy or bzip2 ✓
- ☐ G. Use Server-side encryption (SSE-S3) using an AES-256 encryption key, managed by Amazon S3 or keys managed by AWS Key Management Service (SSE-KMS). ✓
- ☐ H. Use S3 client side encryption

Explanation :

Answer : A, C, E, F, G

A. Yes. Break large files into many smaller files. We recommend using file sizes of 64 MB or larger. Store files for a table in the same folder.

<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>

(<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>)

B. No. Break large files into many smaller files. We recommend using file sizes of 64 MB or larger. Store files for a table in the same folder.

<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>

(<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>)

C. Yes. Break large files into many smaller files. We recommend using file sizes of 64 MB or larger. Store files for a table in the same folder.

<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>

(<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>)

D. No. Break large files into many smaller files. We recommend using file sizes of 64 MB or larger. Store files for a table in the same folder.

<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>

(<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>)

E. Yes, Keep all the files about the same size. If some files are much larger than others, Redshift Spectrum can't distribute the workload evenly

<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>

(<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>)

F. Yes. To reduce storage space, improve performance, and minimize costs, we strongly recommend compressing your data files. Redshift Spectrum recognizes file compression types based on the file extension.

Redshift Spectrum supports the following compression types and extensions:

- gzip – .gz
- Snappy – .snappy
- bzip2 – .bz2

<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>

(<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>)

G. Yes. Redshift Spectrum transparently decrypts data files that are encrypted using the following encryption options:

- Server-side encryption (SSE-S3) using an AES-256 encryption key managed by Amazon S3.
- Server-side encryption with keys managed by AWS Key Management Service (SSE-KMS)

<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>

(<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>)

H. No. Redshift Spectrum doesn't support Amazon S3 client-side encryption

<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>

(<https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html>)

Ask our Experts



QUESTION 25

UNATTEMPTED

STORAGE

KindleYou is a location-based social search mobile app that allows users to like or dislike other users, and allows users to chat if both parties liked each other in the app. It has more than 1 billion customers across the world.

They use DynamoDB to support the mobile application and S3 to host the images and other documents shared between users.

Consider a table named SessionData that tracks the session history of users. Each item in SessionData is identified by a partition key (UserName) and a sort key (SessionId). Additional attributes like UserName, SessionId, CreationTime and ExpirationTime track the session information. The ExpirationTime attribute is set as the Time To Live (TTL) attribute. How the data is managed post TTL? Select 2 options.

- ☐ A. A background job checks the TTL attribute of items to see if they are expired. ✓
- ☐ B. No background job is needed to check TTL. Automatically DynamoDB expires the Item.
- ☐ C. If the epoch time value stored in the attribute is less than the current time, the item is marked as expired and subsequently deleted ✓

- ☐ D. The item is marked as expired and subsequently deleted automatically
- ☐ E. There is a very mild performance impact when deletion is performed, it specially impacts the read traffic
- ☐ F. There is a very mild performance impact when deletion is performed, it specially impacts the write traffic

Explanation :

Answer: A,C

A. Yes. A background job checks the TTL attribute of items to see if they are expired

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/howitworks-ttl.html>
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/howitworks-ttl.html>)

B. No. A background job checks the TTL attribute of items to see if they are expired

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/howitworks-ttl.html>
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/howitworks-ttl.html>)

C. Yes. If the epoch time value stored in the attribute is less than the current time, the item is marked as expired and subsequently deleted

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/howitworks-ttl.html>
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/howitworks-ttl.html>)

D. No. If the epoch time value stored in the attribute is less than the current time, the item is marked as expired and subsequently deleted

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/howitworks-ttl.html>
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/howitworks-ttl.html>)

E. No. This processing of expiry and deletion takes place automatically in the background and does not affect read or write traffic to the table

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/howitworks-ttl.html>
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/howitworks-ttl.html>)

F. No. This processing of expiry and deletion takes place automatically in the background and does not affect read or write traffic to the table

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/howitworks-ttl.html>
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/howitworks-ttl.html>)

Ask our Experts



QUESTION 26

UNATTEMPTED

COLLECTION

HikeHills.com (HH) 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.

HH runs their entire online infrastructure on multiple java based web applications and other web framework applications running on AWS. The HH 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 Streams (KDS) to collect events and transaction logs and process the stream. Multiple departments from HH 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.

HHunderstand that Monitoring plays a major role in monitoring and managing the streaming platform. HH using kinesis agent to process events and uses consumer library to collect and disseminate streams. What kind of monitoring is enabled by kinesis agent? Select 2 options.

- ☐ A. Metrics configured for streams are automatically collected and pushed to CloudWatch every minute
- ☐ B. Metrics configured for streams are automatically collected and pushed to CloudWatch every second
- ☐ C. The agent publishes custom CloudWatch metrics with a namespace of AWSKinesisAgent. ✓
- ☐ D. The agent publishes Bytes sent, number of records attempted, number of records that returned failure and the number of calls to PutRecords that resulted in a service error ✓

Explanation :

Answer: C, D

A.No. Metrics configured for streams are automatically collected and pushed to CloudWatch every minute. These metrics are published by CloudWatch, not by Kinesis Agent.

<https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html> (<https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html>)

B. No. Metrics configured for streams are automatically collected and pushed to CloudWatch every minute. These metrics are published by CloudWatch, not by Kinesis

Agent. <https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html> (<https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html>)

C. Yes. The agent publishes custom CloudWatch metrics with a namespace of AWSKinesisAgent. These metrics help you assess whether the agent is submitting data into Kinesis Data Streams as specified, and whether it is healthy and consuming the appropriate amount of CPU and memory resources on the data producer. <https://docs.aws.amazon.com/streams/latest/dev/agent-health.html> (<https://docs.aws.amazon.com/streams/latest/dev/agent-health.html>)

D. Yes. The agent publishes Bytes sent, number of records attempted, number of records that returned failure and the number of calls to PutRecords that resulted in a service error <https://docs.aws.amazon.com/streams/latest/dev/agent-health.html> (<https://docs.aws.amazon.com/streams/latest/dev/agent-health.html>)

MSP 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. MSP Bank is hosting their existing infrastructure on AWS. MSP 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. The team is looking for a chart to measure values over period of time.

Choose Correct Option.

- ☐ A. Tabular Reports
- ☐ B. Heat Maps
- ☒ C. Line Chart ✓
- ☐ D. Tree Map

Explanation :

Answer: C

A. No. 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.

<https://docs.aws.amazon.com/quicksight/latest/user/tabular.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/tabular.html>)

B. No. Use heat maps to show a measure for the intersection of two dimensions, with color-coding to easily differentiate where values fall in the range. Heat maps can also be used to show the count of values for the intersection of the two dimensions.

<https://docs.aws.amazon.com/quicksight/latest/user/heat-map.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/heat-map.html>)

C. Yes. Use line charts to compare changes in measure values over period of time, for the following scenarios:

- One measure over a period of time, for example gross sales by month.
- Multiple measures over a period of time, for example gross sales and net sales by month.
- One measure for a dimension over a period of time, for example number of flight delays per day by airline.

<https://docs.aws.amazon.com/quicksight/latest/user/line-charts.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/line-charts.html>)

D. No. Use tree maps to visualize one or two measures for a dimension.

Each rectangle on the tree map represents one item in the dimension. Rectangle size represents the proportion of the value for the selected measure that the item represents compared to the whole for the dimension. You can optionally use rectangle color to represent another measure for the item.

Rectangle color represents where the value for the item falls in the range for the measure, with darker colors indicating higher values and lighter colors indicating lower ones.

<https://docs.aws.amazon.com/quicksight/latest/user/tree-map.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/tree-map.html>)

Ask our Experts



QUESTION 28

UNATTEMPTED

ANALYSIS

Allianz Financial Services (AFS) is a banking group offering end-to-end banking and financial solutions in South East Asia through its consumer banking, business banking, Islamic banking, investment finance and stock broking businesses as well as unit trust and asset administration, having served the financial community over the past five decades.

AFS being one the largest banks in the region is planning to improve its segment business by launching a campaign to identify potential customers for various new products launched based on their past behavior? AFS is looking for both batch and real-time predictive analytics.

AFS has lot of databases and applications both on premise and cloud and looking at building conversational interfaces for applications using voice and text. The service shall provide the deep functionality and flexibility of natural language understanding (NLU) and automatic speech recognition (ASR) so you can build highly engaging user experiences with lifelike, conversational interactions, and create new categories of products. Precisely AFS is looking at a chatbots facility

What service can provide this capability? Select 1 option.

- ☐ A. Amazon Comprehend
- ☒ B. Amazon Lex ✓
- ☐ C. Amazon Polly
- ☐ D. Amazon SageMaker

Explanation :

Answer : B

A. No. Amazon Comprehend uses natural language processing (NLP) to extract insights about the content of documents. Amazon Comprehend processes any text file in UTF-8 format. It develops insights by recognizing the entities, key phrases, language, sentiments, and other common elements in a document. Use Amazon Comprehend to create new products based on understanding the structure of documents. For example, using Amazon Comprehend you can search social networking feeds for mentions of products or scan an entire document repository for key phrases. <https://docs.aws.amazon.com/comprehend/latest/dg/what-is.html>

(<https://docs.aws.amazon.com/comprehend/latest/dg/what-is.html>)

B. Yes. Amazon Lex is an AWS service for building conversational interfaces for applications using voice and text. With Amazon Lex, the same conversational engine that powers Amazon Alexa is now available to any developer, enabling you to build sophisticated, natural language chatbots into your new and existing applications. Amazon Lex provides the deep functionality and flexibility of natural language understanding (NLU) and automatic speech recognition (ASR) so you can build highly engaging user experiences with lifelike, conversational interactions, and create new categories of products. <https://docs.aws.amazon.com/lex/latest/dg/what-is.html>

(<https://docs.aws.amazon.com/lex/latest/dg/what-is.html>)

C. No. Amazon Polly is a cloud service that converts text into lifelike speech. You can use Amazon Polly to develop applications that increase engagement and accessibility. Amazon Polly supports multiple languages and includes a variety of lifelike voices, so you can build speech-enabled applications that work in multiple locations and use the ideal voice for your customers. With Amazon Polly, you only pay for the text you synthesize. You can also cache and replay Amazon Polly's generated speech at no additional cost. <https://docs.aws.amazon.com/polly/latest/dg/what-is.html>

(<https://docs.aws.amazon.com/polly/latest/dg/what-is.html>)

D. No. Amazon SageMaker is a fully managed machine learning service. With Amazon SageMaker, data scientists and developers can quickly and easily build and train machine learning models, and then directly deploy them into a production-ready hosted environment. It provides an integrated Jupyter authoring notebook instance for easy access to your data sources for exploration and analysis, so you don't have to manage servers. It also provides common machine learning algorithms that are optimized to run efficiently against extremely large data in a distributed environment <https://docs.aws.amazon.com/sagemaker/latest/dg/whatis.html>

(<https://docs.aws.amazon.com/sagemaker/latest/dg/whatis.html>)

Ask our Experts



QUESTION 29

UNATTEMPTED

ANALYSIS

HikeHills.com (HH) 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.

HH runs their entire online infrastructure on Java based web applications running on AWS. The HH is capturing click stream data and uses a custom-built recommendation engine to recommend products which eventually improve sales, understand customer preferences and already using AWS Streaming capabilities to collect events and transaction logs and process the stream.

HH is planning to use Kinesis Analytics to build SQL capabilities on streaming data. What kind of queries is supported in Kinesis Analytics? Select 2 options.

- ☐ A. Continuous Queries, Queries over a stream executes continuously over streaming data generally supports continuously query a stream and generate alerts. ✓
- ☐ B. Windowed Queries, Queries over a stream executes continuously over streaming data generally supports continuously query a stream and generate alerts
- ☐ C. Continuous Queries, SQL queries in your application code execute continuously over in-application streams which represents unbounded data that flows continuously through your application. Stagger, Tumbling and Sliding are some of the types
- ☐ D. Windowed Queries, SQL queries in your application code execute continuously over in-application streams which represents unbounded data that flows continuously through your application. Stagger, Tumbling and Sliding are some of the types ✓

Explanation :

Answer: A,D

A. 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. <https://docs.aws.amazon.com/kinesisanalytics/latest/dev/continuous-queries-concepts.html>

(<https://docs.aws.amazon.com/kinesisanalytics/latest/dev/continuous-queries-concepts.html>)

B. SQL queries in your application code execute continuously over in-application streams. An in-application stream represents unbounded data that flows continuously through your application. Therefore, to get result sets from this continuously updating input, you often bound queries using a window defined in terms of time or rows. These are also called windowed SQL. You have stagger windows, Tumbling windows and Sliding windows query sub

types <https://docs.aws.amazon.com/kinesisanalytics/latest/dev/windowed-sql.html>

(<https://docs.aws.amazon.com/kinesisanalytics/latest/dev/windowed-sql.html>)

C. 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. <https://docs.aws.amazon.com/kinesisanalytics/latest/dev/continuous-queries-concepts.html>

(<https://docs.aws.amazon.com/kinesisanalytics/latest/dev/continuous-queries-concepts.html>)

D. SQL queries in your application code execute continuously over in-application streams. An in-application stream represents unbounded data that flows continuously through your application. Therefore, to get result sets from this continuously updating input, you often bound queries using a window defined in terms of time or rows. These are also called windowed SQL. You have stagger windows, Tumbling windows and Sliding windows query sub types <https://docs.aws.amazon.com/kinesisanalytics/latest/dev/windowed-sql.html> (<https://docs.aws.amazon.com/kinesisanalytics/latest/dev/windowed-sql.html>)

Ask our Experts



QUESTION 30

UNATTEMPTED

STORAGE

Tiger Investments (TI) is a private equity trust manager specializing in border market investments. The Group is considered a pioneer investor in Southeast Asia's Greater Sub-region and the Caribbean. Tiger Capital creates private equity funds targeting pre-emerging, post-conflict or post-disaster economies that are undergoing transition and are poised for rapid growth. The funds invest commercially in basic businesses, targeting attractive economic and social returns. Tiger Capital invests through a diversity of financial instruments including equity, and debt

TI launched EMR 3.2.1 using EMRFS storage to support their real time data analytics. IT team observed that once objects are added to EMRFS in one operation and then immediately list objects in a subsequent operation, the list and the set of objects processed is incomplete most of the times. This is a continuous problem that TI team is facing mostly when running multi-step sequential steps in extract-transform-load (ETL) data processing pipelines. How can the team track consistency. Select 2 options.

- ☐ A. Change to HDFS Ephemeral storage. This simply solves the issue
- ☐ B. Change to local file system thereby enabling Instance storage. This simply solves the issue
- ☐ C. Enable Consistent View. This allows EMR clusters to check for list and read-after-write consistency ✓
- ☐ D. EMR uses Amazon DynamoDB database to store object metadata and track consistency with Ephemeral Storage
- ☐ E. EMR uses Amazon DynamoDB database to store object metadata and track consistency with EMRFS Storage ✓

Explanation :

Answer : C, E

- A. No. Ephemeral storage is not used for long running EMR clusters
<https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-file-systems.html>
(<https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-file-systems.html>)
- B. No. local file systems is not used for long running EMR clusters
- C. Yes. Consistent view allows EMR clusters to check for list and read-after-write consistency for Amazon S3 objects written by or synced with EMRFS
<https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-consistent-view.html>
(<https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-consistent-view.html>)
- D. No. ephemeral storage does not provide you this facility
- E. Yes. When you create a cluster with consistent view enabled, Amazon EMR uses an Amazon DynamoDB database to store object metadata and track consistency with Amazon S3

Ask our Experts



QUESTION 31

UNATTEMPTED

DATA SECURITY

FlexiToner uses AWS to query 10 years' worth of historical data and get results, with the flexibility to explore data for deeper insights. Movable Ink provides real-time personalization of marketing emails based on a wide range of user, device, and contextual data, driving higher response rates and better customer experiences. Also FlexiToner hosts log files captured from web servers running out of different EC2 machines

FlexiToner has lot of data assets available in structured, semi-structured and unstructured data forms containing emails, logs, structured data from databases in csv files with formats in CSV, LOG, JSON and binary formats like Parquet and ORC. FlexiToner is interested to build a data lake out of all the files stored on S3 and provide Data Lake as a service to users from different departments based on pay per queries run. FlexiToner understands that Athena provides this facility OOTB.

Consider the below structure in S3. When AWS Glue Crawler scans Amazon S3 and detects multiple directories, it uses a heuristic to determine where the root for a table is in the directory structure, and which directories are partitions for the table. What solutions are possible? Select 2 options.

- ☐ A. If the schema for table1 and table2 are similar, and a single data source is set to s3://bucket01/folder1/ in AWS Glue, the crawler may create a single table with two partition columns: one partition column that contains table1 and table2, and a second partition column that contains partition1 through partition5. ✓
- ☐ B. If the schema for table1 and table2 are similar, and uses different data source is set to s3://bucket01/folder1/table1 and s3://bucket01/folder1/table2 in AWS Glue, the crawler may create two tables ✓

- ☐ C. If the schema for table1 and table2 are similar, and uses different data sources is set to s3://bucket01/folder1/table1 and s3://bucket01/folder1/table2 in AWS Glue, the crawler may create only one table
- ☐ D. If the schema for table1 and table2 are similar, and a single data source is set to s3://bucket01/folder1/ in AWS Glue, the crawler may create a single table with single partition column: one partition column that contains table1 and table2

Explanation :

Answer: A, B

A. Yes. When an AWS Glue Crawler scans Amazon S3 and detects multiple directories, it uses a heuristic to determine where the root for a table is in the directory structure, and which directories are partitions for the table. In some cases, where the schema detected in two or more directories is similar, the crawler may treat them as partitions instead of separate tables. One way to help the crawler discover individual tables is to add each table's root directory as a data store for the crawler. If the schema for table1 and table2 are similar, and a single data source is set to s3://bucket01/folder1/ in AWS Glue, the crawler may create a single table with two partition columns: one partition column that contains table1 and table2, and a second partition column that contains partition1 through partition5. <https://docs.aws.amazon.com/athena/latest/ug/glue-best-practices.html> (<https://docs.aws.amazon.com/athena/latest/ug/glue-best-practices.html>)

B. Yes. When an AWS Glue Crawler scans Amazon S3 and detects multiple directories, it uses a heuristic to determine where the root for a table is in the directory structure, and which directories are partitions for the table. In some cases, where the schema detected in two or more directories is similar, the crawler may treat them as partitions instead of separate tables. One way to help the crawler discover individual tables is to add each table's root directory as a data store for the crawler. To have the AWS Glue crawler create two separate tables, set the crawler to have two data sources, s3://bucket01/folder1/table1/ and

s3://bucket01/folder1/table2 <https://docs.aws.amazon.com/athena/latest/ug/glue-best-practices.html> (<https://docs.aws.amazon.com/athena/latest/ug/glue-best-practices.html>)

C. No. When an AWS Glue Crawler scans Amazon S3 and detects multiple directories, it uses a heuristic to determine where the root for a table is in the directory structure, and which directories are partitions for the table. In some cases, where the schema detected in two or more directories is similar, the crawler may treat them as partitions instead of separate tables. One way to help the crawler discover individual tables is to add each table's root directory as a data store for the crawler. To have the AWS Glue crawler create two separate tables, set the crawler to have two data sources, s3://bucket01/folder1/table1/ and

s3://bucket01/folder1/table2 <https://docs.aws.amazon.com/athena/latest/ug/glue-best-practices.html> (<https://docs.aws.amazon.com/athena/latest/ug/glue-best-practices.html>)

D. Yes. When an AWS Glue Crawler scans Amazon S3 and detects multiple directories, it uses a heuristic to determine where the root for a table is in the directory structure, and which directories are partitions for the table. In some cases, where the schema detected in two or more directories is similar, the crawler may treat them as partitions instead of separate tables. One way to help the crawler discover individual tables is to add each table's root directory as a data store for the crawler. If the schema for table1 and table2 are similar, and a single data source is set to s3://bucket01/folder1/ in AWS Glue, the crawler may create a single table with two partition columns: one partition column that contains table1 and table2, and a second partition column that contains partition1 through partition5. <https://docs.aws.amazon.com/athena/latest/ug/glue-best-practices.html> (<https://docs.aws.amazon.com/athena/latest/ug/glue-best-practices.html>)

HikeHills.com (HH) 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.

HH runs their entire online infrastructure on java based web applications running on AWS. The HH 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.

HH has the following requirements to process the data that is being ingested to support their enterprise search built on Elasticsearch-

Load the data (syslog and transformed data) into ES Stream

Capture transformation and delivery failures into same S3 bucket to address audit

Backup the syslog streaming data into S3 bucket Select 3 options.

- ☐ A. Streaming data can directly be delivered into Elasticsearch Domain ✓
- ☐ B. Streaming data is delivered to your S3 bucket first. Kinesis Data Firehose then issues an Amazon Elasticsearch COPY command to load data from your S3 bucket to your Amazon Elasticsearch cluster
- ☐ C. The transformation failures and delivery failures are loaded into processing-failed and errors folders in same S3 bucket ✓
- ☐ D. The transformation failures and delivery failures are loaded into transform-failed and delivery-failed folders in same S3 bucket
- ☐ E. when ES is selected as destination, and Source record S3 backup is enabled, and Backup S3 Bucket is defined, untransformed incoming data can be delivered to a separate S3 bucket ✓
- ☐ F. S3 backups can be managed to bucket policies

Explanation :

Answer: A, C, E

A. Yes. For Amazon ES destinations, streaming data is delivered to your Amazon ES cluster, and it can optionally be backed up to your S3 bucket concurrently.. <https://docs.aws.amazon.com/firehose/latest/dev/what-is-this->

(<https://docs.aws.amazon.com/firehose/latest/dev/what-is-this-service.html#data-flow-diagrams>)
service.html#data-flow-diagrams (<https://docs.aws.amazon.com/firehose/latest/dev/what-is-this-service.html#data-flow-diagrams>)

B. No. For Amazon ES destinations, streaming data is delivered to your Amazon ES cluster, and it can optionally be backed up to your S3 bucket

concurrently. <https://docs.aws.amazon.com/firehose/latest/dev/what-is-this-service.html#data-flow-diagrams> (<https://docs.aws.amazon.com/firehose/latest/dev/what-is-this-service.html#data-flow-diagrams>)
service.html#data-flow-diagrams (<https://docs.aws.amazon.com/firehose/latest/dev/what-is-this-service.html#data-flow-diagrams>)

C. Yes. when S3 is selected as destination, and Source record S3 backup is enabled, untransformed incoming data can be delivered to a separate S3 bucket and errors are delivered to processing-failed and errors folder in S3 bucket <https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html> (<https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html>)
<https://docs.aws.amazon.com/firehose/latest/dev/basic-deliver.html#retry>

(<https://docs.aws.amazon.com/firehose/latest/dev/basic-deliver.html#retry>)

D. No. when S3 is selected as destination, and Source record S3 backup is enabled, untransformed incoming data can be delivered to a separate S3 bucket and errors are delivered to processing-failed and errors folder in S3 bucket <https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html>

<https://docs.aws.amazon.com/firehose/latest/dev/basic-deliver.html#retry> (<https://docs.aws.amazon.com/firehose/latest/dev/basic-deliver.html#retry>)

E. Yes. when S3 is selected as destination, and Source record S3 backup is enabled, untransformed incoming data can be delivered to a separate S3 bucket

<https://docs.aws.amazon.com/firehose/latest/dev/create-destination.html#create-destination-s3> (<https://docs.aws.amazon.com/firehose/latest/dev/create-destination.html#create-destination-s3>)

F. No. Yes. when S3 is selected as destination, and Source record S3 backup is enabled, untransformed incoming data can be delivered to a separate S3 bucket

<https://docs.aws.amazon.com/firehose/latest/dev/create-destination.html#create-destination-s3> (<https://docs.aws.amazon.com/firehose/latest/dev/create-destination.html#create-destination-s3>)

Ask our Experts



QUESTION 33

UNATTEMPTED

VISUALIZATION

MSP 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. MSP Bank is hosting their existing infrastructure on AWS. MSP bank has many organizations internally and they are planning to launch a self-data discovery platform running out of AWS on QuickSight.

Data is spread across multiple data repositories on AWS. Which data repositories cannot be directly connected by QuickSight to build a Dataset? Select 3 options.

- ☐ A. Amazon Aurora
- ☐ B. Amazon Redshift
- ☐ C. Amazon Neptune ✓

- ☐ D. Amazon Redshift Spectrum
- ☐ E. Amazon DynamoDB ✓
- ☐ F. Amazon Elasticsearch ✓
- ☐ G. CSV/TSV files in S3
- ☐ H. XLSX/ELF/CLF files in S3

Explanation :

Answer: C,E,F

The following data sources can be connected through

QuickSight <https://docs.aws.amazon.com/quicksight/latest/user/supported-data-sources.html>
 (https://docs.aws.amazon.com/quicksight/latest/user/supported-data-sources.html)sources.html
 (http://Answer: C,E,F The following data sources can be connected through QuickSight
<https://docs.aws.amazon.com/quicksight/latest/user/supported-data-sources.html>)

Ask our Experts



QUESTION 34

UNATTEMPTED

DATA SECURITY

HikeHills.com (HH) 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.

HH runs their entire online infrastructure on multiple Java-based web applications and other web framework applications running on AWS. The HH is capturing click stream data and uses a custom-built 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 HH 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.

HH uses a VPC to host all their applications and is looking at integration of Kinesis into their web application. What is the best approach of integration to keep the traffic between your Amazon VPC and Kinesis Data Streams from leaving the Amazon network. select 1 option.

- ☒ A. Use interface VPC endpoints to keep traffic between Amazon VPC and Kinesis Data Streams from leaving the Amazon network ✓

- ☐ B. Use internet gateway, NAT device, VPN connection, or AWS Direct Connect connection to keep the traffic between Amazon VPC and Kinesis Data Streams from leaving the Amazon network
- ☐ C. By default in VPC, network traffic is always local between Amazon VPC and Kinesis Data Streams from leaving the Amazon network.
- ☐ D. 0

Explanation :

Answer: A

A. Yes. use an interface VPC endpoint to keep traffic between your Amazon VPC and Kinesis Data Streams from leaving the Amazon network. Interface VPC endpoints don't require an internet gateway, NAT device, VPN connection, or AWS Direct Connect connection. Interface VPC endpoints are powered by AWS PrivateLink, an AWS technology that enables private communication between AWS services using an elastic network interface with private IPs in your Amazon

VPC <https://docs.aws.amazon.com/streams/latest/dev/vpc.html>

(<https://docs.aws.amazon.com/streams/latest/dev/vpc.html>)

B. No. Once we are using internet gateway, NAT device, VPN connection, or AWS Direct Connect connection, the traffic will be leaving the

network <https://docs.aws.amazon.com/streams/latest/dev/vpc.html>

(<https://docs.aws.amazon.com/streams/latest/dev/vpc.html>)

C. No. this is not a default feature. It has to be linked with interface VPC endpoints <https://docs.aws.amazon.com/streams/latest/dev/vpc.html>
(<https://docs.aws.amazon.com/streams/latest/dev/vpc.html>)

Ask our Experts



QUESTION 35

UNATTEMPTED

COLLECTION

Tick-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. Tick-Bank aims to serve as a connection between digital content makers and affiliate dealers, who then promote them to clients. Tick-Bank's technology aids in payments, tax calculations and a variety of customer service tasks. Tick-Bank assists in building perceptibility and revenue making opportunities for entrepreneurs.

Tick-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. Tick-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. Tick-Bank has the following objectives in mind for the solution.

Tick-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. please detail the implementation guidelines. select 2 options.

- ☐ A. each record in a separate Kinesis Data Streams record and make one HTTP request to send it to Kinesis Data Streams ✓
- ☐ B. each HTTP request carries multiple Kinesis Stream records which is sent to Kinesis Data streams
- ☐ C. Batching is implemented as the target implementation
- ☐ D. Batching is not implemented as the target implementation ✓

Explanation :

Answer: A,D

A. Yes. When batching is not implemented, each record is treated as a separate data stream record. In this context batching is not implemented. <https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html> (<https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html>)

B. No. When batching is not implemented, each record is treated as a separate data stream record. In this context batching is not implemented. <https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html> (<https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html>)

C. No. 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. <https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html> (<https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html>)

D. Yes. 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. <https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html> (<https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html>)

Ask our Experts



QUESTION 36

UNATTEMPTED

COLLECTION

KindleYou is a location-based social search mobile app that allows users to like or dislike other users, and allows users to chat if both parties liked each other in the app. It has more than 1 billion customers across the world.

They use DynamoDB to support the mobile application and S3 to host the images and other documents shared between users. KindleYou has a large customer base spread across multiple geographic areas. Customers need to update their profile information while using the application. Being a business critical application, KindleYou wants to monitor and manage the platform efficiently. Please advice. select 3 options.

- ☐ A. Automated Monitoring of DynamoDB through CloudWatch Alarms, CloudWatch Logs, CloudWatch Events, CloudTrail log Monitoring ✓
- ☐ B. Automated Monitoring of DynamoDB through CloudTrail Alarms, CloudTrail Logs, CloudTrail Events, CloudWatch log Monitoring
- ☐ C. Manual Monitoring about recent alerts, total capacity and service health through DynamoDB Dashboard ✓
- ☐ D. Manual Monitoring about current alarms, service health status through CloudWatch ✓

Explanation :

Answer: A,C,D

AWS provides tools that you can use to monitor Amazon DynamoDB. You can configure some of these tools to do the monitoring for you; some require manual intervention.

Automated Monitoring

- CloudWatch Alarms
- CloudWatch Logs
- CloudWatch Events
- CloudTrail log Monitoring

Manual Monitoring

- Amazon DynamoDB Dashboard
- CloudWatch

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/monitoring-automated-manual.html>
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/monitoring-automated-manual.html>)manual.html
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/monitoring-automated-manual.html>)

Ask our Experts



Allianz Financial Services (AFS) is a banking group offering end-to-end banking and financial solutions in South East Asia through its consumer banking, business banking, Islamic banking, investment finance and stock broking businesses as well as unit trust and asset administration, having served the financial community over the past five decades.

AFS being one the largest banks in the region is planning to improve its segment business by launching a campaign to identify potential customers for various new products launched based on their past behavior? AFS is looking for both batch and real-time predictive analytics

AFS understands that they are losing customers to competitive banks and a campaign is launched to address customer retention. AFS has identified a long list of potential customers that have reduced transactions with the bank. AFS look at ML to advise whether a customer would continue with the bank or not? select 3 options.

- ☐ A. Amazon ML uses logistic regression algorithm through Binary classification to solve the business problem ✓
- ☐ B. Amazon ML uses multi-nominal logistic regression algorithm through multi-classclassification to solve the business problem
- ☐ C. Amazon ML uses linear regression algorithm through regression model to solve the business problem
- ☐ D. Amazon ML uses Area Under the (Receiver Operating Characteristic) Curve (AUC) to provide accuracy of the model ✓
- ☐ E. Cross-validation is a technique for evaluating ML models by training several ML models on subsets of the available input data to detect overfitting which eventually fails to generalize the pattern ✓
- ☐ F. Amazon ML uses macro-average F1 score to provide accuracy of the model
- ☐ G. Amazon ML uses standard root mean square error (RMSE) metric to provide accuracy of the model

Explanation :

Answer : A,D, E

A. Yes. ML models for binary classification problems predict a binary outcome

<https://docs.aws.amazon.com/machine-learning/latest/dg/types-of-ml-models.html>

(<https://docs.aws.amazon.com/machine-learning/latest/dg/types-of-ml-models.html>)

B. No. ML models for multiclass classification problems allow you to generate predictions for multiple classes (predict one of more than two outcomes).

<https://docs.aws.amazon.com/machine-learning/latest/dg/types-of-ml-models.html>

(<https://docs.aws.amazon.com/machine-learning/latest/dg/types-of-ml-models.html>)

C. No. ML models for regression problems predict a numeric value.

<https://docs.aws.amazon.com/machine-learning/latest/dg/types-of-ml-models.html>
(<https://docs.aws.amazon.com/machine-learning/latest/dg/types-of-ml-models.html>)

D. Yes. Amazon ML provides an industry-standard accuracy metric for binary classification models called Area Under the (Receiver Operating Characteristic) Curve (AUC).
<https://docs.aws.amazon.com/machine-learning/latest/dg/binary-model-insights.html>
(<https://docs.aws.amazon.com/machine-learning/latest/dg/binary-model-insights.html>)

E. Yes. Cross-validation is a technique for evaluating ML models by training several ML models on subsets of the available input data and evaluating them on the complementary subset of the data. Use cross-validation to detect overfitting
<https://docs.aws.amazon.com/machine-learning/latest/dg/cross-validation.html>
(<https://docs.aws.amazon.com/machine-learning/latest/dg/cross-validation.html>)

F. No. The macro-average F1 score is used to evaluate the predictive accuracy of a multiclass metric.
<https://docs.aws.amazon.com/machine-learning/latest/dg/multiclass-model-insights.html>
(<https://docs.aws.amazon.com/machine-learning/latest/dg/multiclass-model-insights.html>)

G. No. For linear regression tasks, Amazon ML uses the industry standard root mean square error (RMSE) metric.
<https://docs.aws.amazon.com/machine-learning/latest/dg/regression-model-insights.html>
(<https://docs.aws.amazon.com/machine-learning/latest/dg/regression-model-insights.html>)

Ask our Experts



QUESTION 38

UNATTEMPTED

STORAGE

Parson Fortunes Ltd is an Asian-based department store operator with an extensive network of 131 stores, spanning approximately 4.1 million square meters of retail space across cities in India, China, Vietnam, Indonesia and Myanmar.

Parson built a VPC to host their entire enterprise infrastructure on cloud. Parson 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. The files sent from their on premise data center are also hosted into S3 buckets. Parson IT team is well aware of the scalability, performance of AWS services capabilities. Parson hosts their web applications, databases and the data warehouse built on Redshift in VPC. The administrator wants to understand the system tables and views that support the team to understand the metadata of the Redshift architecture. Please advice. select 4 options.

- ☐ A. STL tables are generated from logs that have been persisted to disk to provide a history of the system. These files reside on every node in the data warehouse cluster. The STL tables take the information from the logs and format them into usable tables for system administrators. ✓
- ☐ B. STV tables are virtual tables that contain snapshots of the current system data ✓

- ☐ C. STV tables are generated from logs that have been persisted to disk to provide a history of the system. These files reside on every node in the data warehouse cluster. The STL tables take the information from the logs and format them into usable tables for system administrators.
- ☐ D. STL tables are virtual tables that contain snapshots of the current system data
- ☐ E. System tables and views do not use the same consistency model as regular tables ✓
- ☐ F. System tables and views use the same consistency model as regular tables
- ☐ G. System views that contain any reference to a transient STV table are called SVV views and references to STL tables are called SVL views. ✓

Explanation :

Answer : A, B, E, G

A. Yes. STL system tables are generated from Amazon Redshift log files to provide a history of the system.

These files reside on every node in the data warehouse cluster. The STL tables take the information from the logs and format them into usable tables for system administrators. To manage disk space, the STL log tables only retain approximately two to five days of log history, depending on log usage and available disk space. If you want to retain the log data, you will need to periodically copy it to other tables or unload it to Amazon

S3. https://docs.aws.amazon.com/redshift/latest/dg/c_intro_STL_tables.html
(https://docs.aws.amazon.com/redshift/latest/dg/c_intro_STL_tables.html)

B. Yes, STV tables are actually virtual system tables that contain snapshots of the current system data. https://docs.aws.amazon.com/redshift/latest/dg/c_intro_STV_tables.html
(https://docs.aws.amazon.com/redshift/latest/dg/c_intro_STV_tables.html)

C. no. STV tables are actually virtual system tables that contain snapshots of the current system data. https://docs.aws.amazon.com/redshift/latest/dg/c_intro_STV_tables.html
(https://docs.aws.amazon.com/redshift/latest/dg/c_intro_STV_tables.html)

D. No. STL system tables are generated from Amazon Redshift log files to provide a history of the system.

These files reside on every node in the data warehouse cluster. The STL tables take the information from the logs and format them into usable tables for system administrators. To manage disk space, the STL log tables only retain approximately two to five days of log history, depending on log usage and available disk space. If you want to retain the log data, you will need to periodically copy it to other tables or unload it to Amazon

S3. https://docs.aws.amazon.com/redshift/latest/dg/c_intro_STL_tables.html
(https://docs.aws.amazon.com/redshift/latest/dg/c_intro_STL_tables.html)

E. Yes. There are two types of system tables: STL and STV tables.

STL tables are generated from logs that have been persisted to disk to provide a history of the system. STV tables are virtual tables that contain snapshots of the current system data. They are based on transient in-memory data and are not persisted to disk-based logs or regular tables. System views that contain any reference to a transient STV table are called SVV views. Views containing only references to STL tables are called SVL views. System tables and views do not use the same consistency model as regular tables. It is important to be aware of this issue when querying them, especially for STV tables

and SVV views https://docs.aws.amazon.com/redshift/latest/dg/c_types-of-system-tables-and-views.html (https://docs.aws.amazon.com/redshift/latest/dg/c_types-of-system-tables-and-views.html)

F. No. There are two types of system tables: STL and STV tables.

STL tables are generated from logs that have been persisted to disk to provide a history of the system. STV tables are virtual tables that contain snapshots of the current system data. They are based on transient in-memory data and are not persisted to disk-based logs or regular tables. System views that contain any reference to a transient STV table are called SVV views. Views containing only references to STL tables are called SVL views. System tables and views do not use the same consistency model as regular tables. It is important to be aware of this issue when querying them, especially for STV tables and SVV views https://docs.aws.amazon.com/redshift/latest/dg/c_types-of-system-tables-and-views.html (https://docs.aws.amazon.com/redshift/latest/dg/c_types-of-system-tables-and-views.html)

G. Yes. There are two types of system tables: STL and STV tables.

STL tables are generated from logs that have been persisted to disk to provide a history of the system. STV tables are virtual tables that contain snapshots of the current system data. They are based on transient in-memory data and are not persisted to disk-based logs or regular tables. System views that contain any reference to a transient STV table are called SVV views. Views containing only references to STL tables are called SVL views. System tables and views do not use the same consistency model as regular tables. It is important to be aware of this issue when querying them, especially for STV tables and SVV views https://docs.aws.amazon.com/redshift/latest/dg/c_types-of-system-tables-and-views.html (https://docs.aws.amazon.com/redshift/latest/dg/c_types-of-system-tables-and-views.html)

Ask our Experts



QUESTION 39

UNATTEMPTED

ANALYSIS

Allianz Financial Services (AFS) is a banking group offering end-to-end banking and financial solutions in South East Asia through its consumer banking, business banking, Islamic banking, investment finance and stock broking businesses as well as unit trust and asset administration, having served the financial community over the past five decades.

AFS launched EMR cluster to support their big data analytics requirements. AFS is looking at a data warehouse and analytics environments that provide SQL on Hadoop capabilities.

Which EMR Hadoop ecosystem fulfills the requirements? select 1 option.

- ☒ A. Apache Hive ✓
- ☐ B. Apache HBase
- ☐ C. Apache HCatalog
- ☐ D. Apache Phoenix

Explanation :

Answer : A

A. Yes. Hive is an open-source, data warehouse, and analytic package that runs on top of a Hadoop cluster. Hive scripts use an SQL-like language called Hive QL (query language) that abstracts programming models and supports typical data warehouse interactions. Hive enables you to avoid the complexities of writing Tez jobs based on directed acyclic graphs (DAGs) or MapReduce programs in a lower level computer language, such as Java. Hive extends the SQL paradigm by including serialization formats. You can also customize query processing by creating table schema that matches your data, without touching the data itself. In contrast to SQL (which only supports primitive value types such as dates, numbers, and strings), values in Hive tables are structured elements, such as JSON objects, any user-defined data type, or any function written in

Java. <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hive.html>

(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hive.html>)

B. No. HBase is an open source, non-relational, distributed database developed as part of the Apache Software Foundation's Hadoop project. HBase runs on top of Hadoop Distributed File System (HDFS) to provide non- relational database capabilities for the Hadoop ecosystem. HBase works seamlessly with Hadoop, sharing its file system and serving as a direct input and output to the MapReduce framework and execution engine. HBase also integrates with Apache Hive, enabling SQL-like queries over HBase tables, joins with Hive-based tables, and support for Java Database Connectivity

(JDBC). <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hbase.html>

(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hbase.html>)

C. No. HCatalog is a tool that allows you to access Hive metastore tables within Pig, Spark SQL, and/or custom MapReduce applications. HCatalog has a REST interface and command line client that allows you to create tables or do other

operations. <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hcatalog.html>

(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hcatalog.html>)

D. No. Apache Phoenix is used for OLTP and operational analytics, allowing you to use standard SQL queries and JDBC APIs to work with an Apache HBase backing

store. <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-phoenix.html>

(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-phoenix.html>)

Ask our Experts



QUESTION 40

UNATTEMPTED

PROCESSING

Allianz Financial Services (AFS) is a banking group offering end-to-end banking and financial solutions in South East Asia through its consumer banking, business banking, Islamic banking, investment finance and stock broking businesses as well as unit trust and asset administration, having served the financial community over the past five decades.

AFS uses Redshift on AWS to fulfill the data warehousing needs and uses S3 as the staging area to host files. AFS uses other services like DynamoDB, Aurora, and Amazon RDS on remote hosts to fulfill other needs. The team needs to design Time Series tables. Please advise the best practices to load the time series tables. select 3 options.

- ☐ A. Organize data as a sequence of time-series tables for a fixed retention period and create a UNION ALL view to hide the fact that the data is stored in different tables ✓
- ☐ B. In the sequence, each table should be identical but contain data for different time ranges ✓
- ☐ C. In the sequence, each table should be identical but contain data for different specific column ranges, time based column is not required
- ☐ D. One single table with extending time ranges
- ☐ E. Use DROP TABLE instead of running a large-scale DELETE and a subsequent VACUUM process to reclaim space ✓
- ☐ F. Use DELETE and a subsequent VACUUM instead of running a DROP TABLE to reclaim space

Explanation :

Answer : A, B, E

A. Yes. If your data has a fixed retention period, we strongly recommend that you organize your data as a sequence of time-series tables. In this sequence, each table should be identical but contain data for different time ranges. create a UNION ALL view to hide the fact that the data is stored in different tables https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html) (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html)

B. Yes. If your data has a fixed retention period, we strongly recommend that you organize your data as a sequence of time-series tables. In this sequence, each table should be identical but contain data for different time ranges. create a UNION ALL view to hide the fact that the data is stored in different tables https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html) (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html)

C. No. If your data has a fixed retention period, we strongly recommend that you organize your data as a sequence of time-series tables. In this sequence, each table should be identical but contain data for different time ranges. create a UNION ALL view to hide the fact that the data is stored in different tables https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html) (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html)

D. No. If your data has a fixed retention period, we strongly recommend that you organize your data as a sequence of time-series tables. In this sequence, each table should be identical but contain data for different time ranges. create a UNION ALL view to hide the fact that the data is stored in different

tables https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html)

E. Yes. Create a UNION ALL view to hide the fact that the data is stored in different tables. When you delete old data, simply refine your UNION ALL view to remove the dropped tables. Similarly, as you load new time periods into new tables, add the new tables to the view. https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html)

F. No. Create a UNION ALL view to hide the fact that the data is stored in different tables. When you delete old data, simply refine your UNION ALL view to remove the dropped tables. Similarly, as you load new time periods into new tables, add the new tables to the view. https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-time-series-tables.html)

Ask our Experts



QUESTION 41

UNATTEMPTED

PROCESSING

MindPyramid Limited is a multinational information technology and outsourcing company headquartered in Vizag, India and New Jersey, USA. Founded in 2003, the company employs approximately 2000 employees. The company offers consulting services in cloud computing, big data and analytics. They offer services to major cloud providers including AWS. The team is working with one of the major clients having their infrastructure build on AWS. Currently the client is having lot of performance issues and wants to have a detailed understanding and recommendations. MindPyramid runs Amazon Redshift Advisor to provide recommendations to improve the performance and decrease the operating costs of the Amazon Redshift cluster. Identify the recommendations. select 3 options.

- ☐ A. Addressing uncompressed storage for a single table is a one-time optimization that requires the table to be rebuilt ✓
- ☐ B. Consider moving each actively queried database to a separate dedicated cluster. Using a separate cluster can reduce resource contention and improve query performance ✓
- ☐ C. Increasing the configured slots to match peak workload requirements redistributes underutilized memory to active slots
- ☐ D. Whenever table content changes significantly, update statistics with VACCU

- ☐ E. Modify the WLM configuration to disable SQA. Amazon Redshift uses a machine learning algorithm to analyze each eligible query
- ☐ F. Skip compression analysis using COPY ✓

Explanation :

Answer : A, B, F

A. Yes. rebuild any tables that contain uncompressed columns that are both large and frequently accessed.

<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-> (<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-analysis-recommendation>)analysis-recommendation (<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-analysis-recommendation>)

B. Yes. it enables you to set the size for each cluster for the storage, cost, and performance needs of each workload

<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-> (<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-analysis-recommendation>)analysis-recommendation (<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-analysis-recommendation>)

C. No. Consider reducing the configured slot count for queues where the slots have never been fully utilized

<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-> (<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-analysis-recommendation>)analysis-recommendation (<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-analysis-recommendation>)

D. No. recommend running ANALYZE whenever a significant number of new data rows are loaded into an existing table with COPY or INSERT commands

<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-> (<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-analysis-recommendation>)analysis-recommendation (<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-analysis-recommendation>)

E. No. Modify the WLM configuration to enable SQA. Amazon Redshift uses a machine learning algorithm to analyze each eligible query. Predictions improve as SQA learns from your query patterns

<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-> (<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-analysis-recommendation>)analysis-recommendation (<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-analysis-recommendation>)

F. Yes. Use the column ENCODE parameter when creating any tables that you load using the COPY command or Disable compression altogether by supplying the COMPUPDATE OFF parameter in the COPY command

<https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-> (https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-analysis-recommendation)analysis-recommendation (https://docs.aws.amazon.com/redshift/latest/dg/advisor-recommendations.html#skip-compression-analysis-recommendation)

Ask our Experts



QUESTION 42

UNATTEMPTED

COLLECTION

As a part of the smart city initiatives, Hyderabad (GHMC), 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. GHMC 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 Rekognition for video analytics. The producer API sends the stream of media fragments into Kinesis Video Stream and consumer API access the fragments in the order in which they were added to the stream. What contents are stored when media is added to the stream and what structured data format the media is packed into? select 3 options.

- ☐ A. Kinesis Video Streams stores incoming media data as Kinesis Video Streams chunks ✓
- ☐ B. Kinesis Video Streams stores incoming media data as stream of media fragments
- ☐ C. Each chunk consists of copy of media metadata, fragment, video stream specific metadata like fragment number and server-side and producer-side time stamps ✓
- ☐ D. Each fragment stored has unique fragment number, and producer-side and server-side time stamps for each fragment, as Kinesis Video Streams-specific metadata
- ☐ E. Media data in the fragments is packed into a structured format such as Matroska (MKV) ✓
- ☐ F. Media data in the fragments is packed into a structured format such as Advanced Systems Format (ASF)

Explanation :

Answer:A,C,E

- A. Yes. Kinesis Video Streams stores incoming media data as Kinesis Video Streams chunks <https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html>
- B. No. Kinesis Video Streams stores incoming media data as Kinesis Video Streams chunks. While putting media, into the stream the data is processed as fragments <https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html>
- C. Yes. Each chunk consists of copy of media metadata, fragment, video stream specific metadata like fragment number and server-side and producer-side time stamps <https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html>
- D. No. Each chunk consists of copy of media metadata, fragment, video stream specific metadata like fragment number and server-side and producer-side time stamps. While putting media into stream, Each fragment stored has unique fragment number, and producer-side and server-side time stamps for each fragment, as Kinesis Video Streams-specific metadata <https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html>
- E. Yes. Media data in the fragments is packed into a structured format such as Matroska (MKV) <https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html>
- F. No. Media data in the fragments is packed into a structured format such as Matroska (MKV) <https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html>

Ask our Experts



Marqueguard is a social media monitoring company headquartered in Brighton, England. Marqueguard sells three different products: Analytics, Audiences, and Insights. Marqueguard 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. Marqueguard has access to over 80 million sources.

Marque has a web application which integrates all the data mentioned above and allows users to perform a search large collections of data such as web pages, document files, forum posts, or product information. Also provides capabilities like indexing and searching both structured data and plain text. This includes -

Full text search with language-specific text processing

Boolean search

Prefix searches

Range searches

Term boosting

Faceting

Which AWS service provides these features for the web application? select 1 option.

- ☐ A. Amazon EMR clusters
- ☐ B. Amazon Elasticsearch
- ☒ C. Amazon CloudSearch ✓
- ☐ D. Amazon DynamoDB

Explanation :

Answer: C

A. No. Amazon EMR does not provide search as a managed service. Besides, Amazon EMR is a managed cluster platform that simplifies running big data frameworks, such as Apache Hadoop and Apache Spark, on AWS to process and analyze vast amounts of data. By using these frameworks and related open-source projects, such as Apache Hive and Apache Pig, you can process data for analytics purposes and business intelligence workloads. Additionally, you can use Amazon EMR to transform and move large amounts of data into and out of other AWS data stores and databases, such as Amazon Simple Storage Service (Amazon S3) and Amazon DynamoDB. <https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-what-is-emr.html> (<https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-what-is-emr.html>)

B. No. Amazon Elasticsearch Service (Amazon ES) is a managed service that makes it easy to deploy, operate, and scale Elasticsearch clusters in the AWS Cloud. Elasticsearch is a popular open-source search and analytics engine for use cases such as log analytics, real-time application monitoring, and clickstream analysis. With Amazon ES, you get direct access to the Elasticsearch APIs; existing code and applications work seamlessly with the service. <https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/what-is-amazon-elasticsearch-service.html>

C. Yes. Amazon CloudSearch is a fully managed service in the cloud that makes it easy to set up, manage, and scale a search solution for your website or application. With Amazon CloudSearch you can search large collections of data such as web pages, document files, forum posts, or product information. You can quickly add search capabilities without having to become a search expert or worry about hardware provisioning, setup, and maintenance. As your volume of data and traffic fluctuates, Amazon CloudSearch scales to meet your needs. use Amazon CloudSearch to index and search both structured data and plain text. Amazon CloudSearch features:

- Full text search with language-specific text processing
- Boolean search
- Prefix searches
- Range searches
- Term boosting
- Faceting
- Highlighting
- Autocomplete Suggestions

<https://docs.aws.amazon.com/cloudsearch/latest/developerguide/what-is-cloudsearch.html>

D. No. DynamoDB is a document management database. Though DynamoDB provides search capabilities, it is a managed NoSQL database service. Amazon DynamoDB is a fully managed NoSQL database service that provides fast and predictable performance with seamless scalability. DynamoDB lets you offload the administrative burdens of operating and scaling a distributed database, so that you don't have to worry about hardware provisioning, setup and configuration, replication, software patching, or cluster scaling. Also, DynamoDB offers encryption at rest, which eliminates the operational burden and complexity involved in protecting sensitive data <https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Introduction.html>



Tick-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. Tick-Bank aims to serve as a connection between digital content makers and affiliate dealers, who then promote them to clients. Tick-Bank's technology aids in payments, tax calculations and a variety of customer service tasks. Tick-Bank assists in building perceptibility and revenue making opportunities for entrepreneurs.

Tick-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. Tick-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. Tick-Bank has the following objectives in mind for the solution.

Tick-Bank uses KPL to process the data and KCL library to consume the records. Tick-Bank has multiple data streams supporting various business processes. Since the amount of data

generated by stream is very small, planning to use API operation PutRecords to send multiple Kinesis Data Streams records to one or more shards in your Kinesis data stream. Please detail the specifications of implementation.

Select 3 options.

- ☐ A. Batching of records is part of implementation. ✓
- ☐ B. Batching of records is not a part of implementation.
- ☐ C. Perform aggregation by storing multiple records within a single Kinesis Data Streams record. ✓
- ☐ D. Perform collection by send multiple Kinesis Data Streams records to one or more shards in your Kinesis data stream.
- ☐ E. Aggregation and Collection can coexist and can be turned on or off independently of one another. ✓

Explanation :

Answer: A,C,E

A. Yes. Batching refers to performing a single action on multiple items instead of repeatedly performing the action on each individual item. Batching of records is part of implementation. The KPL supports two types of batching:

- Aggregation – Storing multiple records within a single Kinesis Data Streams record.
- Collection – Using the API operation PutRecords to send multiple Kinesis Data Streams records to one or more shards in your Kinesis data stream.
<https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html>)

B. No. Batching refers to performing a single action on multiple items instead of repeatedly performing the action on each individual item. Batching of records is part of implementation. The KPL supports two types of batching:

- Aggregation – Storing multiple records within a single Kinesis Data Streams record.
- Collection – Using the API operation PutRecords to send multiple Kinesis Data Streams records to one or more shards in your Kinesis data stream.
<https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html>)

C. Yes. Aggregation refers to the storage of multiple records in a Kinesis Data Streams record. Aggregation allows customers to increase the number of records sent per API call, which effectively increases producer throughput. <https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html> (<https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html>)

D. No. Collection refers to batching multiple Kinesis Data Streams records and sending them in a single HTTP request with a call to the API operation PutRecords, instead of sending each Kinesis Data Streams record in its own HTTP request.

This increases throughput compared to using no collection because it reduces the overhead of making many separate HTTP requests. In fact, PutRecords itself was specifically designed for this purpose.

Collection differs from aggregation in that it is working with groups of Kinesis Data Streams records.

The Kinesis Data Streams records being collected can still contain multiple records from the user. <https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html>
(<https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html>)

E. Yes. Aggregation and Collection can coexist and can be turned on or off independently of one another. By default, both are turned on. <https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html> (<https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html>)

Ask our Experts



QUESTION 45

UNATTEMPTED

ANALYSIS

HikeHills.com (HH) 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.

HH runs their entire online infrastructure on Java-based web applications running on AWS. The HH is capturing click stream data and uses a custom-built recommendation engine to recommend products which eventually improve sales, understand customer preferences and already using AWS Streaming capabilities to collect events and transaction logs and process the stream.

HH is using Kinesis Analytics to build SQL querying capability on streaming and planning to use windowed queries to process the data. What kind of window queries need to be used to aggregate data using distinct time-based windows that open and close at regular intervals. Select 1 option.

- ☐ A. Stagger Windows queries
- ☒ B. Tumbling Windows queries ✓
- ☐ C. Sliding windows queries
- ☐ D. Continuous queries

Explanation :

Answer: B

A. 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 <https://docs.aws.amazon.com/kinesisanalytics/latest/dev/stagger-window-concepts.html> (<https://docs.aws.amazon.com/kinesisanalytics/latest/dev/stagger-window-concepts.html>)

B. Tumbling Windows query, A query that aggregates data using distinct time-based windows that open and close at regular intervals. <https://docs.aws.amazon.com/kinesisanalytics/latest/dev/tumbling-window-concepts.html> (<https://docs.aws.amazon.com/kinesisanalytics/latest/dev/tumbling-window-concepts.html>)

C. Sliding windows query, A query that aggregates data continuously, using a fixed time or rowcount interval. <https://docs.aws.amazon.com/kinesisanalytics/latest/dev/sliding-window-concepts.html> (<https://docs.aws.amazon.com/kinesisanalytics/latest/dev/sliding-window-concepts.html>)

D. 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. <https://docs.aws.amazon.com/kinesisanalytics/latest/dev/continuous-queries-concepts.html> (<https://docs.aws.amazon.com/kinesisanalytics/latest/dev/continuous-queries-concepts.html>)

Tick-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. Tick-Bank aims to serve as a connection between digital content makers and affiliate dealers, who then promote them to clients. Tick-Bank's technology aids in payments, tax calculations and a variety of customer service tasks. Tick-Bank assists in building perceptibility and revenue making opportunities for entrepreneurs.

Tick-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. Tick-Bank is looking to enable web-site traffic analytics there by understanding user navigational behavior, preferences and other click related info. Tick-Bank is also looking at improving operations ingesting monitoring logs.

Since the amount of data, that is being processing is very large, Tick-Bank prefers data compression, data transformation when processing and considers Kinesis firehose to process the streams. What kind of producers can produce data into kinesis firehose? select 6 options.

- ☐ A. Write to Kinesis Firehose using KPL Library of kinesis Streams
- ☐ B. Write to Kinesis Firehose using Kinesis Streams ✓
- ☐ C. Write to Kinesis Firehose using AWS SDK ✓
- ☐ D. loudWatch Logs can be used as a Kinesis firehose Source ✓
- ☐ E. CloudWatch Events can be used as a Kinesis firehose Source ✓
- ☐ F. Cloud Advisor can be used as a Kinesis firehose Source
- ☐ G. AWS IOT can be used as a Kinesis firehose Source ✓
- ☐ H. Write to Kinesis Firehose using Kinesis Agent ✓

Explanation :

Answer: B,C,D,E,G, H

Data can be sent to Kinesis firehose stream using Kinesis data stream, the Kinesis Agent, Kinesis Data Firehose API using the AWS SDK. Also AWS CloudWatch Logs, CloudWatch Events, or AWS IoT as Firehose data source <https://docs.aws.amazon.com/firehose/latest/dev/basic-write.html> (<https://docs.aws.amazon.com/firehose/latest/dev/basic-write.html>)

Ask our Experts



MSP 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. MSP Bank is hosting their existing infrastructure on AWS. MSP bank has many segments internally and they are planning to launch a self-data discovery platform running out of AWS on QuickSight. Using QuickSight, dataset is being created based on their existing RDS data source. Multiple tables have been identified which are joined and accessed. A permission file 'permissions.txt' is created as below which outlays rules. Dataset 'transaction_details' dataset of 50 GB size has been created and joined with the permissions file.

UserName	Region	Segment
AlejandroRosalez	EMEA	Enterprise, SMB, Startup
MarthaRivera	US	Enterprise
NikhilJayashankar	US	SMB, Startup
PauloSantos	US	Startup
SaanviSarkar	APAC	Enterprise, SMB
sales-tps@example.com		
ZhangWei	APAC	Enterprise, Startup

To use the permission.txt rules to allow access the transaction_details data, 'Grant access to data set' is chosen

What data does AlejandroRosalez, EMEA and sales-tps@emaple.com (mailto:sales-tps@emaple.com) can see? select 2 options.

- ☐ A. AlejandroRosalez, EMEA can see all EMEA Enterprise, SMB, and Startup data ✓
- ☐ B. AlejandroRosalez, EMEA cannot see any of the EMEA Enterprise, SMB, and Startup data
- ☐ C. sales-tps@emaple.com can see all rows ✓
- ☐ D. sales-tps@emaple.com cannot see any rows

Explanation :

Answer: A,C

A. Yes. AlejandroRosalez, EMEA can see all EMEA Enterprise, SMB, and Startup data

<https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html> (<https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html>) (<https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html>)

B. No. AlejandroRosalez, EMEA can see all of the EMEA Enterprise, SMB, and Startup data

<https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html> (<https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html>) (<https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html>)

C. Yes. sales-tps@emaple.com (mailto:sales-tps@emaple.com) cannot see any rows

<https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html> (<https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html>) (<https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html>)

D. No. sales-tps@emaple.com (mailto:sales-tps@emaple.com) cannot see any rows

<https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html> (<https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html>) (<https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html>)

Ask our Experts



QUESTION 48

UNATTEMPTED

DATA SECURITY

Allianz Financial Services (AFS) is a banking group offering end-to-end banking and financial solutions in South East Asia through its consumer banking, business banking, Islamic banking, investment finance and stock broking businesses as well as unit trust and asset administration, having served the financial community over the past five decades.

AFS uses Redshift on AWS to fulfill the data warehousing needs and uses S3 as the staging area to host files. AFS uses other services like DynamoDB, Aurora, and Amazon RDS on remote hosts to fulfill other needs. AFS want to implement Redshift security end to end. How can this be achieved? select 2 options.

- ☐ A. Access to your Amazon Redshift Management Console is controlled by your AWS account privileges ✓
- ☐ B. Define a cluster security group and associate it with a cluster to control access to specific Amazon Redshift resources ✓
- ☐ C. To encrypt the connection between your SQL client and your cluster, enable cluster encryption when you launch the cluster

- ☐ D. To encrypt the data in all your user-created tables, you can use secure sockets layer (SSL) encryption

Explanation :

Answer : A,B

Amazon Redshift database security is distinct from other types of Amazon Redshift security. In addition to database security, which is described in this section, Amazon Redshift provides these features to manage security:

- Sign-in credentials – Access to your Amazon Redshift Management Console is controlled by your AWS account privileges. For more information, see Sign-In Credentials.
- Access management – To control access to specific Amazon Redshift resources, you define AWS Identity and Access Management (IAM) accounts. For more information, see Controlling Access to Amazon Redshift Resources.
- Cluster security groups – To grant other users inbound access to an Amazon Redshift cluster, you define a cluster security group and associate it with a cluster. For more information, see Amazon Redshift Cluster Security Groups.
- VPC – To protect access to your cluster by using a virtual networking environment, you can launch your cluster in an Amazon Virtual Private Cloud (VPC). For more information, see Managing Clusters in Virtual Private Cloud(VPC).
- Cluster encryption – To encrypt the data in all your user-created tables, you can enable cluster encryption when you launch the cluster. For more information, see Amazon Redshift Clusters.
- SSL connections – To encrypt the connection between your SQL client and your cluster, you can use secure sockets layer (SSL) encryption. For more information, see Connect to Your Cluster Using SSL.
- Load data encryption – To encrypt your table load data files when you upload them to Amazon S3, you can use either server-side encryption or client-side encryption. When you load from server-side encrypted data, Amazon S3 handles decryption transparently. When you load from client-side encrypted data, the Amazon Redshift COPY command decrypts the data as it loads the table. For more information, see Uploading Encrypted Data to Amazon S3.
- Data in transit – To protect your data in transit within the AWS cloud, Amazon Redshift uses hardware accelerated SSL to communicate with Amazon S3 or Amazon DynamoDB for COPY, UNLOAD, backup, and restore operations.https://docs.aws.amazon.com/redshift/latest/dg/c_security-overview.html (https://docs.aws.amazon.com/redshift/latest/dg/c_security-overview.html)

Ask our Experts



QUESTION 49

UNATTEMPTED

ANALYSIS

Allianz Financial Services (AFS) is a banking group offering end-to-end banking and financial solutions in South East Asia through its consumer banking, business banking, Islamic banking, investment finance and stock broking businesses as well as unit trust and asset administration, having served the financial community over the past five decades.

AFS launched EMR cluster to support their big data analytics requirements. AFS has multiple data sources built out of S3, SQL databases, MongoDB, Redis, RDS, other file systems. AFS is looking for fast SQL query engine designed for interactive analytic queries over large datasets from multiple sources

Which EMR Hadoop ecosystem fulfills the requirements? select 1 option.

- ☐ A. Apache Hive
- ☐ B. Apache HBase
- ☐ C. Apache HCatalog
- ☒ D. Apache Presto ✓

Explanation :

Answer : D

A.No. Hive is an open-source, data warehouse, and analytic package that runs on top of a Hadoop cluster. Hive scripts use an SQL-like language called Hive QL (query language) that abstracts programming models and supports typical data warehouse interactions. Hive enables you to avoid the complexities of writing Tez jobs based on directed acyclic graphs (DAGs) or MapReduce programs in a lower level computer language, such as Java. Hive extends the SQL paradigm by including serialization formats. You can also customize query processing by creating table schema that matches your data, without touching the data itself. In contrast to SQL (which only supports primitive value types such as dates, numbers, and strings), values in Hive tables are structured elements, such as JSON objects, any user-defined data type, or any function written in

Java. <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hive.html>

(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hive.html>)

B.No. HBase is an open source, non-relational, distributed database developed as part of the Apache Software Foundation's Hadoop project. HBase runs on top of Hadoop Distributed File System (HDFS) to provide non- relational database capabilities for the Hadoop ecosystem. HBase works seamlessly with Hadoop, sharing its file system and serving as a direct input and output to the MapReduce framework and execution engine. HBase also integrates with Apache Hive, enabling SQL-like queries over HBase tables, joins with Hive-based tables, and support for Java Database Connectivity (JDBC). <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hbase.html>

(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hbase.html>)

C. No. HCatalog is a tool that allows you to access Hive metastore tables within Pig, Spark SQL, and/or custom MapReduce applications. HCatalog has a REST interface and command line client that allows you to create tables or do other

operations. <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hcatalog.html>

(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hcatalog.html>)

D. Yes. Presto is a fast SQL query engine designed for interactive analytic queries over large datasets from multiple sources <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-presto.html>

(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-presto.html>)

Ask our Experts



MSP 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. MSP Bank is hosting their existing infrastructure on AWS. MSP 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 and security. What are the supported methods of managing user accounts in QuickSight Enterprise Edition? select 3 options.

- ☐ A. Add and remove Microsoft Active Directory groups to create and deactivate user accounts ✓
- ☐ B. Federated Logins ✓
- ☐ C. Inviting Users by Email ✓
- ☐ D. Login through IAM

Explanation :

Answer: A, B, C

In Enterprise edition, you can manage users through any of the following:

- Active Directory. You can add and remove Microsoft Active Directory directory groups to create and deactivate user accounts. You can access the directory groups directly or by using the AD Connector.
- Federated logins.
- Inviting users by email.

<https://docs.aws.amazon.com/quicksight/latest/user/managing-users-enterprise.html>
(<https://docs.aws.amazon.com/quicksight/latest/user/managing-users-enterprise.html>)

Ask our Experts



MSP 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. MSP Bank is hosting their existing infrastructure on premise. MSP bank has lot of web applications, databases and data warehouses, NoSQL databases, and other types of data stores. MSP Bank is working with AWS to migrate their existing platform onto AWS.

Currently MSP has the following databases to support OLTP workloads

Microsoft SQL Server

Oracle

Postgres

MySQL

WH Workload

Greenplum.

The team finalized Aurora to migrate OLTP workloads and Greenplum onto Redshift.

What service does AWS propose to convert the schema's from one format to other?
select 1 option.

- ☐ A. AWS Database Migration Service (AWS DMS)
- ☐ B. AWS Data Sync
- ☐ C. AWS Migration Hub
- ☒ D. AWS Schema Conversion ✓

Explanation :

Answer : D

A.No. AWS Database Migration Service (AWS DMS) is a cloud service that makes it easy to migrate relational databases, data warehouses, NoSQL databases, and other types of data stores. You can use AWS DMS to migrate your data into the AWS Cloud, between on-premises instances (through an AWS Cloud setup), or between combinations of cloud and on-premises setups.

With AWS DMS, you can perform one-time migrations, and you can replicate ongoing changes to keep sources and targets in sync. If you want to change database engines, you can use the AWS Schema Conversion Tool (AWS SCT) to translate your database schema to the new platform. You then use AWS DMS to migrate the data. Because AWS DMS is a part of the AWS Cloud, you get the cost efficiency, speed to market, security, and flexibility that AWS services offer.

<https://docs.aws.amazon.com/dms/latest/userguide/Welcome.html>

(<https://docs.aws.amazon.com/dms/latest/userguide/Welcome.html>)

B.No. AWS DataSync is a data transfer service that simplifies, automates, and accelerates moving and replicating data between on-premises storage systems and AWS storage services over the internet or AWS Direct Connect. As a fully managed service, DataSync removes the need to modify applications, develop scripts, or manage infrastructure.

DataSync currently supports data transfer between Network File System (NFS) and Amazon Elastic File System (Amazon EFS), or Amazon Simple Storage Service (Amazon

S3). <https://docs.aws.amazon.com/datasync/latest/userguide/what-is-datasync.html>

(<https://docs.aws.amazon.com/datasync/latest/userguide/what-is-datasync.html>)

C.No. AWS Migration Hub provides a single place to discover your existing servers, plan migrations, and track the status of each application migration. The AWS Migration Hub provides visibility into your application portfolio and streamlines planning and tracking. You can see the status of the servers and databases that make up each of the applications you are migrating regardless of which migration tool you are using. <https://docs.aws.amazon.com/migrationhub/latest/ug/whatishub.html>

(<https://docs.aws.amazon.com/migrationhub/latest/ug/whatishub.html>)

D. Yes. AWS Schema Conversion Tool (AWS SCT) to convert your existing database schema from one database engine to another. You can convert relational OLTP schema, or data warehouse schema. Your converted schema is suitable for an Amazon Relational Database Service (Amazon RDS) MySQL DB instance, an Amazon Aurora DB cluster, an Amazon RDS PostgreSQL DB instance, or an Amazon Redshift cluster. The converted schema can also be used with a database on an Amazon EC2 instance or stored as data on an Amazon S3 bucket https://docs.aws.amazon.com/SchemaConversionTool/latest/userguide/CHAP_Welcome.html (https://docs.aws.amazon.com/SchemaConversionTool/latest/userguide/CHAP_Welcome.html)

Ask our Experts



QUESTION 52

UNATTEMPTED

ANALYSIS

Marqueguard is a social media monitoring company headquartered in Brighton, England. Marqueguard sells three different products: Analytics, Audiences, and Insights. Marqueguard 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. Marqueguard has access to over 80 million sources.

Marqueguard has a large team of data scientists and want to provide fully managed machine learning service data scientists and developers can quickly and easily build and train machine learning models, and then directly deploy them into a production-ready hosted environment.

What service can provide this capability? select 1 option.

- ☐ A. Amazon Comprehend
- ☐ B. Amazon Rekognition
- ☐ C. Amazon Polly
- ☐ D. Amazon SageMaker ✓

Explanation :

Answer : D

A. No. Amazon Comprehend uses natural language processing (NLP) to extract insights about the content of documents. Amazon Comprehend processes any text file in UTF-8 format. It develops insights by recognizing the entities, key phrases, language, sentiments, and other common elements in

a document. Use Amazon Comprehend to create new products based on understanding the structure of documents. For example, using Amazon Comprehend you can search social networking feeds for mentions of products or scan an entire document repository for key phrases. <https://docs.aws.amazon.com/comprehend/latest/dg/what-is.html>
(<https://docs.aws.amazon.com/comprehend/latest/dg/what-is.html>)

B. No. Amazon Rekognition makes it easy to add image and video analysis to your applications. You just provide an image or video to the Rekognition API, and the service can identify objects, people, text, scenes, and activities. It can detect any inappropriate content as well. Amazon Rekognition 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 public safety. <https://docs.aws.amazon.com/rekognition/latest/dg/what-is.html>
(<https://docs.aws.amazon.com/rekognition/latest/dg/what-is.html>)

C. No. Amazon Polly is a cloud service that converts text into lifelike speech. You can use Amazon Polly to develop applications that increase engagement and accessibility. Amazon Polly supports multiple languages and includes a variety of lifelike voices, so you can build speech-enabled applications that work in multiple locations and use the ideal voice for your customers. With Amazon Polly, you only pay for the text you synthesize. You can also cache and replay Amazon Polly's generated speech at no additional cost. <https://docs.aws.amazon.com/polly/latest/dg/what-is.html>
(<https://docs.aws.amazon.com/polly/latest/dg/what-is.html>)

D. Yes. Amazon SageMaker is a fully managed machine learning service. With Amazon SageMaker, data scientists and developers can quickly and easily build and train machine learning models, and then directly deploy them into a production-ready hosted environment. It provides an integrated Jupyter authoring notebook instance for easy access to your data sources for exploration and analysis, so you don't have to manage servers. It also provides common machine learning algorithms that are optimized to run efficiently against extremely large data in a distributed environment <https://docs.aws.amazon.com/sagemaker/latest/dg/whatis.html>
(<https://docs.aws.amazon.com/sagemaker/latest/dg/whatis.html>)

Ask our Experts



QUESTION 53

UNATTEMPTED

STORAGE

KindleYou is a location-based social search mobile app that allows users to like or dislike other users, and allows users to chat if both parties liked each other in the app. It has more than 1 billion customers across the world.

They use DynamoDB to support the mobile application and S3 to host the images and other documents shared between users.

The application tracks # of photos uploaded for a specific user real-time. select 1 option.

- ☒ A. Using Sparse Indexes ✓
- ☐ B. Using Aggregation
- ☐ C. Using Global Secondary Index Overloading

☐ D. Using Global Secondary Index Sharding

Explanation :

Answer: A

A. Yes. Add a Boolean attribute 'Is_Top_Friendly' to the table and add a sparse index. Sparse indexes are useful for queries over a small subsection of a table.

B. No. Aggregation of data for maintaining near real-time aggregations and key metrics on top of rapidly changing data is becoming increasingly valuable to businesses for making rapid decisions

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-gsi-aggregation.html>
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-gsi-aggregation.html>)

C. No. Overloading GSI only addresses adding different fields as to cater different queries

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-gsi-overloading.html>
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-gsi-overloading.html>)

D. No. To enable selective queries across the entire key space, you can use write sharding by adding an attribute containing a (0-N) value to every item that you will use for the global secondary index partition key

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-indexes-gsi-sharding.html>
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-indexes-gsi-sharding.html>)sharding.html
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-indexes-gsi-sharding.html>)

Ask our Experts



QUESTION 54

UNATTEMPTED

STORAGE

MSP 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. MSP Bank is hosting their existing infrastructure on premise.

MSP bank has lot of web applications, databases and data warehouse built on Teradata, NoSQL databases, and other types of data stores. MSP bank has lot of data assets in terms of logs, documents; excel files, CSV files, PDF documents and others. MSP Bank wants to provide seamless integration with data security features between their on-premises IT environment and the AWS storage infrastructure. MSP bank considers AWS Storage Gateway to address the integration.

MSP Bank looking at interface to cost-effectively and durably archive backup data in Glacier which provides a virtual tape infrastructure that scales seamlessly with your business needs and eliminates the operational burden of provisioning, scaling, and maintaining a physical tape infrastructure. select 1 option.

- ☐ A. Volume Gateway
- ☐ B. File Gateway
- ☒ C. Tape Gateway ✓
- ☐ D. All the Above

Explanation :

Answer: B

A. No. A volume gateway provides cloud-backed storage volumes that you can mount as Internet Small Computer System Interface (iSCSI) devices from your on-premises application servers. The gateway supports the following volume configurations:

- **Cached volumes** – You store your data in Amazon Simple Storage Service (Amazon S3) and retain a copy of frequently accessed data subsets locally. Cached volumes offer a substantial cost savings on primary storage and minimize the need to scale your storage on-premises. You also retain low-latency access to your frequently accessed data.
- **Stored volumes** – If you need low-latency access to your entire dataset, first configure your on-premises gateway to store all your data locally. Then asynchronously back up point-in-time snapshots of this data to Amazon S3. This configuration provides durable and inexpensive offsite backups that you can recover to your local data center or Amazon EC2. For example, if you need replacement capacity for disaster recovery, you can recover the backups to Amazon EC2. <https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html> (<https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html>)

B. No. A file gateway supports a file interface into Amazon Simple Storage Service (Amazon S3) and combines a service and a virtual software appliance. By using this combination, store and retrieve objects in Amazon S3 using industry-standard file protocols such as Network File System (NFS) and Server Message Block (SMB). The software appliance, or gateway, is deployed into your on-premises environment as a virtual machine (VM) running on VMware ESXi or Microsoft Hyper-V hypervisor. The gateway provides access to objects in S3 as files or file share mount points. With a file gateway, you can do the following:

- You can store and retrieve files directly using the NFS version 3 or 4.1 protocol.
- You can store and retrieve files directly using the SMB file system version, 2 and 3 protocol.
- You can access your data directly in Amazon S3 from any AWS Cloud application or service.
- You can manage your Amazon S3 data using lifecycle policies, cross-region replication, and versioning. You can think of a file gateway as a file system mount on S3. <https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html> (<https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html>)

C. Yes. With a tape gateway, cost-effectively and durably archive backup data in Glacier. A tape gateway provides a virtual tape infrastructure that scales seamlessly with your business needs and eliminates the operational burden of provisioning, scaling, and maintaining a physical tape infrastructure. <https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html> (<https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html>)


Ask our Experts



Marqueguard is a social media monitoring company headquartered in Brighton, England. Marqueguard sells three different products: Analytics, Audiences, and Insights. Marqueguard 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. Marqueguard has access to over 80 million sources.

Marqueguard hosted their entire infrastructure on AWS and uses Data Pipeline as data integration mechanism. Marqueguard hosts different web services in terms of data includes DynamoDB, RDS, RedShift and S3. When AWS Data Pipeline runs a pipeline, it compiles the pipeline components to create a set of actionable Amazon EC2 instances. Each instance contains all the information for performing a specific task. The complete set of instances is the to-do list of the pipeline. AWS Data Pipeline hands the instances out to task runners to process. please identify the key artefacts. select 2 options.

- ☐ **A. uses only the following instance types of EC2 instances besides default instance types to run pipeline work activities**
 - General Purpose
 - Compute optimized
 - Memory optimized
 - Storage optimized
- ☐ **B. uses only the following instance types of EC2 instances for EMR clusters besides default instance types to run pipeline work activities**
 - General Purpose
 - Compute optimized
 - Memory optimized
 - Storage optimized
- ☐ **C. uses only the following instance types of EC2 instances clusters besides default instance types to run pipeline work activities**
 - General Purpose
 - Compute optimized
 - Memory optimized
 - Storage optimized
 - Accelerated computing

- ☐ D. uses only the following instance types of EC2 instances for EMR clusters besides default instance types to run pipeline work activities

General Purpose

Compute optimized

Memory optimized

Storage optimized

Accelerated computing



Explanation :

Answer: A,D

A. Yes. uses only the following instance types of EC2 instances clusters besides default instance types to run pipeline work activities

- General Purpose
- Compute optimized
- Memory optimized
- Storage

optimized<https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-ec2-supported-instance-types.html>

(<https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-ec2-supported-instance-types.html>)

B. No. uses only the following instance types of EC2 instances for EMR clusters besides default instance types to run pipeline work activities

- General Purpose
- Compute optimized
- Memory optimized
- Storage optimized
- Accelerated

computing<https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-emr-supported-instance-types.html>

(<https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-emr-supported-instance-types.html>)

C. No. uses only the following instance types of EC2 instances clusters besides default instance types to run pipeline work activities

- General Purpose
- Compute optimized
- Memory optimized
- Storage

optimized<https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-ec2-supported-instance-types.html>
(<https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-ec2-supported-instance-types.html>)

D. Yes. uses only the following instance types of EC2 instances for EMR clusters besides default instance types to run pipeline work activities

- General Purpose
- Compute optimized
- Memory optimized
- Storage optimized
- Accelerated

computing<https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-emr-supported-instance-types.html>
(<https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-emr-supported-instance-types.html>)

Ask our Experts



QUESTION 56

UNATTEMPTED

STORAGE

MSP 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. MSP Bank is hosting their existing infrastructure on premise.

MSP bank has lot of web applications, databases and data warehouse built on Teradata, NoSQL databases, and other types of data stores. MSP bank has lot of data assets in terms of logs, documents; excel files, CSV files, PDF documents and others. MSP Bank wants to provide seamless integration with data security features between their on-premises IT environment and the AWS storage infrastructure. MSP bank considers AWS Storage Gateway to address the integration.

MSP Bank looking at interface into Amazon Simple Storage Service (Amazon S3) and combines a service and a virtual software appliance through which the team can store and retrieve objects in Amazon S3 using industry-standard file protocols such as Network File System (NFS) and Server Message Block (SMB). The gateway provides access to objects in S3 as files or file share mount points. Which option of Storage Gateway supports this feature? select 1 option.

- ☐ A. Volume Gateway
- ☒ B. File Gateway ✓
- ☐ C. Tape Gateway
- ☐ D. 0

Explanation :

Answer: B

A. No. A volume gateway provides cloud-backed storage volumes that you can mount as Internet Small Computer System Interface (iSCSI) devices from your on-premises application servers. The gateway supports the following volume configurations:

- **Cached volumes** – You store your data in Amazon Simple Storage Service (Amazon S3) and retain a copy of frequently accessed data subsets locally. Cached volumes offer a substantial cost savings on primary storage and minimize the need to scale your storage on-premises. You also retain low-latency access to your frequently accessed data.
- **Stored volumes** – If you need low-latency access to your entire dataset, first configure your on-premises gateway to store all your data locally. Then asynchronously back up point-in-time snapshots of this data to Amazon S3. This configuration provides durable and inexpensive offsite backups that you can recover to your local data center or Amazon EC2. For example, if you need replacement capacity for disaster recovery, you can recover the backups to Amazon EC2. <https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html> (<https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html>)

B. Yes. A file gateway supports a file interface into Amazon Simple Storage Service (Amazon S3) and combines a service and a virtual software appliance. By using this combination, store and retrieve objects in Amazon S3 using industry-standard file protocols such as Network File System (NFS) and Server Message Block (SMB). The software appliance, or gateway, is deployed into your on-premises environment as a virtual machine (VM) running on VMware ESXi or Microsoft Hyper-V hypervisor. The gateway provides access to objects in S3 as files or file share mount points. With a file gateway, you can do the following:

- You can store and retrieve files directly using the NFS version 3 or 4.1 protocol.
- You can store and retrieve files directly using the SMB file system version, 2 and 3 protocol.
- You can access your data directly in Amazon S3 from any AWS Cloud application or service.
- You can manage your Amazon S3 data using lifecycle policies, cross-region replication, and versioning. You can think of a file gateway as a file system mount on S3. <https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html> (<https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html>)

C. No. With a tape gateway, cost-effectively and durably archive backup data in Glacier. A tape gateway provides a virtual tape infrastructure that scales seamlessly with your business needs and eliminates the operational burden of provisioning, scaling, and maintaining a physical tape infrastructure. <https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html> (<https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html>)

As a part of the smart city initiatives, Hyderabad (GHMC), 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. GHMC 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 Rekognition for video analytics. GHMC is looking at enabling encryption for their streams of data at rest . How can this be enabled? select 2 options.

- ☒ A. Server-side encryption is always enabled on Kinesis Video Streams ✓
- ☐ B. Server-side encryption is always disabled on Kinesis Video Streams. It needs to explicitly enable.
- ☒ C. Server-side encryption using AWS Key Management Service (AWS KMS) Customer Master Key (CMK) allows encrypting your data at rest in Amazon Kinesis Video Streams ✓
- ☐ D. Client-side encryption using AWS Key Management Service (AWS KMS) Client Master Key (CMK) allows encrypting your data at rest in Amazon Kinesis Video Streams
- ☐ E. Client -side encryption is always enabled on Kinesis Video Streams by default
- ☐ F. Server side encryption addresses encryption of data both in transit and at rest

Explanation :

Answer: A, C

A. Yes. Server-side encryption is always enabled on Kinesis Video Streams [https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-\(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-kms.html#creating-using-sse-master-keys-akvs\)kms.html#creating-using-sse-master-keys-akvs](https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-kms.html#creating-using-sse-master-keys-akvs)kms.html#creating-using-sse-master-keys-akvs)

B. No. Server-side encryption is always enabled on Kinesis Video Streams [https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-\(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-kms.html#creating-using-sse-master-keys-akvs\)kms.html#creating-using-sse-master-keys-akvs](https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-kms.html#creating-using-sse-master-keys-akvs)kms.html#creating-using-sse-master-keys-akvs)

- C. Yes. Server-side encryption using AWS Key Management Service (AWS KMS) Customer Master Key (CMK) allows encrypting your data at rest in Amazon Kinesis Video Streams [https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-\(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-kms.html#creating-using-sse-master-keys-akvs\)](https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-kms.html#creating-using-sse-master-keys-akvs)kms.html#creating-using-sse-master-keys-akvs)
- D. No. Server-side encryption using AWS Key Management Service (AWS KMS) Customer Master Key (CMK) allows encrypting your data at rest in Amazon Kinesis Video Streams [https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-\(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-kms.html#creating-using-sse-master-keys-akvs\)](https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-kms.html#creating-using-sse-master-keys-akvs)kms.html#creating-using-sse-master-keys-akvs)
- E. No. Client -side encryption is not enabled on Kinesis Video Streams [https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-\(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-kms.html#creating-using-sse-master-keys-akvs\)](https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-kms.html#creating-using-sse-master-keys-akvs)kms.html#creating-using-sse-master-keys-akvs)
- F. No. Server side encryption addresses encryption of data only at rest [https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-\(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-kms.html#creating-using-sse-master-keys-akvs\)](https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-kms.html#creating-using-sse-master-keys-akvs)kms.html#creating-using-sse-master-keys-akvs)

Ask our Experts



QUESTION 58

UNATTEMPTED

VISUALIZATION

MSP 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. MSP

Bank is hosting their existing infrastructure on AWS. MSP 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. The team is planning to use a Chart to visualize a two or three measures for a dimension in the form of bubbles. select 1 option.

☒ A. Scatter Plot ✓

- ☐ B. Heat Maps
- ☐ C. Pie Chart
- ☐ D. Tree Map

Explanation :

Answer: A

A. Yes. Use scatter plots to visualize two or three measures for a dimension.

Each bubble on the scatter plot represents one item in the dimension. The X and Y axes represent two different measures that apply to the dimension. A bubble appears on the chart at the point where the values for the two measures for an item in the dimension intersect. Optionally, you can also use bubble size to represent an additional measure.

<https://docs.aws.amazon.com/quicksight/latest/user/scatter-plot.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/scatter-plot.html>)

B. No. Use heat maps to show a measure for the intersection of two dimensions, with color-coding to easily differentiate where values fall in the range. Heat maps can also be used to show the count of values for the intersection of the two dimensions.

<https://docs.aws.amazon.com/quicksight/latest/user/heat-map.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/heat-map.html>)

C. No. Use pie charts to compare values for items in a dimension.

Each wedge in a pie chart represents one item in the dimension. Wedge size represents the proportion of the value for the selected measure that the item represents compared to the whole for the dimension. Pie charts are best when precision isn't important and there are few items in the dimension.

<https://docs.aws.amazon.com/quicksight/latest/user/pie-chart.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/pie-chart.html>)

D. No. Use tree maps to visualize one or two measures for a dimension.

Each rectangle on the tree map represents one item in the dimension. Rectangle size represents the proportion of the value for the selected measure that the item represents compared to the whole for the dimension. You can optionally use rectangle color to represent another measure for the item.

Rectangle color represents where the value for the item falls in the range for the measure, with darker colors indicating higher values and lighter colors indicating lower ones.

<https://docs.aws.amazon.com/quicksight/latest/user/tree-map.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/tree-map.html>)

Ask our Experts



QUESTION 59

UNATTEMPTED

STORAGE

MSP 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. MSP Bank is hosting their existing infrastructure on premise.

MSP bank has lot of web applications, databases and data warehouse built on Teradata, NoSQL databases, and other types of data stores. MSP bank has lot of data assets in terms of logs, documents; excel files, CSV files, PDF documents and others. MSP Bank wants to provide seamless integration with data security features between their on-premises IT environment and the AWS storage infrastructure. MSP bank considers AWS Storage Gateway to address the integration.

MSP Bank is looking at gateway interface which provides cloud-backed storage volumes that you can mount as Internet Small Computer System Interface (iSCSI) devices from your on-premises application servers. Also store data in Amazon Simple Storage Service (Amazon S3) and retain a copy of frequently accessed data subsets locally thereby offer a substantial cost savings on primary storage and minimize the need to scale your storage on-premises. How can this be achieved? select 2 options.

- ☒ A. Volume Gateway ✓
- ☒ B. Cached Volumes ✓
- ☐ C. Stored Volumes
- ☐ D. File Gateway

Explanation :

Answer: A,B

A. Yes. A volume gateway provides cloud-backed storage volumes that you can mount as Internet Small Computer System Interface (iSCSI) devices from your on-premises application servers. The gateway supports the following volume configurations:

- Cached volumes – You store your data in Amazon Simple Storage Service (Amazon S3) and retain a copy of frequently accessed data subsets locally. Cached volumes offer a substantial cost savings on primary storage and minimize the need to scale your storage on- premises. You also retain low-latency access to your frequently accessed data.
- Stored volumes – If you need low-latency access to your entire dataset, first configure your on-premises gateway to store all your data locally. Then asynchronously back up point-in- time snapshots of this data to Amazon S3. This configuration provides durable and inexpensive offsite backups that you can recover to your local data center or Amazon EC2. For example, if you need replacement capacity for disaster recovery, you can recover the backups to Amazon EC2. <https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html> (<https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html>)

B. Yes. Cached volumes – You store your data in Amazon Simple Storage Service (Amazon S3) and retain a copy of frequently accessed data subsets locally. Cached volumes offer a substantial cost savings on primary storage and minimize the need to scale your storage on- premises. You also retain low-latency access to your frequently accessed data. <https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html> (<https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html>)

C. No. Stored volumes – If you need low-latency access to your entire dataset, first configure your on-premises gateway to store all your data locally. Then asynchronously back up point-in-time snapshots of this data to Amazon S3. This configuration provides durable and inexpensive offsite backups that you can recover to your local data center or Amazon EC2. For example, if you need replacement capacity for disaster recovery, you can recover the backups to Amazon

EC2 <https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html> (<https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html>)

D. Yes. A file gateway supports a file interface into Amazon Simple Storage Service (Amazon S3) and combines a service and a virtual software appliance. By using this combination, store and retrieve objects in Amazon S3 using industry-standard file protocols such as Network File System (NFS) and Server Message Block (SMB). The software appliance, or gateway, is deployed into your on-premises environment as a virtual machine (VM) running on VMware ESXi or Microsoft Hyper-V hypervisor. The gateway provides access to objects in S3 as files or file share mount points. With a file gateway, you can do the following:

- You can store and retrieve files directly using the NFS version 3 or 4.1 protocol.
 - You can store and retrieve files directly using the SMB file system version, 2 and 3 protocol.
 - You can access your data directly in Amazon S3 from any AWS Cloud application or service.
 - You can manage your Amazon S3 data using lifecycle policies, cross-region replication, and versioning.
- You can think of a file gateway as a file system mount on

S3. <https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html> (<https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html>)

Ask our Experts



QUESTION 60

UNATTEMPTED

PROCESSING

KindleYou is a location-based social search mobile app that allows users to like or dislike other users, and allows users to chat if both parties liked each other in the app. It has more than 1 billion customers across the world.

They use DynamoDB to support the mobile application and S3 to host the images and other documents shared between users.

DynamoDB has a table with 60 partitions and is being heavily accessed by users. There are lots of hot partitions. How can we address writing to different shards to distribute the workload efficiently to both read and write operations? select 2 options.

- ☐ A. distributing loads more evenly across a partition key space is to add a random number to the end of the partition key values to improve writes ✓
- ☐ B. use a number that you can calculate based upon something that you want to query on to improve reads ✓

- ☐ C. distributing loads more evenly across a partition key space is to add a random number to the end of the partition key values to improve reads
- ☐ D. use a number that you can calculate based upon something that you want to query on to improve writes

Explanation :

Answer: A, B

A. Yes. Through random suffixes. Distributing loads more evenly across a partition key space is to add a random number to the end of the partition key values. Then you randomize the writes across the larger space.

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-sharding.html>
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-sharding.html>)sharding.html
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-sharding.html>)

B. Yes. Sharding Using Calculated Suffixes improves reads. use a number that you can calculate based upon something that you want to query on

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-sharding.html>
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-sharding.html>)sharding.html
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-sharding.html>)

C. No. Through random suffixes. Distributing loads more evenly across a partition key space is to add a random number to the end of the partition key values. Then you randomize the writes across the larger space

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-sharding.html>
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-sharding.html>)sharding.html
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-sharding.html>)

D. No. Sharding Using Calculated Suffixes improves reads. use a number that you can calculate based upon something that you want to query on

<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-sharding.html>
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-sharding.html>)sharding.html
(<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-sharding.html>)

Ask our Experts



Parson Fortunes Ltd is an Asian-based department store operator with an extensive network of 131 stores, spanning approximately 4.1 million m2 of retail space across cities in India, China, Vietnam, Indonesia and Myanmar.

Parson 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. Parson 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. What is your suggestion? select 1 option.

- ☒ A. Launch Redshift cluster with node types DS2.xlarge to fulfill the requirements ✓
- ☐ B. Launch Redshift cluster with node types DS2.8xlarge to fulfill the requirements
- ☐ C. Launch Redshift cluster with node types DC2.xlarge to fulfill the requirements
- ☐ D. Launch Redshift cluster with node types DC2.8xlarge to fulfill the requirements

Explanation :

Answer : A

A. 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

<https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes> (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes) nodes (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes)

B. DS2 node types are optimized for large data workloads and use hard disk drive (HDD) storage.

DS2.8xlarge does not fulfill the requirements since it can provide massive parallel processing using multiple nodes. Since cost and performance is also a concern, this is not the right option.

<https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes> (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes) nodes (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes)

C. DC2 node types are optimized for performance-intensive workloads. This is not the requirement

<https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes> (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes) nodes (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes)

D. DC2 node types are optimized for performance-intensive workloads. This is not the requirement

<https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes> (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes) nodes (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes)

Ask our Experts



QUESTION 62

UNATTEMPTED

COLLECTION

HikeHills.com (HH) 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.

HH runs their entire online infrastructure on Java-based web applications running on AWS. The HH is capturing click stream data and uses a custom-built recommendation engine to recommend products which eventually improve sales, understand customer preferences and already using AWS Streaming capabilities to collect events and transaction logs and process the stream.

HH is using Kinesis Analytics to build SQL querying capability on streaming and planning to use windowed queries to process the data. Based on the heavy data in flow, HH needs to ensure proper monitoring and management enabled for Kinesis Analytics applications. What kind of AWS automated Monitoring can be enabled? select 4 options.

- ☐ A. Watch a single metric over a time period that you specify, and perform one or more actions based on the value of the metric relative to a given threshold over a number of time periods using CloudWatch Alarms ✓
- ☐ B. Share log files between accounts, monitor CloudWatch log files in real time by sending them to CloudTrail Logs using AWS CloudWatch Log Monitoring
- ☐ C. Share log files between accounts, monitor CloudTrail log files in real time by sending them to CloudWatch Logs using AWS CloudTrail Log Monitoring ✓
- ☐ D. Monitor, store, and access your log files from AWS CloudTrail or other sources using AWS CloudWatch Logs ✓
- ☐ E. Match events and route them to one or more target functions or streams to make changes, capture state information, and take corrective action using AWS CloudWatch Events ✓
- ☐ F. Match events and route them to one or more target functions or streams to make changes, capture state information, and take corrective action using AWS CloudTrail Events

Explanation :

Answer: A, C,D, E

AWS provides various tools that you can use to monitor Amazon Kinesis Data Analytics. You can configure some of these tools to do the monitoring for you, while some of the tools require manual intervention

- AWS CloudWatch Alarms – Watch a single metric over a time period that you specify, and perform one or more actions based on the value of the metric relative to a given threshold over a number of time periods. The action is a notification sent to an Amazon Simple Notification Service (Amazon SNS) topic or Amazon EC2 Auto Scaling policy.
- AWS CloudWatch Logs – Monitor, store, and access your log files from AWS CloudTrail or other sources.
- AWS CloudWatch Events – Match events and route them to one or more target functions or streams to make changes, capture state information, and take corrective action.
- AWS CloudTrail Log Monitoring – Share log files between accounts, monitor CloudTrail log files in real time by sending them to CloudWatch Logs, write log processing applications in Java, and validate that your log files have not changed after delivery by CloudTrail
<https://docs.aws.amazon.com/kinesisanalytics/latest/dev/monitoring-automated-manual.html> (<https://docs.aws.amazon.com/kinesisanalytics/latest/dev/monitoring-automated-manual.html>)

Ask our Experts



QUESTION 63

UNATTEMPTED

PROCESSING

Tick-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. Tick-Bank aims to serve as a connection between digital content makers and affiliate dealers, who then promote them to clients. Tick-Bank's technology aids in payments, tax calculations and a variety of customer service tasks. Tick-Bank assists in building perceptibility and revenue making opportunities for entrepreneurs.

Tick-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. Tick-Bank is looking to enable web-site traffic analytics there by understanding user navigational behavior, preferences and other click related info. Tick-Bank is also looking at improving operations ingesting monitoring logs. Kinesis agent is used to process the logs and kinesis firehose the process the stream and data is stored into multiple data destinations like S3, ES and Splunk. What is the data delivery frequency? select 3 options.

- ☐ A. The kinesis firehose delivery streams to S3 with buffer size between 1-128 MB and buffer interval between 60-900 seconds ✓
- ☐ B. The kinesis firehose delivery streams to S3 with buffer size between 1-100 MB and buffer interval between 60-900 seconds
- ☐ C. The kinesis firehose delivery streams to ES with buffer size between 1-128 MB and buffer interval between 60-900 seconds
- ☐ D. The kinesis firehose delivery streams to ES with buffer size between 1-100 MB and buffer interval between 60-900 seconds ✓
- ☐ E. The kinesis firehose delivery streams to Splunk with minimum buffer size 1MB and buffer interval 900 seconds
- ☐ F. The kinesis firehose delivery streams to Splunk with buffer size is 5 MB and buffer interval is 60 seconds ✓

Explanation :

Answer: A,D, F

The frequency of data delivery to Amazon S3 is determined by the Amazon S3 Buffer size and Buffer interval value that you configured for your delivery stream. Kinesis Data Firehose buffers incoming data before it delivers it to Amazon S3. You can configure the values for Amazon S3 Buffer size (1–128 MB) or Buffer interval (60–900 seconds)

Kinesis Data Firehose buffers incoming data before delivering it to Amazon ES. You can configure the values for Elasticsearch Buffer size (1–100 MB) or Buffer interval (60–900seconds)

Kinesis Data Firehose buffers incoming data before delivering it to Splunk. The buffer size is 5 MB, and the buffer interval is 60 seconds. <https://docs.aws.amazon.com/firehose/latest/dev/basic-deliver.html#frequency> (<https://docs.aws.amazon.com/firehose/latest/dev/basic-deliver.html#frequency>)

Ask our Experts



QUESTION 64

UNATTEMPTED

ANALYSIS

Allianz Financial Services (AFS) is a banking group offering end-to-end banking and financial solutions in South East Asia through its consumer banking, business banking, Islamic banking, investment finance and stock broking businesses as well as unit trust and asset administration, having served the financial community over the past five decades.

AFS launched EMR cluster to support their big data analytics requirements. AFS has multiple data sources built out of S3, SQL databases, MongoDB, Redis, RDS, other file systems. AFS is looking for a service that supports interactive data exploration and can be accessed through web interface using a SSH tunnel to the EMR master node and a proxy connection

Which EMR Hadoop ecosystem fulfills the requirements?

- ☐ A. Apache Hive
- ☐ B. Apache HBase
- ☐ C. Apache HCatalog
- ☒ D. ApachePresto ✓

Explanation :

Answer : D

A.No. Hive is an open-source, data warehouse, and analytic package that runs on top of a Hadoop cluster. Hive scripts use an SQL-like language called Hive QL (query language) that abstracts programming models and supports typical data warehouse interactions. Hive enables you to avoid the complexities of writing Tez jobs based on directed acyclic graphs (DAGs) or MapReduce programs in a lower level computer language, such as Java. Hive extends the SQL paradigm by including serialization formats. You can also customize query processing by creating table schema that matches your data, without touching the data itself. In contrast to SQL (which only supports primitive value types such as dates, numbers, and strings), values in Hive tables are structured elements, such as JSON objects, any user-defined data type, or any function written in

Java. <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hive.html>
(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hive.html>)

B. No. Apache HBase is an open source, non-relational, distributed database developed as part of the Apache Software Foundation's Hadoop project. HBase runs on top of Hadoop Distributed File System (HDFS) to provide non-relational database capabilities for the Hadoop ecosystem

<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hbase.html>
(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hbase.html>)

C. No. Apache HCatalog is a tool that allows you to access Hive metastore tables within Pig, Spark SQL, and/or custom MapReduce applications. HCatalog has a REST interface and command line client that allows you to create tables or do other operations.

<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hcatalog.html>
(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hcatalog.html>)

D. Yes. Apache Presto is a fast SQL query engine designed for interactive analytic queries over large datasets from multiple sources

<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-presto.html>
(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-presto.html>)

MSP 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. MSP Bank is hosting their existing infrastructure on AWS. MSP 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 identified multiple datasets that need Geospatial data capabilities into visuals and maps. Please advice!

Select 2 options.

- ☒ A. Amazon QuickSight can chart latitude and longitude coordinates. ✓
- ☐ B. Amazon QuickSight can chart only polygon regions.
- ☐ C. Amazon QuickSight can recognize geographic components such as country, state, county, city, and zip code for all geographies.
- ☐ D. Amazon QuickSight can recognize geographic components such as country, state, county, city, and zip code for all geographies except China, India.
- ☒ E. Amazon QuickSight can also create geographic hierarchies that can disambiguate similar entities. ✓

Explanation :

Answer: A,E

A. Yes. Amazon QuickSight can chart latitude and longitude coordinates

<https://docs.aws.amazon.com/quicksight/latest/user/geospatial-data-prep.html>
(<https://docs.aws.amazon.com/quicksight/latest/user/geospatial-data-prep.html>)

B. No. Amazon QuickSight can chart latitude and longitude coordinates, not polygons

<https://docs.aws.amazon.com/quicksight/latest/user/geospatial-data-prep.html>
(<https://docs.aws.amazon.com/quicksight/latest/user/geospatial-data-prep.html>)

C. No. Amazon QuickSight can recognize geographic components such as country, state, county, city, and zip code

<https://docs.aws.amazon.com/quicksight/latest/user/geospatial-data-prep.html>
(<https://docs.aws.amazon.com/quicksight/latest/user/geospatial-data-prep.html>)

D. No. Amazon QuickSight can recognize geographic components such as country, state, county, city, and zip code for all geographies except China, India

<https://docs.aws.amazon.com/quicksight/latest/user/geospatial-data-prep.html>
(<https://docs.aws.amazon.com/quicksight/latest/user/geospatial-data-prep.html>)

E. Yes. Amazon QuickSight can also create geographic hierarchies that can disambiguate similar entities

<https://docs.aws.amazon.com/quicksight/latest/user/geospatial-data-prep.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/geospatial-data-prep.html>)

Ask our Experts



QUESTION 66

UNATTEMPTED

ANALYSIS

MSP 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. MSP Bank is hosting their existing infrastructure on premise.

MSP bank has lot of web applications, databases and data warehouse built on Teradata, NoSQL databases, and other types of data stores. MSP bank has lot of data assets in terms of logs, documents; excel files, CSV files, PDF documents and others. The files of different formats are being loaded into S3 buckets

MSP bank wants to enable interactive query service on the files loaded in S3 that makes it easy to analyze data directly in Amazon Simple Storage Service (Amazon S3) using standard SQL. select 1 option.

- ☐ A. Amazon Aurora
- ☐ B. Amazon RDS
- ☒ C. Amazon Athena ✓
- ☐ D. Amazon Neptune

Explanation :

Answer C

A. No. Amazon Aurora (Aurora) is a fully managed relational database engine that's compatible with MySQL and PostgreSQL. You already know how MySQL and PostgreSQL combine the speed and reliability of high-end commercial databases with the simplicity and cost-effectiveness of open-sourcedatabases. The code, tools, and applications you use today with your existing MySQL and PostgreSQL databases can be used with Aurora. With some workloads, Aurora can deliver up to five times the throughput of MySQL and up to three times the throughput of PostgreSQL without requiring changes to most of your existing applications.

Amazon Aurora supports relational data models and does not support graph model. https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html

(https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html)

B. No. Amazon Relational Database Service (Amazon RDS) is a web service that makes it easier to set up, operate, and scale a relational database in the cloud. It provides cost-efficient, resizable capacity for an industry-standard relational database and manages common database administration tasks.

Amazon RDS supports relational data models and does not support graph

model. <https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html>

(<https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html>)

C. Yes. Amazon Athena is an interactive query service that makes it easy to analyze data directly in Amazon Simple Storage Service (Amazon S3) using standard SQL. With a few actions in the AWS Management Console, you can point Athena at your data stored in Amazon S3 and begin using standard SQL to run ad-hoc queries and get results in seconds.

Athena is serverless, so there is no infrastructure to set up or manage, and you pay only for the queries you run. Athena scales automatically—executing queries in parallel—so results are fast, even with large datasets and complex queries. <https://docs.aws.amazon.com/athena/latest/ug/what-is.html> (<https://docs.aws.amazon.com/athena/latest/ug/what-is.html>)

D. No. Amazon Neptune is a fast, reliable, fully managed graph database service that makes it easy to build and run applications that work with highly connected datasets. The core of Neptune is a purpose-built, high-performance graph database engine that is optimized for storing billions of relationships and querying the graph with milliseconds latency <https://docs.aws.amazon.com/neptune/latest/userguide/intro.html> (<https://docs.aws.amazon.com/neptune/latest/userguide/intro.html>)

Ask our Experts



QUESTION 67

UNATTEMPTED

DATA SECURITY

MSP 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. MSP Bank is hosting their existing infrastructure on premise.

MSP bank has lot of web applications, databases and data warehouse built on Teradata, NoSQL databases, and other types of data stores. MSP bank has lot of data assets in terms of logs, documents; excel files, CSV files, PDF documents and others. The files of different formats are being loaded into S3 buckets. Logs and documents once processed and stored into DWH can be seldom used. As a part of compliance, MSP Bank is looking for a strategy to manage the files. Also MSP wants an approach to optimize the costs of the files stored in long term archives. select 1 option.

- ☐ A. Export back to On-premise
- ☐ B. Use S3, compress the files and delete original files
- ☐ C. Use S3-IA (S3-Infrequent Access) to manage the archives
- ☐ D. Use S3-Glacier (S3-Infrequent Access) to manage the archives ✓

Explanation :

Answer: D

Yes. Glacier is an extremely low-cost storage service that provides durable storage with security features for data archiving and backup. With Glacier, customers can store their data cost effectively for months, years, or even decades. Glacier enables customers to offload the administrative burdens

of operating and scaling storage to AWS, so they don't have to worry about capacity planning, hardware provisioning, data replication, hardware failure detection and recovery, or time-consuming hardware migrations. <https://docs.aws.amazon.com/amazonglacier/latest/dev/introduction.html> (<https://docs.aws.amazon.com/amazonglacier/latest/dev/introduction.html>)

Ask our Experts



QUESTION 68

UNATTEMPTED

VISUALIZATION

Marqueguard is a social media monitoring company headquartered in Brighton, England. Marqueguard sells three different products: Analytics, Audiences, and Insights. Marqueguard 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. Marqueguard has access to over 80 million sources.

Marqueguard is looking at using open source visualization tool to work with logs using Elasticsearch and also looking for a convenient way to use the bulk API to upload data into your Amazon ES domain with the S3 plugin. What tools provide these services?

Select 2 options.

- ☐ A. Kibana is a popular open source visualization tool designed to work with Elasticsearch. Amazon ES provides an installation of Kibana with every Amazon ES domain. You can find a link to Kibana on your domain dashboard on the Amazon ES console ✓
- ☐ B. Logstash provides a convenient way to use the bulk API to upload data into your Amazon ES domain with the S3 plugin. The service also supports all other standard Logstash input plugins that are provided by Elasticsearch ✓
- ☐ C. QuickSight provides open source visualization tool designed to work with Elasticsearch. Amazon ES provides an installation of QuickSight with every Amazon ES domain. You can find a link to QuickSight on your domain dashboard on the Amazon ES console

- ☐ D. Logstash is a popular open source visualization tool designed to work with Elasticsearch. Amazon ES provides an installation of Logstash with every Amazon ES domain. You can find a link to Logstash on your domain dashboard on the Amazon ES console
- ☐ E. AWS Glue acts as the ETL Engine and Glue Data Catalog acts as central metadata repository to process log files and uses its API to upload data into your Amazon ES domain with the S3 plugin. The service also supports all other standard Kibana input plugins that are provided by Elasticsearch
- ☐ F. Kibana provides a convenient way to use the bulk API to upload data into your Amazon ES domain with the S3 plugin. The service also supports all other standard Kibana input plugins that are provided by Elasticsearch

Explanation :

Answer: A,B

A. Yes. Kibana is a popular open source visualization tool designed to work with Elasticsearch. Amazon ES provides an installation of Kibana with every Amazon ES domain. You can find a link to Kibana on your domain dashboard on the Amazon ES console [https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es- \(https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-kibana.html\)kibana.html](https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es- (https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-kibana.html)kibana.html)

(<https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-kibana.html>)

B. Yes. Logstash provides a convenient way to use the bulk API to upload data into your Amazon ES domain with the S3 plugin. The service also supports all other standard Logstash input plugins that are provided by Elasticsearch [https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es- \(https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-kibana.html\) kibana.html](https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es- (https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-kibana.html) kibana.html)

(<https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-kibana.html>)

C. No. Amazon QuickSight is a business analytics service you can use to build visualizations, perform ad hoc analysis, and get business insights from your data. It can automatically discover AWS data sources and also works with your data sources. Amazon QuickSight enables organizations to scale to hundreds of thousands of users, and delivers responsive performance by using a robust in- memory engine (SPICE). <https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>

(<https://docs.aws.amazon.com/quicksight/latest/user/welcome.html>)

D. No. Logstash provides a convenient way to use the bulk API to upload data into your Amazon ES domain with the S3 plugin. The service also supports all other standard Logstash input plugins that are provided by Elasticsearch [https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es- \(https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-kibana.html\) kibana.html](https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es- (https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-kibana.html) kibana.html)

(<https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-kibana.html>)

E. No. AWS Glue is a fully managed ETL (extract, transform, and load) service that makes it simple and cost-effective to categorize your data, clean it, enrich it, and move it reliably between various data stores. <https://docs.aws.amazon.com/glue/latest/dg/what-is-glue.html>

(<https://docs.aws.amazon.com/glue/latest/dg/what-is-glue.html>)

The AWS Glue Data Catalog is persistent metadata store. It is a managed service that lets you store, annotate, and share metadata in the AWS Cloud in the same way you would in an Apache Hive metastore, It provides a uniform repository where disparate systems can store and find metadata to keep track of data in data silos, and use that metadata to query and transform the

data <https://docs.aws.amazon.com/glue/latest/dg/components-overview.html#data-catalog-intro>)
(<https://docs.aws.amazon.com/glue/latest/dg/components-overview.html#data-catalog-intro>)
catalog-intro (<https://docs.aws.amazon.com/glue/latest/dg/components-overview.html#data-catalog-intro>)

F. No. Kibana is a popular open source visualization tool designed to work with Elasticsearch. Amazon ES provides an installation of Kibana with every Amazon ES domain. You can find a link to Kibana on your domain dashboard on the Amazon ES console <https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-kibana.html> (<https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-kibana.html>)

Ask our Experts



QUESTION 69

UNATTEMPTED

STORAGE

MSP 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. MSP Bank is hosting their existing infrastructure on premise.

MSP bank has lot of web applications, databases and data warehouse built on Teradata, NoSQL databases, and other types of data stores. MSP bank has lot of data assets in terms of logs, documents; excel files, CSV files, PDF documents and others.

The Web Application has different user workloads at different parts of the day. MSP Bank is running one of their web application NODE.JS supported by MSSQL relational database. The schema designed is based on 3NF. MSP Bank want to migrate the database on to AWS cloud. Also MSP Bank is looking to build a new database solution which is optimized for storing billions of relationships and querying the storage with milliseconds latency. Which AWS services fulfill the requirement? select 1 option.

- ☐ A. AWS Aurora is a multi-modal database. It supports both relational model and graph model. So AWS Aurora fulfills both the requirements
- ☐ B. AWS RDS is a multi-modal database. It supports both relational model and graph model. So AWS RDS fulfills both the requirements
- ☐ C. AWS RDS fulfills the relational database requirements while AWS DynamoDB supports managing relationships through its graph model
- ☐ D. AWS Aurora fulfills the relational database requirements while AWS Neptune supports managing relationships through its graph model ✓
- ☐ E. AWS DynamoDB fulfills the relational database requirements while AWS Neptune supports managing relationships through its graph model

○ F. AWS DynamoDB supports both relational and graph model.

Explanation :

Answer: D

A. No. Amazon Aurora (Aurora) is a fully managed relational database engine that's compatible with MySQL and PostgreSQL. You already know how MySQL and PostgreSQL combine the speed and reliability of high-end commercial databases with the simplicity and cost-effectiveness of open-sourcedatabases. The code, tools, and applications you use today with your existing MySQL and PostgreSQL databases can be used with Aurora. With some workloads, Aurora can deliver up to five times the throughput of MySQL and up to three times the throughput of PostgreSQL without requiring changes to most of your existing applications.

Amazon Aurora supports relational data models and does not support graph model. https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html (https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html)

B. No. Amazon Relational Database Service (Amazon RDS) is a web service that makes it easier to set up, operate, and scale a relational database in the cloud. It provides cost-efficient, resizable capacity for an industry-standard relational database and manages common database administration tasks.

Amazon RDS supports relational data models and does not support graph model. <https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html> (<https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html>)

C. No. Amazon Relational Database Service (Amazon RDS) is a web service that makes it easier to set up, operate, and scale a relational database in the cloud. It provides cost-efficient, resizable capacity for an industry-standard relational database and manages common database administration tasks.

Amazon RDS supports relational data models and does not support graph model. <https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html> (<https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html>) Amazon DynamoDB is a fully managed NoSQL database service that provides fast and predictable performance with seamless scalability. DynamoDB lets you offload the administrative burdens of operating and scaling a distributed database, so that you don't have to worry about hardware provisioning, setup and configuration, replication, software patching, or cluster scaling. <https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Introduction.html> (<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Introduction.html>)

D. Yes. Amazon Aurora (Aurora) is a fully managed relational database engine that's compatible with MySQL and PostgreSQL. You already know how MySQL and PostgreSQL combine the speed and reliability of high-end commercial databases with the simplicity and cost-effectiveness of open-sourcedatabases. The code, tools, and applications you use today with your existing MySQL and PostgreSQL databases can be used with Aurora. With some workloads, Aurora can deliver up to five times the throughput of MySQL and up to three times the throughput of PostgreSQL without requiring changes to most of your existing applications.

Amazon Aurora supports relational data models https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html (https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html)

Amazon Neptune is a fast, reliable, fully managed graph database service that makes it easy to build and run applications that work with highly connected datasets. The core of Neptune is a purpose-built, high-performance graph database engine that is optimized for storing billions of relationships and querying the graph with milliseconds

latency <https://docs.aws.amazon.com/neptune/latest/userguide/intro.html> (<https://docs.aws.amazon.com/neptune/latest/userguide/intro.html>)

E. No. Amazon DynamoDB is a fully managed NoSQL database service that provides fast and predictable performance with seamless scalability. DynamoDB lets you offload the administrative burdens of operating and scaling a distributed database, so that you don't have to worry about hardware provisioning, setup and configuration, replication, software patching, or cluster scaling. <https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Introduction.html> (<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Introduction.html>)

Amazon Neptune is a fast, reliable, fully managed graph database service that makes it easy to build and run applications that work with highly connected datasets. The core of Neptune is a purpose-built, high-performance graph database engine that is optimized for storing billions of relationships and querying the graph with milliseconds latency <https://docs.aws.amazon.com/neptune/latest/userguide/intro.html> (<https://docs.aws.amazon.com/neptune/latest/userguide/intro.html>)

F. No. Amazon DynamoDB is a fully managed NoSQL database service that provides fast and predictable performance with seamless scalability. DynamoDB lets you offload the administrative burdens of operating and scaling a distributed database, so that you don't have to worry about hardware provisioning, setup and configuration, replication, software patching, or cluster scaling. <https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Introduction.html> (<https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Introduction.html>)

DynamoDB neither supports relational model nor graph model.

Ask our Experts



QUESTION 70

INCORRECT

VISUALIZATION

Allianz Financial Services (AFS) is a banking group offering end-to-end banking and financial solutions in South East Asia through its consumer banking, business banking, Islamic banking, investment finance and stock broking businesses as well as unit trust and asset administration, having served the financial community over the past five decades.

AFS launched EMR cluster to support their big data analytics requirements. AFS has multiple data sources built out of S3, SQL databases, MongoDB, Redis, RDS, other file systems. AFS is looking for a component in EMR that allows to quickly create Jupyter notebooks, attach them to Spark clusters, and then open the Jupyter Notebook editor in the console to remotely run queries and code. Which EMR Hadoop ecosystem fulfills the requirements? select 2 options.

- ☐ A. Apache Hive
- ☐ B. EMR Notebook ✓
- ☒ C. Jupyter Notebook ✓
- ☒ D. JupyterHub ✗

Explanation :

Answer : B, C

A. No. Hive is an open-source, data warehouse, and analytic package that runs on top of a Hadoop cluster. Hive scripts use an SQL-like language called Hive QL (query language) that abstracts programming models and supports typical data warehouse interactions. Hive enables you to avoid the complexities of writing Tez jobs based on directed acyclic graphs (DAGs) or MapReduce programs in a lower level computer language, such as Java. Hive extends the SQL paradigm by including serialization formats. You can also customize query processing by creating table schema that matches your data, without touching the data itself. In contrast to SQL (which only supports primitive value types such as dates, numbers, and strings), values in Hive tables are structured elements, such as JSON objects, any user-defined data type, or any function written in

Java. <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hive.html>

(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hive.html>)

B. Yes. EMR Notebooks is a Jupyter Notebook environment built in to the Amazon EMR console that allows you to quickly create Jupyter notebooks, attach them to Spark clusters, and then open the Jupyter Notebook editor in the console to remotely run queries and code. An EMR notebook is saved in Amazon S3 independently from clusters for durable storage, quick access, and flexibility. You can have multiple notebooks open, attach multiple notebooks to a single cluster, and re-use a notebook on different clusters <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-jupyter-emr->

(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-jupyter-emr-managed-notebooks.html>)managed-notebooks.html

C. Yes. Jupyter Notebook is an open-source web application that you can use to create and share documents that contain live code, equations, visualizations, and narrative text. Amazon EMR offers you two options to work with Jupyter notebooks:

- EMR Notebook
- JupyterHub <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-jupyter.html>
(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-jupyter.html>)

D. No. Jupyter Notebook is an open-source web application that you can use to create and share documents that contain live code, equations, visualizations, and narrative text. JupyterHub allows you to host multiple instances of a single-user Jupyter notebook server. When you create a cluster with JupyterHub, Amazon EMR creates a Docker container on the cluster's master node. JupyterHub, all the components required for Jupyter, and Sparkmagic run within the container. <https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-jupyterhub.html>
(<https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-jupyterhub.html>)

Ask our Experts



Finish Review (<https://www.whizlabs.com/learn/course/aws-bds-practice-tests/quiz/14898>)

Certification

- ➔ Cloud Certification
(<https://www.whizlabs.com/cloud-certification-training-courses/>)
- ➔ Java Certification
(<https://www.whizlabs.com/oracle-java-certifications/>)
- ➔ PM Certification
(<https://www.whizlabs.com/project-management-certifications/>)
- ➔ Big Data Certification
(<https://www.whizlabs.com/big-data-certifications/>)

Mobile App

-  Android Coming Soon
-  iOS Coming Soon

Company

- ➔ Support
(<https://help.whizlabs.com/hc/en-us>)
- ➔ Discussions (<http://ask.whizlabs.com/>)
- ➔ Blog (<https://www.whizlabs.com/blog/>)

Follow us



(<https://www.facebook.com/whizlabs.software/>)



(<https://in.linkedin.com/company/whizlabs-software>)



(<https://twitter.com/whizlabs?lang=en>)



(<https://plus.google.com/+WhizlabsSoftware>)