PRACTICE TEST 1

Attempt 1

Marks Obtained 0 / 70

Your score is 0.0%

Completed on Sunday , 03 February 2019 , 11:37 PM

Time Taken 00 H 00 M 11 S

Result Fail

Domains / Topics wise Quiz Performance Report

S.No.	Topic	Total Questions	Correct	Incorrect	Unattempted
1	Analysis	15	0	0	15
2	Processing	17	0	0	17
3	Storage	18	0	0	18
4	Visualization	5	0	1	4
5	Collection	8	0	0	8
6	Data Security	7	0	0	7

70 Questions	**0** Correct	**1** Incorrect	**69** Unattempted

Show Answers

All

QUESTION 1 UNATTEMPTED ANALYSIS

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. What kind of Producer configuration is needed to retrieve a specific video clips to fulfill on-demand video playback for a specified time range? Select 4 options.

- ☐ **A.** Producer sends a stream of media fragments, each fragment being a self-contained sequence of frames ✔
- ☐ **B.** Producer sends a stream of media frames, each frame being a self-contained sequence of fragments
- ☐ **C.** The frames belonging to a fragment should have no dependency on any frames from other fragments ✔
- ☐ **D.** The fragments belonging to a frame should have no dependency on any fragment from other frames
- ☐ **E.** Kinesis Video Streams assigns a unique frame number, in increasing order.
- ☐ **F.** Kinesis Video Streams assigns a unique fragment number, in increasing order. ✔
- ☐ **G.** Store video stream specific metadata, which includes producer-side and server-side time stamps for each fragment ✔
- ☐ **H.** Store video stream specific metadata, which includes producer-side and server-side time stamps for each frame

Explanation :

Answer: A,C,F,G

A. Yes. Producer sends a stream of media fragments, each fragment being a self-contained sequence of frames

https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html)
B. No. Producer sends a stream of media fragments, each fragment being a self-contained sequence of frames
https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html)
C. Yes. The frames belonging to a fragment should have no dependency on any frames from other fragments
https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html)
D. No. The frames belonging to a fragment should have no dependency on any frames from other fragments
https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html)
E. No. Kinesis Video Streams assigns a unique fragment number, in increasing order
https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html)
F. Yes. Kinesis Video Streams assigns a unique fragment number, in increasing order
https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html)
G. Yes. Store video stream specific metadata, which includes producer-side and server-side time stamps for each fragment
https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html)
H. No. Store video stream specific metadata, which includes producer-side and server-side time stamps for each fragment
https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-it-works-kinesis-video-api-producer-sdk.html)

Ask our Experts

QUESTION 2 UNATTEMPTED ANALYSIS

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 DynamoDB 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

QUESTION 3 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 uses event based streaming, based on Kinesis Stream to address data integration and uses producer library to integrate events.

With growing applications and massively increasing customers and new channels to execute transactions, the performance of the streaming application has drastically reduced and has creating lot of performance issues in the business applications. The administrator is looking to standardize performance and cost generated by shards with Kinesis Streams. Please advise! Select 5 options.

- [ ] **A.** Use single shard to address the integration of all the events from different high performance applications
- [ ] **B.** Use multiple shards to address the integration of all the events from different high performance applications ✔
- [ ] **C.** Reshard by splitting the shards to increase capacity of the stream ✔
- [ ] **D.** Reshard by splitting the shards to reduce the cost of the stream
- [ ] **E.** Reshard by merging the shards to increase capacity of the stream
- [ ] **F.** Reshard by merging the shards to reduce the cost of the stream ✔
- [ ] **G.** Use metrics to determine which are your "hot" or "cold" shards ✔
- [ ] **H.** Use metrics to determine which shard receives much more data, or much less data to understand usage capacity ✔

Explanation :

Answer B,C,F,G,H

A. No. Resharding is necessary with increasing demand to increase the capacity of the application. Multiple shard may increase the costs but eventually provides more capacity to address integration without latency.
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. Resharding is necessary with increasing demand to increase the capacity of the application. Multiple shard may increase the costs but eventually provides more capacity to address integration without latency.
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. Yes. The purpose of resharding in Amazon Kinesis Data Streams is to enable your stream to adapt to changes in the rate of data flow. You split shards to increase the capacity (and cost) of your stream.
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. The purpose of resharding in Amazon Kinesis Data Streams is to enable your stream to adapt to changes in the rate of data flow. You split shards to increase the capacity (and cost) of your stream.

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)
E. No. The purpose of resharding in Amazon Kinesis Data Streams is to enable your stream to adapt to changes in the rate of data flow. You split shards to increase the capacity (and cost) of your stream. You merge shards to reduce the cost (and capacity) of your stream.
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)
F. Yes. The purpose of resharding in Amazon Kinesis Data Streams is to enable your stream to adapt to changes in the rate of data flow. You merge shards to reduce the cost (and capacity) of your stream.
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)
G. Yes. use metrics to determine which are your "hot" or "cold" shards, that is, shards that are receiving much more data, or much less data, than expected. You could then selectively split the hot shards to increase capacity for the hash keys that target those shards. Similarly, you could merge cold shards to make better use of their unused 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)
H. Yes. use metrics to determine which are your "hot" or "cold" shards, that is, shards that are receiving much more data, or much less data, than expected. You could then selectively split the hot shards to increase capacity for the hash keys that target those shards. Similarly, you could merge cold shards to make better use of their unused 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

QUESTION 4 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 is planning to host an EMR cluster to run their analytical workloads which complements their existing ETL jobs built on Data Pipeline. AFS was working with you to identify and understand the configuration needed to support their workload.

Cluster to coordinate the distribution of data and tasks among other nodes for processing both for daily, weekly and monthly jobs
Fixed workload of 4 compute and storage nodes to address their daily workload
2 more compute nodes to support their weekly and monthly workloads.
These nodes supports both Data Pipeline and EMR processing and storage
Please advise on minimum artefacts to also optimize overall TCO. Select 4 options.

- [ ] **A.** 1 or 2 master nodes to coordinate the distribution of data and tasks among other nodes for processing ✔
- [ ] **B.** 1 or 2 primary nodes to coordinate the distribution of data and tasks among other nodes for processing
- [ ] **C.** 4 core nodes with software components that run tasks and store data to support partial workloads (daily/weekly/monthly) ✔
- [ ] **D.** 4 task nodes with software components that run tasks to support partial workloads (daily/weekly/monthly)
- [ ] **E.** 2 task nodes with software components that run tasks to support partial workloads (daily/weekly/monthly) ✔
- [ ] **F.** 4-6 task nodes in overall with software components that run tasks and store data to support full workloads
- [ ] **G.** 4-6 slaves nodes in overall with software components that run tasks and store data to support full workloads ✔
- [ ] **H.** 4-6 data nodes in overall with software components that run tasks and store data to support full workloads

Explanation :

Answer : A, C, E, G

A. Yes. Master node manages the cluster by running software components to coordinate the distribution of data and tasks among other nodes for processing
https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-overview.html#emr-overview-clusters (https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-overview.html#emr-overview-clusters)

B. No. Primary node is not a component of EMR.

C. Yes. Core node is a node with software components that run tasks and store data in the Hadoop Distributed File System (HDFS) on your cluster. This supports both compute and data requirements of daily, weekly and monthly workloads
https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-overview.html#emr-overview-clusters (https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-overview.html#emr-overview-clusters)

D. No. Task node is a node with software components that only runs tasks and does not store data in HDFS. This cannot support data requirements. So may not fulfill weekly jobs and for additional compute for weekly and monthly workloads, this is a overkill by 2 nodes. Besides task nodes provide only compute not storage
https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-overview.html#emr-overview-clusters (https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-overview.html#emr-overview-clusters)

E. Yes. Task node is a node with software components that only runs tasks and does not store data in HDFS. This supports additional compute for weekly and monthly workloads
https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-overview.html#emr-overview-clusters (https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-overview.html#emr-overview-clusters)

F. No. Task node is a node with software components that only runs tasks and does not store data in HDFS. This cannot support data requirements because task nodes provide only compute not storage
https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-overview.html#emr-overview-clusters (https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-overview.html#emr-overview-clusters)

G. Yes. Slave nodes in EMR include both core nodes and task nodes.
https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-overview.html#emr-overview-clusters (https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-overview.html#emr-overview-clusters)

H. No. data node is not a component of EMR

Ask our Experts

QUESTION 5 UNATTEMPTED STORAGE

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 uses event based streaming, based on Kinesis Stream to address data integration and uses producer library to integrate events.
Tick-Bank want to use the data captured for multiple functions which include, Storage of data into S3, which later processed by lambda, load the data to support enterprise search built on ES Service, and Integrate into Data warehouse built on Redshift in near real-time. What is the best approach to consume all the data captured in the stream is shared with all the applications mentioned above which includes S3, Redshift and Elasticsearch (ES). Select 1 option.

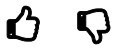
- **A.** Use enhanced Fan out consumers to integrate with above mentioned downstream applications stream
- **B.** Use KCL to integrate with above mentioned downstream applications stream
- **C.** use API to integrate with above mentioned downstream applications stream
- **D.** use Kinesis Firehose to integrate with above mentioned downstream applications stream ✔

Explanation :

Answer: D
If you want to send stream records directly to services such as Amazon Simple Storage Service (Amazon S3), Amazon Redshift, Amazon Elasticsearch Service (Amazon ES), or Splunk, you can

use a Kinesis Data Firehose delivery stream instead of creating a consumer application.
https://docs.aws.amazon.com/streams/latest/dev/amazon-kinesis-consumers.html
(https://docs.aws.amazon.com/streams/latest/dev/amazon-kinesis-consumers.html)

Ask our Experts



QUESTION 6 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 comparison between a key value and its target value
Please advice. Select 1 option.

- A. Bar Charts
- B. Combo Charts
- C. Heat Maps
- D. KPI's ✔

Explanation :

Answer: D
A. No. A bar chart visual type is used to create a single-measure, multi-measure, or clustered bar chart. A single-measure bar chart shows one measure for one dimension, for example average delay time by flight number. A multi-measure bar chart shows two or more measures for one dimension, for example sales total and profit total by automobile mode. A clustered bar chart shows values for a dimension grouped by a related dimension, for example sales totals by automobile model, grouped by car maker.

https://docs.aws.amazon.com/quicksight/latest/user/bar-charts.html
(https://docs.aws.amazon.com/quicksight/latest/user/bar-charts.html)
B. No. On the clustered bar combo chart, bars display for each child dimension, grouped by the parent dimension. On the stacked bar combo chart, one bar displays per parent dimension.
https://docs.aws.amazon.com/quicksight/latest/user/combo-charts.html
(https://docs.aws.amazon.com/quicksight/latest/user/combo-charts.html)
C. No. heat maps to show a measure for the intersection of two dimensions, with color-coding to easily differentiate where values fall in the range
https://docs.aws.amazon.com/quicksight/latest/user/heat-map.html
(https://docs.aws.amazon.com/quicksight/latest/user/heat-map.html)
D. Yes. Use a KPI to visualize a comparison between a key value and its target value.
A KPI displays a value comparison, the two values being compared, and a progress bar. For example, the following KPI shows how closely revenue is meeting its forecast.
https://docs.aws.amazon.com/quicksight/latest/user/kpi.html
(https://docs.aws.amazon.com/quicksight/latest/user/kpi.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 data streams supporting various business processes and uses enhanced fan out consumers to fulfill processing of data. The stream has 2 shards. There are 2 consumers that are using enhanced fan-out to receive data from the stream. Please detail the implementation specifics of consumers. Select 3 options.

- [x] **A.** An enhanced fan-out pipe provides up to 2 MiB/sec of data per shard ✔
- [x] **B.** consumers can be built out KCL library or API ✔
- [ ] **C.** consumers can be built out only out of KCL library
- [x] **D.** KCL automatically subscribes to consumer to all the shards of the stream ✔
- [ ] **E.** KCL automatically subscribes to consumer to the first shard of the shards of the stream

Explanation :

Answer: A, B, D
A. Yes. An enhanced fan-out pipe provides up to 2 MiB/sec of data per shard
https://docs.aws.amazon.com/streams/latest/dev/introduction-to-enhanced-consumers.html (https://docs.aws.amazon.com/streams/latest/dev/introduction-to-enhanced-consumers.html)
B. Yes. consumers can be built out KCL library or API
https://docs.aws.amazon.com/streams/latest/dev/introduction-to-enhanced-consumers.html (https://docs.aws.amazon.com/streams/latest/dev/introduction-to-enhanced-consumers.html)
C. No. consumers can be built out KCL library or API
https://docs.aws.amazon.com/streams/latest/dev/introduction-to-enhanced-consumers.html (https://docs.aws.amazon.com/streams/latest/dev/introduction-to-enhanced-consumers.html)
D. Yes. KCL automatically subscribes to consumer to all the shards of the stream
https://docs.aws.amazon.com/streams/latest/dev/introduction-to-enhanced-consumers.html (https://docs.aws.amazon.com/streams/latest/dev/introduction-to-enhanced-consumers.html)
E. No. KCL automatically subscribes to consumer to all the shards of the stream
https://docs.aws.amazon.com/streams/latest/dev/introduction-to-enhanced-consumers.html (https://docs.aws.amazon.com/streams/latest/dev/introduction-to-enhanced-consumers.html)

Ask our Experts

QUESTION 8 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 building a custom built monitoring platform based on node.js for all the devices in AWS. The platform basically captures the existing persistent representation of the devices. Also the integration of IOT data collected into the application needs to be based on canonical data format like JSON. How can this be achieved?

Select 2 options.

- [ ] **A.** Device Registry is a JSON document used to store and retrieve current state information for a device
- [ ] **B.** Device Shadow is a JSON document used to store and retrieve current state information for a device ✔
- [ ] **C.** Device Shadow service is a JSON document used to store and retrieve current state information for a device
- [ ] **D.** Device Shadow provides persistent representations of your devices in the AWS Cloud
- [ ] **E.** Device Shadow service provides persistent representations of your devices in the AWS Cloud ✔
- [ ] **F.** Device Registry provides persistent representations of your devices in the AWS Cloud

Explanation :

Answer : B, E

A. No. Device registry organizes the resources associated with each device in the AWS Cloud.
https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html
(https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html)

B. Yes. Device Shadow is a JSON document used to store and retrieve current state information for a device
https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html
(https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html)

C. No. device shadow service Provides persistent representations of your devices in the AWS Cloud.
https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html
(https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html)

D. No. Device Shadow is a JSON document used to store and retrieve current state information for a device
https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html
(https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html)

E. Yes, Device shadow service Provides persistent representations of your devices in the AWS Cloud.
https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html
(https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html)

F. No. Device registry organizes the resources associated with each device in the AWS Cloud
https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html
(https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html)

Ask our Experts

QUESTION 9 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 has multiple kinesis data streams supporting various business processes and uses enhanced fan out consumers to fulfill processing of data. Tick-Bank proposes KPL library to produce the data stream while KCL library to consume the data. Lot of kinesis throughput exceptions are thrown. How can this be resolved?Select 2 options.

- [ ] **A.** Increase the read and write throughput of kinesis table pertaining to a specific kinesis application ✔
- [ ] **B.** use enhanced fan-out consumers to process the data
- [ ] **C.** if application uses frequent check-pointing or operates on a stream that is composed of many shards, you might need more throughput. ✔
- [ ] **D.** use shared fan-out consumers to process the data

Explanation :

Answer: A, C
A. Yes. Increase the read and write throughput of kinesis table pertaining to a specific kinesis application to prevent throughput exceptions
https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-ddb.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-ddb.html)
B. No. enhanced fan-out consumers is applied to provide a read rate of 2 MiB/sec per shard
https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-ddb.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-ddb.html)
C.Yes. if application uses frequent check-pointing or operates on a stream that is composed of many shards, you might need more throughput.
https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-ddb.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-ddb.html)
D. No. shared fan-out consumers is applied to provide a maximum read rate of 2 MiB/sec per shard
https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-ddb.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-ddb.html)

Ask our Experts

QUESTION 10 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 Kinesis Producer Library to collect events and transaction logs and process the stream. HH is looking at processing and analyzing streaming data using standard SQL which enables to quickly author and run powerful SQL code against streaming sources using AWS Kinesis Analytics. What kind of use cases can be realized using this capability?Select 3 options.

- [ ] **A.** Calculate metrics over time windows, and then stream values to Amazon S3 or Amazon Redshift through a Kinesis data delivery stream ✔
- [ ] **B.** Send aggregated and processed streaming data results downstream to feed real-time dashboards ✔
- [ ] **C.** Act as Data warehousing ETL Engine which can pre-validate business rules before loading into target applications
- [ ] **D.** Create custom metrics and triggers for use in real-time monitoring, notifications, and alarms ✔
- [ ] **E.** Supports Batch Integration and ingesting data into DWH and RDS databases

Explanation :

Answer: A,B,D
Amazon Kinesis Data Analytics allows processing and analyzing streaming data using standard SQL which enables to quickly author and running powerful SQL code against streaming sources to perform time series analytics, feed real-time dashboards, and create real-time metrics.
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 11 UNATTEMPTED VISUALIZATION

MSP Bank, Limited is a leading varied 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. Team wanted to build some charts on a single dimension, grouped dimensions against a single measure and
multiple measures and their aggregations and summaries based on X and Y dimensions. How can we achieve? Select 2 options.

- [ ] **A.** Bar Charts ✔
- [ ] **B.** Combo Charts ✔
- [ ] **C.** Heat Maps
- [ ] **D.** Line Charts

Explanation :

Answer: A, B
A. Yes. A bar chart visual type is used to create a single-measure, multi-measure, or clustered bar chart. A single-measure bar chart shows one measure for one dimension, for example average

delay time by flight number. A multi-measure bar chart shows two or more measures for one dimension, for example sales total and profit total by automobile mode. A clustered bar chart shows values for a dimension grouped by a related dimension, for example sales totals by automobile model, grouped by car maker.
https://docs.aws.amazon.com/quicksight/latest/user/bar-charts.html
(https://docs.aws.amazon.com/quicksight/latest/user/bar-charts.html)
B. Yes. On the clustered bar combo chart, bars display for each child dimension, grouped by the parent dimension. On the stacked bar combo chart, one bar displays per parent dimension.
https://docs.aws.amazon.com/quicksight/latest/user/bar-charts.html
https://docs.aws.amazon.com/quicksight/latest/user/bar-charts.html
(https://docs.aws.amazon.com/quicksight/latest/user/bar-charts.html)C. No. heat maps to show a measure for the intersection of two dimensions, with color-coding to easily differentiate where values fall in the range
https://docs.aws.amazon.com/quicksight/latest/user/heat-map.html
(https://docs.aws.amazon.com/quicksight/latest/user/heat-map.html)
D. No. Use line charts to compare changes in measure values over period of time

- 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)

Ask our Experts

QUESTION 12 UNATTEMPTED STORAGE

Hymutabs Ltd (Hymutabs) is a global environmental solutions company running its operations in 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
One of the key requirements includes ORDER_FCT and PART_DIM are joined together in most of order related queries. ORDER_FCT has many other dimensions to support analysis.

How would you design the distribution? Select 1 option.

- **A.** Distribute the ORDER_FCT with KEY distribution on its primary KEY (any one of the columns) and PART_DIM with KEY distribution on its PRIMARY KEY
- **B.** Distribute the ORDER_FCT with ALL distribution on its primary KEY (any one of the columns) and PART_DIM with ALL distribution on its PRIMARY KEY
- **C.** Distribute the ORDER_FCT with EVEN distribution on its primary KEY (any one of the columns) and PART_DIM with EVEN distribution on its PRIMARY KEY

- **D.** Distribute the ORDER_FCT and PART_DIM on same key with KEY distribution ✔
- **E.** Distribute the ORDER_FCT and PART_DIM on same key with EVEN distribution

Explanation :

Answer : D
A. KEY DISTRIBUTION distributes the rows are according to the values in one column. Queries initiate lot of redistribution of data of both ORDER_FCT and PART_DIM are not built on same key.
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. 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.
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. EVEN DISTRIBUTION evenly distributes the rows across the slices in a round-robin fashion, regardless of the values in any particular column. EVEN distribution is appropriate when a table does not participate in joins. Definitely not a right approach.
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. KEY DISTRIBUTION distributes the rows are according to the values in one column. With distribution of data on same key in both the tables, there is no change of redistribution. This is the best approach to design.
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. EVEN DISTRIBUTION evenly distributes the rows across the slices in a round-robin fashion, regardless of the values in any
particular column. EVEN distribution is appropriate when a table does not participate in joins. Definitely not a right approach
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 13 UNATTEMPTED STORAGE

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 Producer Library 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 hot shards. 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 332...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 improves performance by unifying hash keys into single shard

Explanation :

Answer: B
Splitting the shards improves performance, also hash keys of shards cannot overlap each other shards
https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-split.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-split.html)

Ask our Experts

QUESTION 14 UNATTEMPTED DATA SECURITY

MSP Bank, Limited is a leading varied 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.
A 'transaction_details' dataset of 50 GB size has been created and will be used by different departments like Sales, Marketing, Business segments like Cards, Accounts, C Level, etc. from the dataset, different organizations access different subsets of data which may or may overlay each other subsets. The team needs a mechanism to enable row-level security, there restricting access to data elements in the dataset. How can this be achieved?Select 4 options.

- [ ] **A.** Creating Data Set Rules for Row-Level Security ✔
- [ ] **B.** Apply row-level permissions by using a file or query that contains data set rules. ✔
- [ ] **C.** choose your permissions data set, and assign permission policies to users and groups ✔
- [ ] **D.** You can assign row-level security only before sharing the transaction_details dataset
- [ ] **E.** You can assign row-level security only after sharing the transaction_details dataset
- [ ] **F.** To apply the data set rules, you add the rules as a permissions data set to your data set. ✔

Explanation :

Answer: A,B,C,F
Restrict access to a data set by configuring row-level security on it and can be done before or after the dataset is shared.

Only the people you shared with can see any of the data. By adding row-level security, you can further control their access.
https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html#create-row-level-security
(https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html#create-row-level-security)

Ask our Experts

QUESTION 15 UNATTEMPTED DATA SECURITY

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 is planning to launch EMR cluster to complement their ETL workloads running on Data Pipeline. The Team is looking for storing persistent data complemented with server-side encryption, read-after-write consistency, and list consistency and enables Data Lake for the enterprise to support analytics. Select 1 option.

- **A.** HDFS Storage launched on master and core nodes with storage reclaimed when the cluster ends
- **B.** EMRFS implementation of HDFS used for reading and writing regular files from Amazon EMR directly to Amazon S3 ✔
- **C.** Master and Core nodes running on EC2 that comes with a preconfigured block of preattached disk storage called an instance store
- **D.** Master and Core nodes running on local file system or local connected disks

Explanation :

Answer: B
A. No. Provides Ephemeral storage can be enabled through HDFS
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. Yes. Provides the convenience of storing persistent data in Amazon S3 for use with Hadoop while also providing features like Amazon S3 server-side encryption, read-after-write consistency, and list consistency
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)
C. No. Each node is created from an EC2 instance that comes with a preconfigured block of pre-attached disk storage called an instance store. Data on instance store volumes persists only during the life of its EC2 instance
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)
D. No. This is same as above defined in option C. The local file system refers to a locally connected disk. When you create a Hadoop cluster, each node is created from an Amazon EC2 instance that comes with a preconfigured block of pre-attached disk storage called an instance store. Data on instance store volumes persists only during the lifecycle of its Amazon EC2 instance.

Ask our Experts

QUESTION 16 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.
The company is looking at data scientists to access interactive query service to access the data (structured, semi-structured, unstructured) loaded in S3 and recommend and provide insights to improve results of customer marketing campaign. The database grows by up to 100 GB per day. To reduce time to insight, optimize costs, and increase flexibility for its analysis, which tool can

provide interactive querying capability for the datasets available in CSV, JSON, or columnar data formats such as Apache Parquet and Apache ORC OOTB with S3? Select 1 option.

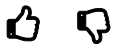
- **A.** Athena helps you analyze unstructured, semi-structured, and structured data stored in Amazon S3 and allow run ad-hoc queries using ANSI SQL, without the need to aggregate or load the data into Athena. ✔
- **B.** QuickSight helps you helps you analyze unstructured, semi-structured, and structured data stored in Amazon S3 and allow run ad-hoc queries using ANSI SQL, without the need to aggregate or load the data into QuickSight
- **C.** AWS Glue helps you analyze unstructured, semi-structured, and structured data stored in Amazon S3 and allow run ad-hoc queries using ANSI SQL, without the need to aggregate or load the data into AWS Glue
- **D.** AWS ML helps you analyze unstructured, semi-structured, and structured data stored in Amazon S3 and allow run ad-hoc queries using ANSI SQL, without the need to aggregate or load the data into AWS ML

Explanation :

Answer: A
A. Yes. 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/when-should-i-use-ate.html (https://docs.aws.amazon.com/athena/latest/ug/when-should-i-use-ate.html)
B. 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)
C. No. AWS Glue is a fully managed extract, transform, and load (ETL) service that makes it easy for customers to prepare and load their data for analytics. Point AWS Glue to your data stored on AWS, and AWS Glue discovers your data and stores the associated metadata (e.g. table definition and schema) in the AWS Glue Data Catalog. Once cataloged, your data is immediately searchable, queryable, and available for ETL.
D. No. Amazon Machine Learning (Amazon ML) is a robust, cloud-based service that makes it easy for developers of all skill levels to use machine learning technology. Amazon ML provides

visualization tools and wizards that guide you through the process of creating machine learning (ML) models without having to learn complex ML algorithms and technology. https://docs.aws.amazon.com/machine-learning/latest/dg/what-is-amazon-machine-learning.html (https://docs.aws.amazon.com/machine-learning/latest/dg/what-is-amazon-machine-learning.html)

Ask our Experts



QUESTION 17 UNATTEMPTED PROCESSING

MSP Bank, Limited is a leading varied 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 kind of tasks are supported by AWS Glue?Select 4 options.

- [ ] **A.** Discovers and catalogs metadata about data stores into AWS Glue catalog ✔
- [ ] **B.** Populates the AWS Glue Data Catalog with table definitions from scheduled crawler programs which classifier logic to infer the schema, format, and data types of the data ✔

- [ ] **C.** Real-time and streaming data ingestion and integration
- [ ] **D.** Generates ETL scripts based on python, scala to transform, flatten, and enrich data from source to target ✔
- [ ] **E.** Generates ETL scripts based on node.js, Shell Scripting to transform, flatten, and enrich data from source to target
- [ ] **F.** Triggers ETL jobs based on a schedule or event and scales resources, as needed, to run jobs ✔

Explanation :

Answer: A, B, D, F

AWS Glue simplifies many tasks when you are building a data warehouse:

- Discovers and catalogs metadata about your data stores into a central catalog. You can process semi-structured data, such as clickstream or process logs.
- Populates the AWS Glue Data Catalog with table definitions from scheduled crawler programs. Crawlers call classifier logic to infer the schema, format, and data types of your data. This metadata is stored as tables in the AWS Glue Data Catalog and used in the authoring process of your ETL jobs.
- Generates ETL scripts to transform, flatten, and enrich your data from source to target.
- Detects schema changes and adapts based on your preferences.
- Triggers your ETL jobs based on a schedule or event. You can initiate jobs automatically to move your data into your data warehouse. Triggers can be used to create a dependency flow between jobs.
- Gathers runtime metrics to monitor the activities of your data warehouse.
- Handles errors and retries automatically.
- Scales resources, as needed, to run your jobs.

https://docs.aws.amazon.com/glue/latest/dg/what-is-glue.html#when-to-use-glue (https://docs.aws.amazon.com/glue/latest/dg/what-is-glue.html#when-to-use-glue)

Ask our Experts

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 clickstream data and use custom-build recommendation engine to recommend products which eventually improve sales, understand customer preferences and already using AWS Kinesis Producer Library 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 -

Apply transformation of syslog data to CSV format
Load the data capture, along with other transformations into Redshift
Capture transformation failures
Capture delivery failures
Backup the syslog streaming data into a separate S3 bucket
Select 3 options.

- [ ] **A.** Streaming data can be directly loaded into Redshift from Kinesis Firehose
- [ ] **B.** Streaming data is delivered to your S3 bucket first. Kinesis Data Firehose then issues an Amazon Redshift COPY command to load data from your S3 bucket to your Amazon Redshift cluster ✔
- [ ] **C.** Streaming data is delivered to your S3 bucket first. Kinesis Data Firehose then issues an Amazon Redshift Export command to load data from your S3 bucket to your Amazon Redshift cluster
- [ ] **D.** The transformation failures and delivery failures are loaded into processing-failed and errors folders in same S3 bucket ✔
- [ ] **E.** The transformation failures and delivery failures are loaded into transform-failed and delivery-failed folders in same S3 bucket
- [ ] **F.** when Redshift 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 ✔
- [ ] **G.** S3 backups can be managed to bucket policies

Explanation :

Answer: B, D, F

A. No. For Amazon Redshift destinations, streaming data is delivered to your S3 bucket first. Kinesis Data Firehose then issues an Amazon Redshift COPY command to load data from your S3 bucket to your Amazon Redshift cluster. If data transformation is enabled, you can optionally back up source data to another Amazon S3 bucket.
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)

B. Yes. For Amazon Redshift destinations, streaming data is delivered to your S3 bucket first. Kinesis Data Firehose then issues an Amazon Redshift COPY command to load data from your S3 bucket to your Amazon Redshift cluster. If data transformation is enabled, you can optionally back up source data to another Amazon S3 bucket.
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)

C. No. For Amazon Redshift destinations, streaming data is delivered to your S3 bucket first. Kinesis Data Firehose then issues an Amazon Redshift COPY command to load data from your S3 bucket to your Amazon Redshift cluster. If data transformation is enabled, you can optionally back up source data to another Amazon S3 bucket.
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)

D.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)

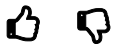
E. No. when S3 is selected as destination, and Source record S3 backup is enabled, untransformed incoming data can be delivered to a separate S3 bucket and errors are delivered to processing-failed and errors folder in S3 bucket
https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html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

F. 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)

G. 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 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 19 UNATTEMPTED COLLECTION

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. Gluebush.com 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. The DWH process the data for further analytics.
Gluebush.com want to extend the capabilities like search, document management, integration into Data Lake built on EMR, etc using the same stream without impacting performance besides above 3 purposes mentioned above.
Please advise key artefacts of end to end architecture. select 1 option.

- **A.** data collection, pre-processing, and writing of data into data streams using KPL and reading of data using shared fan-out consumers using KCL library and writing to downstream applications using connector libraries. New consumer applications need to be added.

- **B.** data collection, pre-processing, and writing of data into data streams using KPL and reading of data using shared enhanced fan-out consumers using KCL library and writing to downstream applications using connector libraries. existing consumer applications would be sufficient.
- **C.** data collection, pre-processing, and writing of data into data streams using Kinesis Agents and reading of data using enhanced fan-out consumers built using kinesis Applications and writing to downstream applications using connector libraries. New consumer applications need to be added. ✔
- **D.** data collection, pre-processing, and writing of data into data streams using Kinesis Agents and reading of data using shared fan-out consumers built using kinesis Applications and writing to downstream applications using connector libraries. Existing consumer applications would be sufficient.
- **E.** data collection, pre-processing, and writing of data into data streams using API's and reading of data using shared fan-out consumers built on API's and writing to downstream applications using connector libraries. Existing consumer applications would be sufficient.

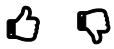
Explanation :

Answer : C
A. Kinesis data stream producer library cannot be used to read log files. we need kinesis agent to collect the information from the log files.
Enhanced Fan-out consumers is mandatory if we need to ensure performance is not degraded with additional consumers. shared fan-out basically distributes the total shard read capacity to any number of consumers and eventually impact performance.
Kinesis data streams producer is any application that puts user data records into a Kinesis data stream. The KPL is an easy-to-use, highly configurable library that helps you write to a Kinesis data stream and acts as an intermediary between your producer application code and the Kinesis Data Streams API actions. KPL library cannot be used to continuously monitors a set of files and sends new data to your stream.
https://docs.aws.amazon.com/streams/latest/dev/developing-producers-with-kpl.html (https://docs.aws.amazon.com/streams/latest/dev/developing-producers-with-kpl.html)
An Amazon Kinesis Data Streams application is a consumer of a stream that commonly runs on a fleet of EC2 instances.
The shard Read throughput is fixed at a total of 2 MiB/sec per shard. If there are multiple consumers reading from the same shard, they all share the throughput
https://docs.aws.amazon.com/streams/latest/dev/building-consumers.html (https://docs.aws.amazon.com/streams/latest/dev/building-consumers.html)
B. No. Kinesis data stream producer library cannot be used to read log files. we need kinesis agent to collect the information from the log files.

Enhanced Fan-out consumers **is** mandatory if we need to ensure performance is not degraded with additional consumers.
Kinesis data streams producer is any application that puts user data records into a Kinesis data stream. The KPL is an easy-to-use, highly configurable library that helps you write to a Kinesis data stream and acts as an intermediary between your producer application code and the Kinesis Data Streams API actions. KPL library cannot be used to continuously monitors a set of files and sends new data to your stream.
https://docs.aws.amazon.com/streams/latest/dev/developing-producers-with-kpl.html (https://docs.aws.amazon.com/streams/latest/dev/developing-producers-with-kpl.html)
An Amazon Kinesis Data Streams application is a consumer of a stream that commonly runs on a fleet of EC2 instances.
Scales as consumers register to use enhanced fan-out. Each consumer registered as enhanced fan-out gets its own read throughput per **shard**, up to 2 MiB/sec, independently of other consumers
https://docs.aws.amazon.com/streams/latest/dev/building-consumers.html (https://docs.aws.amazon.com/streams/latest/dev/building-consumers.html)
C. Yes. we need kinesis agent to collect the information from the log files. Enhanced Fan-out consumers is mandatory if we need to ensure performance is not degraded with additional consumers.
Kinesis Agent is a stand-alone Java application that allows to collect and process data to Kinesis Data Streams. The agent continuously monitors a set of files and sends new data to your stream
https://docs.aws.amazon.com/streams/latest/dev/writing-with-agents.html (https://docs.aws.amazon.com/streams/latest/dev/writing-with-agents.html)
An Amazon Kinesis Data Streams application is a consumer of a stream that commonly runs on a fleet of EC2 instances.
Scales as consumers register to use enhanced fan-out. Each consumer registered as enhanced fan-out gets its own read throughput per shard, up to 2 MiB/sec, independently of other consumers
https://docs.aws.amazon.com/streams/latest/dev/building-consumers.html (https://docs.aws.amazon.com/streams/latest/dev/building-consumers.html)
D. No. we need kinesis agent to collect the information from the log files. Enhanced Fan-out consumers is mandatory if we need to ensure performance is not degraded with additional consumers. shared fan-out basically distributes the total shard read capacity to any number of consumers and eventually impact performance.
Kinesis Agent is a stand-alone Java application that allows to collect and process data to Kinesis Data Streams. The agent continuously monitors a set of files and sends new data to your stream
https://docs.aws.amazon.com/streams/latest/dev/writing-with-agents.html (https://docs.aws.amazon.com/streams/latest/dev/writing-with-agents.html)
An Amazon Kinesis Data Streams application is a consumer of a stream that commonly runs on a fleet of EC2 instances.
The shard Read throughput is fixed at a total of 2 MiB/sec per shard. If there are multiple consumers reading from the same shard, they all share the throughput
https://docs.aws.amazon.com/streams/latest/dev/building-consumers.html

(https://docs.aws.amazon.com/streams/latest/dev/building-consumers.html)
E. No. we need kinesis agent to collect the information from the log files. API's cannot read data from log files.
Enhanced Fan-out consumers is mandatory if we need to ensure performance is not degraded with additional consumers. shared fan-out basically distributes the total shard read capacity to any number of consumers and eventually impact performance.
We can develop producers using the Amazon Kinesis Data Streams API with the AWS SDK for Java but API SDK cannot be used to continuously monitors a set of files and sends new data to your stream.
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)
An Amazon Kinesis Data Streams application is a consumer of a stream that commonly runs on a fleet of EC2 instances.
The shard Read throughput is fixed at a total of 2 MiB/sec per shard. If there are multiple consumers reading from the same shard, they all share the throughput
https://docs.aws.amazon.com/streams/latest/dev/building-consumers.html
(https://docs.aws.amazon.com/streams/latest/dev/building-consumers.html)

Ask our Experts



QUESTION 20 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 is looking at streaming dataflow engine that you can use to run real-time stream processing on high-throughput data sources and support event time semantics for out-of-order events, exactly-once semantics, backpressure control, and APIs optimized for writing both streaming and batch applications. What EMR Hadoop eco-system component can fulfill this requirement? The component also needs to integrate with other AWS services like Kinesis Streams and Elasticsearch. select 1 option.

- A. Apache Hue
- B. Apache Flink ✔
- C. Apache Phoenix
- D. Apache Tez

Explanation :

Answer : B
A. No. 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. Yes. 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

QUESTION 21 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. Please advise why do you think DynamoDB is a right platform to address QuickDialog Business needs.
Select 3 options.

- [ ] **A.** Support key-value and document based data models with a provision to have a flexible schema, so each row can have any number of columns at any point in time which can cater new business requirements at relative ease ✔
- [ ] **B.** Provides two read/write capacity modes for each table: on-demand and provisioned ✔
- [ ] **C.** Uses columnar storage, data compression, and zone maps to reduce the amount of I/O needed to perform queries and uses a massively parallel processing (MPP) architecture to parallelize and distribute SQL operations to take advantage of all available resources
- [ ] **D.** Replicate your data automatically across your choice of AWS Regions and automatically scale capacity to accommodate your workloads so that globally distributed applications can access data locally in the selected regions to get single-digit millisecond read and write performance ✔
- [ ] **E.** utilizes sophisticated algorithms to predict incoming query run times, and assigns them to the optimal queue for the fastest processing

Explanation :

Answer : A,B, D
A. Yes. DynamoDB supports both key-value and document data models. This enables DynamoDB to have a flexible schema, so each row can have any number of columns at any point in time.
https://aws.amazon.com/dynamodb/features/ (https://aws.amazon.com/dynamodb/features/)
B. Yes. DynamoDB provides two read/write capacity modes for each table: on-demand and

provisioned. For workloads that are less predictable for which you are unsure that you will have high utilization, on-demand capacity mode takes care of managing capacity for you, and you only pay for what you consume. Tables using provisioned capacity mode require you to set read and write capacity.
https://aws.amazon.com/dynamodb/features/ (https://aws.amazon.com/dynamodb/features/)
C. No. Amazon Redshift delivers fast query performance on datasets ranging in size from gigabytes to exabytes. Redshift uses columnar storage, data compression, and zone maps to reduce the amount of I/O needed to perform queries. It uses a massively parallel processing (MPP) data warehouse architecture to parallelize and distribute SQL operations to take advantage of all available resources.
https://aws.amazon.com/redshift/features/ (https://aws.amazon.com/redshift/features/)
D. Yes. DynamoDB global tables replicate your data automatically across your choice of AWS Regions and automatically scale capacity to accommodate your workloads. With global tables, your globally distributed requests can access data locally in the selected regions to get single-digit millisecond read and write performance
E. No. this feature is provided by Redshift. Amazon Redshift uses machine learning to deliver high throughout based on your workloads. Redshift utilizes sophisticated algorithms to predict incoming query run times, and assigns them to the optimal queue for the fastest processing.
https://aws.amazon.com/redshift/features/ (https://aws.amazon.com/redshift/features/)

Ask our Experts

QUESTION 22 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 Producer Library 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

perform Resharding. There are 2 shards SHARD 1 with a hash key range of 276...381, SHARD 2 with a hash key range of 382...454. Post re-sharding how does Kinesis Stream behave?Select 5 options.

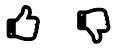
- [ ] **A.** Data records that were flowing to the parent shards are re-routed to flow to the child shards based on the hash key values that the data-record partition keys map to ✔
- [ ] **B.** The parent shards disappear when the reshard occurs and move all the records in the parent shard to child shards
- [ ] **C.** data records that were in the parent shards before the reshard remain in those shards ✔
- [ ] **D.** Before a reshard operation, a parent shard is in the OPEN state ✔
- [ ] **E.** After a reshard operation, the parent shard transitions to a CLOSED state ✔
- [ ] **F.** After a reshard operation, the parent shard still in OPEN state
- [ ] **G.** After a reshard operation, the child shard still in CLOSED state
- [ ] **H.** After the stream's retention period has expired, parent SHARD moves to EXPIRED state ✔

Explanation :

Answer: A, C, D, E, H
A. Yes. Data records that were flowing to the parent shards are re-routed to flow to the child shards based on the hash key values that the data-record partition keys map to
https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html)
B. No. The parent shards does not disappear when the reshard occurs
https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html)
C. Yes. data records that were in the parent shards before the reshard remain in those shards
https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html)
D. Yes. Before a reshard operation, a parent shard is in the OPEN state
https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-

resharding.html)
E. Yes. After a reshard operation, the parent shard transitions to a CLOSED state
https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html)
F. No. After a reshard operation, the parent shard will be in CLOSED state
https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html)
G. No. After a reshard operation, the child shard will be in OPEN state
https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html)
H. Yes. After the stream's retention period has expired, parent SHARD moves to EXPIRED state
https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-after-resharding.html)

Ask our Experts



QUESTION 23 UNATTEMPTED STORAGE

Hymutabs Ltd (Hymutabs) is a global environmental solutions company running its operations in 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
LOCATION_DIM is a newly identified table and has around 2.8 million rows and size increases 4% every month.
Hymutabs has very limited number of suppliers. The administrator just left the company for good and also not available during design meetings. There is a urgent need to deploy LOCATION_DIM. Being a new table and no workload requirements, the team is not aware of what approach need to be taken. What is the safest approach for distribution? select 1 option.

- **A.** DISTRIBUTE LOCATION_DIM with KEY DISTRIBUTION style
- **B.** DISTRIBUTE LOCATION_DIM with EVEN DISTRIBUTION style
- **C.** DISTRIBUTE LOCATION_DIM with ALL DISTRIBUTION style
- **D.** DISTRIBUTE LOCATION_DIM without ANY DISTRIBUTION style ✔

Explanation :

Answer : D
A. KEY DISTRIBUTION distributes the rows are according to the values in one column. This is the perfect solution with distribution key on same keys.
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. 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 (s) does not participate in joins. 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. 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 design cannot be applied for large 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. when the administrator is not available and we are not sure of the workload, it is always better to go with no no distribution style because RedShift handles the distribution mechanism. https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-best-dist-key.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-best-dist-key.html)

Ask our Experts

QUESTION 24 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 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. Also Parson Fortunes use other AWS services like Athena, and EMR. How can this data be accessed through Redshift Spectrum? Select 3 options.

- [ ] **A.** Redshift Spectrum accesses external databases in Athena and EMR using external schema and external tables ✔
- [ ] **B.** Amazon Redshift external schema references an external database in an external data catalog ✔
- [ ] **C.** Amazon Redshift external schema references an external database in an internal data catalog
- [ ] **D.** For external schemas, Amazon Redshift needs authorization to access the data catalog in Athena and the data files in Amazon S3 using IAM Roles and policies ✔
- [ ] **E.** For external schemas, Amazon Redshift needs authorization to access the data catalog in Athena, but not the data files in Amazon S3 using IAM Roles and policies

Explanation :

Answer : A, B, D
A. Yes. Redshift Spectrum accesses external databases in Athena and EMR using external schema and external tables
https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-external-schemas.html (https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-external-schemas.html)
B. Yes. Amazon Redshift external schema references an external database in an external data catalog. You can create the external database in Amazon Redshift, in Amazon Athena, or in an Apache Hive metastore, such as Amazon EMR.
https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-external-schemas.html (https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-external-schemas.html)
C. No. Amazon Redshift external schema references an external database in an external data catalog. You can create the external database in Amazon Redshift, in Amazon Athena, or in an Apache Hive metastore, such as Amazon EMR.
https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-external-schemas.html (https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-external-schemas.html)
D. Yes. Amazon Redshift needs authorization to access the data catalog in Athena and the data files in Amazon S3 on your behalf. To provide that authorization, you first create an AWS Identity and Access Management (IAM) role. Then you attach the role to your cluster and provide Amazon Resource Name (ARN) for the role in the Amazon Redshift CREATE EXTERNAL SCHEMA statement.
https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-external-schemas.html (https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-external-schemas.html)
E. No. Amazon Redshift needs authorization to access the data catalog in Athena and the data files in Amazon S3 on your behalf. To provide that authorization, you first create an AWS Identity and Access Management (IAM) role. Then you attach the role to your cluster and provide

Amazon Resource Name (ARN) for the role in the Amazon Redshift CREATE EXTERNAL SCHEMA statement.
https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-external-schemas.html (https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-external-schemas.html)

Ask our Experts

QUESTION 25 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 clickstream 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 has 2 different Kinesis Streams that are being configured as 2 in-application streams to be accessed by Kinesis Analytics application. The data joined between the 2 in-application streams need to go through multiple transformations and conversions. What are the key artefacts of streaming SQL capabilities provided by Kinesis analytics? Select 3 options.

- [ ] **A.** in-application stream and kinesis data stream are same
- [ ] **B.** in-application stream and kinesis firehose stream are same
- [ ] **C.** More in-application streams are needed to store intermediate query results ✔
- [ ] **D.** in-application streams are needed cannot store intermediate query results
- [ ] **E.** Insert data into the in-application stream using a pump ✔
- [ ] **F.** in-application stream is analogous to publish/subscribe messaging paradigm ✔

☐ **G.** in-application stream is analogous to DB SQL paradigm

Explanation :

Answer: C,E,F

A. No. In-application streams exist only in the context of an Amazon Kinesis Data Analytics application. Kinesis data streams and Kinesis Data Firehose delivery streams exist independent of your application.
https://docs.aws.amazon.com/kinesisanalytics/latest/dev/streams-pumps.html (https://docs.aws.amazon.com/kinesisanalytics/latest/dev/streams-pumps.html)

B. No. In-application streams exist only in the context of an Amazon Kinesis Data Analytics application. Kinesis data streams and Kinesis Data Firehose delivery streams exist independent of your application.
https://docs.aws.amazon.com/kinesisanalytics/latest/dev/streams-pumps.html (https://docs.aws.amazon.com/kinesisanalytics/latest/dev/streams-pumps.html)

C. Yes. more in-application streams as needed to store intermediate query results
https://docs.aws.amazon.com/kinesisanalytics/latest/dev/streams-pumps.html (https://docs.aws.amazon.com/kinesisanalytics/latest/dev/streams-pumps.html)

D. No. more in-application streams as needed to store intermediate query results
https://docs.aws.amazon.com/kinesisanalytics/latest/dev/streams-pumps.html (https://docs.aws.amazon.com/kinesisanalytics/latest/dev/streams-pumps.html)

E. Yes. Insert data into the in-application stream using a pump which is a continuous insert query running that inserts data from one in-application stream to another in-application stream
https://docs.aws.amazon.com/kinesisanalytics/latest/dev/streams-pumps.html (https://docs.aws.amazon.com/kinesisanalytics/latest/dev/streams-pumps.html)

F. Yes. In-application stream is analogous to implementing a publish/subscribe messaging paradigm. In this paradigm, the data row, including the time of creation and time of receipt, can be processed, interpreted, and forwarded by a cascade of streaming SQL statements
https://docs.aws.amazon.com/kinesisanalytics/latest/dev/streams-pumps.html (https://docs.aws.amazon.com/kinesisanalytics/latest/dev/streams-pumps.html)

G. No. In-application stream is analogous to implementing a publish/subscribe messaging paradigm. In this paradigm, the data row, including the time of creation and time of receipt, can be processed, interpreted, and forwarded by a cascade of streaming SQL statements
https://docs.aws.amazon.com/kinesisanalytics/latest/dev/streams-pumps.html (https://docs.aws.amazon.com/kinesisanalytics/latest/dev/streams-pumps.html)

Ask our Experts

QUESTION 26 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 uses event based streaming, based on Kinesis Stream to address data integration and uses producer library to integrate events. Also Tick-Bank is looking at data transformations which will be processed using Kinesis Firehose.
The enterprise architecture team is trying to understand the default configuration parameters of kinesis firehose. Please advice. Select 3 options.

- [ ] **A.** The maximum size of kinesis firehose delivery stream record size is 1000 KB ✔
- [ ] **B.** the maximum size of kinesis firehose delivery stream record size is 1 MB
- [ ] **C.** The buffer size of Kinesis Firehose stream is between 1-128 MB, buffer interval can be 60-900 seconds ✔
- [ ] **D.** The buffer size of Kinesis Firehose stream is between 1-1024 MB, buffer interval can be 5-900 seconds
- [ ] **E.** Kinesis Firehose supports only GZIP, Snappy, Zip data compression, and no data compression ✔
- [ ] **F.** Kinesis Firehose supports only GZIP, Snappy and Zip data compression. Data compression is enabled by default

Explanation :

Answer: A, C, E
The maximum size of kinesis firehose delivery stream record size is 1000 KB, while The buffer size of Kinesis Firehose stream is between 1-128 MB and buffer interval can be 60-900 seconds and Kinesis Firehose supports only GZIP, Snappy, Zip data compression, and no data compression.
https://docs.aws.amazon.com/firehose/latest/dev/create-configure.html
(https://docs.aws.amazon.com/firehose/latest/dev/create-configure.html)

Ask our Experts

QUESTION 27 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 understands that the performance of MapReduce is not competent based on the SLA's and is looking for an alternative framework instead of MapReduce which runs by creating complex directed acyclic graph (DAG) of tasks for processing data
Which EMR Hadoop ecosystem fulfills the requirements? select 1 option.

- A. Apache Hue
- B. Apache Flink
- C. Apache Phoenix
- D. Apache Tez ✔

Explanation :

Answer : D

A. No. 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. Yes. 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

QUESTION 28 UNATTEMPTED ANALYSIS

PMG Group Malaysia is a Chinese group of companies best known for its book retailing and online retailing services as well as being involved in the printing, publishing and supply of books and library services in China and Taiwan. PMG Bookstores currently has 65 outlets in China and 6 in Taiwan.
The management team has been strengthened to improve its customer service and its range of books. Steps have been taken to upgrade the computer system to improve the efficiency of PMG Bookstores' inventory control and customer service delivery. PMG Bookstores continues to seek choice locations for new

outlets in China. PMG Group hosts their web application to sell the books and improve web sales. The application is built on AWS running out EC2 and RDS. PMG Group understand recently an avalanche of negative reviews about some of the books released in market and is interested to know whether the reviews provided are by customers or bots. PMG identifies AWS ML to provide a quick turnaround. Please advice. 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-class classification 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 29 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. FlexiToner understands that Athena provides this facility OOTB.
FlexiToner has a group of big data professionals specialized in Hadoop and want

to understand the underlying hadoop components under the hood so that the team can easily understand the platform and use it for further analytics. Please advise.

Select 2 options.

- [ ] **A.** Athena uses Hive to execute DML statements
- [ ] **B.** Athena uses Hive to execute the DDL statements that create and modify schema ✔
- [ ] **C.** Athena uses Presto to execute the DDL statements that create and modify schema
- [ ] **D.** Athena uses Presto to execute DML statements ✔

Explanation :

Answer: B,D
A. No. Athena uses Presto to execute DML statements
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. Athena uses Hive to execute the DDL statements that create and modify schema
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. Athena uses Hive to execute the DDL statements that create and modify schema
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. Athena uses Presto to execute DML statements
https://docs.aws.amazon.com/athena/latest/ug/glue-best-practices.html
(https://docs.aws.amazon.com/athena/latest/ug/glue-best-practices.html)

Ask our Experts

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.
HH understand that Monitoring plays a major role in monitoring and managing the streaming platform. HH using producer library to process events and consumer library to collect and disseminate. What kind of monitoring can be enabled using CloudWatch? Select 4 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.** Metrics are archived for two weeks; after that period, the data is discarded ✔
- [ ] **D.** Metrics are archived for one week; after that period, the data is discarded
- [ ] **E.** Basic stream level data is captured every minute at no additional charge ✔
- [ ] **F.** Basic stream level data is captured every second at no additional charge
- [ ] **G.** Shard-level data is sent every second at no additional charge
- [ ] **H.** Shard-level data is sent every minute for an additional cost. ✔

Explanation :

Answer: A, C, E, H
A. Yes. Metrics configured for streams are automatically collected and pushed to CloudWatch every minute

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
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. Metrics are archived for two weeks; after that period, the data is discarded
https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html (https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html)
D. No. Metrics are archived for two weeks; after that period, the data is discarded
https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html (https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html)
E. Yes. Basic stream level data is captured every minute at no additional charge. This includes basic information about stream.
https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html (https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html)
F. No. Basic stream level data is captured every minute at no additional charge. This includes basic information about stream.
https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html (https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html)
G. No. Enhanced Shard-level Metrics are being sent to CloudWatch every minute. These metrics are not enabled by default. There is a charge for enhanced metrics emitted from Kinesis
https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html (https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html)
H. Yes. Enhanced Shard-level Metrics are being sent to CloudWatch every minute. These metrics are not enabled by default. There is a charge for enhanced metrics emitted from Kinesis
https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html (https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html)

Ask our Experts

QUESTION 31 UNATTEMPTED DATA SECURITY

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. EMRFS Consistency View is enabled. EMRFS on S3 Encryption needs to be enabled to align to enterprise security guidelines and Consistency notifications. Please advise.

Select 2 options.

- ☐ **A.** S3 server-side encryption with KMS Key Management (SSE-KMS)
- ☐ **B.** AWS KMS Customer Master Keys (CMKs) for EMRFS Encryption ✔
- ☐ **C.** S3 client-side encryption with custom materials provider ✔
- ☐ **D.** server-side encryption with custom materials provider

Explanation :

Answer : B, C
A. No. S3 server-side encryption with KMS Key Management (SSE-KMS) is not available when using Amazon EMR release version 4.4 or earlier
B. Yes. Enable CloudWatch metrics and Amazon SQS messages in EMRFS for Amazon S3 eventual consistency issues
https://docs.aws.amazon.com/emr/latest/ManagementGuide/emrfs-configure-sqs-cw.html (https://docs.aws.amazon.com/emr/latest/ManagementGuide/emrfs-configure-sqs-cw.html)
C. Yes. S3 client-side encryption with custom materials provider can be enabled by providing the Class name and the JAR location
D. No does not support S3 server side encryption with custom materials provider

Ask our Experts

QUESTION 32 UNATTEMPTED COLLECTION

JP Industries Limited is Indian home appliances company based in Hyderabad, Telangana, India. It manufactures and markets various Home Appliances and Non-stick cookware's under the brand name of JP. It is also manufactures of consumer durable products, laptops, desktops, mobile handsets and other devices. Off late JP had undergone throughlot of thefts of their assets and JP is looking at building asset tracking through AWS IOT thereby allowing enterprise to easily locate and monitor key assets, including along the supply chain (e.g. raw materials, final products and containers) to optimize logistics, maintain inventory levels, prevent quality issues and detect theft.
IT teams are working on authentication of mobile applications, web and desktop applications and IOT devices. Suggest appropriate authentication mechanisms for each of the devices listed above. Select 3 options.

- [ ] **A.** AWS IoT devices use X.509 certificates for authentication ✔
- [ ] **B.** AWS IoT devices use IAM groups, users, roles for authentication
- [ ] **C.** Mobile Applications use Amazon Cognito Identities for authentication ✔
- [ ] **D.** Mobile Applications use Federated identities for authentication
- [ ] **E.** Web and desktop applications use IAM or federated entities for authentication ✔
- [ ] **F.** Web and desktop applications use x.509 certificates for authentication

Explanation :

Answer : A, C, E
A. Yes. AWS IoT devices use X.509 certificates
https://docs.aws.amazon.com/iot/latest/developerguide/iot-authentication.html
(https://docs.aws.amazon.com/iot/latest/developerguide/iot-authentication.html)
B. No. AWS IOT devices cannot be authenticated using IAM groups, users, roles for authentication
https://docs.aws.amazon.com/iot/latest/developerguide/iot-authentication.html
(https://docs.aws.amazon.com/iot/latest/developerguide/iot-authentication.html)
C. Yes. Mobile Applications use Amazon Cognito Identities for authentication
https://docs.aws.amazon.com/iot/latest/developerguide/iot-authentication.html
(https://docs.aws.amazon.com/iot/latest/developerguide/iot-authentication.html)

D. No. Mobile Applications cannot use Federated identities for authentication
https://docs.aws.amazon.com/iot/latest/developerguide/iot-authentication.html
(https://docs.aws.amazon.com/iot/latest/developerguide/iot-authentication.html)
E. Yes. Web and desktop applications use IAM or federated entities for authentication
https://docs.aws.amazon.com/iot/latest/developerguide/iot-authentication.html
(https://docs.aws.amazon.com/iot/latest/developerguide/iot-authentication.html)
F. No. Web and desktop applications cannot use x.509 certificates for authentication
https://docs.aws.amazon.com/iot/latest/developerguide/iot-authentication.html
(https://docs.aws.amazon.com/iot/latest/developerguide/iot-authentication.html)

Ask our Experts

QUESTION 33 UNATTEMPTED PROCESSING

MSP Bank, Limited is a leading varied 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 enabling auditing to track the records of actions taken by a user, role, or an AWS service in Amazon QuickSight. Also the team need to capture the logs and storage it for long term archival to address compliance. Please advice. Select 3 options.

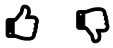
- [ ] **A.** Amazon QuickSight is integrated with AWS CloudTrail which provides a record of actions taken by a user, role, or an AWS service in Amazon QuickSight ✔
- [ ] **B.** Amazon QuickSight is integrated with AWS CloudWatch which provides a record of actions taken by a user, role, or an AWS service in Amazon QuickSight
- [ ] **C.** when CloudTrail is enabled, you can enable continuous delivery of CloudTrail events to an Amazon S3 bucket, including events for Amazon QuickSight ✔

- [ ] **D.** when CloudWatch is enabled, you can enable continuous delivery of CloudWatch events to an Amazon S3 bucket, including events for Amazon QuickSight
- [ ] **E.** If you don't configure a trail, you can still view the most recent events in the CloudTrail console in Event history ✔
- [ ] **F.** If you don't configure a log, you can still view the most recent events in the CloudWatch console in Event history

Explanation :

Answer: A,C,E
Amazon QuickSight is integrated with AWS CloudTrail. This service provides a record of actions taken by a user, role, or an AWS service in Amazon QuickSight. The calls captured include calls from the Amazon QuickSight console. If you create a trail, you can enable continuous delivery of CloudTrail events to an Amazon S3 bucket, including events for Amazon QuickSight. If you don't configure a trail, you can still view the most recent events in the CloudTrail console in Event history. Using the information collected by CloudTrail, you can determine the request that was made to Amazon QuickSight, the IP address from which the request was made, who made the request, when it was made, and additional details
https://docs.aws.amazon.com/quicksight/latest/user/logging-using-cloudtrail.html (https://docs.aws.amazon.com/quicksight/latest/user/logging-using-cloudtrail.html)

Ask our Experts



QUESTION 34 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 is trying to train some of their new hires on DynamoDB and asked your help to define the different read/write capacity modes for processing reads and writes of the table. Please advice. Select 2 options.

- [ ] **A.** On-demand Mode enables billing option capable of serving thousands of requests per second without capacity planning thereby offering pay-per-request pricing for read and write requests. Mostly used for unknown workloads, and unpredictable application traffic ✔
- [ ] **B.** Provisioned Mode enables billing option capable of serving thousands of requests per second without capacity planning thereby offering pay-per-request pricing for read and write requests. Mostly used for unknown workloads, and unpredictable application traffic
- [ ] **C.** On-demand Mode enables to specify the number of reads and writes per second that is required for the application and also allow use auto scaling to adjust your table's provisioned capacity automatically in response to traffic changes. Mostly used for known workloads, and predictable application traffic
- [ ] **D.** Provisioned Mode enables to specify the number of reads and writes per second that is required for the application and also allow use auto scaling to adjust your table's provisioned capacity automatically in response to traffic changes. Mostly used for known workloads, and predictable application traffic ✔

Explanation :

Answer: A,D
A. On-demand Mode enables billing option capable of serving thousands of requests per second without capacity planning thereby offering pay-per-request pricing for read and write requests
B. Provisioned Mode enables to specify the number of reads and writes per second that is required for the application and also allow use auto scaling to adjust your table's provisioned capacity automatically in response to traffic changes
C. On-demand Mode enables billing option capable of serving thousands of requests per second without capacity planning thereby offering pay-per-request pricing for read and write requests
D. Provisioned Mode enables to specify the number of reads and writes per second that is required for the application and also allow use auto scaling to adjust your table's provisioned capacity automatically in response to traffic changes

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 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.
Parson want to securely integrate the data between Redshift and S3 in a secured way and not being accessed through internet for both Redshift spectrum integration and loading staging data directly from S3 buckets using copy and unload commands.
Please advise next steps. Select 4 options.

- [ ] **A.** Create a VPC endpoint to create a managed connection between your Amazon Redshift cluster in a VPC and Amazon S3 in same region ✔
- [ ] **B.** Create a VPC endpoint to create a managed connection between your Amazon Redshift cluster in a VPC and Amazon S3 in any region
- [ ] **C.** Disable Enhanced VPC Routing when you create a cluster
- [ ] **D.** Attach endpoint policy to your endpoint to more closely manage access to your data ✔
- [ ] **E.** Enable Enhanced VPC Routing when you create a cluster ✔
- [ ] **F.** Enable Domain Name Service (DNS) resolution in your VPC ✔
- [ ] **G.** Disable Domain Name Service (DNS) resolution in your VPC

Explanation :

Answer : A, D, E, F
A. Yes. Use a VPC endpoint to create a managed connection between your Amazon Redshift cluster in a VPC and Amazon Simple Storage Service (Amazon S3). When you do, COPY and UNLOAD traffic between your cluster and your data on Amazon S3 stays in your Amazon VPC

and both S3 and Redshift should be in the same region
https://docs.aws.amazon.com/redshift/latest/mgmt/enhanced-vpc-working-with-endpoints.html (https://docs.aws.amazon.com/redshift/latest/mgmt/enhanced-vpc-working-with-endpoints.html)
B. No. Use a VPC endpoint to create a managed connection between your Amazon Redshift cluster in a VPC and Amazon Simple Storage Service (Amazon S3). When you do, COPY and UNLOAD traffic between your cluster and your data on Amazon S3 stays in your Amazon VPC and both S3 and Redshift should be in the same region
https://docs.aws.amazon.com/redshift/latest/mgmt/enhanced-vpc-working-with-endpoints.html (https://docs.aws.amazon.com/redshift/latest/mgmt/enhanced-vpc-working-with-endpoints.html)
C. No. Enable Enhanced VPC Routing when you create a cluster
https://docs.aws.amazon.com/redshift/latest/mgmt/enhanced-vpc-working-with-endpoints.html (https://docs.aws.amazon.com/redshift/latest/mgmt/enhanced-vpc-working-with-endpoints.html)
D. Yes. Attach an endpoint policy to your endpoint to more closely manage access to your data.
https://docs.aws.amazon.com/redshift/latest/mgmt/enhanced-vpc-working-with-endpoints.html (https://docs.aws.amazon.com/redshift/latest/mgmt/enhanced-vpc-working-with-endpoints.html)
E. Yes. Enable Enhanced VPC Routing when you create a cluster
https://docs.aws.amazon.com/redshift/latest/mgmt/enhanced-vpc-working-with-endpoints.html (https://docs.aws.amazon.com/redshift/latest/mgmt/enhanced-vpc-working-with-endpoints.html)
F. Yes. Enable Domain Name Service (DNS) resolution in your VPC
https://docs.aws.amazon.com/redshift/latest/mgmt/enhanced-vpc-enabling-cluster.html (https://docs.aws.amazon.com/redshift/latest/mgmt/enhanced-vpc-enabling-cluster.html)
G. No. Enable Domain Name Service (DNS) resolution in your VPC
https://docs.aws.amazon.com/redshift/latest/mgmt/enhanced-vpc-enabling-cluster.html (https://docs.aws.amazon.com/redshift/latest/mgmt/enhanced-vpc-enabling-cluster.html)

Ask our Experts

QUESTION 36 UNATTEMPTED VISUALIZATION

MSP Bank, Limited is a leading varied 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 a dataset that they want to delete, on which some analysis is already generated. They submitted the deletion. What happens next? Select 3 options.

- [ ] **A.** You receive a warning if you have any analyses that use the data set you have chosen for deletion ✔
- [ ] **B.** delete both the dataset and the dependent analysis
- [ ] **C.** data set deletion does not delete the dependent analyses ✔
- [ ] **D.** Next time you open any of those analyses, you are prompted to select a new data set for any visuals that were based on the deleted data set. ✔
- [ ] **E.** QuickSight is working on keeping the analysis, even if the dataset is deleted. The latest feature will come live in March 2019 as per documentation

Explanation :

Answer: A, C, D
A. Yes. You receive a warning if you have any analyses that use the data set you have chosen for deletion
https://docs.aws.amazon.com/quicksight/latest/user/delete-a-data-set.html (https://docs.aws.amazon.com/quicksight/latest/user/delete-a-data-set.html)
B. No. only the dataset is deleted. The analysis stays live.
https://docs.aws.amazon.com/quicksight/latest/user/delete-a-data-set.html (https://docs.aws.amazon.com/quicksight/latest/user/delete-a-data-set.html)
C. Yes. data set deletion does not delete the dependent analyses
https://docs.aws.amazon.com/quicksight/latest/user/delete-a-data-set.html (https://docs.aws.amazon.com/quicksight/latest/user/delete-a-data-set.html)
D. Yes. Next time you open any of those analyses, you are prompted to select a new data set for any visuals that were based on the deleted data set.
https://docs.aws.amazon.com/quicksight/latest/user/delete-a-data-set.html (https://docs.aws.amazon.com/quicksight/latest/user/delete-a-data-set.html)
E. No. this feature is already available in QuickSight
https://docs.aws.amazon.com/quicksight/latest/user/delete-a-data-set.html (https://docs.aws.amazon.com/quicksight/latest/user/delete-a-data-set.html)

Ask our Experts

QUESTION 37 UNATTEMPTED VISUALIZATION

MSP Bank, Limited is a leading varied 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 visualize one or two measures for a dimension with rectangle representing each dimension and size of the rectangle representing the proportion of the value for the selected measure that the item represents compared to the whole for the dimension. Please advice. select 1 option.

- A. Tabular Reports
- B. Heat Maps
- C. Line Chart
- D. Tree Maps ✔

Explanation :

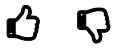
Answer: D
A. No. Use tabular reports to see a customized table view of your data.
To create a table visual, choose at least one field of any data type. You can add as many columns as you need. Plus, you can add calculated columns.
https://docs.aws.amazon.com/quicksight/latest/user/tabular.html (https://docs.aws.amazon.com/quicksight/latest/user/tabular.html)
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. 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. Yes. 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 38 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.
DynamoDB has a table with 60 partitions and is being heavily accessed by users. There are lots of hot partitions. To better accommodate uneven access patterns, how will DynamoDB use its capacity to prevent throttling. select 1 option.

- A. using Burst Capacity effectively
- B. Using Adaptive Capacity ✔

- C. Design Partition Keys to distribute workload evenly
- D. Using Write Sharding to Distribute Workloads Evenly

Explanation :

Answer: B
A. No. DynamoDB provides some flexibility in your per-partition throughput provisioning by providing burst capacity. DynamoDB reserves a portion of that unused capacity for later bursts of throughput to handle usage spikes.
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-design.html#bp-partition-key-partitions-adaptive
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-design.html#bp-partition-key-partitions-adaptive)
B. Yes. Adaptive Capacity enables your application to continue reading and writing to hot partitions without being throttled, provided that traffic does not exceed your table's total provisioned capacity or the partition maximum capacity
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-design.html#bp-partition-key-partitions-adaptive
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-design.html#bp-partition-key-partitions-adaptive)

Ask our Experts

QUESTION 39 UNATTEMPTED STORAGE

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 monitoring and managing the platform by ensuring reliability, availability, and performance of Kinesis Video Streams. Also GHMC need to track event logs of the actions taken by a user, role, or an AWS service by enabling continuous delivery in Amazon Kinesis Video Streams. Where can the log data be stored? Select 3 options.

- [ ] **A.** Amazon CloudWatch collects and processes raw data from Kinesis Video Streams into readable, near real-time metrics ✔
- [ ] **B.** Amazon CloudTrail collects and processes raw data from Kinesis Video Streams into readable, near real-time metrics
- [ ] **C.** Amazon CloudTrail captures all API calls for Amazon Kinesis Video Streams as events ✔
- [ ] **D.** Amazon CloudWatch captures all API calls for Amazon Kinesis Video Streams as events
- [ ] **E.** Enable continuous delivery of CloudTrail events, Event logs are stored in S3 with data being in log format ✔
- [ ] **F.** Enable continuous delivery of CloudWatch events, Event logs are stored in DynamoDB with data being in log format

Explanation :

Answer: A, C, E
A. Yes. Amazon CloudWatch collects and processes raw data from Kinesis Video Streams into readable, near real-time metrics
https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/monitoring-cloudwatch.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/monitoring-cloudwatch.html)
B. No. Amazon CloudWatch collects and processes raw data from Kinesis Video Streams into readable, near real-time metrics
https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/monitoring-cloudwatch.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/monitoring-cloudwatch.html)
C. Yes. Amazon CloudTrail captures all API calls for Amazon Kinesis Video Streams as events
https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/monitoring-cloudtrail.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/monitoring-cloudtrail.html)
D. No. Amazon CloudTrail captures all API calls for Amazon Kinesis Video Streams as events
https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/monitoring-cloudtrail.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/monitoring-cloudtrail.html)

E. Yes. Enable continuous delivery of CloudTrail events, Event logs are stored in S3 with data being in log format
https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/monitoring-cloudtrail.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/monitoring-cloudtrail.html)
F. No. Enable continuous delivery of CloudTrail events, Event logs are stored in S3 with data being in log format
https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/monitoring-cloudtrail.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/monitoring-cloudtrail.html)

Ask our Experts

QUESTION 40 UNATTEMPTED ANALYSIS

PMG Group Malaysia is a Chinese group of companies best known for its book retailing and online retailing services as well as being involved in the printing, publishing and supply of books and library services in China and Taiwan. PMG Bookstores currently has 65 outlets in China and 6 in Taiwan.
The management team has been strengthened to improve its customer service and its range of books. Steps have been taken to upgrade the computer system to improve the efficiency of PMG Bookstores' inventory control and customer service delivery. PMG Bookstores continues to seek choice locations for new outlets in China. PMG Group hosts their web application to sell the books and improve web sales. The application is built on AWS running out EC2 and RDS. PMG Group has lot of existing customers. They launched a campaign to sell new products based on the customer's interest that can upscale the business. Management want to know the top 50 books and their quantity every day that will be sold. Please advise. 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-class classification 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 : B, E and F
A. No. 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. Yes. 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. No. 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. Yes. 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 41 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 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.
HH want to enable end to end encryption for the Kinesis Stream. How can this be achieved? Select 4 options.

- [ ] **A.** Server-side encryption can be enabled through KMS which provides the master keys ✔
- [ ] **B.** Server side encryption can use a CMK for Kinesis that is managed by AWS ✔
- [ ] **C.** Client side encryption can use a user-specified CMK
- [ ] **D.** Client side encryption can use a master key imported into the AWS KMS service.
- [ ] **E.** Server side encryption can use a user-specified CMK ✔
- [ ] **F.** Server side encryption can use a master key imported into the AWS KMS service. ✔

Explanation :

Answer: A, B, E, F

A. Yes. Server-side encryption can be enabled through KMS which provides the master keys
https://docs.aws.amazon.com/streams/latest/dev/what-is-sse.html
(https://docs.aws.amazon.com/streams/latest/dev/what-is-sse.html)

B. Yes. Server side encryption can use a CMK for Kinesis that is managed by AWS
https://docs.aws.amazon.com/streams/latest/dev/creating-using-sse-master-keys.html
(https://docs.aws.amazon.com/streams/latest/dev/creating-using-sse-master-keys.html)

C. No. Client side encryption is not enabled for kinesis
https://docs.aws.amazon.com/streams/latest/dev/what-is-sse.html
(https://docs.aws.amazon.com/streams/latest/dev/what-is-sse.html)

D. No. Client side encryption is not enabled for kinesis
https://docs.aws.amazon.com/streams/latest/dev/what-is-sse.html
(https://docs.aws.amazon.com/streams/latest/dev/what-is-sse.html)

E. Yes. Server side encryption can use a user-specified CMK
https://docs.aws.amazon.com/streams/latest/dev/creating-using-sse-master-keys.html
(https://docs.aws.amazon.com/streams/latest/dev/creating-using-sse-master-keys.html)

F. Yes. Server side encryption can use a master key imported into the AWS KMS service
https://docs.aws.amazon.com/streams/latest/dev/creating-using-sse-master-keys.html
(https://docs.aws.amazon.com/streams/latest/dev/creating-using-sse-master-keys.html)

Ask our Experts

QUESTION 42 UNATTEMPTED PROCESSING

MassOrigins is a online social media and social networking service company. As of June 2018, it had an estimated 2 million users. MassOrigins for Business, the company's advertising portal, has an estimated 2,000 clients
MassOrigins hosted their entire infrastructure on AWS and uses Data Pipeline as data integration mechanism. Lot of scheduled pipelines are created to address integration with different data repositories like DynamoDB, S3, RDS and Redshift databases. What are the key items of scheduled Pipeline? Select 4 options.

- [ ] **A.** Pipeline components represent the business logic of the pipeline and are represented by the different sections of a pipeline definition ✔

- [ ] **B.** When AWS Data Pipeline runs a pipeline; it compiles the pipeline components to create a set of actionable instances. ✔
- [ ] **C.** When AWS Data Pipeline runs a pipeline; it compiles the pipeline components to create a set of task runners
- [ ] **D.** AWS Data Pipeline retries a failed operation. It continues to do so until the task reaches the maximum number of allowed retry attempts. ✔
- [ ] **E.** AWS Data Pipeline does not retry a failed operation.
- [ ] **F.** AWS Data Pipeline hands the task runners out to instances to process.
- [ ] **G.** AWS Data Pipeline hands the instances out to task runners to process. ✔

Explanation :

Answer: A, B, D, G

A. Yes. Pipeline Components — Pipeline components represent the business logic of the pipeline and are represented by the different sections of a pipeline definition. Pipeline components specify the data sources, activities, schedule, and preconditions of the workflow. They can inherit properties from parent components.
https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-how-tasks-scheduled.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-how-tasks-scheduled.html)

B. Yes. When AWS Data Pipeline runs a pipeline, it compiles the pipeline components to create a set of actionable instances. Each instance contains all the information for performing a specific task
https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-how-tasks-scheduled.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-how-tasks-scheduled.html)

C. No. When AWS Data Pipeline runs a pipeline, it compiles the pipeline components to create a set of actionable instances. Each instance contains all the information for performing a specific task
https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-how-tasks-scheduled.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-how-tasks-scheduled.html)

D. Yes. To provide robust data management, AWS Data Pipeline retries a failed operation. It continues to do so until the task reaches the maximum number of allowed retry attempts. Attempt objects track the various attempts, results, and failure reasons if applicable.
https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-how-tasks-scheduled.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-how-tasks-scheduled.html)

E. No. To provide robust data management, AWS Data Pipeline retries a failed operation. It

continues to do so until the task reaches the maximum number of allowed retry attempts. Attempt objects track the various attempts, results, and failure reasons if applicable. https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-how-tasks-scheduled.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-how-tasks-scheduled.html)

F. No. AWS Data Pipeline hands the instances out to task runners to process. https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-how-tasks-scheduled.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-how-tasks-scheduled.html)

G. Yes. AWS Data Pipeline hands the instances out to task runners to process. https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-how-tasks-scheduled.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-how-tasks-scheduled.html)

Ask our Experts

QUESTION 43 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 is planning to launch EMR cluster to complement their ETL workloads running on Data Pipeline.

The team is looking for storage configuration that supports storing temporary data that is continually changing, such as buffers, caches, scratch data, and other temporary content. Select 2 options

- [ ] **A.** HDFS Storage launched on master and core nodes with storage reclaimed when the cluster ends

- [ ] **B.** EMRFS implementation of HDFS used for reading and writing regular files from Amazon EMR directly to Amazon S3
- [ ] **C.** Master and Core nodes running on EC2 that comes with a preconfigured block of preattached disk storage called an instance store ✔
- [ ] **D.** Master and Core nodes running on local file system or local connected disks ✔

Explanation :

Answer: C,D
A. No. Provides Ephemeral storage can be enabled through HDFS
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. Provides the convenience of storing persistent data in Amazon S3 for use with Hadoop while also providing features like Amazon S3 server-side encryption, read-after-write consistency, and list consistency
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)
C. Yes. Each node is created from an EC2 instance that comes with a preconfigured block of pre-attached disk storage called an instance store. Data on instance store volumes persists only during the life of its EC2 instance
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)
D. Yes. This is same as above defined in option C. The local file system refers to a locally connected disk. When you create a Hadoop cluster, each node is created from an Amazon EC2 instance that comes with a preconfigured block of pre-attached disk storage called an instance store. Data on instance store volumes persists only during the lifecycle of its Amazon EC2 instance.
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)

Ask our Experts

QUESTION 44 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 clickstream 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 is using kinesis analytics to build SQL querying capability on streaming and planning to use windowed Queries to process the data. What kind of windows queries need to be used to that aggregates data continuously, using a fixed time or rowcount interval for e.g. after 1 minute or after 2000 rows. select 1 option.

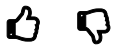
- **A.** Stagger Windows queries
- **B.** Tumbling Windows queries
- **C.** Sliding windows queries ✔
- **D.** Continuous queries

Explanation :

Answer: C
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)

Ask our Experts



QUESTION 45 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 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, Tick-Bank wants to collect and store multiple records into a stream before it is processed to different streams. Since the number of clicks are massive, Tick-Bank planning to use sharding. Please detail the specifications of implementation. Select 3 options.

- [x] **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://www.google.com/url?q=https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html&sa=D&ust=1547628312927000)

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. No. 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. Yes. 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 46 UNATTEMPTED STORAGE

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 in their DWH built on Redshift, and wants to understand the design best practices of Redshift from MindPyramid team. Please suggest the best practices in improving the queries. Select 4 options.

- ☐ **A.** Use CASE Expression to perform complex aggregations instead of selecting from the same table multiple times ✔
- ☐ **B.** Avoid using functions in query predicates ✔
- ☐ **C.** Avoid using predicates to restrict the dataset
- ☐ **D.** use a WHERE clause to restrict the dataset ✔
- ☐ **E.** Use sort keys in the GROUP BY clause to improve aggregations ✔
- ☐ **F.** do not use subqueries in cases where one table in the query is used only for predicate conditions and the subquery returns a minor number of rows

Explanation :

Answer : A, B, D, E

A. Yes. Use a CASE Expression to perform complex aggregations instead of selecting from the same table multiple times
https://docs.aws.amazon.com/redshift/latest/dg/c_designing-queries-best-practices.html (https://docs.aws.amazon.com/redshift/latest/dg/c_designing-queries-best-practices.html)

B. Yes. Using them can drive up the cost of the query by requiring large numbers of rows to resolve the intermediate steps of the query.
https://docs.aws.amazon.com/redshift/latest/dg/c_designing-queries-best-practices.html (https://docs.aws.amazon.com/redshift/latest/dg/c_designing-queries-best-practices.html)

C. No. Use predicates to restrict the dataset as much as possible.
https://docs.aws.amazon.com/redshift/latest/dg/c_designing-queries-best-practices.html (https://docs.aws.amazon.com/redshift/latest/dg/c_designing-queries-best-practices.html)

D. Yes. The query planner can then use row order to help determine which records match the criteria, so it skips scanning of large numbers of disk blocks
https://docs.aws.amazon.com/redshift/latest/dg/c_designing-queries-best-practices.html (https://docs.aws.amazon.com/redshift/latest/dg/c_designing-queries-best-practices.html)

E. Yes. A query might qualify for one-phase aggregation when its GROUP BY list contains only sort key columns, one of which is also the distribution key. The sort key columns in the GROUP BY list must include the first sort key, then other sort keys that you want to use in sort key order
https://docs.aws.amazon.com/redshift/latest/dg/c_designing-queries-best-practices.html (https://docs.aws.amazon.com/redshift/latest/dg/c_designing-queries-best-practices.html)

F. No. we should use them for performance
https://docs.aws.amazon.com/redshift/latest/dg/c_designing-queries-best-practices.html (https://docs.aws.amazon.com/redshift/latest/dg/c_designing-queries-best-practices.html)

Ask our Experts

QUESTION 47 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 will be stored in S3.
Tick-Bank wants to standardize the logs and need to apply data transformation so that all the logs captured will be in same format using lambda blueprints. The record format conversion need to be applied on the outputs of the standardized logs transformed through blueprints. How can this be done? Select 2 options.

- [x] **A.** Parses and converts Apache log lines to JSON objects ✔
- [ ] **B.** Parses and converts Apache log lines to CSV formats
- [x] **C.** Parses and converts Syslog lines to JSON objects, using predefined JSON field names. ✔
- [ ] **D.** Parses and converts Syslog lines to CSV format

Explanation :

Answer: A,C
The Apache logs and syslogs can be transformed from log format to JSON and CSV formats. But for record-format-conversion, the input data format need to be in JSON format
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/record-format-conversion.html (https://docs.aws.amazon.com/firehose/latest/dev/record-format-conversion.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 is looking for an AWS Service that can integrate with multiple database and web applications that can convert text to life like speech. The service shall be capable to understand input text or speech.
What service can provide this capability? Select 1 option.

- A. Amazon Comprehend
- B. Amazon Rekognition
- C. Amazon Polly ✔
- D. Amazon SageMaker

Explanation :

Answer : C
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. Yes. 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 49 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 is using Kinesis Producer Library to process the data into stream and KCL to collect the data. What is Kinesis Data Stream user record? Select 3 options

- [ ] **A.** KPL user record is a blob of data that has particular meaning to the user. Examples include a JSON blob representing a UI event on a website, or a log entry from a web server.
- [ ] **B.** A Kinesis Data Streams record is an instance of the Record data structure defined by the Kinesis Data Streams service API ✔
- [ ] **C.** Kinesis Data Streams record contains a partition key, sequence number, and a blob of data. ✔
- [ ] **D.** KPL user record and Kinesis Data stream record are same
- [ ] **E.** KPL user record and Kinesis Data stream record are different ✔

Explanation :

Answer: B, C, E
A. No. This is different from Kinesis Data stream record. KPL user record is a blob of data that has particular meaning to the user. Examples include a JSON blob representing a UI event on a website, or a log entry from a web server
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. Yes. A Kinesis Data Streams record is an instance of the Record data structure defined by the Kinesis Data Streams service API. Kinesis Data Streams record contains a partition key, sequence number, and a blob of data.
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. Kinesis Data Streams record is an instance of the Record data structure defined by the Kinesis Data Streams service API. Kinesis Data Streams record contains a partition key, sequence number, and a blob of data.
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. KPL user record and Kinesis Data stream record are different
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. KPL user record and Kinesis Data stream record are different
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 50 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 has built their entire infrastructure on AWS which includes web applications built on EC2, Files and logs on S3, databases on Amazon RDS, DynamoDB and DWH on Redshift. AFS built the data sources and now working on understanding the ML capabilities and establishing the guidelines for building the Amazon ML model insights. Please advice. Select 4 options.

- [ ] **A.** Different types of linear regression is implemented as Binary classification, and multiclass classification respectively and logistic regression is implemented as regression model
- [ ] **B.** Prediction accuracy metric defines the overall success of the model ✔
- [ ] **C.** Different types of logistic regression is implemented as Binary classification, and multiclass classification respectively and linear regression is implemented as regression model ✔
- [ ] **D.** Prevent overfitting the models by using cross validation techniques ✔
- [ ] **E.** Visualizations explore the accuracy of your model beyond the prediction accuracy metric ✔
- [ ] **F.** overfitting (through hyper parameters) identifies the right model which can be generalized later

Explanation :

Answer : B. C, D,E
A. No. Different types of logistic regression is implemented as Binary classification, and multiclass classification respectively and linear regression is implemented as regression model
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. Yes. A prediction accuracy metric to report on the overall success of the model
https://docs.aws.amazon.com/machine-learning/latest/dg/ml-model-insights.html (https://docs.aws.amazon.com/machine-learning/latest/dg/ml-model-insights.html)
C. Yes. Different types of logistic regression is implemented as Binary classification, and multiclass classification respectively and linear regression is implemented as regression model
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. Prevent overfitting which will make lose generalization
https://docs.aws.amazon.com/machine-learning/latest/dg/evaluating_models.html#overfitting (https://docs.aws.amazon.com/machine-learning/latest/dg/evaluating_models.html#overfitting)
E. Yes. Visualizations to help explore the accuracy of your model beyond the prediction accuracy metric
https://docs.aws.amazon.com/machine-learning/latest/dg/ml-model-insights.html (https://docs.aws.amazon.com/machine-learning/latest/dg/ml-model-insights.html)
F. No. Prevent overfitting which will make lose generalization
https://docs.aws.amazon.com/machine-learning/latest/dg/evaluating_models.html#overfitting (https://docs.aws.amazon.com/machine-learning/latest/dg/evaluating_models.html#overfitting)

Ask our Experts

QUESTION 51 UNATTEMPTED STORAGE

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 metadata management tool that allows you to access Hive

metastore tables within Pig, Spark SQL, and/or custom MapReduce applications. The component is similar to Glue Data Catalog.
Which EMR Hadoop ecosystem fulfills the requirements? Select 1 option.

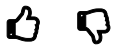
- A. Apache Hive
- B. Apache HBase
- C. Apache HCatalog ✔
- D. Apache Phoenix

Explanation :

Answer : 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. 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. Yes. 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 52 UNATTEMPTED STORAGE

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 uses load the data from different data sources using COPY command. What data sources does not support loading of data using COPY command? Select 1 option.

- A. Amazon S3
- B. Amazon EMR
- C. Amazon ES ✔
- D. Amazon DynamoDB

Explanation :

Answer : C
Yes. The COPY command loads data in parallel from Amazon S3, Amazon EMR, Amazon DynamoDB, or multiple data sources on remote hosts. COPY loads large amounts of data much more efficiently than using INSERT statements, and stores the data more effectively as well.
Integration of COPY command with Elasticsearch is not supported.
https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-copy.htmlz
(https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-copy.htmlz)

Ask our Experts

QUESTION 53 UNATTEMPTED STORAGE

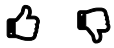
YokeFleece is a social media intelligence company headquartered in Bangalore, India. The company offers services across the social media monitoring space, combining a 'software as a service' solution Sensorly and human consultancy to help clients improve their digital marketing and PR strategies.
Sensorly allows users to monitor online conversations in real-time, providing treatment of more than 300 million sources from blogs, forums, online media, and social networks such as Facebook, Sina Weibo, Twitter, Instagram, Wechat and YouTube. Functionality includes sentiment identification, named entities and topic extraction, image analysis, logo detection, geolocation, and audience demographic insight.
YokeFleece hosted their entire infrastructure on AWS and uses Data Pipeline as data integration mechanism. The Team is designing new Data Pipeline activities. The team is interested to know what kind of databases are available and what kind of data nodes are used to as input or output to Data pipeline activity. Select 2 options.

- [ ] **A.** DynamoDBDataNode, SqlDataNode, RedshiftDataNode and S3DataNode are some of the Data nodes ✔
- [ ] **B.** EMRDataNode, SQLDbDataNode, EC2DataNode, ElasticsearchDataNode, NeptuneDataNode and CloudSearchDataNode are some of the Data nodes
- [ ] **C.** DynamoDBDatabase, RedshiftDatabase, S3Database and SqlDatabase are databases
- [ ] **D.** EMRDatabase, SqlDatabase, EC2Database, ElasticsearchDatabase, NeptuneDatabase, and CloudSearchDatabase are databases
- [ ] **E.** RdsDatabase, RedshiftDatabase, JdbcDatabase are databases ✔

Explanation :

Answer: A,E
A. Yes. DynamoDBDataNode, SqlDataNode, RedshiftDataNode and S3DataNode are some of the Data nodes
https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-concepts-datanodes.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-concepts-datanodes.html)
B. No. DynamoDBDataNode, SqlDataNode, RedshiftDataNode and S3DataNode are some of the Data nodes
https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-concepts-datanodes.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-concepts-datanodes.html)
C. No. RdsDatabase, RedshiftDatabase, JdbcDatabase are databases
https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-concepts-databases.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-concepts-databases.html)
D. No. RdsDatabase, RedshiftDatabase, JdbcDatabase are databases
https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-concepts-databases.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-concepts-databases.html)
E. Yes. RdsDatabase, RedshiftDatabase, JdbcDatabase are databases
https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-concepts-databases.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-concepts-databases.html)

Ask our Experts



QUESTION 54 UNATTEMPTED STORAGE

ConsumersHalt (CH) is an Indian department collection chain. There are 63 branches across 32 towns in India, with clothing, accessories, bags, shoes, jewelry, scents, faces, health and exquisiteness products, home furnishing and decor products.
CH runs their existing operations and analytics infrastructure out of AWS which includes S3, EC2, Auto Scaling, CDN and also Redshift. The Redshift platform is being used for advanced analytics, real time analytics and being actively used for past 2 years. Suddenly performance issues are occurring in the application and

administrator being a superuser needs to provide a list of reports in terms of current and historical performance of the cluster. What types of tables/views can help access the performance related info for diagnosis.

Select 3 options.

- [ ] **A.** STL system tables are generated from Amazon Redshift log files to provide a history of the system. They serve logging. ✔
- [ ] **B.** STL tables are actually virtual system tables that contain snapshots of the current system data. They serve snapshots.
- [ ] **C.** STV system tables are generated from Amazon Redshift log files to provide a history of the system. They serve logging.
- [ ] **D.** STV tables are actually virtual system tables that contain snapshots of the current system data. They serve snapshots. ✔
- [ ] **E.** System views contain full data found in several of the STL and STV system tables.
- [ ] **F.** The system catalogs store schema metadata, such as information about tables and columns. ✔

Explanation :

Answer : A, D, F
A. Yes. STL system tables are generated from Amazon Redshift log files to provide a history of the system.
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. No. STL system tables are generated from Amazon Redshift log files to provide a history of the system.
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)
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. 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)

E. No. System tables contain only subset of data
https://docs.aws.amazon.com/redshift/latest/dg/c_intro_system_views.html
(https://docs.aws.amazon.com/redshift/latest/dg/c_intro_system_views.html)
F. Yes. The system catalogs store schema metadata, such as information about tables and columns. System catalog tables have a PG prefix.
https://docs.aws.amazon.com/redshift/latest/dg/c_intro_catalog_views.html
(https://docs.aws.amazon.com/redshift/latest/dg/c_intro_catalog_views.html)

Ask our Experts

QUESTION 55 UNATTEMPTED ANALYSIS

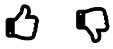
Allianz Financial Services (AFS) is a banking group offering end-to-end banking and financial solutions in South East Asia through its consumer banking, business banking, Islamic banking, investment finance and stock broking businesses as well as unit trust and asset administration, having served the financial community over the past five decades.
AFS 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. In Redshift, There is a large dimension table that needs to be upserted. How can this be achieved?
Select 2 options.

- [ ] **A.** Use UPSERT operation to perform upserts to records
- [ ] **B.** Efficiently update and insert new data by loading your data into a intermediate table first ✔
- [ ] **C.** Use INSERT_UPDATE operation to perform upserts to records
- [ ] **D.** 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. ✔

Explanation :

Answer : B, D
A. No. Amazon Redshift doesn't support a single merge statement (update or insert, also known as an upsert) to insert and update data from a single data source
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)
B. Yes. Efficiently update and insert new data by loading your data into a staging table first
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)
C. No. Amazon Redshift doesn't support a single merge statement (update or insert, also known as an upsert) to insert and update data from a single data source
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)
D. Yes. 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 56 UNATTEMPTED PROCESSING

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 considers Elastic Transcoder to perform video conversion. What kind of capabilities can be achieved using elastic transcoder? Select 4 options.

- [x] **A.** Provides a management console to create Jobs, Pipelines, and Presets, ✔
- [x] **B.** Enables a RESTful web service that uses HTTPS as the protocol and JavaScript Object Notation (JSON) as the message format. ✔
- [ ] **C.** Enables a RESTless web service that uses HTTPS as the protocol and JavaScript Object Notation (JSON) as the message format.
- [x] **D.** Allows application code make requests directly to the Elastic Transcoder ✔
- [x] **E.** must write the necessary code to sign and authenticate your requests ✔
- [ ] **F.** authentication and authorization is implicitly built into the platform

Explanation :

Answer: A,B,D,E
Elastic Transcoder provides the following features
https://docs.aws.amazon.com/elastictranscoder/latest/developerguide/accessing.html
(https://docs.aws.amazon.com/elastictranscoder/latest/developerguide/accessing.html)

Ask our Experts

QUESTION 57 UNATTEMPTED PROCESSING

IOT-Manage is a IT consulting and managed services wo offers implementation and managed services on IOT to customers. IOT-Manage is working with one of the leading manufacturing company in building a IOT management platform for their devices. All the messages captured need to be stored in DynamoDB in document format.

Please identify what different artefacts and their functions mentioned below need to be used to complete the setup. Select 4 options

- [ ] **A.** Gateway to securely and efficiently communicate with AWS IoT ✔
- [ ] **B.** Message broker to securely and efficiently communicate with AWS IoT
- [ ] **C.** Group Registry to organize the resources associated with each device in the AWS Cloud
- [ ] **D.** Rules engine to provide integration services with AWS DynamoDB ✔
- [ ] **E.** Message broker to provide a secure mechanism for devices and AWS IoT applications to publish and receive messages from each other ✔
- [ ] **F.** Shadow service to provide persistent representations of your devices in the AWS Cloud ✔

Explanation :

Answer : A, D, E, F
A. Yes. Gateway enables devices to securely and efficiently communicate with AWS IoT.
https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html
(https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html)
B. No. Message broker provides a secure mechanism for devices and AWS IoT applications to publish and receive messages from each other.
https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html
(https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html)
C. No. Groups allow you to manage several devices at once by categorizing them into groups.
https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html
(https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html)
D. Yes. Provides message processing and integration with DynamoDB
https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html
(https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html)
E. Yes, Message broker provides a secure mechanism for devices and AWS IoT applications to publish and receive messages from each other.
https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html
(https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html)

F. Yes. Shadow service provides persistent representations of your devices in the AWS Cloud
https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html
(https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html)

Ask our Experts

QUESTION 58 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 multiple java based web applications and other web framework applications running on AWS. The HH is capturing clickstream data and use custom-build recommendation engine to recommend products which eventually improve sales, understand customer preferences and already using AWS Kinesis Streams (KDS) to collect events and transaction logs and process the stream. Multiple departments from 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. To understand the network flow behavior based on every 15 minutes, HH is looking at aggregating data based on the VPC logs for analytics. VPC Flow Logs have a capture window of approximately 10 minutes.
What kind of queries can be used to capture aggregates based on each client for every 15 mins using Amazon Kinesis Data Analytics. Select 1 option.

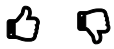
- A. Stagger Windows queries ✔
- B. Tumbling Windows queries
- C. Sliding windows queries
- D. Continuous queries

Explanation :

Answer: A
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.
VPC Flow Logs have a capture window of approximately 10 minutes. But they can have a capture window of up to 15 minutes if you're aggregating data on the client. Stagger windows are ideal for

aggregating these logs for analysis.
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)

Ask our Experts



QUESTION 59 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 is looking for a managed service that makes it easy to deploy, operate, and scale “Search Service” clusters in the AWS Cloud which supports open-source text search and analytics engine for use cases such as log

analytics, real-time application monitoring, and clickstream analysis and also integrate with web applications seamlessly.
Which service would provide you the facility to perform text search and analytics? Select 1 option.

- A. Amazon EMR clusters
- B. Amazon Elasticsearch Cluster ✔
- C. Amazon DynamoDB
- D. Amazon RedShift

Explanation :

Answer: B
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. Yes. 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 (https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/what-is-amazon-elasticsearch-service.html)
C. 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

(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/Introduction.html)
D. No. Redshift offers a Data warehouse as a service. The Amazon Redshift service manages all of the work of setting up, operating, and scaling a data warehouse. These tasks include provisioning capacity, monitoring and backing up the cluster, and applying patches and upgrades to the Amazon Redshift engine.
https://docs.aws.amazon.com/redshift/latest/mgmt/overview.html
(https://docs.aws.amazon.com/redshift/latest/mgmt/overview.html)

Ask our Experts

QUESTION 60 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 is running a campaign to improve the revenue by recommending new products to customer and upscale the business. AFS depend on ML to advise recommendations of new products that would suit a specific customer based on his previous history. 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-class classification 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 : B,E, F
A. No. 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. Yes. 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. No. 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. Yes. 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 61 UNATTEMPTED STORAGE

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 centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services on Hadoop.
Which EMR Hadoop ecosystem fulfills the requirements? Select 1 option.

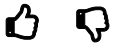
- A. Apache Hive
- B. Apache HBase
- C. Apache HCatalog
- D. Apache ZooKeeper ✔

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. Apache ZooKeeper is a centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services.
https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-zookeeper.html (https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-zookeeper.html)

Ask our Experts



QUESTION 62 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 clickstream data and use custom-build recommendation engine to recommend products which eventually improve sales, understand customer preferences and already using AWS Kinesis Streams API and Agents to collect events and transaction logs and process the stream. The event/log size is around 12 Bytes.
HH is using Kinesis firehose to apply data conversion to store the data in a

standardized format into S3 and uses Lambda Blueprints to process data conversion. What kind of Lambda blueprints are available to process transformations? Select 3 options.

- [ ] **A.** General Firehose Processing which contains the data transformation and status model used for any custom transformation logic ✔
- [ ] **B.** Parses and converts Apache log lines to JSON or CSV format ✔
- [ ] **C.** Parses and converts Syslog lines to JSON or CSV format ✔
- [ ] **D.** Parses and converts Apache log lines to Apache Parquet or Apache ORC format
- [ ] **E.** Parses and converts Syslog lines to Apache Parquet or Apache ORC format

Explanation :

Answer: A,B,C
Kinesis Data Firehose provides the following Lambda blueprints that you can use to create a Lambda function for data transformation.

- General Firehose Processing — Contains the data transformation and status model described in the previous section. Use this blueprint for any custom transformation logic.
- Apache Log to JSON — Parses and converts Apache log lines to JSON objects, using predefined JSON field names.
- Apache Log to CSV — Parses and converts Apache log lines to CSV format.
- Syslog to JSON — Parses and converts Syslog lines to JSON objects, using predefined JSON field names.
- Syslog to CSV — Parses and converts Syslog lines to CSV format.
- Kinesis Data Firehose Process Record Streams as source — Accesses the Kinesis Data Streams records in the input and returns them with a processing status.
- Kinesis Data Firehose CloudWatch Logs Processor — Parses and extracts individual log events from records sent by CloudWatch Logs subscription filters.

https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html#lambda-blueprints (https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html#lambda-blueprints)

Ask our Experts

QUESTION 63 UNATTEMPTED ANALYSIS

AdvanceU is an online learning platform founded by Cambridge professors that offers courses, specializations, and degrees. AdvanceU works with universities and other organizations to offer online courses, specializations, and degrees in a variety of subjects, such as engineering, humanities, medicine, biology, social sciences, mathematics, business, computer science, digital marketing, data science, and others. AdvanceU had more than 33 million registered users and more than 2,400 courses Every course has lot of videos, audios, image content materials that would be offered to customers. AdvanceU want to offer the text downloaded for a better understanding of the audio files and the audio track from a video recording to create closed captioning for the video that are shared to customers. What service can provide this capability? Select 1 option.

- A. Amazon Comprehend
- B. Amazon Transcribe ✔
- C. Amazon Polly
- D. Amazon SageMaker

Explanation :

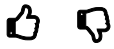
Answer : B

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 Transcribe uses advanced machine learning technologies to recognize speech in audio files and transcribe them into text. You can use Amazon Transcribe to convert audio to text and to create applications that incorporate the content of audio files. For example, you can transcribe the audio track from a video recording to create closed captioning for the video. https://docs.aws.amazon.com/transcribe/latest/dg/what-is-transcribe.html (https://docs.aws.amazon.com/transcribe/latest/dg/what-is-transcribe.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 64 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 number of friends of all users, but need to populate TOP 10 customers who has highest number of friends. 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)

Ask our Experts

QUESTION 65 UNATTEMPTED DATA SECURITY

KindleYou 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. KindleYou wants to ensure that the data in CustomerProfile table is
encrypted at rest. How is it possible? Select 5 options.

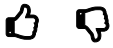
- [ ] **A.** Encrypts the data in the table using 256-bit Advanced Encryption Standard to secure the data from unauthorized access to the underlying storage ✔
- [ ] **B.** Encryption at rest integrates with AWS Key Management Service (AWS KMS) for managing the encryption key that is used to encrypt your table ✔
- [ ] **C.** AWS Owned CMK is used to encrypt the table is owned by DynamoDB ✔
- [ ] **D.** AWS Managed CMK to encrypt the table is owned by DynamoDB
- [ ] **E.** AWS Managed CMK is managed by AWS KMS ✔
- [ ] **F.** AWS Owned CMK is managed by AWS KMS
- [ ] **G.** Auditing of encryption and decryption of AWS owned CMK can be accessed using AWS CloudTrial
- [ ] **H.** Auditing of encryption and decryption of AWS managed CMK can be accessed using AWS CloudTrial ✔

Explanation :

Answer: A,B,C,E,H A. Yes. Amazon DynamoDB encryption at rest encrypts your data using 256-bit Advanced Encryption Standard (AES256), which helps secure your data from unauthorized access to the underlying storage
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html (https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html)
B. Yes. Encryption at rest integrates with AWS Key Management Service (AWS KMS) for managing the encryption key that is used to encrypt your tables.
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html (https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html)

C. Yes. AWS owned CMK – Default encryption type. The key is owned by DynamoDB
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html)
D. No. AWS managed CMK -The key is stored in your account and is managed by AWS KMS
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html)
E. Yes. The key is stored in your account and is managed by AWS KMS
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html)
F. No. The key is owned by DynamoDB
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html)
G. No. You can audit the encryption and decryption of your DynamoDB table by examining the DynamoDB API calls to AWS KMS using AWS CloudTrail.
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html)
H. Yes. You can audit the encryption and decryption of your DynamoDB table by examining the DynamoDB API calls to AWS KMS using AWS CloudTrail.
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/encryption.howitworks.html)

Ask our Experts



QUESTION 66 UNATTEMPTED COLLECTION

MSP Bank, Limited is a leading varied 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. What service does AWS propose to address migration of databases? Select 1 option.

- **A.** AWS Database Migration Service (AWS DMS) ✔
- **B.** AWS Data Sync

- **C.** AWS Migration Hub
- **D.** AWS Schema Conversion

Explanation :

Answer : A

A. Yes. 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 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. 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 67 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 database security. How can this be achieved? Select 4 options.

- [ ] **A.** Groups are collections of users that can be collectively assigned privileges for easier security maintenance ✔
- [ ] **B.** By default, privileges are granted only to the object owner ✔
- [ ] **C.** Roles are collections of users that can be collectively assigned privileges for easier security maintenance
- [ ] **D.** Amazon Redshift database users are named user accounts that can connect to a database. A user account is granted privileges explicitly, by having those privileges assigned directly to the account, or implicitly, by being a member of a group that is granted privileges ✔
- [ ] **E.** Schemas are collections of database tables and other database objects. Schemas are similar to operating system directories, except that schemas cannot be nested. ✔
- [ ] **F.** Databases are collections of tables and other schema objects. Databases are similar to operating system directories, except that databases cannot be nested.

Explanation :

Answer : A,B, D, E

Amazon Redshift database security by controlling which users have access to which database objects. Access to database objects depends on the privileges that you grant to user accounts or groups. The following guidelines summarize how database security works:

- By default, privileges are granted only to the object owner.
- Amazon Redshift database users are named user accounts that can connect to a database. A user account is granted privileges explicitly, by having those privileges assigned directly to the account, or implicitly, by being a member of a group that is granted privileges.
- Groups are collections of users that can be collectively assigned privileges for easier security maintenance.
- Schemas are collections of database tables and other database objects. Schemas are similar to operating system directories, except that schemas cannot be nested. Users can be granted access to a single schema or to multiple schemas.

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

Ask our Experts

QUESTION 68 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 is planning to understand the relationships between customers and store billions of relationships and querying the graph with milliseconds latency thereby efficiently navigate between highly connected datasets to evaluate use cases such as recommendation engines, knowledge graphs, etc. which AWS Service provides you the support. Select 1 option.

- A. DynamoDB also supports managing relationships at large scale
- B. Amazon Neptune supports managing relationships at large scale ✔
- C. Amazon EMR supports managing relationships at large scale
- D. Amazon Redshift supports managing relationships at large scale

Explanation :

Answer : B

A. 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)

B. Yes. 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 C. No.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)

D. No. The Amazon Redshift service manages all of the work of setting up, operating, and scaling a data warehouse. These tasks include provisioning capacity, monitoring and backing up the cluster, and applying patches and upgrades to the Amazon Redshift engine
https://docs.aws.amazon.com/redshift/latest/mgmt/overview.html (https://docs.aws.amazon.com/redshift/latest/mgmt/overview.html)

Ask our Experts

QUESTION 69 UNATTEMPTED DATA SECURITY

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 using Elasticsearch to address use cases related to search and analytics engine such as log analytics, real-time application monitoring, and clickstream analysis. Marqueguard enables encryption of data at rest, a security feature that helps prevent unauthorized access to your data. When encryption of data is enabled, what all aspects of domain are encrypted? Select 3 options.

- [x] **A.** Automated snapshots ✔
- [x] **B.** Elasticsearch logs ✔
- [ ] **C.** Manual snapshots
- [x] **D.** Swap files ✔
- [ ] **E.** Slow logs and error logs

Explanation :

Answer: A,B,D
Amazon ES domains offer encryption of data at rest, a security feature that helps prevent unauthorized access to your data. The feature uses AWS Key Management Service (AWS KMS) to store and manage your encryption keys. If enabled, it encrypts the following aspects of a domain:

- Indices
- Automated snapshots
- Elasticsearch logs
- Swap files
- All other data in the application directory The following are not encrypted when you enable encryption of data at rest, but you can take additional steps to protect them:
- Manual snapshots
- Slow logs and error logs https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/aes-bp.html (https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/aes-bp.html)

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 web application to create and share documents that contain live code, equations, visualizations, and narrative text. Also allows to host multiple instances of a single-user Jupyter notebook server and Amazon EMR creates a Docker container on the cluster's master node and sparkmagic as key component to run within the container Which EMR Hadoop ecosystem fulfills the requirements? Select 2 options.

- [ ] **A.** Apache Hive
- [x] **B.** EMR Notebook ✔
- [ ] **C.** Jupyter Notebook
- [ ] **D.** JupyterHub ✔

Explanation :

Answer: B, 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. 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-managed-notebooks.html (https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-jupyter-emr-managed-notebooks.html)
C. 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. 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. 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. 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/14897)

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/)

Company

- Support (https://help.whizlabs.com/hc/en-us)
- Discussions (http://ask.whizlabs.com/)
- Blog (https://www.whizlabs.com/blog/)

Mobile App

- Android Coming Soon
- iOS Coming Soon

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)