PRACTICE TEST 3

Attempt	1	**Completed on**	Sunday , 03 February 2019 , 11:40 PM
Marks Obtained	1 / 70	**Time Taken**	00 H 00 M 11 S
Your score is	1.43%	**Result**	Fail

| Domains / Topics wise Quiz Performance Report

S.No.	Topic	Total Questions	Correct	Incorrect	Unattempted
1	Processing	13	0	0	13
2	Analysis	14	1	0	13
3	Data Security	12	0	0	12
4	Collection	11	0	0	11
5	Storage	14	0	0	14
6	Visualization	6	0	0	6

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

Show Answers | All |

QUESTION 1   **UNATTEMPTED**   **PROCESSING**

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 has built a large ecosystem of analytics in AWS based on different services like DynamoDB to address document Management, Elasticsearch to address search, S3 to provide persistent storage, notifications using SNS service. The IOT data collected needs to be integrated with different services and filtered to different services based on SQL Queries against the data collected OOTB. How can we enable this? select 1 option.

- ○ **A.** Kinesis and Redshift can provide integration and SQL querying capabilities and fulfill the above requirement.

- ○ **B.** IOT Rules address filtering and integration with the above services mentioned. ✔

- ○ **C.** Lambda and Redshift can provide integration and SQL querying capabilities and fulfill the above requirement.

- ○ **D.** Lambda and RDS can provide integration and SQL querying capabilities and fulfill the above requirement.

- ○ **E.** IOT Message Broker address filtering and integration with the above services mentioned

---

Explanation :

Answer : B

A. No. Even though the solution can be implemented by the combo, IOT rules provide the solution OOTB.

https://docs.aws.amazon.com/iot/latest/developerguide/iot-rules.html
(https://docs.aws.amazon.com/iot/latest/developerguide/iot-rules.html)

B. Yes. IOT Rules provide the solution required

https://docs.aws.amazon.com/iot/latest/developerguide/iot-rules.html
(https://docs.aws.amazon.com/iot/latest/developerguide/iot-rules.html)

C. No. Even though the solution can be implemented by the combo, IOT rules provide the solution OOTB.

https://docs.aws.amazon.com/iot/latest/developerguide/iot-rules.html
(https://docs.aws.amazon.com/iot/latest/developerguide/iot-rules.html)

D. No. Even though the solution can be implemented by the combo, IOT rules provide the solution OOTB.

https://docs.aws.amazon.com/iot/latest/developerguide/iot-rules.html
(https://docs.aws.amazon.com/iot/latest/developerguide/iot-rules.html)

E. he AWS IoT message broker is a publish/subscribe broker service that enables the sending and receiving of messages to and from AWS IoT. This does not fulfil the purpose.

https://docs.aws.amazon.com/iot/latest/developerguide/iot-message-broker.html
(https://docs.aws.amazon.com/iot/latest/developerguide/iot-message-broker.html)

Ask our Experts

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

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-demandcapacity 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

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.

There are multiple requirements to fulfill being a large e-learning platform

Use pipelines to manage the conversion of videos into different formats
Transform videos into different formats using a single job (address reusability) and add new formats with simple changes
Pre-defined templates to convert videos into specific formats
Notify administrators apprised of the status of a job
How can this be achieved quickly with a faster turnaround time for each pipeline of the video already available in S3? select 1 option.

○   **A. AWS Data Pipeline service alone fulfills the requirement**

○ **B.** AWS Data Pipeline along with Kinesis Video Streams can be a prominent solution to address the requirement

○ **C.** AWS Elastic Transcoder fulfills the requirement ✔

○ **D.** AWS Elemental Media Package Fulfills the requirement

○ **E.** AWS Elemental Media Connect allows to connect to various Media and also support conversion of videos

---

Explanation :

Answer: C

A. No. AWS Data Pipeline is a web service that you can use to automate the movement and transformation of data. With AWS Data Pipeline, you can define data-driven workflows, so that tasks can be dependent on the successful completion of previous tasks. AWS Pipeline is used to process structured and unstructured data (JSON,
XML) https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/what-is-(https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/what-is-datapipeline.html)
datapipeline.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/what-is-datapipeline.html)

B. No. AWS Data Pipeline is a web service that you can use to automate the movement and transformation of data. With AWS Data Pipeline, you can define data-driven workflows, so that tasks can be dependent on the successful completion of previous tasks. AWS Pipeline is used to process structured and unstructured data (JSON,
XML) https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/what-is-(https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/what-is-datapipeline.html)
datapipeline.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/what-is-datapipeline.html)
Amazon Kinesis Video Streams is a fully managed AWS service that you can use to stream live video from devices to the AWS Cloud, or build applications for real-timevideo processing or batch-oriented video analytics. This is not used for
conversion. https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/what-is-kinesis-(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/what-is-kinesis-video.html)video.html
(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/what-is-kinesis-video.html)

C. Yes. Amazon Elastic Transcoder lets you convert media files that you have stored in Amazon Simple Storage Service (Amazon S3) into media files in the formats required by consumer playback devices. For example, you can convert large, high-quality digital media files into formats that users can play back on mobile devices, tablets, web browsers, and connected televisions. AWS Elastic Transcoder has Jobs, Pipelines, Presets and Notifications to address the above
requirements https://docs.aws.amazon.com/elastictranscoder/latest/developerguide/introduction.html
(https://docs.aws.amazon.com/elastictranscoder/latest/developerguide/introduction.html)

D. AWS Elemental MediaPackage is a just-in-time video packaging and origination service that runs in the AWS Cloud. With MediaPackage, you can deliver highly secure, scalable, and reliable video streams to a wide variety of playback devices and content delivery networks
(CDNs) https://docs.aws.amazon.com/mediapackage/latest/ug/what-is.html
(https://docs.aws.amazon.com/mediapackage/latest/ug/what-is.html)

E. No. AWS Elemental Media Connect is a high-quality transport service for live video. Today, broadcasters and content owners rely on satellite networks or fiber connections to send their high-value content into the cloud or to transmit it to partners for distribution. https://docs.aws.amazon.com/mediaconnect/latest/ug/what-is.html (https://docs.aws.amazon.com/mediaconnect/latest/ug/what-is.html)

Ask our Experts

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. Security of the data captures plays a major role in the end to end implementation. Different stakeholders have access to different datasets, live streams, play backs, archived videos, etc. How can the IT Teams control access to videos streams? select 3 options.

- [ ] **A.** Manage access to AWS Kinesis Video Streams and resources securely through IAM Users and Groups ✔
- [ ] **B.** Manage access to AWS Kinesis Video Streams and resources securely through IAM Policies
- [ ] **C.** Manage Actions, effects, resources through IAM Policies ✔
- [ ] **D.** Grant other IAM Accounts access to Video streams through polices and Roles ✔

Explanation :

Answer: A, C,D
A. Yes. Manage access to AWS Kinesis Video Streams and resources securely through IAM Users and Groups https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-iam.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-iam.html)
B. No. Manage access to AWS Kinesis Video Streams and resources securely through IAM Users and Groups https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-iam.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-iam.html)

C. Yes. Manage Actions, effects, resources through IAM
Policies https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-iam.html
(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-iam.html)
D. Yes. Grant other IAM Accounts access to Video streams through polices and
Roles https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-iam.html
(https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-iam.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.

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

HHobserved couple of issues in the implementation and want to quickly fix the coding

Some of the events and transaction logs are missing and it's adversely impacting the recommendation
Need for accountancy of failures
How can we solve the issues? select 3 options.

- [ ] **A.** Write to a minimum of 2 streams to ensure replicas of event/log data is always available even if a record could not be inserted into a stream

- [ ] **B.** Writes to the Kinesis data stream with an automatic and configurable retry mechanism ✔

- [ ] **C.** Maintain event/log information in temporary memory/storage of the application/session

- [ ] **D.** Compare the stream records using GetRecords method with temporary memory/storage of the application/session

- [ ] **E.** examine failures using the Future objects that are returned from addUserRecord method ✔

F. Increase the time-to-live on records if records could not be inserted after RecordMaxBufferedTime ✔

Explanation :

Answer : B, E, F

A. Writing to more than 1 stream is needed if the requirement drives the demand for 2 or more different streams. There is no guarantee that the records would be written in case even into the second stream.

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

B. Validate the transaction by checking the successful insert into the stream by embedding automatic and configurable retry mechanism

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)

C. Kinesis Streams provide capabilities to use Future objects to validate UserRecords. No need to complicate the code by storing in memory/transient storage

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

D. Kinesis Streams provide capabilities to use Future objects to validate UserRecords. No need to complicate the code by storing in memory/transient storage

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

E. Kinesis Streams provide capabilities to use Future objects to validate UserRecords. No need to complicate the code by storing in memory/transient storage

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

F. time-to-live records need to be increases if the UserRecords could not inserted into stream in time.

https://github.com/awslabs/amazon-kinesis-producer/blob/master/java/amazon-kinesis-producer-
(https://github.com/awslabs/amazon-kinesis-producer/blob/master/java/amazon-kinesis-producer-
sample/default_config.properties)sample/default_config.properties
(https://github.com/awslabs/amazon-kinesis-producer/blob/master/java/amazon-kinesis-producer-
sample/default_config.properties)

Ask our Experts   👍 👎

QUESTION 6   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 is using Kinesis Streams to address the data integration from social media channels and using multiple shards to process the data. Currently they are following 5 twitter handles and more than 20 advertisements on Facebook. The architecture team want to optimize the costs and start low, before they grow big. The team understands there are around 200 records captured per second from Facebook and 100 tweets from twitter with a average size of 20 bytes . Please advise your architecture building blocks. select 3 options.

- ☐ **A.** A single shard can ingest up to 1 MiB of data per second ✔
- ☐ **B.** A single shard can process 10,000 records per second for writes
- ☐ **C.** Number of shards to process Facebook and Twitter are 4 and 2 respectively for reads ✔
- ☐ **D.** Number of shards to process Facebook and Twitter are 1 and 1 respectively for reads
- ☐ **E.** The maximum size of the data payload of a record before base64-encoding is up to 2 MiB.
- ☐ **F.** GetRecords can retrieve up to 1 MiB of data per call from a single shard and up to 5,000 records per call
- ☐ **G.** Each shard can support up to five read transactions per second ✔
- ☐ **H.** Each read transaction can provide up to 5,000 records with an upper limit of 5 MiB per transaction.

Explanation :

Answer: A, C, G

A. Yes. A single shard can ingest up to 1 MiB of data per second (including partition keys) or 1,000 records per second for writes. Similarly, if you scale your stream to 5,000 shards, the stream can ingest up to 5 GiB per second or 5 million records per second. https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html)limits.html (https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html)

B. No. A single shard can ingest up to 1 MiB of data per second (including partition keys) or 1,000 records per second for writes. Similarly, if you scale your stream to 5,000 shards, the stream can ingest up to 5 GiB per second or 5 million records per

second. https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-
(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html)limits.html
(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html)

C. Yes. Number of shards to process Facebook and Twitter are 4 and 2
respectively https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-
(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html)limits.html
(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html)

D. No. Number of shards to process Facebook and Twitter are 4 and 2
respectively https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-
(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html)limits.html
(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html)

E. No. The maximum size of the data payload of a record before base64- encoding is up to 1
MiB. https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-
(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html) limits.html
(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html)

F. No. GetRecords can retrieve up to 10 MiB of data per call from a single shard, and up to 10,000
records per call. Each call to GetRecords is counted as one read
transaction. https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-
(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html)limits.html
(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html)

G. Yes. Each shard can support up to five read transactions per
second https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-
(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html)limits.html
(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html)

H. No. Each read transaction can provide up to 10,000 records with an upper limit of 10 MiB per
transaction. https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-
(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html) limits.html
(https://docs.aws.amazon.com/streams/latest/dev/service-sizes-and-limits.html)

Ask our Experts

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. Tick-Bank is interested to track the Amazon Kinesis Data Streams Application State. How can this be achieved. select 2 options.

- ☐ **A.** KCL uses a unique Amazon DynamoDB table to keep track of the application's state ✔
- ☐ **B.** CloudWatch provides information to track the application state
- ☐ **C.** KCL uses the name of the Amazon Kinesis Data Streams application to create the name of the table, each application name must be unique. ✔
- ☐ **D.** 0

Explanation :

Answer: A,C

A.Yes. KCL uses a unique Amazon DynamoDB table to keep track of the application's state https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-(https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-ddb.html)ddb.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-ddb.html)

B. No. KCL uses a unique Amazon DynamoDB table to keep track of the application's state https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-(https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-ddb.html)ddb.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-ddb.html)

C. Yes. KCL uses the name of the Amazon Kinesis Data Streams application to create the name of the table, each application name must be unique https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-(https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-ddb.html) ddb.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-record-processor-ddb.html)

Ask our Experts 👍 👎

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

They use DynamoDB to support the mobile application and S3 to host the images

and other documents shared between users.

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

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

Explanation :

Answer: B, D,F

A. No. the metrics are published to CloudWatch

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoSc
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWc
aling.HowItWorks
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWc

B. Yes. the metrics are published to CloudWatch

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoSc
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWc
aling.HowItWorks
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWc

C. No cloudwatch triggers the alarm

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWo
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWc

D. Yes. the alarm on the AWS Management Console and receive notifications using Amazon Simple Notification Service

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoSc
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWc
aling.HowItWorks
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWc

E. No. the user is notified using SNS

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWo
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWc

F. The CloudWatch alarm invokes Application Auto Scaling to evaluate your scaling policy using SNS which issues an UpdateTable request to adjust your table's provisioned throughput.
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWo
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/AutoScaling.html#AutoScaling.HowItWo

Ask our Experts   👍 👎

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

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

Security plays a major role in FlexiToner and wants to enable right policies to restrict access to Athena operations. What kind of permissions is needed to run queries in Athena? select 4 options.

- ☐ **A.** Users must have appropriate permissions for the Athena actions ✔
- ☐ **B.** Users must have appropriate permissions for The Amazon S3 locations where the underlying data is stored that you are going to query in Athena ✔
- ☐ **C.** Users must have appropriate permissions for Athena to access data from the encrypted query results
- ☐ **D.** Users must have appropriate permissions for AWS Glue to Presto and Hive which are internal components in Athena
- ☐ **E.** Users must have appropriate permissions for the resources that you store in AWS Glue Data Catalog, such as databases and tables that are going to be queried in Athena. ✔
- ☐ **F.** Users must have appropriate permissions for The encrypted metadata in the AWS Glue Data Catalog ✔

Explanation :

Answer: A, B,E,F

A. Yes. To run queries in Athena, you must have the appropriate permissions for:

- The Athena actions.

- The Amazon S3 locations where the underlying data is stored that you are going to query in Athena.

- The resources that you store in AWS Glue Data Catalog, such as databases and tables, that you are going to query in Athena.

- The encrypted metadata in the AWS Glue Data Catalog (if you migrated to using that metadata in Athena and the metadata is encrypted).https://docs.aws.amazon.com/athena/latest/ug/access.html (https://docs.aws.amazon.com/athena/latest/ug/access.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.

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

HHhas the following requirements to process the data that is being ingested -

Apply transformation of syslog data to JSON and CSV format and store it into different buckets to support different processing needs
Capture transformation failures into same S3 bucket to address audit
Backup the syslog streaming data into S3 bucket
How can this be achieved? select 3 options.

- [ ] **A.** Data transformation from syslog to JSON and CSV can be performed through Lambda blueprints ✔

- [ ] **B.** Data transformation from syslog to JSON is performed through Lambda, but transformation to CSV is performed implicitly by Kinesis Firehose

C. Data transformation from syslog to CSV is performed through Lambda, but transformation to JSON is performed implicitly by Kinesis Firehose

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

E. S3 backups can be managed to bucket policies

F. Data Transformation failures are delivered to processing-failed folder ✔

G. Data Transformation failures are delivered to transform-failed folder

---

Explanation :

Answer: A, D, F

A. Yes. Kinesis Data Firehose provides Lambda blueprints that you can use to create a Lambda function for data transformation https://docs.aws.amazon.com/firehose/latest/dev/data- (https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html#lambda-blueprints) transformation.html#lambda-blueprints (https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html#lambda-blueprints)

B. No. Kinesis Data Firehose provides Lambda blueprints that you can use to create a Lambda function for data transformation https://docs.aws.amazon.com/firehose/latest/dev/data- (https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html#lambda-blueprints) transformation.html#lambda-blueprints (https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html#lambda-blueprints)

C. No. Kinesis Data Firehose provides Lambda blueprints that you can use to create a Lambda function for data transformation https://docs.aws.amazon.com/firehose/latest/dev/data- (https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html#lambda-blueprints) transformation.html#lambda-blueprints (https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html#lambda-blueprints)

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

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

F. Yes. If data transformation fails, the unsuccessfully processed records are delivered to your S3 bucket in the processing-failed folder https://docs.aws.amazon.com/firehose/latest/dev/data- (https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html#data-transformation-failure-handling)transformation.html#data-transformation-failure-handling (https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html#data-transformation-failure-handling)

G. No. If data transformation fails, the unsuccessfully processed records are delivered to your S3 bucket in the processing-failed folder https://docs.aws.amazon.com/firehose/latest/dev/data-(https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html#data-transformation-failure-handling)transformation.html#data-transformation-failure-handling (https://docs.aws.amazon.com/firehose/latest/dev/data-transformation.html#data-transformation-failure-handling)

Ask our Experts

MSP Bank, Limited is a leading Japanese monetary institution that provides a full range of financial products and services to both institutional and individual customers. It is headquartered in Tokyo. MSP Bank is hosting their existing infrastructure on AWS. MSP bank has many organizations internally and they are planning to launch a self-data discovery platform running out of AWS on QuickSight.

Using QuickSight, dataset is being created based on their existing RDS data source. Multiple tables have been identified which needs to be joined and accessed. Team is evaluating if they can join this data from another Aurora database. Please advise the team what can be achieved.

select 4 options.

- ☐ **A.** Both data sets are based on the same SQL database data source. ✔

- ☐ **B.** Both data sets are not required to be on the same SQL database data source

- ☐ **C.** To join tables from different data sources, create the join before importing to Amazon QuickSight. ✔

- ☐ **D.** QuickSight provides facility to join tables from different data sources

- ☐ **E.** QuickSight allows configure join type ✔

- ☐ **F.** If you chose a table and made changes to the fields (for example, changing a field name or adding a calculated field), these changes are discarded when you add tables using the join interface. ✔

- ☐ **G.** If you chose a table and made changes to the fields (for example, changing a field name or adding a calculated field), these changes are accepted even when you add tables using the join interface.

Explanation :

Answer: A,C, E,F
A. Yes. Both data sets are based on the same SQL database data source.

https://docs.aws.amazon.com/quicksight/latest/user/joining-tables.html
(https://docs.aws.amazon.com/quicksight/latest/user/joining-tables.html)
B. No. Both data sets are based on the same SQL database data source.
https://docs.aws.amazon.com/quicksight/latest/user/joining-tables.html
(https://docs.aws.amazon.com/quicksight/latest/user/joining-tables.html)
C. Yes. To join tables from different data sources, create the join before importing to Amazon
QuickSight
https://docs.aws.amazon.com/quicksight/latest/user/joining-tables.html
(https://docs.aws.amazon.com/quicksight/latest/user/joining-tables.html)
D. No. QuickSight does not provide this facility
https://docs.aws.amazon.com/quicksight/latest/user/joining-tables.html
(https://docs.aws.amazon.com/quicksight/latest/user/joining-tables.html)
E. Yes. QuickSight allows configure join type, can be inner, outer, left, right
https://docs.aws.amazon.com/quicksight/latest/user/joining-tables.html
(https://docs.aws.amazon.com/quicksight/latest/user/joining-tables.html)
F. Yes. If you chose a table and made changes to the fields (for example, changing a field name or
adding a calculated field), these changes are discarded when you add tables using the join interface.
https://docs.aws.amazon.com/quicksight/latest/user/joining-tables.html
(https://docs.aws.amazon.com/quicksight/latest/user/joining-tables.html)
G. No. If you chose a table and made changes to the fields (for example, changing a field name or
adding a calculated field), these changes are discarded when you add tables using the join interface.
https://docs.aws.amazon.com/quicksight/latest/user/joining-tables.html
(https://docs.aws.amazon.com/quicksight/latest/user/joining-tables.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.

HHruns 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 KPL to collect events and
transaction logs and process the stream.

HHIT 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 make better
use of their unused capacity of the shards. How can they achieve that? select 2
options.

☐   A. Merge the hot shards to make better use of their unused capacity

B. Merge the shards that receive more data to make better use of their unused capacity

C. Merge the cold shards to merge cold shards to make better use of their unused capacity ✔

D. Merge the shards that receive less data to merge cold shards to make better use of their unused capacity ✔

Explanation :

Answer: C,D

A. No. Hot shards or shards with more data is generally split to improve the performance of the stream https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-(https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html)resharding-strategies.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html)

B. No. Hot shards or shards with more data is generally split to improve the performance of the stream https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-(https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html)resharding-strategies.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html)

C. Yes. Merge the cold shards to merge cold shards to make better use of their unused capacity https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-(https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html)resharding-strategies.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html)

D. Yes Merge the shards that receive less data to merge cold shards to make better use of their unused capacity https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-(https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html)resharding-strategies.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-strategies.html)

Ask our Experts 👍 👎

QUESTION 13   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
Most of the sales queries in involve a subset of the customer dimension. Please advise your distribution styles. select 1 option.

- ○ **A.** DISTRIBUTE SALES_FCT and CUST_DIM on SAME KEY with KEY DISTRIBUTION ✔
- ○ **B.** DISTRIBUTE SALES_FCT and CUST_DIM on SAME KEY with EVEN DISTRIBUTION
- ○ **C.** DISTRIBUTE SALES_FCT and CUST_DIM on SAME KEY with ALL DISTRIBUTION
- ○ **D.** DISTRIBUTE SALES_FCT and CUST_DIM on DIFFERENT KEYS with KEY DISTRIBUTION
- ○ **E.** DISTRIBUTE SALES_FCT and CUST_DIM on SAME KEY by not specifying DISTSTYLE

Explanation :

Answer : A

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. KEY DISTRIBUTION distributes the rows are according to the values in one column. With distribution key on different keys, this initiates lot of data copy between nodes and 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)
E. Redshift decides the distribution based on the statistics. Not a right design approach to build the solution
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 14   UNATTEMPTED   DATA SECURITY

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

They use DynamoDB to support the mobile application and S3 to host the images and other documents shared between users. KindleYou has a large customer base spread across multiple geographic areas. Customers need to update their profile information while using the application. Propose a solution that can be easily implemented and provides full consistency.

Select 1 Option

○   **A. Use global tables, a fully managed solution across multiple regions, multi-master databases ✔**

○ **B.** Create CustomerProfile table in a region, create replication copies in different AWS regions and enable replication through AWS Kinesis Data Streams

○ **C.** Create CustomerProfile table in a region, create replication copies in different AWS regions and enable replication through AWS Data Pipeline

○ **D.** Create CustomerProfile table in a region, create replication copies in different AWS regions and enable replication through AWS Kinesis Data Firehose

---

Explanation :

Answer: A

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

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

B. No. Replication with Kinesis Streams is possible, but does not provide full consistency. But again producers using KPL and Consumers need to be configured

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

C. No. Replication with Data Pipeline is based on pipelines and schedule. It cannot be real-time and does not provide full consistency

https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/what-is-datapipeline.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/what-is-datapipeline.html)

D. No. Replication with Kinesis Firehose runs with a lag. It cannot be real-time and does not provide full consistency

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

Ask our Experts

👍 👎

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

Parson built a VPC to host their entire enterprise infrastructure on cloud. Parson has large assets of data around 20 TB's of structured data and 45 TB of unstructured data and is planning to host their data warehouse on AWS and unstructured data storage on S3. The files sent from their on premise data center are also hosted into S3 buckets. Parson IT team is well aware of the scalability, performance of AWS services capabilities. Parson hosts their web applications, databases and the data warehouse built on Redshift in VPC

The structured, semi-structured and unstructured formats are stored in S3 in various buckets. This data be joined and queried along with data in Redshift using Redshift Spectrum. What kind of below features supported by Redshift Spectrum? select 2 options.

☐ **A.** Query the data in its original format directly from Amazon S3 in same region ✔

☐ **B.** Redshift Spectrum supports the following structured and semi-structured data formats like AVRO, PARQUET, TEXTFILE, SEQUENCEFILE, RCFILE, others ✔

☐ **C.** Redshift Spectrum supports only CSV and TXT formats

☐ **D.** Files in S3 cannot have compression for Redshift Spectrum to access the data

---

Explanation :

Answer : A, B

A. Yes. If the files are formatted in a format that Redshift Spectrum supports and located in an Amazon S3 bucket that your cluster can access, you can query the data in its original format directly from Amazon S3. The Amazon S3 bucket with the data files and the Amazon Redshift cluster must be in the same AWS Region. https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html (https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html)

B. Yes. Redshift Spectrum supports the following structured and semistructured data formats:

- AVRO
- PARQUET
- TEXTFILE
- SEQUENCEFILE
- RCFILE
- RegexSerDe
- Optimized row columnar (ORC)
- Grok
- OpenCSV
- Ion
- JSON

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

C. No. Redshift Spectrum supports the following structured and semistructured data formats:

- AVRO
- PARQUET
- TEXTFILE
- SEQUENCEFILE
- RCFILE

- RegexSerDe

- Optimized row columnar (ORC)

- Grok

- OpenCSV

- Ion

- JSON

https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html
(https://docs.aws.amazon.com/redshift/latest/dg/c-spectrum-data-files.html)
D. No. Redshift Spectrum supports the following compression types and extensions:

- gzip – .gz

- Snappy – .snappy

- bzip2 – .bz2

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

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 launched EMR cluster to support their big data analytics requirements. AFS is planning to build an application running on EMR which supports both OLTP and operational analytics allowing you to use standard SQL queries and JDBC APIs to work with an Apache HBase backing store.

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

○ **A.** Apache Hue

○ **B.** Apache Flink

○ **C.** Apache Phoenix ✔

○ **D.** Apache HBase

Explanation :

Answer : C

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-timestream 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. Yes. 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. HBase is an open source, non-relational, distributed database developed as part of the Apache Software Foundation's Hadoop project. HBase runs on top of Hadoop Distributed File System (HDFS) to provide non- relational database capabilities for the Hadoop ecosystem https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hbase.html (https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hbase.html)

Ask our Experts

QUESTION 17   UNATTEMPTED   ANALYSIS

MSP Bank, Limited is a leading Japanese monetary institution that provides a full range of financial products and services to both institutional and individual customers. It is headquartered in Tokyo. MSP Bank is hosting their existing infrastructure on 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 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 tool can help? select 1 option.

○ **A.** AWS Athena acts as the ETL Engine and Glue Data Catalog acts as central metadata repository

○ **B.** AWS Glue acts as the ETL Engine and Glue Data Catalog acts as central metadata repository ✔

○ **C.** AWS Kinesis acts as the Data Processing Engine and Glue Data Catalog can be easily configured as central data repository

○ **D.** AWS Glue Catalog acts as the ETL Engine and central metadata repository

---

Explanation :

Answer B

A. No. Amazon Athena is an interactive query service that makes it easy to analyze data directly in Amazon Simple Storage Service (Amazon S3) using standard SQL. Athena is serverless, so there is no infrastructure to set up or manage, and you pay only for the queries you run. Athena scales automatically—executing queries in parallel—so results are fast, even with large datasets and complex queries. https://docs.aws.amazon.com/athena/latest/ug/what-is.html (https://docs.aws.amazon.com/athena/latest/ug/what-is.html)

The AWS Glue Data Catalog is persistent metadata store. It is a managed service that lets you store, annotate, and share metadata in the AWS Cloud in the same way you would in an Apache Hive metastore, It provides a uniform repository where disparate systems can store and find metadata to keep track of data in data silos, and use that metadata to query and transform the data https://docs.aws.amazon.com/glue/latest/dg/components-(https://docs.aws.amazon.com/glue/latest/dg/components-overview.html#data-catalog-intro) overview.html#data-catalog-intro (https://docs.aws.amazon.com/glue/latest/dg/components-overview.html#data-catalog-intro)

B.Yes. AWS Glue is a fully managed ETL (extract, transform, and load) service that makes it simple and cost-effective to categorize your data, clean it, enrich it, and move it reliably between various data stores. https://docs.aws.amazon.com/glue/latest/dg/what-is-glue.html (https://docs.aws.amazon.com/glue/latest/dg/what-is-glue.html)

The AWS Glue Data Catalog is persistent metadata store. It is a managed service that lets you store, annotate, and share metadata in the AWS Cloud in the same way you would in an Apache Hive metastore, It provides a uniform repository where disparate systems can store and find metadata to keep track of data in data silos, and use that metadata to query and transform the data https://docs.aws.amazon.com/glue/latest/dg/components-(https://docs.aws.amazon.com/glue/latest/dg/components-overview.html#data-catalog-intro) overview.html#data-catalog-intro (https://docs.aws.amazon.com/glue/latest/dg/components-overview.html#data-catalog-intro)

C. No. Amazon Kinesis Data Streams (KDS) is a massively scalable and durable real-time data streaming service. KDS can continuously capture gigabytes of data per second from hundreds of thousands of sources such as website clickstreams, database event streams, financial transactions, social media feeds, IT logs, and location-tracking events. The data collected is available in milliseconds to enable real-time analytics use cases such as real- time dashboards, real-time anomaly detection, dynamic pricing, and more. https://aws.amazon.com/kinesis/data-streams/ (https://aws.amazon.com/kinesis/data-streams/)

The AWS Glue Data Catalog is persistent metadata store. It is a managed service that lets you store, annotate, and share metadata in the AWS Cloud in the same way you would in an Apache Hive metastore, It provides a uniform repository where disparate systems can store and find metadata to

keep track of data in data silos, and use that metadata to query and transform the
data https://docs.aws.amazon.com/glue/latest/dg/components-
(https://docs.aws.amazon.com/glue/latest/dg/components-overview.html#data-catalog-intro)
overview.html#data-catalog-intro (https://docs.aws.amazon.com/glue/latest/dg/components-
overview.html#data-catalog-intro)

D. No. The AWS Glue Data Catalog is persistent metadata store. It is a managed service that lets you
store, annotate, and share metadata in the AWS Cloud in the same way you would in an Apache Hive
metastore, It provides a uniform repository where disparate systems can store and find metadata to
keep track of data in data silos, and use that metadata to query and transform the
data https://docs.aws.amazon.com/glue/latest/dg/components-
(https://docs.aws.amazon.com/glue/latest/dg/components-overview.html#data-catalog-intro)
overview.html#data-catalog-intro (https://docs.aws.amazon.com/glue/latest/dg/components-
overview.html#data-catalog-intro)

Ask our Experts

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

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

Security plays a major role in FlexiToner and wants to enable end to end data
encryption for all the data that is being accessed through Athena for both data in S3
and also encrypted results. select 3 options.

- [ ] **A.** Server side encryption (SSE) with an Amazon S3-managed key (SSE-S3) for
  encrypted datasets in Amazon S3 and for encrypted query results ✔

- [ ] **B.** Server side encryption (SSE) with an Amazon S3-managed key (SSE-S3) for
  encrypted datasets in Amazon S3

- [ ] **C.** Client-side encryption (CSE) with a AWS KMS customer managed key (CSE-
  KMS) for encrypted query results

☐ **D.** Server-side encryption (SSE) with a AWS Key Management Service customer managed key (SSE-KMS) for encrypted datasets in Amazon S3 ✔

☐ **E.** Server-side encryption (SSE) with a AWS Key Management Service customer managed key (SSE-KMS) for encrypted datasets in Amazon S3 and for encrypted query results

☐ **F.** Client-side encryption (CSE) with a AWS KMS customer managed key (CSE-KMS) for encrypted datasets in Amazon S3 and for encrypted query results ✔

Explanation :

Answer: A, D,F

A. Yes. Athena supports the following Amazon S3 encryption options, both for encrypted datasets in Amazon S3 and for encrypted query results:

- Server side encryption (SSE) with an Amazon S3-managed key (SSE-S3)

- Server-side encryption (SSE) with a AWS Key Management Service customer managed key (SSE-KMS).

- Client-side encryption (CSE) with a AWS KMS customer managed key (CSE-KMS)
https://docs.aws.amazon.com/athena/latest/ug/encryption.html#encryption-options-S3-and-(https://docs.aws.amazon.com/athena/latest/ug/encryption.html#encryption-options-S3-and-Athena) Athena (https://docs.aws.amazon.com/athena/latest/ug/encryption.html#encryption-options-S3-and-Athena)

Ask our Experts   👍 👎

MindPyramid Limited is a multinational information technology and outsourcing company headquartered in Vizag, India and New Jersey, USA. Founded in 2003, the company employs approximately 2000 employees. The company offers consulting services in cloud computing, big data and analytics. They offer services to major cloud providers including AWS. The team is working with one of the major clients having their infrastructure build on AWS. Currently the client is having lot of performance issues and wants to understand the design best practices from MindPyramid team. Please suggest the best practices in terms of designing tables. select 2 options.

☐ **A.** choose the optimal sort key to address querying of most recent data, frequent range filtering or equality filtering and joining of multiple tables ✔

☐ **B.** Define constraints to ensure uniqueness, primary-key, and foreign-key constraints.

☐   **C.**   understand and define compression encoding techniques manually

☐   **D.**   Define best distribution key to distribute the fact table and one dimension table on their common columns, large dimensions with high cardinality in the filtered result set and handle uniform distribution for small dimension tables ✔

☐   **E.**   use the maximum column size for convenience of design of large tables

☐   **F.**   varchar and char store data more efficiently than the date data types. Use char and varchar more frequently

☐   **G.**   Splitting of the data that is being loaded into multiple files

---

Explanation :

Answer : A, D

A. Yes. Amazon Redshift query optimizer uses sort order when it determines optimal query plans. https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-sort-key.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-sort-key.html)

B. No. Amazon Redshift does not enforce unique, primary-key, and foreign-key constraints. https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-defining-constraints.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-defining-constraints.html)

C. No. You can specify compression encodings when you create a table, but mostly, automatic compression produces the best results https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-auto-compression.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-auto-compression.html)

D. Yes. the query optimizer redistributes the rows to the compute nodes as needed to perform any joins and aggregations through optimal distribution techniques 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)

E. No. Use the Smallest Possible Column Size https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-smallest-column-size.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-smallest-column-size.html)

F. No. Use Date/Time Data Types for Date Columns https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-timestamp-date-columns.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-timestamp-date-columns.html)

G. This is a best practice for loading data into tables, not for design. https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-multiple-files.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-multiple-files.html)

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 extracted the datasets from their SQL databases on premises, logs from web applications and loaded the datasets into S3 for further predictive analytics.

AFS is looking at machine learning capabilities to enabled Natural Language Processing capabilities to extract insights about the content of documents by recognizing the entities, key phrases, language, sentiments, and other common elements in documents.

What service can provide this capability? select 1 option.

○ **A.** Amazon Comprehend ✔

○ **B.** Amazon Rekognition

○ **C.** Amazon Polly

○ **D.** Amazon SageMaker

Explanation :

Answer : A

A. Yes. Amazon Comprehend uses natural language processing (NLP) to extract insights about the content of documents. Amazon Comprehend processes any text file in UTF-8 format. It develops insights by recognizing the entities, key phrases, language, sentiments, and other common elements in a document. Use Amazon Comprehend to create new products based on understanding the structure of documents. For example, using Amazon Comprehend you can search social networking feeds for mentions of products or scan an entire document repository for key phrases. https://docs.aws.amazon.com/comprehend/latest/dg/what-is.html (https://docs.aws.amazon.com/comprehend/latest/dg/what-is.html)

B. No. Amazon Rekognition makes it easy to add image and video analysis to your applications. You just provide an image or video to the Rekognition API, and the service can identify objects, people, text, scenes, and activities. It can detect any inappropriate content as well. Amazon Rekognition also provides highly accurate facial analysis and facial recognition. You can detect, analyze, and compare faces for a wide variety of use cases, including user verification, cataloging, people counting, and public safety. https://docs.aws.amazon.com/rekognition/latest/dg/what-is.html (https://docs.aws.amazon.com/rekognition/latest/dg/what-is.html)

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

D. 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-readyhosted 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

👍 👎

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.

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

HHIT team identified lot of performance issues with the Kinesis Stream and based on the metrics captured, identified hot and cold shards. IT team wants to effectively remove the unused capacity. There are 3 shards SHARD 1 with a hash key range of 276...381, SHARD 2 with a hash key range of 382...454 and SHARD 3 with a hash key range of 455...510.

Shard 1 and Shard 3 are cold shards while Shard 2 is a hot shard. What re-sharding strategy needs to be applied and how can it be applied? select 1 option.

○ **A.** DELETE SHARD 1, SHARD 3 use only SHARD 2

○ **B.** DELETE SHARD 1, SHARD 2, SHARD 3, redesign the SHARDING strategy

C. MERGE SHARD 1, SHARD2 and SHARD3 into 1 and split shards into 2 shards, SHARD123A with hash keys 276...420, SHARD123B as 421...510

D. MERGE 3 SHARDS into 1 SHARD with hash keys between 276..510 which remove un-usability ✔

Explanation :

Answer: D

Merge unused SHARDS into 1. SHARD Adjacency is applicable only for neighbor shards which implies if the union of the hash key ranges for the two shards forms a contiguous set with no gaps https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-(https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-merge.html) resharding-merge.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-using-sdk-java-resharding-merge.html)

Ask our Experts 👍 👎

MSP Bank, Limited is a leading Japanese monetary institution that provides a full range of financial products and services to both institutional and individual customers. It is headquartered in Tokyo. MSP Bank is hosting their existing infrastructure on AWS. MSP bank has many segments internally and they are planning to launch a self-data discovery platform running out of AWS on QuickSight.

Using QuickSight, dataset is being created based on their existing RDS data source. Multiple tables have been identified which are joined and accessed. A permission file 'permissions.txt' is created as below which outlays rules. Dataset 'transaction_details' dataset of 50 GB size has been created and joined with the permissions file.

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

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

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

Select 2 options.

- ☐ **A.** AlejandroRosalez, EMEA can see all EMEA Enterprise, SMB, and Startup data.

- ☐ **B.** AlejandroRosalez, EMEA cannot see any of the EMEA Enterprise, SMB, and Startup data. ✔

- ☐ **C.** sales-tps@emaple.com can see all rows.

- ☐ **D.** sales-tps@emaple.com cannot see any rows. ✔

Explanation :

Answer: B,D
A. No. AlejandroRosalez, EMEA cannot see any EMEA Enterprise, SMB, and Startup data
https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-(https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html)security.html (https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html)
B. Yes. AlejandroRosalez, EMEA cannot see any of the EMEA Enterprise, SMB, and Startup data
https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-(https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html)security.html (https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html)
C. No. sales-tps@emaple.com (mailto:sales-tps@emaple.com) can see all rows
https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html
D. Yes. sales-tps@emaple.com (mailto:sales-tps@emaple.com) can see all rows
https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-(https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html)security.html (https://docs.aws.amazon.com/quicksight/latest/user/restrict-access-to-a-data-set-using-row-level-security.html)

Ask our Experts   👍 👎

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

Parson built a VPC to host their entire enterprise infrastructure on cloud. Parson has large assets of data around 20 TB's of structured data and 45 TB of unstructured data and is planning to host their data warehouse on AWS and unstructured data storage on S3. The files sent from their on premise data center are also hosted into S3 buckets. Parson IT team is well aware of the scalability, performance of AWS services capabilities. Parson hosts their web applications, databases and the data warehouse built on Redshift in VPC. The administrator want to improve the query improve and planning to implement Workload Management (WLM) in Redshift. Suggest some of the best practices! select 4 options.

- [ ] **A.** Configure up to eight query queues and set the number of queries that can run in each of those queues concurrently ✔

- [ ] **B.** Set up rules to route queries to particular queues based on the user running the query or labels ✔

- [ ] **C.** Configure the amount of CPU allocated to each queue, so that large queries run in queues with more CPU than other queues

- [ ] **D.** Configure the WLM timeout property to limit short-running queries.

- [ ] **E.** Short query acceleration (SQA) prioritizes selected short-running queries ahead of longer- running queries in a dedicated space ✔

- [ ] **F.** if you enable Short query acceleration (SQA), long-running queries contend with short queries for slots in a queue

- [ ] **G.** If you enable Short query acceleration (SQA), you can reduce or eliminate workload management (WLM) queues that are dedicated to running short queries ✔

Explanation :

Answer : A, B, E, G

A. Yes. Configure up to eight query queues and set the number of queries that can run in each of those queues concurrently. You can set up rules to route queries to particular queues based on the user running the query or labels that you specify. You can also configure the amount of memory allocated to each queue, so that large queries run in queues with more memory than other queues. You can also configure the WLM timeout property to limit long-

running queries. https://docs.aws.amazon.com/redshift/latest/dg/cm-c-implementing-
(https://docs.aws.amazon.com/redshift/latest/dg/cm-c-implementing-workload-management.html)
workload-management.html (https://docs.aws.amazon.com/redshift/latest/dg/cm-c-implementing-
workload-management.html)

B. Yes. Configure up to eight query queues and set the number of queries that can run in each of those queues concurrently. You can set up rules to route queries to particular queues based on the user running the query or labels that you specify. You can also configure the amount of memory allocated to each queue, so that large queries run in queues with more memory than other queues. You can also configure the WLM timeout property to limit long-

running queries. https://docs.aws.amazon.com/redshift/latest/dg/cm-c-implementing-

(https://docs.aws.amazon.com/redshift/latest/dg/cm-c-implementing-workload-management.html)
workload-management.html (https://docs.aws.amazon.com/redshift/latest/dg/cm-c-implementing-workload-management.html)

C. No. Configure up to eight query queues and set the number of queries that can run in each of those queues concurrently. You can set up rules to route queries to particular queues based on the user running the query or labels that you specify. You can also configure the amount of memory allocated to each queue, so that large queries run in queues with more memory than other queues. You can also configure the WLM timeout property to limit long- running queries. https://docs.aws.amazon.com/redshift/latest/dg/cm-c-implementing-
(https://docs.aws.amazon.com/redshift/latest/dg/cm-c-implementing-workload-management.html)
workload-management.html (https://docs.aws.amazon.com/redshift/latest/dg/cm-c-implementing-workload-management.html)

D. No. Configure up to eight query queues and set the number of queries that can run in each of those queues concurrently. You can set up rules to route queries to particular queues based on the user running the query or labels that you specify. You can also configure the amount of memory allocated to each queue, so that large queries run in queues with more memory than other queues. You can also configure the WLM timeout property to limit long-
running queries. https://docs.aws.amazon.com/redshift/latest/dg/cm-c-implementing-
(https://docs.aws.amazon.com/redshift/latest/dg/cm-c-implementing-workload-management.html)
workload-management.html (https://docs.aws.amazon.com/redshift/latest/dg/cm-c-implementing-workload-management.html)

E. Yes. Short query acceleration (SQA) prioritizes selected short-running queries ahead of longer-running queries. SQA executes short-running queries in a dedicated space, so that SQA queries aren't forced to wait in queues behind longer queries. SQA only prioritizes queries that are short-running and are in a user-defined queue. With SQA, short-running queries begin running more quickly and users see results sooner.

If you enable SQA, you can reduce or eliminate workload management (WLM) queues that are dedicated to running short queries. In addition, long-running queries don't need to contend with short queries for slots in a queue, so you can configure your WLM queues to use fewer query slots. When you use lower concurrency, query throughput is increased and overall system performance is improved for most workloads. https://docs.aws.amazon.com/redshift/latest/dg/wlm-short-query-
(https://docs.aws.amazon.com/redshift/latest/dg/wlm-short-query-acceleration.html)
acceleration.html (https://docs.aws.amazon.com/redshift/latest/dg/wlm-short-query-acceleration.html)

F. No. Short query acceleration (SQA) prioritizes selected short-running queries ahead of longer-running queries. SQA executes short-running queries in a dedicated space, so that SQA queries aren't forced to wait in queues behind longer queries. SQA only prioritizes queries that are short-running and are in a user-defined queue. With SQA, short-running queries begin running more quickly and users see results sooner.

If you enable SQA, you can reduce or eliminate workload management (WLM) queues that are dedicated to running short queries. In addition, long-running queries don't need to contend with short queries for slots in a queue, so you can configure your WLM queues to use fewer query slots. When you use lower concurrency, query throughput is increased and overall system performance is improved for most workloads. https://docs.aws.amazon.com/redshift/latest/dg/wlm-short-query-
(https://docs.aws.amazon.com/redshift/latest/dg/wlm-short-query-acceleration.html)
acceleration.html (https://docs.aws.amazon.com/redshift/latest/dg/wlm-short-query-acceleration.html)

G. Yes. Short query acceleration (SQA) prioritizes selected short-running queries ahead of longer-running queries. SQA executes short-running queries in a dedicated space, so that SQA queries aren't forced to wait in queues behind longer queries. SQA only prioritizes queries that are short-running and are in a user-defined queue. With SQA, short-running queries begin running more quickly and users see results sooner.

If you enable SQA, you can reduce or eliminate workload management (WLM) queues that are dedicated to running short queries. In addition, long-running queries don't need to contend with short queries for slots in a queue, so you can configure your WLM queues to use fewer query slots. When you use lower concurrency, query throughput is increased and overall system performance is improved for most workloads. https://docs.aws.amazon.com/redshift/latest/dg/wlm-short-query-(https://docs.aws.amazon.com/redshift/latest/dg/wlm-short-query-acceleration.html) acceleration.html (https://docs.aws.amazon.com/redshift/latest/dg/wlm-short-query-acceleration.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.

HHruns 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 KPL to collect events and transaction logs and process the stream. The event/log size is around 12 bytes.

HHhas the following requirements to process the data that is being ingested -

Load the streaming data into Splunk cluster
Capture transformation failures into same S3 bucket to address audit
Backup the syslog streaming data into S3 bucket
select 3 options

- [ ] **A.** Streaming data can directly be delivered into ElasticSearch Domain ✔

- [ ] **B.** Streaming data is delivered to your S3 bucket first. Kinesis Data Firehose then issues an Amazon ElasticSearch COPY command to load data from your S3 bucket to your Amazon ElasticSearch cluster

- [ ] **C.** The transformation failures and delivery failures are loaded into processing-failed and errors folders in same S3 bucket ✔

☐ **D.** The transformation failures and delivery failures are loaded into transform-failed and delivery-failed folders in same S3 bucket

☐ **E.** when ES is selected as destination, and Source record S3 backup is enabled, and Backup S3 Bucket is defined, untransformed incoming data can be delivered to a separate S3 bucket ✔

☐ **F.** S3 backups can be managed to bucket policies

Explanation :

Answer: A, C, E

A. Yes. For Splunk destinations, streaming data is delivered to Splunk cluster, and it can optionally be backed up to your S3 bucket concurrently.. https://docs.aws.amazon.com/firehose/latest/dev/what-is-this- (https://docs.aws.amazon.com/firehose/latest/dev/what-is-this-service.html#data-flow-diagrams)service.html#data-flow-diagrams (https://docs.aws.amazon.com/firehose/latest/dev/what-is-this-service.html#data-flow-diagrams)

B. No. For Splunk destinations, streaming data is delivered to your Splunk cluster, and it can optionally be backed up to your S3 bucket concurrently. https://docs.aws.amazon.com/firehose/latest/dev/what-is-this- (https://docs.aws.amazon.com/firehose/latest/dev/what-is-this-service.html#data-flow-diagrams) service.html#data-flow-diagrams (https://docs.aws.amazon.com/firehose/latest/dev/what-is-this-service.html#data-flow-diagrams)

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

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

E. Yes. when S3 is selected as destination, and Source record S3 backup is enabled, untransformed incoming data can be delivered to a separate S3 bucket https://docs.aws.amazon.com/firehose/latest/dev/create- destination.html#create-destination-s3 (https://docs.aws.amazon.com/firehose/latest/dev/create- destination.html#create-destination-s3)

F. No.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 👍 👎

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.

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

HHprefers to use Kinesis Analytics to process streams using SQL queries and then send the feeds to downstream applications. What kind of downstream applications are supported? select 3 options.

- ☐ **A.** Allow AWS Lambda to access the data feed directly ✔
- ☐ **B.** Send data feed to Amazon Kinesis Data Streams directly ✔
- ☐ **C.** Send data feed to Amazon Kinesis Firehose directly ✔
- ☐ **D.** Send data feed to Amazon Redshift directly
- ☐ **E.** Send data feed to Amazon Elasticsearch directly
- ☐ **F.** Send data feed to AWS S3 directly

Explanation :

Answer: A,B,C
Kinesis Data Analytics supports Amazon Kinesis Data Firehose (Amazon S3, Amazon Redshift, and Amazon Elasticsearch Service), AWS Lambda, and Amazon Kinesis Data Streams as destinations. 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 👍 👎

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

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

Team understands the Global tables are the right solution to implement replication. The solution is implemented in 2 additional replica copies. CloudWatch is used to monitor the global tables. Over time, the team identifies the following issues:

a)Updates from one replica are not propagating to other replica tables in a timely manner.

b)The number of item updates that are written to one replica table, but that have not yet been written to another replica in the global table

Why do you think so and what metrics and action can help resolve the issues? select 3 options.

- ☐ **A.** ReplicationLatency and PendingReplicationCount metrics provide information about the issue ✔

- ☐ **B.** PendingReplicationLatency and ReplicationCount metrics provide information about the issue

- ☐ **C.** To resolve ReplicationLatency, verify that the read capacity units (RCUs) and write capacity units (WCUs) are identical for each of the replica tables ✔

- ☐ **D.** To resolve ReplicationCount, verify that the provisioned write capacity settings are sufficient for your current workload.

- ☐ **E.** To resolve PendingReplicationLatency, verify that the read capacity units (RCUs) and write capacity units (WCUs) are identical for each of the replica tables

- ☐ **F.** To resolve PendingReplicationCount, verify that the provisioned write capacity settings are sufficient for your current workload. ✔

Explanation :

Answer:A, C,F

A. Yes. ReplicationLatency and PendingReplicationCount metrics provide the information about the monitoring and performance of replica of the global tables using cloudwatch

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_monitoring. html

(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_monitoring. html B. No. ReplicationLatency and PendingRepli)

B. No. ReplicationLatency and PendingRepli

(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_monitoring. html B. No. ReplicationLatency and PendingRepli)cationCount metrics provide the information about the monitoring and performance of replica of the global tables

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_monitoring. html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_monitoring.html)
C. Yes. Verify whether the read capacity units (RCUs) and write capacity units (WCUs) are identical for each of the replica tables. In addition, the WCU settings you choose should follow the recommendations in Best Practices and Requirements for Managing Capacity.
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_monitoring. html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_monitoring.html)
D. No. there is no metric called ReplicationCount
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_monitoring. html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_monitoring.html)
E. No. there is no metric called PendingReplicationLatency
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_monitoring. html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_monitoring.html)
F. Yes. , verify that the provisioned write capacity settings are sufficient for your current workload.
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_monitoring. html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/globaltables_monitoring.html)

Ask our Experts

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 Investments 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 Investments invests through a diversity of financial instruments including equity, and debt

TI is planning to launch a EMR and evaluating different file system options how the storage can be configured. Please advise. select 3 options.

- [ ] **A.** Enable Ephemeral storage using HDFS by distributing the data it stores across instances in the cluster, storing multiple copies of data on different instances to ensure that no data is lost if an individual instance fails ✔

- [ ] **B.** EMRFS to directly access the data stored in S3 ✔

☐ **C.** Local file system storage when each node is created from an Amazon EC2 instance that comes with a preconfigured block of pre-attached disk storage called an instance store ✔

☐ **D.** Enable Ephemeral storage when each node is created from an Amazon EC2 instance that comes with a preconfigured block of pre-attached disk storage

☐ **E.** Enable local file system storage using HDFS by distributing the data it stores across instances in the cluster, storing multiple copies of data on different instances to ensure that no data is lost if an individual instance fails

---

Explanation :

Answer : A, B, C

A. Yes. Ephemeral storage can be enabled through HDFS

B. Yes. EMRFS extends Hadoop to add the ability to directly access data stored in Amazon S3 as if it were a file system like HDFS

C. Yes. local file system storage is created when each node is created from an Amazon EC2 instance that comes with a preconfigured block of pre-attached disk storage called an instance store

D. No. Ephemeral storage can be enabled only through HDFS

E. No. local file system storage is created when each node is created from an Amazon EC2 instance that comes with a preconfigured block of pre-attached disk storage called an instance store

---

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.

HHruns 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 data streams (KDS) KPL 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. HH is looking at generating costs and system usage by different departments. How can this be achieved? select 1 option.

○ **A.** Use tags to categorize and track the AWS KDS costs generated by each department ✔

○ **B.** Use unique names for streams and track the AWS costs generated by each department

○ **C.** Define different subscriptions for each account and allocate same administrator to manage all the subscriptions for each department.

○ **D.** Define different User ID's to process the stream and calculate costs based on different users. This is easy to implement.

Explanation :

Answer: A

Use tags to categorize and track your AWS costs. When you apply tags to your AWS resources, including streams, your AWS cost allocation report includes usage and costs aggregated by tags https://docs.aws.amazon.com/streams/latest/dev/tagging.html (https://docs.aws.amazon.com/streams/latest/dev/tagging.html)

Ask our Experts 👍 👎

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

The camera's write stream into Kinesis Video Stream securely and eventually consumed by applications for custom video processing, on-demand video playback and also consumed by AWS Rekognition for video analytics. The data is consumed to fulfill 2 requirements

Kinesis Video Stream is accessed for live playback and to view archived videos (minimum latency of 5-10 seconds) to inspect by the Investigation team
Kinesis Video Stream is accessed for live streaming by Operations and Monitoring team to monitor (low latency, almost near real-time)
How can this be achieved? select 2 options.

☐ **A.** Operations and Monitoring team use application using HLS (HTTP Live Streaming) to access videos to monitor live situation

B. Investigation team use application using HLS (HTTP Live Streaming) to access videos to inspect by replay videos and view identified archived videos ✔

C. Operations and Monitoring team use application using 'GetMedia' API, along with Stream Parser Library to access videos to monitor live situation ✔

D. Investigation team use application using 'GetMedia' API, along with Stream Parser Library to access videos to inspect by replay videos and view identified archived videos

---

Explanation :

Answer: B,C

A.No. HTTP Live Streaming (HLS) is an industry-standard HTTP-based media streaming communications protocol. You can use HLS to view an Amazon Kinesis Video Stream, either for live playback or to view archived video. You can use HLS for live playback. Latency is typically between 3 and 5 seconds, but it can be between 1 and 10 seconds, depending on the use case, player, and network conditions https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-hls.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-hls.html)

B.Yes. HTTP Live Streaming (HLS) is an industry-standard HTTP-based media streaming communications protocol. You can use HLS to view an Amazon Kinesis Video Stream, either for live playback or to view archived video. You can use HLS for live playback. Latency is typically between 3 and 5 seconds, but it can be between 1 and 10 seconds, depending on the use case, player, and network conditions https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-hls.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-hls.html)

C.Yes. You use the GetMedia API to build your own applications to process Kinesis Video Streams. GetMedia is a real-time API with low latency. If you want to create a player that uses GetMedia, you have to build it yourself.

For information about how to develop an application that displays a Kinesis Video Stream using GetMedia https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-hls.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-hls.html)

D. No. You use the GetMedia API to build your own applications to process Kinesis Video Streams. GetMedia is a real-time API with low latency. If you want to create a player that uses GetMedia, you have to build it yourself.

For information about how to develop an application that displays a Kinesis Video Stream using GetMedia https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-hls.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-hls.html)

Ask our Experts 👍 👎

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

The tool's coverage includes blogs, news sites, forums, videos, reviews, images and social networks such as Twitter and Facebook. Users can search data by using Text and Image Search, and use charting, categorization, sentiment analysis and other features to provide further information and analysis. Marqueguard has access to over 80 million sources.

Marqueguard hosted their entire infrastructure on AWS and uses Data Pipeline as data integration mechanism. Marqueguard runs multiple web applications to manage the platform built on DynamoDB and backups are written into S3 using server side encryption. The IT Team is planning to restore backup of a DynamoDB table available in S3 to a development environment. Define steps. select 4 options.

- ☐ **A.** Defines a data node using DynamoDB, which is specified as an input to EMRActivity which launches EMR Cluster ✔

- ☐ **B.** Defines a data node using DynamoDB, which is specified as an input to EC2Activity which launches EC2 Server

- ☐ **C.** Defines a data node using Amazon S3 aka S3DataNode ✔

- ☐ **D.** Enable server side encryption on S3DataNode, by default, s3EncryptionType set to NONE

- ☐ **E.** A precondition to check whether Amazon S3 objects with the given prefix is present. ✔

- ☐ **F.** Use EmrActivity to process the restoration ✔

- ☐ **G.** Use Ec2Activity to process the restoration

Explanation :

Answer: A,C,E,F
A. Yes. Defines a data node using DynamoDB, which is specified as an input to EMRActivity which launches EMR Cluster https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-
(https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-part1.html)importexport-ddb-part1.html
(https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-part1.html)

B. No. Defines a data node using DynamoDB, which is specified as an input to EMRActivity which launches EMR Cluster https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp- (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-part1.html)importexport-ddb-part1.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-part1.html)

C. Yes. Defines a data node using Amazon S3 aka S3DataNode. It is already server side encrypted https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp- (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-part1.html)importexport-ddb-part1.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-part1.html)

D. No. Defines a data node using Amazon S3 aka S3DataNode. It is already server side encrypted https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp- (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-part1.html)importexport-ddb-part1.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-part1.html)

E. Yes. A precondition to check whether Amazon S3 objects with the given prefix is present https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp- (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-part1.html)importexport-ddb-part1.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-part1.html)

F. Yes. Use EmrActivity to process the restoration https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp- (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-part1.html)importexport-ddb-part1.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-part1.html)

G. No. Use EmrActivity to process the restoration https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp- (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-part1.html)importexport-ddb-part1.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/dp-importexport-ddb-part1.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 launched EMR cluster to support their big data analytics requirements. AFS is looking at a non-relational database that runs on top of Hadoop Distributed File System (HDFS) to provide non-relational database capabilities for the Hadoop ecosystem This supports region servers to process the data

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

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

Explanation :

Answer : B

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. HBase is an open source, non-relational, distributed database developed as part of the Apache Software Foundation's Hadoop project. HBase runs on top of Hadoop Distributed File System (HDFS) to provide non- relational database capabilities for the Hadoop ecosystem. HBase works seamlessly with Hadoop, sharing its file system and serving as a direct input and output to the MapReduce framework and execution engine. HBase also integrates with Apache Hive, enabling SQL-like queries over HBase tables, joins with Hive-based tables, and support for Java Database Connectivity (JDBC). https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hbase.html (https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hbase.html)

C. No. HCatalog is a tool that allows you to access Hive metastore tables within Pig, Spark SQL, and/or custom MapReduce applications. HCatalog has a REST interface and command line client that allows you to create tables or do other operations. https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hcatalog.html (https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-hcatalog.html)

D. No. Apache Phoenix is used for OLTP and operational analytics, allowing you to use standard SQL queries and JDBC APIs to work with an Apache HBase backing store. https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-phoenix.html (https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-phoenix.html)

Ask our Experts

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. The queries run by users do not fully use the per-partition's throughput. In less than 3 minutes, a heavy load of queries flow in and this happens occasionally. Sometimes there are many background tasks that are running in background. How does DynamoDB handle the workload? select 1 option.

○   **A.** using Burst Capacity effectively ✔

○   **B.** using Adaptive Capacity

○   **C.** 0

○   **D.** 0

Explanation :

Answer: A

A.Yes. 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-(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-design.html#bp-partition-key-partitions-adaptive)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. No. 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-(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-design.html#bp-partition-key-partitions-adaptive)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   👍 👎

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

Since the amount of data, that is being processing is very large, Tick-Bank prefers data compression, data transformation when processing and considers Kinesis firehose to process the streams. Tick-Bank is considering record-format conversion. The logs are captured into S3 and further integrated with AWS Glue to provide analytics.

What formats are supported when record format conversion in enabled in Firehose. select 2 options.

- ☐ A. Record format conversion from JSON to Avro format
- ☐ B. Record format conversion from JSON to Parquet format ✔
- ☐ C. Record format conversion from JSON to Apache ORC format ✔
- ☐ D. Record format conversion from Apache Log to Avro format
- ☐ E. Record format conversion from Apache Log to Parquet format
- ☐ F. Record format conversion from Apache Log to Apache ORC format

---

Explanation :

Answer:B,C
Amazon Kinesis Data Firehose can convert the format of your input data from JSON to Apache Parquet or Apache ORC before storing the data in Amazon S3. Parquet and ORC are columnar data formats that save space and enable faster queries compared to row-oriented formats like JSON. https://docs.aws.amazon.com/firehose/latest/dev/record-format-(https://docs.aws.amazon.com/firehose/latest/dev/record-format-conversion.html) conversion.html (https://docs.aws.amazon.com/firehose/latest/dev/record-format-conversion.html)

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-serveapplication" 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 wants provide Image and text analysis capabilities to the applications which includes identify objects, people, text, scenes, and activities and also also provides highly accurate facial analysis and facial recognition.

What service can provide this capability? select 1 option.

- ○ **A.** Amazon Comprehend
- ○ **B.** Amazon Rekognition ✔
- ○ **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 Rekognition makes it easy to add image and video analysis to your applications. You just provide an image or video to the Rekognition API, and the service can identify objects, people, text, scenes, and activities. It can detect any inappropriate content as well. Amazon Rekognition also provides highly accurate facial analysis and facial recognition. You can detect, analyze, and compare

faces for a wide variety of use cases, including user verification, cataloging, people counting, and public safety. https://docs.aws.amazon.com/rekognition/latest/dg/what-is.html (https://docs.aws.amazon.com/rekognition/latest/dg/what-is.html)

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

D. 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-readyhosted 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 35   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.

HHruns 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 data streams (KDS) to collect events and transaction logs and process the stream. Multiple departments from HH use different streams to address real-time integration and induce analytics into their applications and uses

Kinesis as the backbone of real-time data integration across the enterprise.

HHunderstand that Monitoring plays a major role in monitoring and managing the streaming platform. HH using kinesis agent to process events and uses consumer library KCL o collect and disseminate streams. What kind of monitoring is enabled by KCL Library? select 3 options.

☐ **A.** Metrics configured for streams are automatically collected and pushed to CloudWatch every minute

☐ **B.** Per-KCL-Application Metrics, The metrics that are aggregated across all KCL workers within the scope of the application, as defined by the Amazon CloudWatch namespace. ✔

☐ **C.** Per-Worker Metrics, These metrics are aggregated across all record processors consuming data from a Kinesis data stream, such as an Amazon EC2 instance. ✔

☐ **D.** The KCL publishes Bytes sent, number of records attempted, number of records that returned failure and the number of calls to PutRecords that resulted in a service error

☐ **E.** Per-Shard Metrics, The metrics are aggregated across a single record processor. ✔

Explanation :

Answer: B,C, E

A. No. These metrics are published by Kinesis Streams to cloud
watch. https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-
(https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-cloudwatch.html)
cloudwatch.html (https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-
cloudwatch.html)

B. Yes. This is one of the key metrics provided by KCL library.. Per-KCL-Application Metrics, The
metrics that are aggregated across all KCL workers within the scope of the application, as defined by
the Amazon CloudWatch namespace https://docs.aws.amazon.com/streams/latest/dev/monitoring-
with-kcl.html (https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-kcl.html)

C. Yes. This is one of the key metrics provided by KCL library. Per-Worker Metrics, These metrics are
aggregated across all record processors consuming data from a Kinesis data stream, such as an
Amazon EC2 instance https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-kcl.html
(https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-kcl.html)

D. No. These metrics are published by Kinesis
Agent. https://docs.aws.amazon.com/streams/latest/dev/agent-health.html
(https://docs.aws.amazon.com/streams/latest/dev/agent-health.html)

E. Yes. This is one of the key metrics provided by KCL library. Per-Shard Metrics, The metrics are
aggregated across a single record
processor. https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-kcl.html
(https://docs.aws.amazon.com/streams/latest/dev/monitoring-with-kcl.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.

HHruns 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.

HHis 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 aggregate data using keyed time-based windows that open as data arrives. 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-basedwindows 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-(https://docs.aws.amazon.com/kinesisanalytics/latest/dev/stagger-window-
concepts.html) 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-(https://docs.aws.amazon.com/kinesisanalytics/latest/dev/tumbling-window-
concepts.html) 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-(https://docs.aws.amazon.com/kinesisanalytics/latest/dev/sliding-window-concepts.html)
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- (https://docs.aws.amazon.com/kinesisanalytics/latest/dev/continuous-queries-concepts.html) concepts.html (https://docs.aws.amazon.com/kinesisanalytics/latest/dev/continuous-queries-concepts.html)

Ask our Experts 👍 👎

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

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

There are around more than 85 core components of DynamoDB are created. Identify the DynamoDB components. select 4 options.

- [ ] **A.** Table is similar to a collection in other databases (notation) and it stores data ✔
- [ ] **B.** Collection is similar to a Table in other databases (notation) and it stores data
- [ ] **C.** Items are similar to Rows and Attributes are similar to columns (notation) in other databases ✔
- [ ] **D.** Rows are similar to Items and Columns are similar to Attributes (notation) in other databases
- [ ] **E.** Indexes are made up of partition keys and sort keys ✔
- [ ] **F.** Indexes are made up of primary keys and sort keys
- [ ] **G.** Streams captures data modification events in DynamoDB tables ✔
- [ ] **H.** Streams uniquely identifies each item in the table that can be replicated only as a data stream to other table

Explanation :

Answer: A,C,E,G
A. Yes. DynamoDB stores data in tables

https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html)
B. No. tables, items, and attributes are the core components that you work with
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html)
C Yes. An item is a group of attributes that is uniquely identifiable among all of the other items
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html)
D No. An item is a group of attributes that is uniquely identifiable among all of the other items
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html)
E Yes. Indexes are made up of partition keys and sort keys
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html)
F No. Indexes are made up of partition keys and sort keys
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html)
G. Yes. Streams captures data modification events in DynamoDB tables
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html)
H. No. Streams captures data modification events in DynamoDB tables
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/HowItWorks.CoreComponents.html)

Ask our Experts

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 infrastructure out of AWS and inviting AWS engineers to understand the end to end architecture of the solution. The meeting was scheduled with the AWS meeting and AWS teams shared lot of artefacts for users to understand. Internally the team is preparing presentation to the security team since they need to present Redshift end to end security features

Please advise whether the options listed above helps administrators to present their solution to security team.

Select 3 options.

☐ **A.** Users inbound access to an Amazon Redshift cluster can be managed through cluster encryption.

☐ **B.** Users inbound access to an Amazon Redshift cluster can be managed through security groups. ✔

☐ **C.** Access Management to Amazon Redshift cluster can be managed through IAM. ✔

☐ **D.** Access Management to Amazon Redshift cluster can be managed through Redshift security module.

☐ **E.** Encryption can be enabled for data in transit, data loaded, cluster and SSL connections. ✔

Explanation :

Answer : B, C and E

A. No. cluster encryption encrypts the data in all user-created tables, and can be enabled when cluster is launched

https://docs.aws.amazon.com/redshift/latest/dg/c_security-overview.html
(https://docs.aws.amazon.com/redshift/latest/dg/c_security-overview.html)

B. Yes. users inbound access to an Amazon Redshift cluster, you define a cluster security group and associate it with a cluster

https://docs.aws.amazon.com/redshift/latest/dg/c_security-overview.html
(https://docs.aws.amazon.com/redshift/latest/dg/c_security-overview.html)

C.Yes. To control access to specific Amazon Redshift resources, define AWS Identity and Access Management (IAM) accounts

https://docs.aws.amazon.com/redshift/latest/dg/c_security-overview.html
(https://docs.aws.amazon.com/redshift/latest/dg/c_security-overview.html)

D. No. To control access to specific Amazon Redshift resources, define AWS Identity and Access Management (IAM) accounts

https://docs.aws.amazon.com/redshift/latest/dg/c_security-overview.html
(https://docs.aws.amazon.com/redshift/latest/dg/c_security-overview.html)

E. Yes

https://docs.aws.amazon.com/redshift/latest/dg/c_security-overview.html
(https://docs.aws.amazon.com/redshift/latest/dg/c_security-overview.html)

Ask our Experts 👍 👎

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-serveapplication" 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 uses multiple web services like S3, Redshift, others to address various business requirements and Elasticsearch to perform search. What services provide integration of streaming data with Elasticsearch? select 2 options.

☐ **A.** Amazon S3 ✔

☐ **B.** Amazon Firehose ✔

☐ **C.** Amazon CloudTrial

☐ **D.** Amazon Glacier

Explanation :

Answer: A, B
A. Yes. load streaming data into your Amazon Elasticsearch Service domain from many different sources. Some sources, like Amazon Kinesis Data Firehose and Amazon CloudWatch Logs, have built-in support for Amazon ES. Others, like Amazon S3, Amazon Kinesis Data Streams, and Amazon DynamoDB, use AWS Lambda functions as event handlers https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-(https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-aws-integrations.html)aws-integrations.html (https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-aws-integrations.html)

Ask our Experts   👍 👎

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 KPL user record? select 2 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: A, E

A. Yes. 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. No. This is different from KPL user record. 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. No. This is different from KPL user record. 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)

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 👍 👎

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 are the best practices to load using COPY command to copy the data from S3? select 3 options.

- ☐ **A.** Split your data into multiple files and upload to S3 ✔
- ☐ **B.** Run a single COPY command to load the table. ✔
- ☐ **C.** multiple concurrent COPY commands to load one table from multiple files resulting in parallel load
- ☐ **D.** use VACUUM process at the end of the load, if the table has a sort column defined ✔

Explanation :

Answer : A,B,D
A. The COPY command leverages the Amazon Redshift massively parallel processing (MPP) architecture to read and load data in parallel from files in an Amazon S3 bucket. You can take maximum advantage of parallel processing by splitting your data into multiple files and by setting distribution keys on your tables https://docs.aws.amazon.com/redshift/latest/dg/t_Loading-data-from-S3.html (https://docs.aws.amazon.com/redshift/latest/dg/t_Loading-data-from-S3.html)
B. Yes. Amazon Redshift automatically loads in parallel from multiple data files.
If you use multiple concurrent COPY commands to load one table from multiple files, Amazon Redshift is forced to perform a serialized load. This type of load is much slower and requires a VACUUM process at the end if the table has a sort column defined https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-single-copy- (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-single-copy-command.html) command.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-single-copy-command.html)

C. No. Amazon Redshift automatically loads in parallel from multiple data files.
If you use multiple concurrent COPY commands to load one table from multiple files, Amazon
Redshift is forced to perform a serialized load. This type of load is much slower and requires a
VACUUM process at the end if the table has a sort column
defined https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-single-copy-
(https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-single-copy-command.html)
command.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-single-copy-
command.html)
D. Yes. Amazon Redshift automatically loads in parallel from multiple data files.
If you use multiple concurrent COPY commands to load one table from multiple files, Amazon
Redshift is forced to perform a serialized load. This type of load is much slower and requires a
VACUUM process at the end if the table has a sort column
defined https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-single-copy-
(https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-single-copy-command.html)
command.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-single-copy-
command.html)

Ask our Experts   👍 👎

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.

There is a DynamoDB table that uses a composite primary key with UserID as the
partition key and MessageID as the sort key. The data comes from different users as
different files which is collected, processed will be uploaded to DynamoDB. Each file
is based on 1 partition key.

The administrator observes that the load performance is too slow. How can we
improve

performance of Distributing Write Activity Efficiently During Data Upload? select 1
option.

○   **A.** Segregate each user's data into different files and upload. This will improve
        partition load performance

○   **B.** Collate each user's data into same/multiple files and upload, but the order of load
        shall be based on partition key. This will improve partition load performance

C. distribute your upload work by using the partition key to load one item from each partition key value, then another item from each partition key value, and so on:

D. distribute your upload work by using the sort key to load one item from each partition key value, then another item from each partition key value, and so on: ✔

Explanation :

Answer: D
Yes. Distribute your upload work by using the sort key to load one item from each partition key value, then another item from each partition key value, and so on. This improves the Distributing Write Activity Efficiently During Data Upload
https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-data-
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-data-
upload.html)upload.html
(https://docs.aws.amazon.com/amazondynamodb/latest/developerguide/bp-partition-key-data-
upload.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 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 distributed processing framework and programming model that helps you do machine learning, stream processing, or graph analytics using Amazon EMR clusters

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

A. Apache Hive

B. Apache HBase

C. pache HCatalog

D. Apache Spark ✔

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 Spark is a distributed processing framework and programming model that helps you do machine learning, stream processing, or graph analytics using Amazon EMR clusters. Similar to Apache Hadoop, Spark is an open-source,distributed processing system commonly used for big data workloads. However, Spark has several notable differences from Hadoop MapReduce. Spark has an optimized directed acyclic graph (DAG) execution engine and actively caches data in-memory, which can boost performance, especially for certain algorithms and interactive queries. https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-spark.html (https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-spark.html)

Ask our Experts

QUESTION 44   UNATTEMPTED   DATA SECURITY

Shell Labs is an American manufacturer of smart home products including thermostats, smoke detectors, and security systems. They use IOT integration to manage the product lifecycle. They build their IOT platform on AWS. Currently the team is working on setting up authorization for the devices which includes setting up AWS IOT policies and IAM policies. Please identify the list of IOT Policy actions that can be enabled to manage operations. select 4 options.

☐ **A.** permission to connect to the AWS IoT message broker ✔

☐ **B.** permission to publish on an MQTT topic ✔

☐ **C.** permission to receive a message from AWS IoT ✔

☐ **D.** permission to subscribe to a topic filter ✔

☐ **E.** permission to create thing

☐ **F.** permission to cancel job execution

Explanation :

Answer : A, B, C, D

A. Yes. This is an IOT policy https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html (https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html)

B. Yes. This is an IOT policy https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html (https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html)

C. Yes. This is an IOT policy https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html (https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html)

D. Yes. This is an IOT policy https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html (https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html)

E. No. This is an IAM policy https://docs.aws.amazon.com/iot/latest/developerguide/iam-policies.html (https://docs.aws.amazon.com/iot/latest/developerguide/iam-policies.html)

F. No. This is an IAM policy https://docs.aws.amazon.com/iot/latest/developerguide/iam-policies.html (https://docs.aws.amazon.com/iot/latest/developerguide/iam-policies.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 is looking to collect and process the log files in near real time that are generated from thousands of applications in their AWS cloud. AFS is also collecting stock pricing information from stock price publishing data providers and using the information to recommend stocks to customers. AFS is looking at querying streams and using Kinesis Analytics application to process all the stocks for recommendation if price changes greater than 10 percent. What kind of Queries will help fulfill the requirement? select 1 option.

○　**A.**　Stagger Windows queries

○　**B.**　Tumbling Windows queries

○　**C.**　Sliding windows queries

○　**D.**　Continuous queries　✔

Explanation :

Answer: D

A. Stagger windows query, A query that aggregates data using keyed time-basedwindows 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-
(https://docs.aws.amazon.com/kinesisanalytics/latest/dev/stagger-window-
concepts.html) 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-
(https://docs.aws.amazon.com/kinesisanalytics/latest/dev/tumbling-window-
concepts.html) 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-
(https://docs.aws.amazon.com/kinesisanalytics/latest/dev/sliding-window-concepts.html)
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- (https://docs.aws.amazon.com/kinesisanalytics/latest/dev/continuous-queries-
concepts.html) concepts.html
(https://docs.aws.amazon.com/kinesisanalytics/latest/dev/continuous-queries-concepts.html)

Ask our Experts　　　　　　　　　　　　　　　　　　　　　👍　👎

QUESTION 46　　　　**UNATTEMPTED**　　　　　　　　　　**VISUALIZATION**

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

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

The whole of the enterprise infrastructure of FlexiToner is built on AWS and the team want to understand with what services do the Athena provide integration OOTB with? select 3 options.

- ☐ **A.** Using Athena with CloudTrail logs is a powerful way to enhance your analysis of AWS service activity. ✔
- ☐ **B.** Use Athena to query Amazon QuickSight
- ☐ **C.** Use Athena to query Amazon CloudFront ✔
- ☐ **D.** Query the VPC logs in Athena to investigate network traffic patterns and identify threats and risks across your Amazon VPC network. ✔

> **Explanation :**
>
> Answer: A, C,D
> A. Yes. You can query data from other AWS services in Athena. Athena leverages several AWS services like CloudFront, CloudTrial, VPC, IAM and
> others https://docs.aws.amazon.com/athena/latest/ug/athena-aws-service-
> (https://docs.aws.amazon.com/athena/latest/ug/athena-aws-service-integrations.html)
> integrations.html (https://docs.aws.amazon.com/athena/latest/ug/athena-aws-service-
> integrations.html)

Ask our Experts 👍 👎

MSP Bank, Limited is a leading Japanese monetary institution that provides a full range of financial products and services to both institutional and individual customers. It is headquartered in Tokyo. MSP Bank is hosting their existing infrastructure on AWS. MSP bank has many segments internally and they are planning to launch a self-data discovery platform running out of AWS on QuickSight.

Using QuickSight, multiple datasets are created and multiple analyses are generated respectively. The Team is working on visuals. They are looking for a customized table view of data. Please help. select 1 option.

○　**A.** Tabular Reports ✔

○　**B.** Heat Maps

○　**C.** Pie Chart

○　**D.** Tree Map

Explanation :

Answer: A

A. Yes. 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 pie charts to compare values for items in a dimension.

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

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

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

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

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

Ask our Experts

👍 👎

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 is defining data sources. Please help identify the tasks.

Select 3 options.

☐ **A.** Data of files in S3, tables, views and collections in databases are the data sources.

☐ **B.** Amazon ML datasources can be created only for RDS, Redshift and S3. ✔

☐ **C.** Amazon ML provides only 2 options to split the datasets, sequential and random split.

☐ **D.** Metadata of files in S3, tables, views and collections in databases are the data sources. ✔

☐ **E.** Amazon ML datasources can be created on any of the above data sources.

☐ **F.** AttributeType includes Binary, Categorical, Numeric and Text datatypes. ✔

☐ **G.** AttributeType includes Ranking, Categorization, Word Prominence, Count number and Count percentage.

Explanation :

Answer : B, D, F

A. No. This is called schema which is composed of all attributes in the input data and their corresponding data types

https://docs.aws.amazon.com/machine-learning/latest/dg/creating-a-data-schema-for-amazon-ml.html (https://docs.aws.amazon.com/machine-learning/latest/dg/creating-a-data-schema-for-amazon-ml.html)

B. Yes. Data Sources can be created only for RDS, Redshift and S3.

https://docs.aws.amazon.com/machine-learning/latest/dg/creating-and-using-datasources.html (https://docs.aws.amazon.com/machine-learning/latest/dg/creating-and-using-datasources.html)

C. No. Amazon ML Amazon ML provides three options for splitting your data:

• Pre-split the data - Split the data into two data input locations, before uploading them to Amazon Simple Storage Service (Amazon S3) and creating two separate datasources with them.

• Amazon ML sequential split – configure Amazon ML to split your data sequentially when creating the training and evaluation datasources.

• Amazon ML random split - Configure Amazon ML to split your data using a seeded random method when creating the training and evaluation datasources.

D.Yes. Data source objects contain metadata about your input data. for a data source, Amazon ML reads your input data, computes descriptive statistics on its attributes, and stores the statistics, a schema, and other information as part of the data source object

https://docs.aws.amazon.com/machine-learning/latest/dg/creating-and-using-datasources.html (https://docs.aws.amazon.com/machine-learning/latest/dg/creating-and-using-datasources.html)

E. No. Data Source cannot be created on DynamoDB

https://docs.aws.amazon.com/machine-learning/latest/dg/creating-and-using-datasources.html (https://docs.aws.amazon.com/machine-learning/latest/dg/creating-and-using-datasources.html)

F. Yes. AttributeType includes Binary, Categorical, Numeric and Text datatypes
https://docs.aws.amazon.com/machine-learning/latest/dg/creating-a-data-schema-for-amazon-
(https://docs.aws.amazon.com/machine-learning/latest/dg/creating-a-data-schema-for-amazon-
ml.html#assigning-data-types)ml.html#assigning-data-types
(https://docs.aws.amazon.com/machine-learning/latest/dg/creating-a-data-schema-for-amazon-
ml.html#assigning-data-types)
G. No. these define distribution of text attributes. https://docs.aws.amazon.com/machine-
learning/latest/dg/data-insights.html (https://docs.aws.amazon.com/machine-
learning/latest/dg/data-insights.html)

Ask our Experts

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

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

Tick-Bank uses KPL to process the data and KCL library to consume the records. Thousands of events are being generated every second. To optimize the costs Tick-Bank proposes to group all the user records generated during a specific period and process the data as a stream and want to optimize the costs and processing capability. Please detail the implementation guidelines. select 2 options.

- [ ] **A.** each record in a separate Kinesis Data Streams record and make one HTTP request to send it to Kinesis Data Streams

- [ ] **B.** each HTTP request carries multiple Kinesis Stream records which is sent to kinesis Data streams ✔

- [ ] **C.** Batching is implemented as the target implementation ✔

- [ ] **D.** Batching is not implemented as the target implementation

Explanation :

Answer: B,C

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

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

C. Yes. Batching refers to performing a single action on multiple items instead of repeatedly performing the action on each individual item. In this context, the "item" is a record, and the action is sending it to Kinesis Data Streams. In a non-batchingsituation, you would place each record in a separate Kinesis Data Streams record and make one HTTP request to send it to Kinesis Data Streams. With batching, each HTTP request can carry multiple records instead of just one. https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html)

D. No. Batching refers to performing a single action on multiple items instead of repeatedly performing the action on each individual item. In this context, the "item" is a record, and the action is sending it to Kinesis Data Streams. In a non-batchingsituation, you would place each record in a separate Kinesis Data Streams record and make one HTTP request to send it to Kinesis Data Streams. With batching, each HTTP request can carry multiple records instead of just one. https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html (https://docs.aws.amazon.com/streams/latest/dev/kinesis-kpl-concepts.html)

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 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 data modeling team is working on designing the tables on Redshift and want to adapt best practices for querying. Please advice. select 4 options.

- [ ] **A.** Choose the Best SORT key ✔
- [ ] **B.** Choose the Best Distribution Style ✔
- [ ] **C.** Specify compression encodings when table is created

D. Use Automatic Compression ✔

E. Define primary key and foreign key constraints between tables wherever appropriate, even though they are only informational ✔

F. Use CHAR/VARCHAR for Date Columns

Explanation :

Answer : A,B, D,E

A. Yes. Amazon Redshift stores your data on disk in sorted order according to the sort key. The Amazon Redshift query optimizer uses sort order when it determines optimal query plans.
If recent data is queried most frequently, specify the timestamp column as the leading column for the sort key. Queries are more efficient because they can skip entire blocks that fall outside the time range.
If you do frequent range filtering or equality filtering on one column, specify that column as the sort key. Amazon Redshift can skip reading entire blocks of data for that column. It can do so because it tracks the minimum and maximum column values stored on each block and can skip blocks that don't apply to the predicate range.
If you frequently join a table, specify the join column as both the sort key and the distribution key. Doing this enables the query optimizer to choose a sort merge join instead of a slower hash join. Because the data is already sorted on the join key, the query optimizer can bypass the sort phase of the sort merge join. https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-sort-(https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-sort-key.html) key.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-sort-key.html)

B. Yes. the query optimizer redistributes the rows to the compute nodes as needed to perform any joins and aggregations. The goal in selecting a table distribution style is to minimize the impact of the redistribution step by locating the data where it needs to be before the query is executed.
Distribute the fact table and one dimension table on their common columns.
Choose the largest dimension based on the size of the filtered dataset. Choose a column with high cardinality in the filtered result set. Change some dimension tables to use ALL distribution. https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-best-dist-(https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-best-dist-key.html)key.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-best-dist-key.html)

C. No. Let COPY Choose Compression Encodings. Automatic compression balances overall performance when choosing compression encodings. Range- restricted scans might perform poorly if sort key columns are compressed much more highly than other columns in the same query. As a result, automatic compression chooses a less efficient compression encoding to keep the sort key columns balanced with other columns https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-auto- (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-auto-compression.html)compression.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-use-auto-compression.html)

D. Yes. Let COPY Choose Compression Encodings. Automatic compression balances overall performance when choosing compression encodings. Range- restricted scans might perform poorly if sort key columns are compressed much more highly than other columns in the same query. As a result, automatic compression chooses a less efficient compression encoding to keep the sort key columns balanced with other columns https://docs.aws.amazon.com/redshift/latest/dg/c_best-

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

E. Yes. Define primary key and foreign key constraints between tables wherever appropriate. Even though they are informational only, the query optimizer uses those constraints to generate more efficient query plans.

Do not define primary key and foreign key constraints unless your application enforces the constraints. Amazon Redshift does not enforce unique, primary- key, and foreign-key constraints https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-defining-(https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-defining-constraints.html) constraints.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-defining-constraints.html)

F. No. Amazon Redshift stores DATE and TIMESTAMP data more efficiently than CHAR or VARCHAR, which results in better query performance. Use the DATE or TIMESTAMP data type, depending on the resolution you need, rather than a character type when storing date/time information https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-timestamp-(https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-timestamp-date-columns.html) date-columns.html (https://docs.aws.amazon.com/redshift/latest/dg/c_best-practices-timestamp-date-columns.html)

Ask our Experts   👍 👎

QUESTION 51   UNATTEMPTED   DATA SECURITY

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. Along with the stream, different modes of streaming metadata are sent along with the stream. There are 2 scenarios that need to be fulfilled.

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

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

How can this be achieved? select 2 options.

- [ ] **A.** Requirement 1 can be fulfilled by sending Nonpersistent data ✔
- [ ] **B.** Requirement 2 can be fulfilled by sending Nonpersistent dat
- [ ] **C.** Requirement 1 can be fulfilled by sending Persistent data
- [ ] **D.** Requirement 2 can be fulfilled by sending Persistent data ✔
- [ ] **E.** Both Requirement 1 and Requirement 2 can be fulfilled by sending Nonpersistent data
- [ ] **F.** Both Requirement 1 and Requirement 2 can be fulfilled by sending Persistent data

Explanation :

Answer: A, D

A. Yes. Requirement 1 can be fulfilled by sending Nonpersistent data https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-meta.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-meta.html)

B. No. Requirement 2 cannot be fulfilled by sending Nonpersistent data https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-meta.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-meta.html)

C. No. Requirement 1 cannot be fulfilled by sending Persistent data https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-meta.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-meta.html)

D. Yes. Requirement 2 can be fulfilled by sending Persistent data https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-meta.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-meta.html)

E. No. Requirement 1 can be fulfilled by sending Nonpersistent data while Requirement 2 can be fulfilled by sending Persistent data https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-meta.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-meta.html)

F. No. Requirement 1 cannot be fulfilled by sending Persistent data while Requirement 2 can be fulfilled by sending Persistent data https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-meta.html (https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/how-meta.html)

Ask our Experts

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 want to launch a campaign to identify potential customers and recommend new products based on the customer's interest that can upscale the business. 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, 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

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

MSP bank has lot of web applications, databases and data warehouse built on
Teradata, NoSQL databases, and other types of data stores. MSP bank has lot of data
assets in terms of logs, documents; excel files, CSV files, PDF documents and others.

MSP Bank is running one of their web application NODE.JS supported by MSSQL
relational database. The schema designed is based on 3NF. The Web Application has
different user workloads at different parts of the day.

MSP Bank is looking to keep the migration simple. As a part of the migration, MSP bank
needs some inputs on migration on the following

Migration of MySQL database to a managed relational database with performance of
up to 5 times the existing implementation
Database to support both read and write workloads, with READ queries do not
impact write query performance
select 2 options.

A. Migrate MySQL Server to Amazon Aurora database MYSQL version with Aurora supporting read replicas which enables segregation of read and write workloads to primary and read replicas and provides 5 times performance compared to standard MySQL ✔

B. Migrate MySQL Server to Amazon Aurora database PostgreSQL version, as PostgreSQL is more performant compared to MySQL and provides 5 times performance compared to standard MySQL

C. Migrate MySQL Server to Amazon RDS database MYSQL version, with RDS supporting read replicas which enables segregation of read and write workloads to primary and read replicas and provides 5 times performance compared to standard MySQL

D. Migrate MySQL Server to MySQL cluster built on EC2 machines, configure cluster to support read replicas which enables segregation of read and write workloads to primary and read replicas

E. Use Data Migration Service (DMS) to address migration of MySQL database on premise to database on AWS cloud ✔

F. Use Data Migration Service (DMS) and Schema Conversion Tool (SCT) to address migration of MySQL database on premise to database on AWS cloud

---

Explanation :

Answer: A, E

A. Yes. Migration of MySQL on premise to Aurora database. Amazon Aurora (Aurora) is a fully managed relational database engine that's compatible with MySQL and PostgreSQL. Aurora can deliver up to five times the throughput of MySQL.

An Amazon Aurora DB cluster consists of one or more DB instances and a cluster volume that manages the data for those DB instances. An Aurora cluster volume is a virtual database storage volume that spans multiple Availability Zones, with each Availability Zone having a copy of the DB cluster data. Two types of DB instances make up an Aurora DB cluster:

• Primary DB instance – Supports read and write operations, and performs all of the data modifications to the cluster volume. Each Aurora DB cluster has one primary DB instance.

• Aurora Replica – Connects to the same storage volume as the primary DB instance and supports only read operations. Each Aurora DB cluster can have up to 15 Aurora Replicas in addition to the primary DB instance. Maintain high availability by locating Aurora Replicas in separate Availability Zones. Aurora automatically fails over to an Aurora Replica in case the primary DB instance becomes unavailable. You can specify the failover priority for Aurora Replicas. Aurora Replicas can also offload read workloads from the primary DB instance.

Since Migration from MySQL to MySQL is a straight forward migration, migration will be simple. https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html (https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html) B. No. Migration of MySQL to PostgreSQL is not encouraged because the migration of data between different databases makes migration complex and does not guarantee 5 times performance in terms of implementation.

An Amazon Aurora DB cluster consists of one or more DB instances and a cluster volume that manages the data for those DB instances. An Aurora cluster volume is a virtual database storage volume that spans multiple Availability Zones, with each Availability Zone having a copy of the DB cluster data. Two types of DB instances make up an Aurora DB cluster:

- Primary DB instance – Supports read and write operations, and performs all of the data modifications to the cluster volume. Each Aurora DB cluster has one primary DB instance.

- Aurora Replica – Connects to the same storage volume as the primary DB instance and supports only read operations. Each Aurora DB cluster can have up to 15 Aurora Replicas in addition to the primary DB instance. Maintain high availability by locating Aurora Replicas in separate Availability Zones. Aurora automatically fails over to an Aurora Replica in case the primary DB instance becomes unavailable. You can specify the failover priority for Aurora Replicas. Aurora Replicas can also offload read workloads from the primary DB instance.

Since Migration from MySQL to MySQL is a straight forward migration, migration will be simple. https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html (https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html)
C. No. Migration of MySQL to RDS is not a right option since it does not guarantee 5 times performance improvement compared to standard implementation. Amazon Relational Database Service (Amazon RDS) is a web service that makes it easier to set up, operate, and scale a relational database in the cloud. It provides cost-efficient, resizable capacity for an industry- standard relational database and manages common database administration
tasks. https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html (https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html)
D. No. Migration of MySQL to cluster built on EC2 instances is not a managed service. This makes the setup and migration very complex. Instead the team can go ahead with Aurora which provides 5 times performance compared to standard implementation, a managed service and also segregates read and write workloads through replica servers
An Amazon Aurora DB cluster consists of one or more DB instances and a cluster volume that manages the data for those DB instances. An Aurora cluster volume is a virtual database storage volume that spans multiple Availability Zones, with each Availability Zone having a copy of the DB cluster data. Two types of DB instances make up an Aurora DB cluster:

- Primary DB instance – Supports read and write operations, and performs all of the data modifications to the cluster volume. Each Aurora DB cluster has one primary DB instance.

- Aurora Replica – Connects to the same storage volume as the primary DB instance and supports only read operations. Each Aurora DB cluster can have up to 15 Aurora Replicas in addition to the primary DB instance. Maintain high availability by locating Aurora Replicas in separate Availability Zones. Aurora automatically fails over to an Aurora Replica in case the primary DB instance becomes unavailable. You can specify the failover priority for Aurora Replicas. Aurora Replicas can also offload read workloads from the primary DB
instance.https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html (https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html)

E. Yes. Since Migration of data from MySQL to Aurora MySQL is a standard Migration, DMS would fulfill the requirement. There is no schema conversion involved in the implementation https://docs.aws.amazon.com/dms/latest/userguide/Welcome.html (https://docs.aws.amazon.com/dms/latest/userguide/Welcome.html)
F. No. Since migration is between same databases, schema conversion is not needed. Use the AWS Schema Conversion Tool (AWS SCT) to translate your database schema to the new platform. Use AWS DMS to migrate the data 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 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 👍 👎

Shell Labs is an American manufacturer of smart home products including thermostats, smoke detectors, and security systems. They use IOT integration to manage the product lifecycle. They build their IOT platform on AWS. Currently the team is working on setting up authorization for the devices which includes setting up AWS IOT policies and IAM policies. Please identify the list of IOT Policy actions that can be enabled to manage operations. select 4 options.

- [ ] **A.** permission to connect to the AWS IoT message broker ✔
- [ ] **B.** permission to publish on an MQTT topic ✔
- [ ] **C.** permission to receive a message from AWS IoT ✔
- [ ] **D.** permission to subscribe to a topic filter ✔
- [ ] **E.** permission to create thing
- [ ] **F.** permission to cancel job execution

Explanation :

Answer : A, B, C, D
A. Yes. This is an IOT policy
https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html
(https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html)
B. Yes. This is an IOT policy
https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html
(https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html)
C. Yes. This is an IOT policy
https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html
(https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html)
D. Yes. This is an IOT policy
https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html
(https://docs.aws.amazon.com/iot/latest/developerguide/policy-actions.html)
E. No. This is an IAM policy
https://docs.aws.amazon.com/iot/latest/developerguide/iam-policies.html
(https://docs.aws.amazon.com/iot/latest/developerguide/iam-policies.html)
F. No. This is an IAM policy

Ask our Experts   👍 👎

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

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

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

- ○ **A.** Volume Gateway
- ○ **B.** File Gateway ✔
- ○ **C.** Tape Gateway
- ○ **D.** 0

Explanation :

Answer: B
A. No. A volume gateway provides cloud-backed storage volumes that you can mount as Internet Small Computer System Interface (iSCSI) devices from your on-premises application servers. The gateway supports the following volume configurations:
oCached volumes – You store your data in Amazon Simple Storage Service (Amazon S3) and retain a copy of frequently accessed data subsets locally. Cached volumes offer a substantial cost savings on primary storage and minimize the need to scale your storage on- premises. You also retainlow-latency access to your frequently accessed data.

oStored volumes – If you need low-latency access to your entire dataset, first configure your on-premises gateway to store all your data locally. Then asynchronously back up point-in- time snapshots of this data to Amazon S3. This configuration provides durable and inexpensive offsite backups that you can recover to your local data center or Amazon EC2. For example, if you need replacement capacity for disaster recovery, you can recover the backups to Amazon EC2. https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html (https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html)

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

- You can store and retrieve files directly using the NFS version 3 or 4.1 protocol.

- You can store and retrieve files directly using the SMB file system version, 2 and 3 protocol.

- You can access your data directly in Amazon S3 from any AWS Cloud application or service.

- You can manage your Amazon S3 data using lifecycle policies, cross-region replication, and versioning. You can think of a file gateway as a file system mount on S3. https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html (https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html)

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

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.

HHruns 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.

HHis using kinesis analytics to build SQL querying capability on streaming and planning to use different types of queries to process the data. HH need to ensure proper authentication and authorization control for kinesis analytics application needs to be enabled. How can this be achieved? select 2 options.

- ☐ **A.** Authentication and Access to AWS resources using following identities like root user, IAM User, and IAM role thereby managing federated user access, AWS service access and Applications running on Amazon EC2 ✔

- ☐ **B.** Access Control using following identities like root user, IAM User, and IAM role thereby managing federated user access, AWS service access and Applications running on Amazon EC2

- ☐ **C.** Authentication and Access to AWS resources through Permissions, policies, Actions and Resources

- ☐ **D.** Access Control through Permissions, policies, Actions and Resources ✔

Explanation :

Answer: A, D
Access to Amazon Kinesis Data Analytics requires credentials. Those credentials must have permissions to access AWS resources, such as an Amazon Kinesis Data Analytics application or an Amazon Elastic Compute Cloud (Amazon EC2) instance.

- Authentication - root user, IAM User, and IAM role thereby managing federated user access, AWS service access and Applications running on Amazon EC2

- Access Control - through Permissions, policies, Actions and Resourceshttps://docs.aws.amazon.com/kinesisanalytics/latest/dev/authentication-and-access-(https://docs.aws.amazon.com/kinesisanalytics/latest/dev/authentication-and-access-control.html)control.html (https://docs.aws.amazon.com/kinesisanalytics/latest/dev/authentication-and-access-control.html)

Ask our Experts 👍 👎

QUESTION 57   UNATTEMPTED   STORAGE

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

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

performance of AWS services capabilities. Parson is currently using running their DWH, on-premises on Teradata and is concerned on the overall costs of the DWH on AWS. They want to initially migrate the platform onto AWS to address their performance intensive workloads in place. Also they have business needs around real-time data integration, data driven analytics as a roadmap in next 6 months. Currently the number of users accessing the application would be around 100. What is your suggestion? select 1 option.

○ **A.** Launch Redshift cluster with node types DS2.xlarge to fulfill the requirements

○ **B.** Launch Redshift cluster with node types DS2.8xlarge to fulfill the requirements

○ **C.** Launch Redshift cluster with node types DC2.xlarge to fulfill the requirements ✔

○ **D.** Launch Redshift cluster with node types DC2.8xlarge to fulfill the requirements

---

Explanation :

Answer : C

A. Incorrect. DS2 node types are optimized for large data workloads and use hard disk drive (HDD) storage. This is not the requirement
https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and- (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes) nodes (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes)

B. Incorrect. DS2 node types are optimized for large data workloads and use hard disk drive (HDD) storage. This is not the requirement
https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and- (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes) nodes (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes)

C. Correct. DC2 node types are optimized for performance-intensive workloads. DC2.xlarge fulfills the requirements since it provides massive parallel processing using multiple nodes. Based on the amount of data loaded, this is the right option
https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and- (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes) nodes (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes)

D. Incorrect. DC2 node types are optimized for performance-intensive workloads. DC2.8xlarge does not fulfill the requirements since it can provide massive parallel processing using multiple nodes. Since cost and performance is also a concern, this is not the right option.
https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and- (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes) nodes (https://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#rs-about-clusters-and-nodes)

Ask our Experts 👍 👎

MSP Bank, Limited is a leading Japanese monetary institution that provides a full range of financial products and services to both institutional and individual customers. It is headquartered in Tokyo. MSP Bank is hosting their existing infrastructure on AWS. MSP bank has many segments internally and they are planning to launch a self-data discovery platform running out of AWS on QuickSight.

Using QuickSight, multiple datasets are created and multiple analyses are generated respectively. The Team is working on visuals and looking to measure for the intersection of two dimensions, with color- coding to easily differentiate where values fall in the range Please select 1 option.

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

Explanation :

Answer: C

A.No. A bar chart visual type is used to create a single-measure, multi-measure, or clustered bar chart. A single-measurebar 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. Yes. 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 👍 👎

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.

Which EMR Hadoop ecosystem fulfills the requirements? select 1 option

○ **A.** Apache Hive

○ **B.** Apache Zepplin

○ **C.** Jupyter Notebook ✔

○ **D.** Apache Presto

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. Use Apache Zeppelin as a notebook for interactive data exploration. Zepplin can be accessed through web interface using a SSH tunnel to the EMR master node and a proxy connection https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-zeppelin.html (https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-zeppelin.html)

C. Yes. Jupyter Notebook is an open-source web application that you can use to create and share documents that contain live code, equations, visualizations, and narrative text. Amazon EMR offers you two options to work with Jupyter notebooks:

- EMR Notebook

- JupyterHub

https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-jupyter.html (https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-jupyter.html)
D. No. Presto is a fast SQL query engine designed for interactive analytic queries over large datasets from multiple sources https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-presto.html (https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-presto.html)

Ask our Experts   👍 👎

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-Bankhas the following objectives in mind for the solution.

Tick-Bank has multiple data streams supporting various business processes. To optimize the costs, Tick-Bank wants to use batching methods (aggregation and collection) appropriate library to process the data. Which library supports aggregation mechanism. select 1 option.

○   **A.** KPL (Kinesis Producer Library) supports batching of user records, both aggregation and collection ✔

○   **B.** Kinesis Agent supports batching of user records, both aggregation and collection

○   **C.** Amazon Kinesis Data Streams API supports batching of user records, both aggregation and collection

○   **D.** 0

Explanation :

Answer: A

A. Yes. KPL supports batching of records. Batching refers to performing a single action on multiple items instead of repeatedly performing the action on each individual item.

In this context, the "item" is a record, and the action is sending it to Kinesis Data Streams. In a non-batching situation, you would place each record in a separate Kinesis Data Streams record and make one HTTP request to send it to Kinesis Data Streams. With batching, each HTTP request can carry multiple records instead of just one.

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.

The two types of KPL batching are designed to 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)

B. No. Develop producers using the Amazon Kinesis Data Streams API with the AWS SDK for Java. Batching is not supported/ https://docs.aws.amazon.com/streams/latest/dev/developing-producers-with- (https://docs.aws.amazon.com/streams/latest/dev/developing-producers-with-sdk.html) sdk.html (https://docs.aws.amazon.com/streams/latest/dev/developing-producers-with-sdk.html)

C. Kinesis Agent is a stand-alone Java software application that offers an easy way to collect and send data to Kinesis Data Streams. The agent continuously monitors a set of files and sends new data to your stream. The agent handles file rotation, checkpointing, and retry upon failures. It delivers all of your data in a reliable, timely, and simple manner. It also emits Amazon CloudWatch metrics to help you better monitor and troubleshoot the streaming process. Batching is not supported. https://docs.aws.amazon.com/streams/latest/dev/writing-with-agents.html (https://docs.aws.amazon.com/streams/latest/dev/writing-with-agents.html)

Ask our Experts

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 Investments 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 Investments 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. ETL team is working with IT operations team to get notifications on Consistency Views in terms of Eventual consistency. What needs to be done to monitor and manage notifications? select 1 option.

○ **A.** Enable CloudTrial metrics and Amazon SNS messages in EMRFS for Amazon S3 eventual consistency issues.

○ **B.** Enable CloudWatch metrics and Amazon SNS messages in EMRFS for Amazon S3 eventual consistency issues. ✔

○ **C.** Enable CloudWatch metrics and Amazon SQS messages in EMRFS for Amazon S3 eventual consistency issues.

○ **D.** Enable CloudTrial metrics and Amazon SQS messages in EMRFS for Amazon S3 eventual consistency issues.

Explanation :

Answer : B

A.No

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 (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.No

D.No

Ask our Experts   👍 👎

GiantCampaign is a technology-driven media measurement company. The GiantCampaign dashboard is used primarily by music industry professionals such as concert promoters, venues, radio programmers, managers, agents, and marketers to access information about the popularity of artists and songs across radio airplay, online streaming, social activity, sales, and live events.

GiantCampaign hosted their entire infrastructure on AWS and uses Data Pipeline as data integration mechanism which includes data integration, processing and store data. AWS Pipeline works with AWS compute services to address transformation and works with AWS storage services to store data. Please identify the relevant services.

Select 2 options.

- ☐ **A.** AWS Data Pipeline works with Amazon DynamoDB, Amazon RDS, Amazon Redshift, and Amazon S3 to provide compute for transformation of data

- ☐ **B.** AWS Data Pipeline works with Amazon EC2, and Amazon EMR to provide storage of data

- ☐ **C.** AWS Data Pipeline works with Amazon DynamoDB, Amazon RDS, Amazon Redshift, and Amazon S3 to provide storage for data ✔

- ☐ **D.** AWS Data Pipeline works with Amazon EC2, and Amazon EMR to provide compute for transformation of data ✔

Explanation :

Answer: C, D

A. No. AWS Data Pipeline works with Amazon DynamoDB, Amazon RDS, Amazon Redshift, and Amazon S3 to provide storage for

data https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/datapipelin (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/datapipeline-related-services.html)e-related-services.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/datapipeline-related-services.html)

B. No. AWS Data Pipeline works with Amazon EC2, and Amazon EMR to provide compute for transformation of

data https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/datapipelin (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/datapipeline-related-services.html)e-related-services.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/datapipeline-related-services.html)

C.Yes. AWS Data Pipeline works with Amazon DynamoDB, Amazon RDS, Amazon Redshift, and Amazon S3 to provide storage for

datahttps://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/datapipelin (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/datapipeline-related-services.html)e-related-services.html (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/datapipeline-related-services.html)

D.Yes. AWS Data Pipeline works with Amazon EC2, and Amazon EMR to provide compute for transformation of

datahttps://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/datapipelin (https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/datapipeline-related-

services.html)e-related-services.html
(https://docs.aws.amazon.com/datapipeline/latest/DeveloperGuide/datapipeline-related-
services.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.

HHruns 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 Streams API and Agents to collect events and transaction logs and process the stream. The event/log size is around 12 Bytes.

The data transformation and conversion is applied using kinesis firehose and stored in S3. Since the data captured is sensitive, HH is looking to enable encryption end to end for all the data sources. The data processed through kinesis firehose is captured from Kinesis Stream and other data sources though kinesis agent, firehose API, etc. How can this be enabled? select 2 options.

- [ ] **A.** Server-Side Encryption with Kinesis Data Streams as the Data Source ✔
- [ ] **B.** Client-Side Encryption with Kinesis Data Streams as the Data Source
- [ ] **C.** Client-Side Encryption with Direct PUT or Other Data Sources
- [ ] **D.** Server-Side Encryption with Direct PUT or Other Data Sources ✔

Explanation :

Answer: A,D
Enable server-side data encryption when you use Amazon Kinesis Data Firehose. The data encryption can be enabled through two modes.

- Server-Side Encryption with Kinesis Data Streams as the Data Source using AWS Key Management Service (AWS KMS) key before storing the data at rest

- Server-Side Encryption with Direct PUT or Other Data Sources using StartDeliveryStreamEncryption operation. https://docs.aws.amazon.com/firehose/latest/dev/encryption.html
(https://docs.aws.amazon.com/firehose/latest/dev/encryption.html)

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

Management drives marketing team to recommend new products to Customers and upscale revenue. Management is setting up forecasts and want to know how much more revenue can be generated by each Marketing team in the region end of every quarter. IT team is providing information about Marketing team, their customers and prospects, previous actuals and other information.

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, 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

MSP Bank, Limited is a leading Japanese monetary institution that provides a full range of financial products and services to both institutional and individual customers. It is headquartered in Tokyo. MSP Bank is hosting their existing infrastructure on AWS. MSP bank has many segments internally and they are planning to launch a self-data discovery platform running out of AWS on QuickSight.

Using QuickSight, multiple datasets are created and multiple analyses are generated respectively. The Team identified multiple datasets both dependent and not dependent on SPICE that needs to be refreshed. Team is evaluating the options. Please Advice! select 6 options.

- [ ] **A.** Refresh a SPICE Data Set from the Your Data Sets Page ✔
- [ ] **B.** Refresh a SPICE Data Set During Data Preparation ✔
- [ ] **C.** Refresh a SPICE Data Set During Data Analysis
- [ ] **D.** Refresh a SPICE Data Set on a Schedule. ✔
- [ ] **E.** To Refresh file-based data, you must delete and recreate the data set ✔
- [ ] **F.** To refresh file-based data, you must copy the new dataset onto the old data set
- [ ] **G.** To refresh data from a database, reopen your data set ✔
- [ ] **H.** To refresh data from a database, reopen the visualization ✔

Explanation :

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

A. Yes. Refresh a SPICE Data Set from the Your Data Sets Page is allowed
https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html#refresh-
(https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html#refresh-spice-
data) (https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html#refresh-
spice-data)spice-data (https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-
data.html#refresh-spice-data)

B. Yes. Refresh a SPICE Data Set During Data Preparation is allowed.
https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html
(https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html)

C. No. Refresh a SPICE Data Set During Data Analysis is not allowed. Instead, Refresh a Data Set During
Data Preparation is allowed
https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html
(https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html)

D. Yes. Refresh a SPICE Data Set on a Schedule.
https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html
(https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html)

E. Yes. To refresh file-based data, you must delete and recreate the data set
https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html
(https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html)

F. No. To refresh file-based data, you must delete and recreate the data set
https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html
(https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html)

G. Yes. To refresh data from a database, reopen your data set or the visualization you created.
https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html
(https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html)

H.Yes. To refresh data from a database, reopen your data set or the visualization you
created.https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html
(https://docs.aws.amazon.com/quicksight/latest/user/refreshing-imported-data.html)

Ask our Experts 👍 👎 ⌃

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

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

MSP Bank is looking at gateway interface which provides cloud-backed storage volumes that you can mount as Internet Small Computer System Interface (iSCSI) devices from your on-premises application servers. Also need low-latency access to your entire dataset thereby configuring on- premises gateway to store all your data locally, then asynchronously back up point-in-time snapshots of this data to Amazon S3? How can this be achieved? select 2 options.

- ☐ **A.** Volume Gateway ✔
- ☐ **B.** Cached Volumes
- ☐ **C.** Stored Volumes ✔
- ☐ **D.** Tape Gateway

Explanation :

Answer: A,C

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

- Cached volumes – You store your data in Amazon Simple Storage Service (Amazon S3) and retain a copy of frequently accessed data subsets locally. Cached volumes offer a substantial cost savings on primary storage and minimize the need to scale your storage on- premises. You also retain low-latency access to your frequently accessed data.

- Stored volumes – If you need low-latency access to your entire dataset, first configure your on-premises gateway to store all your data locally. Then asynchronously back up point-in- time snapshots of this data to Amazon S3. This configuration provides durable and inexpensive offsite backups that you can recover to your local data center or Amazon EC2. For example, if you need replacement capacity for disaster recovery, you can recover the backups to Amazon EC2. https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html (https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html)

B. No. Cached volumes – You store your data in Amazon Simple Storage Service (Amazon S3) and retain a copy of frequently accessed data subsets locally. Cached volumes offer a substantial cost savings on primary storage and minimize the need to scale your storage on-premises. You also retain low-latency access to your frequently accessed data. https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html (https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html)

C. Yes. Stored volumes – If you need low-latency access to your entire dataset, first configure your on-premises gateway to store all your data locally. Then asynchronously back up point-in- time snapshots of this data to Amazon S3. This configuration provides durable and inexpensive offsite backups that you can recover to your local data center or Amazon EC2. For example, if you need replacement capacity for disaster recovery, you can recover the backups to Amazon EC2 https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html (https://docs.aws.amazon.com/storagegateway/latest/userguide/WhatIsStorageGateway.html)

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

Ask our Experts

MSP Bank, Limited is a leading Japanese monetary institution that provides a full range of financial products and services to both institutional and individual customers. It is headquartered in Tokyo. MSP Bank is hosting their existing infrastructure on premise. MSP bank has lot of web applications, databases and data warehouses, NoSQL databases, and other types of data stores. MSP Bank is working with AWS to migrate their existing platform onto AWS. What service does AWS propose to discover your existing servers, plan migrations, and track the status of each application/database migration? select 1 option.

- ○ **A.** AWS Database Migration Service (AWS DMS)
- ○ **B.** AWS Data Sync
- ○ **C.** AWS Migration Hub ✔
- ○ **D.** AWS Schema Conversion

Explanation :

Answer : C

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 (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.Yes. 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. No. 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 68   UNATTEMPTED   STORAGE

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

MSP bank has lot of web applications, databases and data warehouse built on Teradata, NoSQL databases, and other types of data stores. MSP bank has lot of data assets in terms of logs, documents; excel files, CSV files, PDF documents and others.

The Web Application has different user workloads at different parts of the day. MSP Bank is running one of their web application NODE.JS supported by MongoDB Database. The schema designed is document based. The team wants to migrate the platform on to AWS. Which NoSQL Managed service provides the document management capability? select 1 option.

○ **A.** Amazon Aurora Database, being a multi-modal database support document models and NoSQL requirements

○ **B.** Amazon RDS Database, being a multi-modal database support document models and NoSQL requirements

○ **C.** Amazon DynamoDB Database, being a document database support document models and NoSQL requirements ✔

○ **D.** Amazon Neptune Database, being a graph database support document models and NoSQL requirements

---

Explanation :

Answer: C

A. No. Amazon Aurora (Aurora) is a fully managed relational database engine that's compatible with MySQL and PostgreSQL. You already know how MySQL and PostgreSQL combine the speed and reliability of high-end commercial databases with the simplicity and cost-effectiveness of open-sourcedatabases. The code, tools, and applications you use today with your existing MySQL and PostgreSQL databases can be used with Aurora. With some workloads, Aurora can deliver up to five times the throughput of MySQL and up to three times the throughput of PostgreSQL without requiring changes to most of your existing applications.

Amazon Aurora supports relational data models and does not support graph model. https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html (https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/CHAP_AuroraOverview.html)

B. No. Amazon Relational Database Service (Amazon RDS) is a web service that makes it easier to set up, operate, and scale a relational database in the cloud. It provides cost-efficient, resizable capacity for an industry-standard relational database and manages common database administration tasks.

Amazon RDS supports relational data models and does not support graph model. https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html (https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html)

C. 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)

D. No. Amazon Neptune is a fast, reliable, fully managed graph database service that makes it easy to build and run applications that work with highly connected datasets. The core of Neptune is a purpose-built, high-performance graph database engine that is optimized for storing billions of

relationships and querying the graph with milliseconds
latency https://docs.aws.amazon.com/neptune/latest/userguide/intro.html
(https://docs.aws.amazon.com/neptune/latest/userguide/intro.html)

Ask our Experts   👍 👎

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-serveapplication" 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 uses Elasticsearch to perform search on all the datasets. The team finds that that the performance of the search application is very slow. How the performance and availability of Elasticsearch can be improved? select 3 options

- ☑ **A.** choosing number of shards (increasing number of shards improves performance) ✔

- ☑ **B.** choosing instance types and sizes (increasing instance size improves performance) ✔

- ☑ **C.** choosing number of shards (decreasing number of shards improves performance)

- ☑ **D.** choosing instance types and sizes(decreasing instance size improves performance)

- ☑ **E.** choosing relevant storage for different types of indexes (long lived/rolling) ✔

Explanation :

Answer: A,B,E
A. Yes. Instance sizes, workloads, indexes, shards always improves the performance and availability https://docs.aws.amazon.com/elasticsearch-
(https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/aes-bp.html)
service/latest/developerguide/aes-bp.html (https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/aes-bp.html)
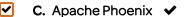
Allianz Financial Services (AFS) is a banking group offering end-to-end banking and financial solutions in South East Asia through its consumer banking, business banking, Islamic banking, investment finance and stock broking businesses as well as unit trust and asset administration, having served the financial community over the past five decades.

AFS launched EMR cluster to support their big data analytics requirements. AFS is planning to build an application running on EMR which supports both OLTP and operational analytics allowing you to use standard SQL queries and JDBC APIs to work with an Apache HBase backing store. Also data transfer tool between Amazon S3, Hadoop, HDFS, and RDBMS databases.

Which EMR Hadoop ecosystem fulfills the requirements? select 2 options.

- ☐ **A.** Apache Hue
- ☐ **B.** pache Flink
- ☑ **C.** Apache Phoenix ✔
- ☑ **D.** Apache Sqoop ✔

Explanation :

Answer : C,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-timestream 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. Yes. 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 Sqoop is a tool for transferring data between Amazon S3, Hadoop, HDFS, and RDBMS databases. https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-sqoop.html (https://docs.aws.amazon.com/emr/latest/ReleaseGuide/emr-sqoop.html)

Ask our Experts

Finish Review (https://www.whizlabs.com/learn/course/aws-bds-practice-tests/quiz/14899)

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

f
(https://www.facebook.com/whizlabs.software/)

in
(https://in.linkedin.com/company/whizlabs-software)

🐦
(https://twitter.com/whizlabs?lang=en)

G+
(https://plus.google.com/+WhizlabsSoftware)