**QUESTION 18**

A company needs to configure a real-time data ingestion architecture for its application.

The company needs an API, a process that transforms data as the data is streamed, and a storage solution for the data.

Which solution will meet these requirements with the LEAST operational overhead?

1. Deploy an Amazon EC2 instance to host an API that sends data to an Amazon Kinesis data stream.

Create an Amazon Kinesis Data Firehose delivery stream that uses the Kinesis data stream as a data source.

Use AWS Lambda functions to transform the data.

Use the Kinesis Data Firehose delivery stream to send the data to Amazon S3.

1. Deploy an Amazon EC2 instance to host an API that sends data to AWS Glue. Stop source/destination checking on the EC2 instance.

Use AWS Glue to transform the data and to send the data to Amazon S3.

1. Configure an Amazon API Gateway API to send data to an Amazon Kinesis data stream. Create an Amazon Kinesis Data Firehose delivery stream that uses the Kinesis data stream as a data source.

Use AWS Lambda functions to transform the data.

Use the Kinesis Data Firehose delivery stream to send the data to Amazon S3.

1. Configure an Amazon API Gateway API to send data to AWS Glue.

Use AWS Lambda functions to transform the data. Use AWS Glue to send the data to Amazon S3.

**Answer:** C

**QUESTION 74**

A company produces batch data that comes from different databases.

The company also produces live stream data from network sensors and application APIs. The company needs to consolidate all the data into one place for business analytics.

The company needs to process the incoming data and then stage the data in different Amazon S3 buckets.

Teams will later run onetime queries and import the data into a business intelligence tool to show key performance indicators (KPIs).

Which combination of steps will meet these requirements with the LEAST operational overhead? (Choose TWO)

1. Use Amazon Athena foe one-time queries.

Use Amazon QuickSight to create dashboards for KPIs.

1. Use Amazon Kinesis Data Analytics for one-time queries. Use Amazon QuickSight to create dashboards for KPIs.

1. Create custom AWS Lambda functions to move the individual records from me databases to an Amazon Redshift duster.
2. Use an AWS Glue extract transform, and toad (ETL) job to convert the data into JSON format.

Load the data into multiple Amazon OpenSearch Service (Amazon Elasticsearch Service) dusters.

1. Use blueprints in AWS Lake Formation to identify the data that can be ingested into a data lake.

Use AWS Glue to crawl the source extract the data and load the data into Amazon S3 in Apache Parquet format.

**Answer:** AB

**QUESTION 132**

A company hosts more than 300 global websites and applications.

The company requires a platform to analyze more than 30 TB of clickstream data each day.

What should a solutions architect do to transmit and process the clickstream data?

1. Design an AWS Data Pipeline to archive the data to an Amazon S3 bucket and run an Amazon EMR duster with the data to generate analytics.
2. Create an Auto Scaling group of Amazon EC2 instances to process the data and send it to an Amazon S3 data lake for Amazon Redshift to use tor analysis.
3. Cache the data to Amazon CloudFron.

Store the data in an Amazon S3 bucket.

When an object is added to the S3 bucket, run an AWS Lambda function to process the data tor analysis.

1. Collect the data from Amazon Kinesis Data Streams.

Use Amazon Kinesis Data Firehose to transmit the data to an Amazon S3 data lake Load the data in Amazon Redshift for analysis

**Answer:** D

**QUESTION 132**

A company hosts more than 300 global websites and applications.

The company requires a platform to analyze more than 30 TB of clickstream data each day.

What should a solutions architect do to transmit and process the clickstream data?

1. Design an AWS Data Pipeline to archive the data to an Amazon S3 bucket and run an Amazon EMR duster with the data to generate analytics.
2. Create an Auto Scaling group of Amazon EC2 instances to process the data and send it to an Amazon S3 data lake for Amazon Redshift to use tor analysis.
3. Cache the data to Amazon CloudFron.

Store the data in an Amazon S3 bucket.

When an object is added to the S3 bucket, run an AWS Lambda function to process the data tor analysis.

1. Collect the data from Amazon Kinesis Data Streams.

Use Amazon Kinesis Data Firehose to transmit the data to an Amazon S3 data lake Load the data in Amazon Redshift for analysis

**Answer:** D

**QUESTION 185**

A company is using a fleet of Amazon EC2 instances to ingest data from on-premises data sources.

The data is in JSON format and ingestion rates can be as high as 1 MB/s.

When an EC2 instance is rebooted, the data in-flight is lost.

The company's data science team wants to query ingested data near-real time.

Which solution provides near-real-time data querying that is scalable with minimal data loss?

1. Publish data to Amazon Kinesis Data Streams.

Use Kinesis Data Analytics to query the data.

1. Publish data to Amazon Kinesis Data Firehose with Amazon Redshift as the destination.

Use Amazon Redshift to query the data.

1. Store ingested data in an EC2 instance store.

Publish data to Amazon Kinesis Data Firehose with Amazon S3 as the destination.

Use Amazon Athena to query the data.

1. Store ingested data in an Amazon Elastic Block Store (Amazon EBS) volume.

Publish data to Amazon ElastiCache for Redis.

Subscribe to the Redis channel to query the data.

**Answer:** B

**Explanation:**

Kinesis data streams consists of shards.

The more througput is needed, the more shards you add, the less throughput, the more shards you remove, so it's scalable. Each shard can handle up to 1MB/s of writes.

However Kinesis data streams stores ingested data for only 1 to 7 days so there is a chance of data loss.

Additionally, Kinesis data analytics and kinesis data streams are both for real-time ingestion and analytics.

Firehouse on the other hand is also scalable and processes data in near real time as per the requirement. It also transfers data into Redshift which is a data warehouse so data won't be lost.

Redshift also has a SQL interface for performing queries for data analytics.

**QUESTION 186**

A company is developing a mobile game that streams score updates to a backend processor and then posts results on a leaderboard.

A solutions architect needs to design a solution that can handle large traffic spikes, process the mobile game updates in order of receipt, and store the processed updates in a highly available database.

The company also wants to minimize the management overhead required to maintain the solution.

What should the solutions architect do to meet these requirements?

1. Push score updates to Amazon Kinesis Data Streams.

Process the updates in Kinesis Data Streams with AWS Lambda.

Store the processed updates in Amazon DynamoDB.

1. Push score updates to Amazon Kinesis Data Streams.

Process the updates with a fleet of Amazon EC2 instances set up for Auto Scaling.

Store the processed updates in Amazon Redshift.

1. Push score updates to an Amazon Simple Notification Service (Amazon SNS) topic. Subscribe an AWS Lambda function to the SNS topic to process the updates.

1. Store the processed updates in a SQL database running on Amazon EC2.
2. Push score updates to an Amazon Simple Queue Service (Amazon SQS) queue.

Use a fleet of Amazon EC2 instances with Auto Scaling to process the updates in the SOS queue.

Store the processed updates in an Amazon RDS Multi-AZ DB instance.

**Answer:** A

**Explanation:**

Keywords to focus on would be highly available database - DynamoDB would be a better choice for leaderboard.

**QUESTION 185**

A company is using a fleet of Amazon EC2 instances to ingest data from on-premises data sources.

The data is in JSON format and ingestion rates can be as high as 1 MB/s.

When an EC2 instance is rebooted, the data in-flight is lost.

The company's data science team wants to query ingested data near-real time.

Which solution provides near-real-time data querying that is scalable with minimal data loss?

1. Publish data to Amazon Kinesis Data Streams.

Use Kinesis Data Analytics to query the data.

1. Publish data to Amazon Kinesis Data Firehose with Amazon Redshift as the destination.

Use Amazon Redshift to query the data.

1. Store ingested data in an EC2 instance store.

Publish data to Amazon Kinesis Data Firehose with Amazon S3 as the destination.

Use Amazon Athena to query the data.

1. Store ingested data in an Amazon Elastic Block Store (Amazon EBS) volume.

Publish data to Amazon ElastiCache for Redis.

Subscribe to the Redis channel to query the data.

**Answer:** B

**Explanation:**

Kinesis data streams consists of shards.

The more througput is needed, the more shards you add, the less throughput, the more shards you remove, so it's scalable.

Each shard can handle up to 1MB/s of writes.

However Kinesis data streams stores ingested data for only 1 to 7 days so there is a chance of data loss.

Additionally, Kinesis data analytics and kinesis data streams are both for real-time ingestion and analytics.

Firehouse on the other hand is also scalable and processes data in near real time as per the requirement. It also transfers data into Redshift which is a data warehouse so data won't be lost.

Redshift also has a SQL interface for performing queries for data analytics.

**QUESTION 181**

A company runs an online marketplace web application on AWS.

The application serves hundreds of thousands of users during peak hours.

The company needs a scalable, near-realtime solution to share the details of millions of financial transactions with several other internal applications.

Transactions also need to be processed to remove sensitive data before being stored in a document database for low-latency retrieval.

What should a solutions architect recommend to meet these requirements?

1. Store the transactions data into Amazon DynamoDB.

Set up a rule in DynamoDB to remove sensitive data from every transaction upon write.

Use DynamoDB Streams to share the transactions data with other applications

1. Stream the transactions data into Amazon Kinesis Data.

Firehose to store data in Amazon DynamoDB and Amazon S3.

Use AWS Lambda integration with Kinesis Data Firehose to remove sensitive data.

Other applications can consume the data stored in Amazon S3.

1. Stream the transactions data into Amazon Kinesis Data Streams.

Use AWS Lambda integration to remove sensitive data from every transaction and then store the transactions data in Amazon DynamoDB.

Other applications can consume the transactions data off the Kinesis data stream.

1. Store the batched transactions data in Amazon S3 as files.

Use AWS Lambda to process every file and remove sensitive data before updating the files in Amazon S3.

The Lambda function then stores the data in Amazon DynamoDB.

Other applications can consume transaction files stored in Amazon S3.

**Answer:** C

**Explanation:**

The destination of your Kinesis Data Firehose delivery stream.

Kinesis Data Firehose can send data records to various destinations, including Amazon Simple Storage Service (Amazon S3), Amazon Redshift, Amazon OpenSearch Service, and any HTTP endpoint that is owned by you or any of your third-party service providers.

The following are the supported destinations:

* Amazon OpenSearch Service
* Amazon S3
* Datadog
* Dynatrace
* Honeycomb
* HTTP Endpoint
* Logic Monitor
* MongoDB Cloud
* New Relic
* Splunk \* Sumo Logic

https://docs.aws.amazon.com/firehose/latest/dev/create-name.html

https://aws.amazon.com/kinesis/data-streams/

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.

**QUESTION 230**

A solutions architect is optimizing a website for an upcoming musical event.

Videos of the performances will be streamed in real time and then will be available on demand. The event is expected to attract a global online audience.

Which service will improve the performance of both the real-lime and on-demand streaming?

1. Amazon CloudFront.

1. AWS Global Accelerator.
2. Amazon Route 53.
3. Amazon S3 Transfer Acceleration.

**Answer:** A

**Explanation:**

You can use CloudFront to deliver video on demand (VOD) or live streaming video using any HTTP origin. One way you can set up video workflows in the cloud is by using CloudFront together with AWS Media Services.

<https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/on-demand-streamingvideo.html>