1. **True or False:** Amazon Kinesis processes and analyzes data as it arrives, and there is no need to wait until all data is collected before the processing can begin.

   ◉ True

   ○ False

   > ✓ **Correct**
   > Amazon Kinesis collects, processes, and analyzes real-time data—such as video, audio, application logs, website clickstreams, and Internet of Things (IoT) telemetry data—for machine learning, analytics, and other applications. For more information, see the *Diving Deep on Amazon Kinesis* reading.

2. A company wants to transfer files into and out of Amazon Simple Storage Service (Amazon S3) storage by using AWS Transfer Family. Several users in the company need permissions to access a specific object storage bucket that hosts the files from AWS Transfer Family. What is the BEST way to provide the needed bucket-access permissions to these users?

   ○ AWS account root user

   ○ AWS Identity and Access Management (IAM) user

   ◉ AWS Identity and Access Management (IAM) role

   ○ Access keys

   > ✓ **Correct**
   > An IAM user can assume a role to temporarily take on different permissions for a specific task. An IAM role does not have any credentials (password or access keys) that are associated with it. Instead of being uniquely associated with one person, a role can be assumed by anyone who needs it. For more information, see the *Batch Data Ingestion with AWS Transfer Family* video.

3. What is the most common way of categorizing data in terms of structure?

1 / 1 point

- ◉ Structured data, unstructured data, and semi-structured data
- ○ Ready data, not-ready data, and semi-ready data
- ○ The good data, the bad data, and the ugly data
- ○ Development data, quality assurance (QA) data, and production data

> ✓ **Correct**
> Data that is categorized in structured and semi-structured formats have some consistency that makes it easier for computer systems to consume without further modification. In contrast, unstructured data contains content that does not have a predefined data model. For more information, see the *Understanding Data Structure and When to Process Data* video in week 3.

4. A solutions architect is tasked with transporting data from their organization to AWS by using the AWS Snow Family. The organization wants to transport data that needs less than 8 terabytes (TB) of usable storage. Which Snow Family device should the solutions architect recommend for this use case?

1 / 1 point

- ○ AWS Snowball
- ○ AWS Glue
- ○ AWS Snowmobile
- ◉ AWS Snowcone

> ✓ **Correct**
> AWS Snowcone is the most compact and portable device in the AWS Snow Family. It is designed to move up to 8 TB of data. For more information, see the *Batch Data Ingestion with AWS Services* video.

5. Which statement about the AWS Glue crawler is TRUE?

1 / 1 point

- ◉ An AWS Glue crawler can scan a data store, such as an Amazon Simple Storage Service (Amazon S3) bucket, and use the data from the data store to create or update tables in the AWS Glue Data Catalog.

○ An AWS Glue crawler runs Structured Query Language (SQL) queries to analyze data directly in Amazon Simple Storage Service (Amazon S3).

○ An AWS Glue crawler collects and catalogs data from databases and object storage, moves the data into a new Amazon Simple Storage Service (Amazon S3) data lake, and classifies the data by using machine learning algorithms.

○ An AWS Glue crawler performs interactive log analytics, real-time application monitoring, website search, and more.

⊘ **Correct**
AWS Glue crawlers scan various data stores to automatically infer schemas and populate the AWS Glue Data Catalog with corresponding table definitions and statistics. The catalog that AWS Glue generates can be used by Amazon Athena, Amazon Redshift, Amazon Redshift Spectrum, Amazon EMR, and third-party analytics tools that use a standard Apache Hive metastore catalog. For more information, see the *Using Glue Crawlers* video in week 3.

6. **True or False:** AWS future-proofs data lakes with a standardized storage solution that has capabilities to grow and scale with an organization's needs.

1 / 1 point

⦿ True

○ False

⊘ **Correct**
It is important that data lakes can non-disruptively evolve as needed. By building data lakes on AWS, customers can evolve their business around data assets. They can then use these data assets to quickly drive more business value and competitive differentiation with virtually no limits. For more information, see the *Importance of Data Cataloging* reading in week 2.

7. Which statements about the Amazon Kinesis Family are true? (Choose TWO.)

1 / 1 point

☐ Amazon Kinesis Data Streams stores data only in the JSON format.

☑ The Amazon Kinesis Family can ingest a high volume of small bits of data that are being processed in real time.

> ⊘ **Correct**
> With Amazon Kinesis, customers can collect, process, and analyze real-time, streaming data to get timely insights and react quickly to new information. For more information, see the *Data Streaming Ingestion with Kinesis Services* video.

☑ By writing data consumers, customers can move data that is ingested into Amazon Kinesis Data Streams to an Amazon Simple Storage Service (Amazon S3) bucket with minimum modification.

> ⊘ **Correct**
> A *consumer* is an application that processes all data from a Kinesis data stream. One reason for writing custom consumers is to move data that is ingested by Kinesis Data Streams to Amazon S3. For more information, see the *Data Streaming Ingestion with Kinesis Services* video.

☐ Amazon Kinesis Data Analytics loads data streams into AWS databases.

☐ Amazon Kinesis Data Analytics provides an option to author non-Structured Query Language (SQL) code to process and analyze streaming data.

8. A company is receiving large amounts of streaming data from mobile devices, websites, servers, and sensors. They want to run analytics on the data that they are receiving. Which service should the company use for this use case?

1 / 1 point

○ Account monitoring with AWS CloudTrail

○ Log monitoring with Amazon CloudWatch

◉ Log analysis with Amazon Kinesis Family

○ Log analysis with Amazon Pinpoint

✓ Correct

One of the most common forms of data analysis with big data is log analytics. For more information, see the *Reviewing the Ingestion Part in Data Lake Architectures* video in week 3.