Module 4 - Quiz 3

1. A company needs a scalable and Hadoop-compatible service to process a large amount of historical data to identify trends and patterns. Which Google Cloud Platform is best suited to help this company?

- Looker
- Dataproc
- BigQuery
- Dataflow

Feedback:

- - Looker: Looker is primarily used for data visualization and reporting, not for processing large datasets.
- - Dataproc: Correct. Dataproc is a scalable, Hadoop-compatible service ideal for batch processing and analytics.
- - BigQuery: BigQuery is excellent for querying large datasets but not specifically designed for Hadoop compatibility.
- - Dataflow: Dataflow is used for stream and batch data processing but not Hadoop-compatible.

2. A cloud team has made the decision to use Google Cloud Storage as a data management tool. During a meeting, a company executive asks why the team made this decision. Which of the following statements supports the data team's decision?

- Google Cloud Storage allows the cloud team to create data integrations, run high-volume data pipelines, and continuously integrate user data.
- Google Cloud Storage allows the cloud team to move the company's application databases, analytics, and data science tools into one place.
- Google Cloud Storage allows the cloud team to stream data from the company's various applications into one database.
- Google Cloud Storage allows the cloud team to write SQL queries to join data and clean the data to ensure it is complete.

Feedback:

- - Google Cloud Storage allows the cloud team to create data integrations, run high-volume data pipelines, and continuously integrate user data.: This describes Cloud Data Fusion, not Google Cloud Storage.
- - Google Cloud Storage allows the cloud team to move the company's application databases, analytics, and data science tools into one place.: Correct. Cloud Storage serves as a central repository for various data sources and tools.

- - Google Cloud Storage allows the cloud team to stream data from the company's various applications into one database.: This describes Dataflow, not Cloud Storage.
- - Google Cloud Storage allows the cloud team to write SQL queries to join data and clean the data to ensure it is complete.: This describes BigQuery, not Cloud Storage.

3. A company wants to create a real-time data pipeline to process and analyze streaming data from its Internet of Things, or IoT devices. The company also needs to be able to store and query the data for historical analysis. Which Google Cloud Platform is best suited to help this company?

- Dataflow
- BigQuery
- Looker
- Dataproc

Feedback:

- - Dataflow: Correct. Dataflow is designed for real-time and batch data processing, ideal for IoT pipelines.
- - BigQuery: BigQuery is used for querying and analyzing data, not for building real-time pipelines.
- - Looker: Looker is a visualization tool, not suitable for data pipeline creation.
- - Dataproc: Dataproc is used for batch processing and machine learning, not real-time streaming.

4. A nonprofit organization needs help with large-scale analytics on both structured and unstructured data. The organization wants to use an open-source tool for batch processing, querying, streaming, and machine learning. Which data analytic tool should the nonprofit use?

- BigQuery
- Looker
- Dataflow
- Dataproc

Feedback:

- - BigQuery: BigQuery is great for structured data analysis but not for open-source batch processing.
- - Looker: Looker is used for visualization and reporting, not for large-scale analytics.
- - Dataflow: Dataflow is good for stream and batch processing but not focused on open-source tools.
- - Dataproc: Correct. Dataproc supports open-source tools like Hadoop and Spark for large-scale analytics.

5. A streaming video content company seeks advice from a cloud professional about creating a data pipeline. The company wants a recommendation for a data management tool that can build a pipeline to stream data from various applications and computers into a single database. Which cloud data solution should the cloud professional suggest?

- Dataflow
- BigQuery
- Looker
- Dataproc

Feedback:

- - Dataflow: Correct. Dataflow is ideal for building streaming and batch data pipelines.
- - BigQuery: BigQuery is used for querying and analyzing data, not for building pipelines.
- - Looker: Looker is a visualization tool, not suitable for data pipeline creation.
- - Dataproc: Dataproc is used for batch processing and machine learning, not for streaming pipelines.