

Data Engineering with Google Cloud Platform Training

Multiple Choice Questions

1. What is the main benefit of using Google Cloud Dataflow?
 - A. Dataflow is a real-time data analysis tool
 - B. Dataflow provides batch processing capabilities only
 - C. Dataflow allows efficient real-time and batch data processing
 - D. Dataflow is only used for data storage
2. Which of the following tools is integrated into Google Cloud Data Studio for real-time data visualization?
 - A. TensorFlow
 - B. BigQuery
 - C. Google Cloud Storage
 - D. Google Analytics
3. Which SQL operation does BigQuery support for combining two or more tables?
 - A. Merge
 - B. Group by
 - C. Select
 - D. Inner Join
4. What is a key feature of Google BigQuery?
 - A. It supports real-time streaming and querying of data.
 - B. It requires manual scaling for large datasets.
 - C. It only supports batch processing.
 - D. It is a non-cloud-based solution.
5. Which of the following is NOT a type of machine learning?
 - A. Supervised Learning
 - B. Unsupervised Learning
 - C. Reinforcement Learning
 - D. Collaborative Learning
6. What is the purpose of using side inputs in Apache Beam?
 - A. To merge data streams
 - B. To inject small datasets into the main processing pipeline
 - C. To increase the size of datasets
 - D. To filter unnecessary data

7. In the context of Spark, what is a DataFrame?
 - A. A distributed collection of data organized into named columns
 - B. A data structure that holds arbitrary data
 - C. A machine learning model
 - D. A type of RDD
8. What is the role of TensorFlow in machine learning?
 - A. It is used for data cleaning.
 - B. It is a framework for building and training machine learning models.
 - C. It is a cloud storage service.
 - D. It is a data visualization tool.
9. What service does Google Cloud offer for real-time data analytics?
 - A. BigQuery
 - B. Cloud Pub/Sub
 - C. Cloud Dataproc
 - D. Cloud Pub/Sub
10. Which of the following is a characteristic of Data Lakehouse?
 - A. It stores only structured data
 - B. It supports both operational and advanced analytics
 - C. It is limited to batch data processing
 - D. It cannot handle real-time data
11. Which tool can you use in GCP for serverless machine learning model deployment?
 - A. Vertex AI
 - B. Cloud Storage
 - C. Cloud Functions
 - D. Google Cloud Pub/Sub
12. Which feature of Apache Beam ensures parallel processing of large datasets?
 - A. Side Inputs
 - B. CoGroupByKey
 - C. DoFn
 - D. GroupByKey
13. What is the primary function of BigQuery in GCP?
 - A. To serve as a data visualization tool
 - B. To manage data security
 - C. To provide machine learning capabilities
 - D. To store and process data at large scale

14. Which data processing framework is used by Google Cloud Dataflow?
 - A. Apache Beam
 - B. Apache Spark
 - C. Apache Kafka
 - D. Apache Hive
15. What is a key characteristic of Data Lake in the data architecture?
 - A. It uses relational databases for data storage
 - B. It is only suitable for structured data
 - C. It stores large volumes of unstructured data
 - D. It does not support real-time analytics
16. Which SQL clause in BigQuery allows you to group rows based on similar values?
 - A. SELECT
 - B. GROUP BY
 - C. WHERE
 - D. ORDER BY
17. What type of machine learning model is used for classification tasks?
 - A. Linear Regression
 - B. Decision Trees
 - C. K-means Clustering
 - D. Logistic Regression
18. Which Google Cloud tool provides scalable analytics on structured data?
 - A. BigQuery
 - B. Cloud SQL
 - C. Cloud Spanner
 - D. Google Data Studio
19. Which feature does Apache Pig offer for large-scale data processing?
 - A. Automatic data scaling
 - B. Integration with Cloud Pub/Sub
 - C. Real-time data processing
 - D. A scripting language called Pig Latin
20. What is the main advantage of using Data Studio?
 - A. It offers real-time data integration
 - B. It supports advanced machine learning features
 - C. It provides a SQL-based querying tool
 - D. It is primarily used for data storage