Google GCP-Data Engineer Exam

Volume: 172 Questions

Question No:1
When training a machine learning model on Cloud ML Engine on a distributed scaled tier, what types of machines are part of that distributed resource? (Choose all that apply)

A. Host

B. Worker

C. Master

D. Parameter server

Answer: B,C,D

Explanation: You can have multiple workers, which divide up the work of training the model. Parameter servers coordinate shared model states between the workers.

Question No:2
You are working on a project with two compliance requirements. The first requirement states that your developers should be able to see the Google Cloud Platform billing charges for only their projects. The second requirement states that your finance team members can set budgets and view the current charges for all projects in the organization. The finance team should not be able to view the project contents. You want to set permissions. What should you do?

A. Add the finance team to the Viewer role for the Project. Add the developers to the Security Reviewer role for each of the billing accounts.

B. Add the developers and finance managers to the Viewer role for the Project.

C. Add the finance team members to the default IAM Owner role. Add the developers to a custom role that allows them to see their spending only.

D. Add the finance team members to the Billing Administrator role for each of the billing accounts that they need to manage. Add the developers to the Viewer role for the Project.

Answer: D

Explanation: Primitive roles are far too broad for this requirement.
https://cloud.google.com/iam/docs/understanding-roles

Google GCP-Data Engineer Exam

Question No:3
When using Cloud ML Engine to train machine learning models, how are online predictions different from batch predictions? (Choose all that apply)

A. Online prediction results are written to Cloud Storage as output.

B. Batch predictions are used to reduce latency in serving predictions.

C. Online predictions are returned in the response message.

D. Batch predictions are optimized to handle a high volume of prediction examples while running on more complex models.

Answer: C,D

Explanation: Online predictions create near real-time feedback with small, inline predictions.

Question No:4
What is the difference between a deep and wide neural network? What would you use a deep AND wide neural network for? (Choose all that apply)

A. Wide models are used for generalizations. Deep models are for memorization.

B. Deep and wide models are ideal for solving regression problems.

C. Wide models are used for memorization. Deep models are for generalization

D. Deep and wide models are ideal for a recommendation application.

Answer: C,D

Question No:5
While conducting BigQuery queries against a large table with many columns, you notice in the details section that you have a very large purple bar in the first stage of your query execution. How can you troubleshoot this to increase performance and reduce costs? (Choose all that apply)

A. Restrict the number of columns in your SELECT field for those needed. This will reduce read times on your query.

B. Partition or separate your large table into smaller pieces. Conduct a query against your

smaller (or partitioned) tables to reduce read times.

C. Reduce the number of read operations by adding a LIMIT clause to your query.

D. Reduce the number of write operations by optimizing the complexity of your query functions.

Answer: A,B

Explanation: The purple bar indicates the number of read operations. Limiting columns read will reduce the read time of your query.

Question No:6
You need to choose a structure storage option for storing very large amounts of data with the following properties and requirements:
. The data has a single key
. You need very low latency Which solution should you choose?

A. Bigtable

B. Datastore

C. Cloud SQL

D. BigQuery

Answer: A

Explanation: Bigtable uses a single key and has very low latency (in milliseconds). It is the best choice.
Cloud SQL holds at most 10TB of data and is not a high performance single key storage type.

Question No:7
You created a job which runs daily to import highly sensitive data from an on-premises location to Cloud Storage. You also set up a streaming data insert into Cloud Storage via a Kafka node that is running on a Compute Engine instance. You need to encrypt the data at rest and supply your own encryption key. Your key should not be stored in the Google Cloud. What should you do?

A. Upload your own encryption key to Cloud Key Management Service, and use it to encrypt your data in your Kafka node hosted on Compute Engine.

B. Create a dedicated service account, and use encryption at rest to reference your data stored in Cloud Storage and Compute Engine data as part of your API service calls.

C. Upload your own encryption key to Cloud Key Management Service, and use it to encrypt your data in Cloud Storage. Use your uploaded encryption key and reference it as part of your API service calls to encrypt your data in the Kafka node hosted on Compute Engine.

D. Supply your own encryption key, and reference it as part of your API service calls to encrypt your data in Cloud Storage and your Kafka node hosted on Compute Engine.

Answer: D

Question No:8
You are creating a machine learning model to predict the likelihood of fraud from credit card transaction data. What type of learning model problem is this?

A. Categorical

B. Regression

C. Hyperparameter

D. Clustering

Answer: A

Explanation: Categorical is for a set of finite categories, such as 'yes' or 'no'. Fraud is a yes/no output, so this fits.
Hyperparameters are for adjusting learning models

Question No:9
You have data stored in a Cloud Storage dataset and also in a BigQuery dataset. You need to secure the data and provide 3 different types of access levels for your Google Cloud Platform users: administrator, read/write, and read-only. You want to follow Google-recommended practices. What should you do?

A. At the Organization level, add your administrator user accounts to the Owner role, add your read/write user accounts to the Editor role, and add your read-only user accounts to the Viewer role.

B. Create 3 custom IAM roles with appropriate policies for the access levels needed for Cloud

Storage and BigQuery. Add your users to the appropriate roles.

C. At the Project level, add your administrator user accounts to the Owner role, add your read/write user accounts to the Editor role, and add your read-only user accounts to the Viewer role.

D. Use the appropriate pre-defined IAM roles for each of the access levels needed for Cloud Storage and BigQuery. Add your users to those roles for each of the services.

Answer: D

Explanation: It is best practice to use pre-created roles over custom roles with associated policies when they match your requirements, which they do in this scenario.
The principle of least privilege favors using pre-defined roles for granular access. It is also best practice to use pre-created roles over custom roles with associated policies when they match your requirements.

Question No:10
You are developing an application that will only recognize and tag specific business to business product logos in images. What is the best method to accomplish this task?

A. Use the Cloud Vision API to recognize logos in the images.

B. Create a custom machine learning model to recognize specific logos in photos, then train it on Cloud ML Engine.

C. Train your model on Kubernetes Engine to scale training as quickly as possible.

D. Use the Cloud Vision API to recognize all logos in images, then use the Cloud Natural Language API to recognize specific logos by name.

Answer: B

Explanation: Option B: This is a custom use case that Cloud Vision API is unable to do; therefore you would need to create a custom learning model on Cloud ML Engine.
Option C: Cloud ML Engine is the better service to use for this task.

Question No:11
You currently have a Bigtable instance you've been using for development running a development instance type, using HDD's for storage. You are ready to upgrade your development instance to a production instance for increased performance. You also want to upgrade your