Google

Exam Questions Professional-Machine-Learning-Engineer

Google Professional Machine Learning Engineer

NEW QUESTION 1
You want to rebuild your ML pipeline for structured data on Google Cloud. You are using PySpark to conduct data transformations at scale, but your pipelines are taking over 12 hours to run. To speed up development and pipeline run time, you want to use a serverless tool and SQL syntax. You have already moved your raw data into Cloud Storage. How should you build the pipeline on Google Cloud while meeting the speed and processing requirements?

A. Use Data Fusion's GUI to build the transformation pipelines, and then write the data into BigQuery
B. Convert your PySpark into SparkSQL queries to transform the data and then run your pipeline on Dataproc to write the data into BigQuery.
C. Ingest your data into Cloud SQL convert your PySpark commands into SQL queries to transform the data, and then use federated queries from BigQuery for machine learning
D. Ingest your data into BigQuery using BigQuery Load, convert your PySpark commands into BigQuery SQL queries to transform the data, and then write the transformations to a new table

Answer: B

NEW QUESTION 2
Your organization wants to make its internal shuttle service route more efficient. The shuttles currently stop at all pick-up points across the city every 30 minutes between 7 am and 10 am. The development team has already built an application on Google Kubernetes Engine that requires users to confirm their presence and shuttle station one day in advance. What approach should you take?

A. 1. Build a tree-based regression model that predicts how many passengers will be picked up at each shuttle station.* 2. Dispatch an appropriately sized shuttle and provide the map with the required stops based on the prediction.
B. 1. Build a tree-based classification model that predicts whether the shuttle should pick up passengers at each shuttle station.* 2. Dispatch an available shuttle and provide the map with the required stops based on the prediction
C. 1. Define the optimal route as the shortest route that passes by all shuttle stations with confirmed attendance at the given time under capacity constraints.* 2 Dispatch an appropriately sized shuttle and indicate the required stops on the map
D. 1. Build a reinforcement learning model with tree-based classification models that predict the presenceof passengers at shuttle stops as agents and a reward function around a distance-based metric* 2. Dispatch an appropriately sized shuttle and provide the map with the required stops based on the simulated outcome.

Answer: D

NEW QUESTION 3
You recently designed and built a custom neural network that uses critical dependencies specific to your organization's framework. You need to train the model using a managed training service on Google Cloud. However, the ML framework and related dependencies are not supported by AI Platform Training. Also, both your model and your data are too large to fit in memory on a single machine. Your ML framework of choice uses the scheduler, workers, and servers distribution structure. What should you do?

A. Use a built-in model available on AI Platform Training
B. Build your custom container to run jobs on AI Platform Training
C. Build your custom containers to run distributed training jobs on AI Platform Training
D. Reconfigure your code to a ML framework with dependencies that are supported by AI Platform Training

Answer: C

NEW QUESTION 4
You need to train a computer vision model that predicts the type of government ID present in a given image using a GPU-powered virtual machine on Compute Engine. You use the following parameters:
• Optimizer: SGD
• Image shape = 224x224
• Batch size = 64
• Epochs = 10
• Verbose = 2
During training you encounter the following error: ResourceExhaustedError: out of Memory (oom) when allocating tensor. What should you do?

A. Change the optimizer
B. Reduce the batch size
C. Change the learning rate
D. Reduce the image shape

Answer: A

NEW QUESTION 5
You are an ML engineer at a large grocery retailer with stores in multiple regions. You have been asked to create an inventory prediction model. Your models features include region, location, historical demand, and seasonal popularity. You want the algorithm to learn from new inventory data on a daily basis. Which algorithms should you use to build the model?

A. Classification
B. Reinforcement Learning
C. Recurrent Neural Networks (RNN)
D. Convolutional Neural Networks (CNN)

Answer: B

NEW QUESTION 6
Your team has been tasked with creating an ML solution in Google Cloud to classify support requests for one of your platforms. You analyzed the requirements and decided to use TensorFlow to build the classifier so that you have full control of the model's code, serving, and deployment. You will use Kubeflow pipelines for the ML platform. To save time, you want to build on existing resources and use managed services instead of building a completely new model. How should you

build the classifier?

A. Use the Natural Language API to classify support requests
B. Use AutoML Natural Language to build the support requests classifier
C. Use an established text classification model on Al Platform to perform transfer learning
D. Use an established text classification model on Al Platform as-is to classify support requests

Answer: D

NEW QUESTION 7
You developed an ML model with Al Platform, and you want to move it to production. You serve a few thousand queries per second and are experiencing latency issues. Incoming requests are served by a load balancer that distributes them across multiple Kubeflow CPU-only pods running on Google Kubernetes Engine (GKE). Your goal is to improve the serving latency without changing the underlying infrastructure. What should you do?

A. Significantly increase the max_batch_size TensorFlow Serving parameter
B. Switch to the tensorflow-model-server-universal version of TensorFlow Serving
C. Significantly increase the max_enqueued_batches TensorFlow Serving parameter
D. Recompile TensorFlow Serving using the source to support CPU-specific optimizations Instruct GKE to choose an appropriate baseline minimum CPU platform for serving nodes

Answer: A

NEW QUESTION 8
You are building a real-time prediction engine that streams files which may contain Personally Identifiable Information (Pll) to Google Cloud. You want to use the Cloud Data Loss Prevention (DLP) API to scan the files. How should you ensure that the Pll is not accessible by unauthorized individuals?

A. Stream all files to Google CloudT and then write the data to BigQuery Periodically conduct a bulk scan of the table using the DLP API.
B. Stream all files to Google Cloud, and write batches of the data to BigQuery While the data is being written to BigQuery conduct a bulk scan of the data using the DLP API.
C. Create two buckets of data Sensitive and Non-sensitive Write all data to the Non-sensitive bucket Periodically conduct a bulk scan of that bucket using the DLP API, and move the sensitive data to the Sensitive bucket
D. Create three buckets of data: Quarantine, Sensitive, and Non-sensitive Write all data to the Quarantine bucket.
E. Periodically conduct a bulk scan of that bucket using the DLP API, and move the data to either the Sensitive or Non-Sensitive bucket

Answer: A

NEW QUESTION 9
You are building a linear regression model on BigQuery ML to predict a customer's likelihood of purchasing your company's products. Your model uses a city name variable as a key predictive component. In order to train and serve the model, your data must be organized in columns. You want to prepare your data using the least amount of coding while maintaining the predictable variables. What should you do?

A. Create a new view with BigQuery that does not include a column with city information
B. Use Dataprep to transform the state column using a one-hot encoding method, and make each city a column with binary values.
C. Use Cloud Data Fusion to assign each city to a region labeled as 1, 2, 3, 4, or 5r and then use that number to represent the city in the model.
D. Use TensorFlow to create a categorical variable with a vocabulary list Create the vocabulary file, and upload it as part of your model to BigQuery ML.

Answer: C

NEW QUESTION 10
You are an ML engineer at a bank that has a mobile application. Management has asked you to build an ML-based biometric authentication for the app that verifies a customer's identity based on their fingerprint. Fingerprints are considered highly sensitive personal information and cannot be downloaded and stored into the bank databases. Which learning strategy should you recommend to train and deploy this ML model?

A. Differential privacy
B. Federated learning
C. MD5 to encrypt data
D. Data Loss Prevention API

Answer: B

NEW QUESTION 10
You recently joined a machine learning team that will soon release a new project. As a lead on the project, you are asked to determine the production readiness of the ML components. The team has already tested features and data, model development, and infrastructure. Which additional readiness check should you recommend to the team?

A. Ensure that training is reproducible
B. Ensure that all hyperparameters are tuned
C. Ensure that model performance is monitored
D. Ensure that feature expectations are captured in the schema

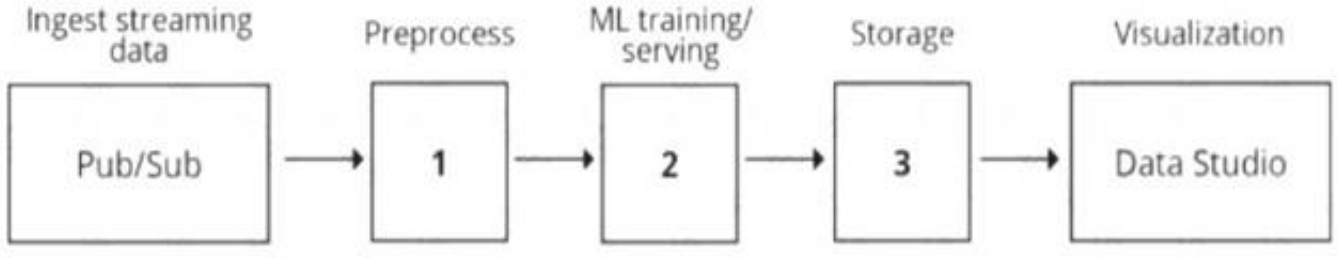
Answer: B

NEW QUESTION 11
You are responsible for building a unified analytics environment across a variety of on-premises data marts. Your company is experiencing data quality and security challenges when integrating data across the servers, caused by the use of a wide range of disconnected tools and temporary solutions. You need a fully managed, cloud-native data integration service that will lower the total cost of work and reduce repetitive work. Some members on your team prefer a codeless interface for building Extract, Transform, Load (ETL) process. Which service should you use?

A. Dataflow
B. Dataprep
C. Apache Flink
D. Cloud Data Fusion

Answer: D

NEW QUESTION 16
You are building an ML model to detect anomalies in real-time sensor data. You will use Pub/Sub to handle incoming requests. You want to store the results for analytics and visualization. How should you configure the pipeline?



A. 1 = Dataflow, 2 - Al Platform, 3 = BigQuery
B. 1 = DataProc, 2 = AutoML, 3 = Cloud Bigtable
C. 1 = BigQuery, 2 = AutoML, 3 = Cloud Functions
D. 1 = BigQuery, 2 = Al Platform, 3 = Cloud Storage

Answer: C

NEW QUESTION 19
You have trained a text classification model in TensorFlow using Al Platform. You want to use the trained model for batch predictions on text data stored in BigQuery while minimizing computational overhead. What should you do?

A. Export the model to BigQuery ML.
B. Deploy and version the model on Al Platform.
C. Use Dataflow with the SavedModel to read the data from BigQuery
D. Submit a batch prediction job on Al Platform that points to the model location in Cloud Storage.

Answer: A

NEW QUESTION 23
You are designing an ML recommendation model for shoppers on your company's ecommerce website. You will use Recommendations Al to build, test, and deploy your system. How should you develop recommendations that increase revenue while following best practices?

A. Use the "Other Products You May Like" recommendation type to increase the click-through rate
B. Use the "Frequently Bought Together' recommendation type to increase the shopping cart size for each order.
C. Import your user events and then your product catalog to make sure you have the highest quality event stream
D. Because it will take time to collect and record product data, use placeholder values for the product catalog to test the viability of the model.

Answer: C

NEW QUESTION 28
As the lead ML Engineer for your company, you are responsible for building ML models to digitize scanned customer forms. You have developed a TensorFlow model that converts the scanned images into text and stores them in Cloud Storage. You need to use your ML model on the aggregated data collected at the end of each day with minimal manual intervention. What should you do?

A. Use the batch prediction functionality of Al Platform
B. Create a serving pipeline in Compute Engine for prediction
C. Use Cloud Functions for prediction each time a new data point is ingested
D. Deploy the model on Al Platform and create a version of it for online inference.

Answer: D

NEW QUESTION 32
You are developing models to classify customer support emails. You created models with TensorFlow Estimators using small datasets on your on-premises system, but you now need to train the models using large datasets to ensure high performance. You will port your models to Google Cloud and want to minimize code refactoring and infrastructure overhead for easier migration from on-prem to cloud. What should you do?

A. Use Al Platform for distributed training
B. Create a cluster on Dataproc for training
C. Create a Managed Instance Group with autoscaling
D. Use Kubeflow Pipelines to train on a Google Kubernetes Engine cluster.

Answer: D

NEW QUESTION 34
You are training an LSTM-based model on Al Platform to summarize text using the following job submission script:

```
gcloud ai-platform jobs submit training $JOB_NAME \
        --package-path $TRAINER_PACKAGE_PATH \
        --module-name $MAIN_TRAINER_MODULE \
        --job-dir $JOB_DIR \
        --region $REGION \
        --scale-tier basic \
        -- \
        --epochs 20 \
        --batch_size=32 \
        --learning_rate=0.001 \
```

You want to ensure that training time is minimized without significantly compromising the accuracy of your model. What should you do?

A. Modify the 'epochs' parameter
B. Modify the 'scale-tier' parameter
C. Modify the batch size' parameter
D. Modify the 'learning rate' parameter

Answer: A

NEW QUESTION 35
You have a demand forecasting pipeline in production that uses Dataflow to preprocess raw data prior to model training and prediction. During preprocessing, you employ Z-score normalization on data stored in BigQuery and write it back to BigQuery. New training data is added every week. You want to make the process more efficient by minimizing computation time and manual intervention. What should you do?

A. Normalize the data using Google Kubernetes Engine
B. Translate the normalization algorithm into SQL for use with BigQuery
C. Use the normalizer_fn argument in TensorFlow's Feature Column API
D. Normalize the data with Apache Spark using the Dataproc connector for BigQuery

Answer: B

NEW QUESTION 40
You work for an online retail company that is creating a visual search engine. You have set up an end-to-end ML pipeline on Google Cloud to classify whether an image contains your company's product. Expecting the release of new products in the near future, you configured a retraining functionality in the pipeline so that new data can be fed into your ML models. You also want to use Al Platform's continuous evaluation service to ensure that the models have high accuracy on your test data set. What should you do?

A. Keep the original test dataset unchanged even if newer products are incorporated into retraining
B. Extend your test dataset with images of the newer products when they are introduced to retraining
C. Replace your test dataset with images of the newer products when they are introduced to retraining.
D. Update your test dataset with images of the newer products when your evaluation metrics drop below a pre-decided threshold.

Answer: C

NEW QUESTION 44
Your data science team needs to rapidly experiment with various features, model architectures, and hyperparameters. They need to track the accuracy metrics for various experiments and use an API to query the metrics over time. What should they use to track and report their experiments while minimizing manual effort?

A. Use Kubeflow Pipelines to execute the experiments Export the metrics file, and query the results using the Kubeflow Pipelines API.
B. Use Al Platform Training to execute the experiments Write the accuracy metrics to BigQuery, and query the results using the BigQueryAPI.
C. Use Al Platform Training to execute the experiments Write the accuracy metrics to Cloud Monitoring, and query the results using the Monitoring API.
D. Use Al Platform Notebooks to execute the experiment
E. Collect the results in a shared Google Sheetsfile, and query the results using the Google Sheets API

Answer: A

NEW QUESTION 47
You built and manage a production system that is responsible for predicting sales numbers. Model accuracy is crucial, because the production model is required to keep up with market changes. Since being deployed to production, the model hasn't changed; however the accuracy of the model has steadily deteriorated. What issue is most likely causing the steady decline in model accuracy?

A. Poor data quality
B. Lack of model retraining
C. Too few layers in the model for capturing information
D. Incorrect data split ratio during model training, evaluation, validation, and test

Answer: D

NEW QUESTION 51
You are an ML engineer in the contact center of a large enterprise. You need to build a sentiment analysis tool that predicts customer sentiment from recorded

phone conversations. You need to identify the best approach to building a model while ensuring that the gender, age, and cultural differences of the customers who called the contact center do not impact any stage of the model development pipeline and results. What should you do?

A. Extract sentiment directly from the voice recordings
B. Convert the speech to text and build a model based on the words
C. Convert the speech to text and extract sentiments based on the sentences
D. Convert the speech to text and extract sentiment using syntactical analysis

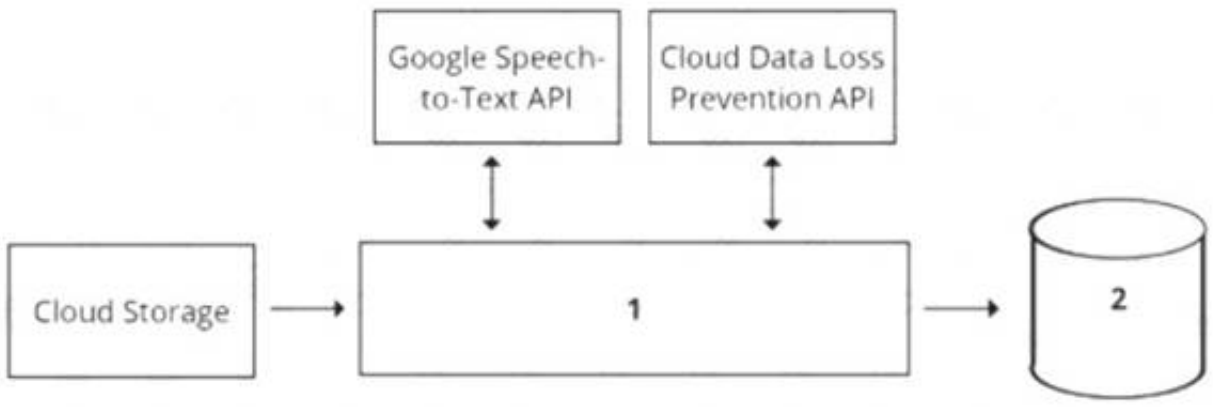
Answer: C

NEW QUESTION 56
You work on a growing team of more than 50 data scientists who all use AI Platform. You are designing a strategy to organize your jobs, models, and versions in a clean and scalable way. Which strategy should you choose?

A. Set up restrictive I AM permissions on the Al Platform notebooks so that only a single user or group can access a given instance.
B. Separate each data scientist's work into a different project to ensure that the jobs, models, and versions created by each data scientist are accessible only to that user.
C. Use labels to organize resources into descriptive categorie
D. Apply a label to each created resource so that users can filter the results by label when viewing or monitoring the resources
E. Set up a BigQuery sink for Cloud Logging logs that is appropriately filtered to capture information about Al Platform resource usage In BigQuery create a SQL view that maps users to the resources they are using.

Answer: B

NEW QUESTION 57
Your organization's call center has asked you to develop a model that analyzes customer sentiments in each call. The call center receives over one million calls daily, and data is stored in Cloud Storage. The data collected must not leave the region in which the call originated, and no Personally Identifiable Information (PII) can be stored or analyzed. The data science team has a third-party tool for visualization and access which requires a SQL ANSI-2011 compliant interface. You need to select components for data processing and for analytics. How should the data pipeline be designed?



A. 1 = Dataflow, 2 = BigQuery
B. 1 = Pub/Sub, 2 = Datastore
C. 1 = Dataflow, 2 = Cloud SQL
D. 1 = Cloud Function, 2 = Cloud SQL

Answer: D

NEW QUESTION 59
.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons and Answers in PDF Format

Professional-Machine-Learning-Engineer Practice Exam Features:

* Professional-Machine-Learning-Engineer Questions and Answers Updated Frequently

* Professional-Machine-Learning-Engineer Practice Questions Verified by Expert Senior Certified Staff

* Professional-Machine-Learning-Engineer Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* Professional-Machine-Learning-Engineer Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year