

➤ **Vendor: Microsoft**

➤ **Exam Code: DP-100**

➤ **Exam Name: Designing and Implementing a Data Science Solution on Azure**

➤ **New Updated Questions from [Braindump2go](#) (Updated in [Oct/2020](#))**

Visit Braindump2go and Download Full Version DP-100 Exam Dumps

QUESTION 191

You use Azure Machine Learning designer to create a real-time service endpoint. You have a single Azure Machine Learning service compute resource. You train the model and prepare the real-time pipeline for deployment. You need to publish the inference pipeline as a web service. Which compute type should you use?

- A. a new Machine Learning Compute resource
- B. Azure Kubernetes Services
- C. HDInsight
- D. the existing Machine Learning Compute resource
- E. Azure Databricks

Answer: B

Explanation:

Azure Kubernetes Service (AKS) can be used real-time inference.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target>

QUESTION 192

You plan to run a script as an experiment using a Script Run Configuration. The script uses modules from the scipy library as well as several Python packages that are not typically installed in a default conda environment. You plan to run the experiment on your local workstation for small datasets and scale out the experiment by running it on more powerful remote compute clusters for larger datasets. You need to ensure that the experiment runs successfully on local and remote compute with the least administrative effort. What should you do?

- A. Do not specify an environment in the run configuration for the experiment. Run the experiment by using the default environment.
- B. Create a virtual machine (VM) with the required Python configuration and attach the VM as a compute target. Use this compute target for all experiment runs.
- C. Create and register an Environment that includes the required packages. Use this Environment for all experiment runs.
- D. Create a config.yaml file defining the conda packages that are required and save the file in the experiment folder.
- E. Always run the experiment with an Estimator by using the default packages.

Answer: C

Explanation:

[DP-100 Exam Dumps](#) [DP-100 Exam Questions](#) [DP-100 PDF Dumps](#) [DP-100 VCE Dumps](#)

<https://www.braindump2go.com/dp-100.html>

If you have an existing Conda environment on your local computer, then you can use the service to create an environment object. By using this strategy, you can reuse your local interactive environment on remote runs.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-environments>

QUESTION 193

You write a Python script that processes data in a comma-separated values (CSV) file.

You plan to run this script as an Azure Machine Learning experiment.

The script loads the data and determines the number of rows it contains using the following code:

```
from azureml.core import Run
import pandas as pd

run = Run.get_context()
data = pd.read_csv('./data.csv')
rows = (len(data))
# record row_count metric here
...
```

You need to record the row count as a metric named row_count that can be returned using the get_metrics method of the Run object after the experiment run completes.

Which code should you use?

- A. run.upload_file('row_count', './data.csv')
- B. run.log('row_count', rows)
- C. run.tag('row_count', rows)
- D. run.log_table('row_count', rows)
- E. run.log_row('row_count', rows)

Answer: B

Explanation:

Log a numerical or string value to the run with the given name using log(name, value, description=").

Logging a metric to a run causes that metric to be stored in the run record in the experiment. You can log the same metric multiple times within a run, the result being considered a vector of that metric.

Example: run.log("accuracy", 0.95)

Incorrect Answers:

E: Using log_row(name, description=None, **kwargs) creates a metric with multiple columns as described in kwargs. Each named parameter generates a column with the value specified. log_row can be called once to log an arbitrary tuple, or multiple times in a loop to generate a complete table.

Example: run.log_row("Y over X", x=1, y=0.4)

Reference:

<https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.core.run>

QUESTION 194

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You create an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:

```
/data/2018/Q1.csv
/data/2018/Q2.csv
/data/2018/Q3.csv
/data/2018/Q4.csv
/data/2019/Q1.csv
```

All files store data in the following format:

```
id, f1, f2, I
```


1,1,2,0
2,1,1,1
3,2,1,0
4,2,2,1

You run the following code:

```
data_store = Datastore.register_azure_blob_container(workspace=ws,  
    datastore_name= 'data_store',  
    container_name= 'quarterly_data',  
    account_name= 'companydata',  
    account_key='NRPxk8duxbM3...'  
    create_if_not_exists=False)
```

You need to create a dataset named training_data and load the data from all files into a single data frame by using the following code:

```
data_frame = training_data.to_pandas_dataframe()
```

Solution: Run the following code:

```
from azureml.core import Dataset  
paths = (data_store, 'data/**/*.csv')  
training_data = Dataset.Tabular.from_delimited_files(paths)
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Define paths with two file paths instead.

Use Dataset.Tabular_from_delimited as the data isn't cleansed.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-register-datasets>

QUESTION 195

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You create an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:

/data/2018/Q1.csv
/data/2018/Q2.csv
/data/2018/Q3.csv
/data/2018/Q4.csv
/data/2019/Q1.csv

All files store data in the following format:

```
id, f1, f2, I  
1,1,2,0  
2,1,1,1  
3,2,1,0  
4,2,2,1
```

You run the following code:


```
data_store = Datastore.register_azure_blob_container(workspace=ws,  
datastore_name= 'data_store',  
container_name= 'quarterly_data',  
account_name= 'companydata',  
account_key='NRPxk8duxbM3...'  
create_if_not_exists=False)
```

You need to create a dataset named training_data and load the data from all files into a single data frame by using the following code:

```
data_frame = training_data.to_pandas_dataframe()
```

Solution: Run the following code:

```
from azureml.core import Dataset  
paths = [(data_store, 'data/2018/*.csv'), (data_store, 'data/2019/*.csv')]  
training_data = Dataset.File.from_files(paths)
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Use two file paths.

Use Dataset.Tabular_from_delimited, instead of Dataset.File.from_files as the data isn't cleansed.

Note:

A FileDataset references single or multiple files in your datastores or public URLs. If your data is already cleansed, and ready to use in training experiments, you can download or mount the files to your compute as a FileDataset object.

A TabularDataset represents data in a tabular format by parsing the provided file or list of files. This provides you with the ability to materialize the data into a pandas or Spark DataFrame so you can work with familiar data preparation and training libraries without having to leave your notebook. You can create a TabularDataset object from .csv, .tsv, .parquet, .jsonl files, and from SQL query results.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-register-datasets>

QUESTION 196

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You create an Azure Machine Learning service datastore in a workspace. The datastore contains the following files:

```
/data/2018/Q1.csv  
/data/2018/Q2.csv  
/data/2018/Q3.csv  
/data/2018/Q4.csv  
/data/2019/Q1.csv
```

All files store data in the following format:

```
id, f1, f2, I  
1, 1, 2, 0  
2, 1, 1, 1  
3, 2, 1, 0  
4, 2, 2, 1
```

You run the following code:


```
data_store = Datastore.register_azure_blob_container(workspace=ws,  
datastore_name= 'data_store',  
container_name= 'quarterly_data',  
account_name= 'companydata',  
account_key='NRPxk8duxbM3...'  
create_if_not_exists=False)
```

You need to create a dataset named training_data and load the data from all files into a single data frame by using the following code:

```
data_frame = training_data.to_pandas_dataframe()
```

Solution: Run the following code:

```
from azureml.core import Dataset  
paths = [(data_store, 'data/2018/*.csv'), (data_store, 'data/2019/*.csv')]  
training_data = Dataset.Tabular.from_delimited_files(paths)
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

Use two file paths.

Use Dataset.Tabular.from_delimited_files as the data isn't cleansed.

Note:

A TabularDataset represents data in a tabular format by parsing the provided file or list of files. This provides you with the ability to materialize the data into a pandas or Spark DataFrame so you can work with familiar data preparation and training libraries without having to leave your notebook. You can create a TabularDataset object from .csv, .tsv, .parquet, .jsonl files, and from SQL query results.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-register-datasets>

QUESTION 197

You plan to use the Hyperdrive feature of Azure Machine Learning to determine the optimal hyperparameter values when training a model.

You must use Hyperdrive to try combinations of the following hyperparameter values:

- learning_rate: any value between 0.001 and 0.1
- batch_size: 16, 32, or 64

You need to configure the search space for the Hyperdrive experiment.

Which two parameter expressions should you use? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. a choice expression for learning_rate
- B. a uniform expression for learning_rate
- C. a normal expression for batch_size
- D. a choice expression for batch_size
- E. a uniform expression for batch_size

Answer: BD

Explanation:

B: Continuous hyperparameters are specified as a distribution over a continuous range of values.

Supported distributions include:

uniform(low, high) - Returns a value uniformly distributed between low and high

D: Discrete hyperparameters are specified as a choice among discrete values. choice can be:

one or more comma-separated values

a range object

any arbitrary list object

[DP-100 Exam Dumps](#) [DP-100 Exam Questions](#) [DP-100 PDF Dumps](#) [DP-100 VCE Dumps](#)

<https://www.braindump2go.com/dp-100.html>

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters>

QUESTION 198

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You train a classification model by using a logistic regression algorithm.

You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions.

You need to create an explainer that you can use to retrieve the required global and local feature importance values.

Solution: Create a MimicExplainer.

Does the solution meet the goal?

A. Yes

B. No

Answer: B

Explanation:

Instead use Permutation Feature Importance Explainer (PFI).

Note 1: Mimic explainer is based on the idea of training global surrogate models to mimic blackbox models.

A global surrogate model is an intrinsically interpretable model that is trained to approximate the predictions of any black box model as accurately as possible. Data scientists can interpret the surrogate model to draw conclusions about the black box model.

Note 2: Permutation Feature Importance Explainer (PFI): Permutation Feature Importance is a technique used to explain classification and regression models. At a high level, the way it works is by randomly shuffling data one feature at a time for the entire dataset and calculating how much the performance metric of interest changes. The larger the change, the more important that feature is. PFI can explain the overall behavior of any underlying model but does not explain individual predictions.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability>

QUESTION 199

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You train a classification model by using a logistic regression algorithm.

You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions.

You need to create an explainer that you can use to retrieve the required global and local feature importance values.

Solution: Create a TabularExplainer.

Does the solution meet the goal?

A. Yes

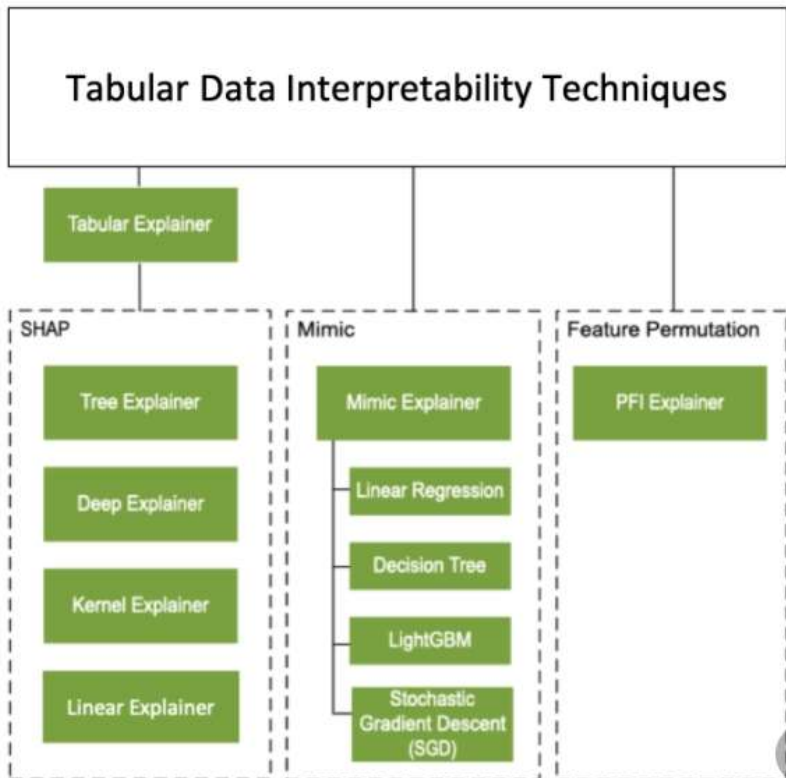
B. No

Answer: B

Explanation:

Instead use Permutation Feature Importance Explainer (PFI).

Note 1:



Note 2: Permutation Feature Importance Explainer (PFI): Permutation Feature Importance is a technique used to explain classification and regression models. At a high level, the way it works is by randomly shuffling data one feature at a time for the entire dataset and calculating how much the performance metric of interest changes. The larger the change, the more important that feature is. PFI can explain the overall behavior of any underlying model but does not explain individual predictions.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability>

QUESTION 200

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You train a classification model by using a logistic regression algorithm.

You must be able to explain the model's predictions by calculating the importance of each feature, both as an overall global relative importance value and as a measure of local importance for a specific set of predictions.

You need to create an explainer that you can use to retrieve the required global and local feature importance values.

Solution: Create a PFIExplainer.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

Permutation Feature Importance Explainer (PFI): Permutation Feature Importance is a technique used to explain classification and regression models. At a high level, the way it works is by randomly shuffling data one feature at a time for the entire dataset and calculating how much the performance metric of interest changes. The larger the change, the more important that feature is. PFI can explain the overall behavior of any underlying model but does not explain individual predictions.

Reference:

[DP-100 Exam Dumps](#) [DP-100 Exam Questions](#) [DP-100 PDF Dumps](#) [DP-100 VCE Dumps](#)

<https://www.braindump2go.com/dp-100.html>

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability>

QUESTION 201

You create a machine learning model by using the Azure Machine Learning designer. You publish the model as a real-time service on an Azure Kubernetes Service (AKS) inference compute cluster. You make no change to the deployed endpoint configuration.

You need to provide application developers with the information they need to consume the endpoint.

Which two values should you provide to application developers? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. The name of the AKS cluster where the endpoint is hosted.
- B. The name of the inference pipeline for the endpoint.
- C. The URL of the endpoint.
- D. The run ID of the inference pipeline experiment for the endpoint.
- E. The key for the endpoint.

Answer: CE

Explanation:

Deploying an Azure Machine Learning model as a web service creates a REST API endpoint. You can send data to this endpoint and receive the prediction returned by the model.

You create a web service when you deploy a model to your local environment, Azure Container Instances, Azure Kubernetes Service, or field-programmable gate arrays (FPGA). You retrieve the URI used to access the web service by using the Azure Machine Learning SDK. If authentication is enabled, you can also use the SDK to get the authentication keys or tokens.

Example:

```
# URL for the web service
```

```
scoring_uri = '<your web service URI>'
```

```
# If the service is authenticated, set the key or token
```

```
key = '<your key or token>'
```

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-consume-web-service>

QUESTION 202

You use Azure Machine Learning designer to create a training pipeline for a regression model.

You need to prepare the pipeline for deployment as an endpoint that generates predictions asynchronously for a dataset of input data values.

What should you do?

- A. Clone the training pipeline.
- B. Create a batch inference pipeline from the training pipeline.
- C. Create a real-time inference pipeline from the training pipeline.
- D. Replace the dataset in the training pipeline with an Enter Data Manually module.

Answer: C

Explanation:

You must first convert the training pipeline into a real-time inference pipeline. This process removes training modules and adds web service inputs and outputs to handle requests.

Incorrect Answers:

A: Use the Enter Data Manually module to create a small dataset by typing values.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-deploy>

<https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/enter-data-manually>

QUESTION 203

You retrain an existing model.

You need to register the new version of a model while keeping the current version of the model in the registry.

What should you do?

[DP-100 Exam Dumps](#) **[DP-100 Exam Questions](#) **[DP-100 PDF Dumps](#) **[DP-100 VCE Dumps](#)******

<https://www.braindump2go.com/dp-100.html>

- A. Register a model with a different name from the existing model and a custom property named version with the value 2.
- B. Register the model with the same name as the existing model.
- C. Save the new model in the default datastore with the same name as the existing model. Do not register the new model.
- D. Delete the existing model and register the new one with the same name.

Answer: B

Explanation:

Model version: A version of a registered model. When a new model is added to the Model Registry, it is added as Version 1. Each model registered to the same model name increments the version number.

Reference:

<https://docs.microsoft.com/en-us/azure/databricks/applications/mlflow/model-registry>

QUESTION 204

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You create a model to forecast weather conditions based on historical data.

You need to create a pipeline that runs a processing script to load data from a datastore and pass the processed data to a machine learning model training script.

Solution: Run the following code:

```
data_store = Datastore.get(ws, "ml-data")
data_input = DataReference(
    datastore = data_store,
    data_reference_name = "training_data",
    path_on_datastore = "train/data.txt")
data_output = PipelineData("processed_data", datastore=datastore)
process_step = PythonScriptStep(script_name="process.py",
    arguments=["- -data", data_input], outputs=[data_output],
    compute_target=aml_compute, source_directory=process_directory)
train_step = PythonScriptStep(script_name="train.py",
    arguments=["- -data", data_output], inputs=[data_output],
    compute_target=aml_compute, source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps = [process_step, train_step])
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

The two steps are present: process_step and train_step
Data_input correctly references the data in the data store.

Note:

Data used in pipeline can be produced by one step and consumed in another step by providing a PipelineData object as an output of one step and an input of one or more subsequent steps.

PipelineData objects are also used when constructing Pipelines to describe step dependencies. To specify that a step requires the output of another step as input, use a PipelineData object in the constructor of both steps.

For example, the pipeline train step depends on the process_step_output output of the pipeline process step:

[DP-100 Exam Dumps](#) [DP-100 Exam Questions](#) [DP-100 PDF Dumps](#) [DP-100 VCE Dumps](#)

<https://www.braindump2go.com/dp-100.html>


```
from azureml.pipeline.core import Pipeline, PipelineData
from azureml.pipeline.steps import PythonScriptStep
datastore = ws.get_default_datastore()
process_step_output = PipelineData("processed_data", datastore=datastore) process_step =
PythonScriptStep(script_name="process.py", arguments=["--data_for_train", process_step_output],
outputs=[process_step_output],
compute_target=aml_compute,
source_directory=process_directory)
train_step = PythonScriptStep(script_name="train.py",
arguments=["--data_for_train", process_step_output],
inputs=[process_step_output],
compute_target=aml_compute,
source_directory=train_directory)
pipeline = Pipeline(workspace=ws, steps=[process_step, train_step]) Reference:
https://docs.microsoft.com/en-us/python/api/azureml-pipeline-core/azureml.pipeline.core.pipelinedata?view=azure-ml-py
```

QUESTION 205

You run an experiment that uses an AutoMLConfig class to define an automated machine learning task with a maximum of ten model training iterations. The task will attempt to find the best performing model based on a metric named accuracy.

You submit the experiment with the following code:

```
from azureml.core.experiment import Experiment
automl_experiment = Experiment(ws, 'automl_experiment')
automl_run = automl_experiment.submit(automl_config, show_output=True)
```

You need to create Python code that returns the best model that is generated by the automated machine learning task. Which code segment should you use?

- A. `best_model = automl_run.get_details()`
- B. `best_model = automl_run.get_metrics()`
- C. `best_model = automl_run.get_file_names()[1]`
- D. `best_model = automl_run.get_output()[1]`

Answer: D

Explanation:

The `get_output` method returns the best run and the fitted model.

Reference:

<https://notebooks.azure.com/azureml/projects/azureml-getting-started/html/how-to-use-azureml/automated-machine-learning/classification/auto-ml-classification.ipynb>