DP-100 Q&As

Designing and Implementing a Data Science Solution on Azure

Pass Microsoft DP-100 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.leads4pass.com/dp-100.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft Official Exam Center

QUESTION 1

HOTSPOT

You have an Azure blob container that contains a set of TSV files. The Azure blob container is registered as a datastore for an Azure Machine Learning service workspace. Each TSV file uses the same data schema.

You plan to aggregate data for all of the TSV files together and then register the aggregated data as a dataset in an Azure Machine Learning workspace by using the Azure Machine Learning SDK for Python.

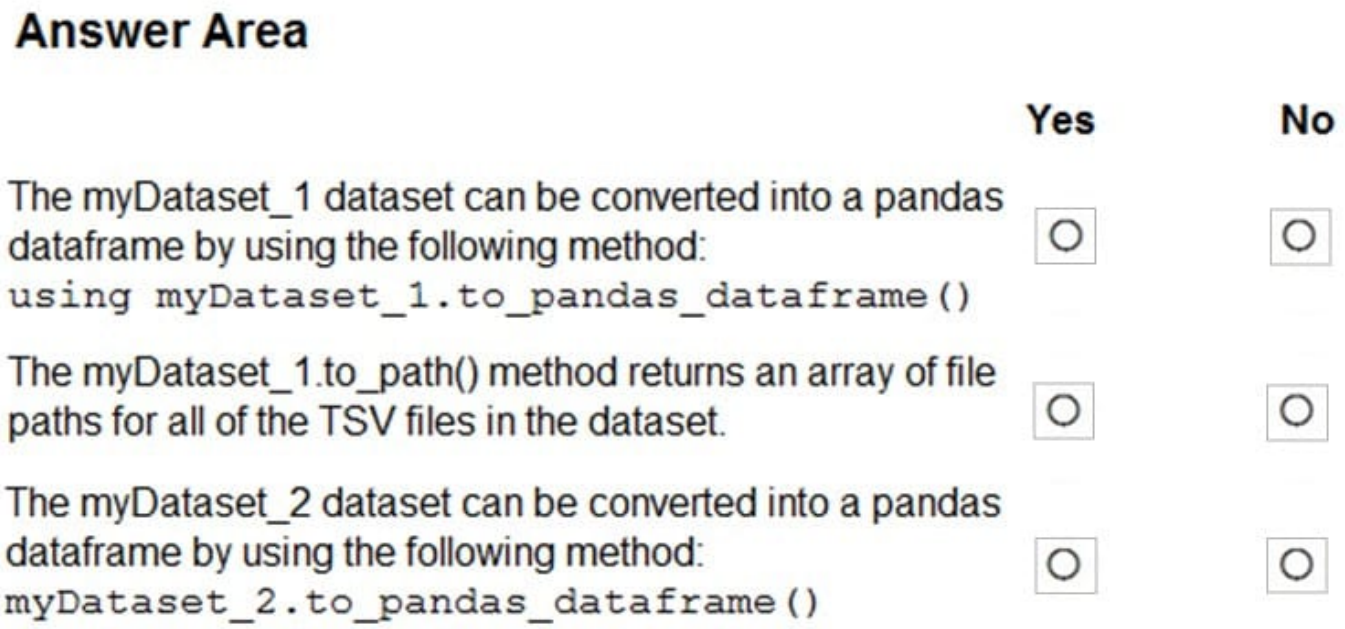
You run the following code.

```
from azureml.core.workspace import Workspace
from azureml.core.datastore import Datastore
from azureml.core.dataset import Dataset
import pandas as pd
datastore_paths = (datastore, './data/*.tsv')
myDataset_1 = Dataset.File.from_files(path=datastore_paths)
myDataset_2 = Dataset.Tabular.from_delimited_files(path=datastore_paths, separator='\t')
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:

Answer Area

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: `using myDataset_1.to_pandas_dataframe()`	○	○
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.	○	○
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: `myDataset_2.to_pandas_dataframe()`	○	○

Box 1: No

FileDataset references single or multiple files in datastores or from public URLs. The TSV files need to be parsed.

Box 2: Yes

to_path() gets a list of file paths for each file stream defined by the dataset.

Box 3: Yes

TabularDataset.to_pandas_dataframe loads all records from the dataset into a pandas DataFrame.

TabularDataset represents data in a tabular format created by parsing the provided file or list of files.

Note: TSV is a file extension for a tab-delimited file used with spreadsheet software. TSV stands for Tab Separated Values. TSV files are used for raw data and can be imported into and exported from spreadsheet software. TSV files are

essentially text files, and the raw data can be viewed by text editors, though they are often used when moving raw data between spreadsheets.

Reference:

https://docs.microsoft.com/en-us/python/api/azureml-core/azureml.data.tabulardataset

QUESTION 2

You have been tasked with employing a machine learning model, which makes use of a PostgreSQL database and needs GPU processing, to forecast prices.

You are preparing to create a virtual machine that has the necessary tools built into it.

You need to make use of the correct virtual machine type.

Recommendation: You make use of a Data Science Virtual Machine (DSVM) Windows edition.

Will the requirements be satisfied?

A. Yes

B. No

Correct Answer: A

In the DSVM, your training models can use deep learning algorithms on hardware that\\'s based on graphics processing units (GPUs).

PostgreSQL is available for the following operating systems: Linux (all recent distributions), 64-bit installers available for macOS (OS X) version 10.6 and newer – Windows (with installers available for 64-bit version; tested on latest versions and back to Windows 2012 R2.

Reference: https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview

QUESTION 3

You manage an Azure Machine Learning workspace. The workspace includes an Azure Machine Learning Kubernetes compute target configured as an Azure Kubernetes Service (AKS) cluster named AKS1. AKS1 is configured to enable the

targeting of different nodes to train workloads.

You must run a command job on AKS1 by using the Azure ML Python SDK v2. The command job must select different types of compute nodes. The compute node types must be specified by using a command parameter.

You need to configure the command parameter.

Which parameter should you use?

A. environment

B. compute

C. limits

D. instance_type

Correct Answer: D

command

Create a Command object which can be used inside dsl.pipeline as a function and can also be created as a standalone command job.

Parameters include:

instance_type

Optional type of VM used as supported by the compute target.

Incorrect:

*

environment

the environment to use for this command

*

compute

the name of the compute where the command job is executed(will not be used if the command is used as a component/function)

Reference:

https://learn.microsoft.com/en-us/python/api/azure-ai-ml/azure.ai.m

QUESTION 4

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.

Correct Answer: CE

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 = \\'\\'

If the service is authenticated, set the key or token

key = \\'\\'

Reference:

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

QUESTION 5

You need to visually identify whether outliers exist in the Age column and quantify the outliers before the outliers are removed. Which three Azure Machine Learning Studio modules should you use? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. Create Scatterplot

B. Summarize Data

C. Clip Values

D. Replace Discrete Values

E. Build Counting Transform

Correct Answer: ABC

B: To have a global view, the summarize data module can be used. Add the module and connect it to the data set that needs to be visualized.

A: One way to quickly identify Outliers visually is to create scatter plots.

C: The easiest way to treat the outliers in Azure ML is to use the Clip Values module. It can identify and optionally replace data values that are above or below a specified threshold.

You can use the Clip Values module in Azure Machine Learning Studio, to identify and optionally replace data values that are above or below a specified threshold. This is useful when you want to remove outliers or replace them with a mean, a constant, or other substitute value.

References: https://blogs.msdn.microsoft.com/azuredev/2017/05/27/data-cleansing-tools-in-azure-machine-learning/

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/clip-values Question Set 3

QUESTION 6

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.

Correct Answer: C

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 7

You train and register an Azure Machine Learning model.

You plan to deploy the model to an online endpoint.

You need to ensure that applications will be able to use the authentication method with a non-expiring artifact to access the model.

Solution: Create a Kubernetes online endpoint and set the value of its auth_mode parameter to aml_token. Deploy the model to the online endpoint.

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B

Correct Solution: Create a managed online endpoint and set the value of its auth_mode parameter to key. Deploy the model to the online endpoint.

Authentication mode: The authentication method for the endpoint. Choose between key-based authentication and Azure Machine Learning token-based authentication. A key doesn\\'t expire, but a token does expire.

Reference:

https://learn.microsoft.com/en-us/azure/machine-learning/how-to-deploy-online-endpoints

---

QUESTION 8

You need to select an environment that will meet the business and data requirements. Which environment should you use?

A. Azure HDInsight with Spark MLlib

B. Azure Cognitive Services

C. Azure Machine Learning Studio

D. Microsoft Machine Learning Server

Correct Answer: D

QUESTION 9

You are in the process of constructing a deep convolutional neural network (CNN). The CNN will be used for image classification.

You notice that the CNN model you constructed displays hints of overfitting.

You want to make sure that overfitting is minimized, and that the model is converged to an optimal fit.

Which of the following is TRUE with regards to achieving your goal?

A. You have to add an additional dense layer with 512 input units, and reduce the amount of training data.

B. You have to add L1/L2 regularization, and reduce the amount of training data.

C. You have to reduce the amount of training data and make use of training data augmentation.

D. You have to add L1/L2 regularization, and make use of training data augmentation.

E. You have to add an additional dense layer with 512 input units, and add L1/L2 regularization.

Correct Answer: B

B: Weight regularization provides an approach to reduce the overfitting of a deep learning neural network model on the training data and improve the performance of the model on new data, such as the holdout test set. Keras provides a weight regularization API that allows you to add a penalty for weight size to the loss function.

Three different regularizer instances are provided; they are:

L1: Sum of the absolute weights.

L2: Sum of the squared weights.

L1L2: Sum of the absolute and the squared weights.

Because a fully connected layer occupies most of the parameters, it is prone to overfitting. One method to reduce overfitting is dropout. At each training stage, individual nodes are either "dropped out" of the net with probability 1-p or kept with

probability p, so that a reduced network is left; incoming and outgoing edges to a dropped-out node are also removed.

By avoiding training all nodes on all training data, dropout decreases overfitting.

Reference:

https://machinelearningmastery.com/how-to-reduce-overfitting-in-deep-learning-with-weight-regularization/

https://en.wikipedia.org/wiki/Convolutional_neural_network

QUESTION 10

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 have a Python script named train.py in a local folder named scripts. The script trains a regression model by using scikit-learn. The script includes code to load a training data file which is also located in the scripts folder.

You must run the script as an Azure ML experiment on a compute cluster named aml-compute.

You need to configure the run to ensure that the environment includes the required packages for model training. You have instantiated a variable named aml-compute that references the target compute cluster.

Solution: Run the following code:

```
from azureml.train.dnn import TensorFlow
sk_est = TensorFlow(source_directory='./scripts',
 compute_target=aml-compute,
 entry_script='train.py')
```

Does the solution meet the goal?

A. Yes

B. No

Correct Answer: B

The scikit-learn estimator provides a simple way of launching a scikit-learn training job on a compute target. It is implemented through the SKLearn class, which can be used to support single-node CPU training.

Example:

from azureml.train.sklearn import SKLearn

}

estimator = SKLearn(source_directory=project_folder,

compute_target=compute_target,

entry_script=\\'train_iris.py\\'

)

Reference:

https://docs.microsoft.com/en-us/azure/machine-learning/how-to-train-scikit-learn

QUESTION 11

You create a new Azure subscription. No resources are provisioned in the subscription.

You need to create an Azure Machine Learning workspace.

What are three possible ways to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. Run Python code that uses the Azure ML SDK library and calls the Workspace.create method with name, subscription_id, resource_group, and location parameters.

B. Use an Azure Resource Management template that includes a Microsoft.MachineLearningServices/ workspaces resource and its dependencies.

C. Use the Azure Command Line Interface (CLI) with the Azure Machine Learning extension to call the az group create function with --name and --location parameters, and then the az ml workspace create function, specifying ? and ? parameters for the workspace name and resource group.

D. Navigate to Azure Machine Learning studio and create a workspace.

E. Run Python code that uses the Azure ML SDK library and calls the Workspace.get method with name, subscription_id, and resource_group parameters.

Correct Answer: BCD

B: You can use an Azure Resource Manager template to create a workspace for Azure Machine Learning.

Example:

{"type": "Microsoft.MachineLearningServices/workspaces", ...

C: You can create a workspace for Azure Machine Learning with Azure CLI Install the machine learning extension.

Create a resource group: az group create --name --location

To create a new workspace where the services are automatically created, use the following command: az ml workspace create -w -g

D: You can create and manage Azure Machine Learning workspaces in the Azure portal.

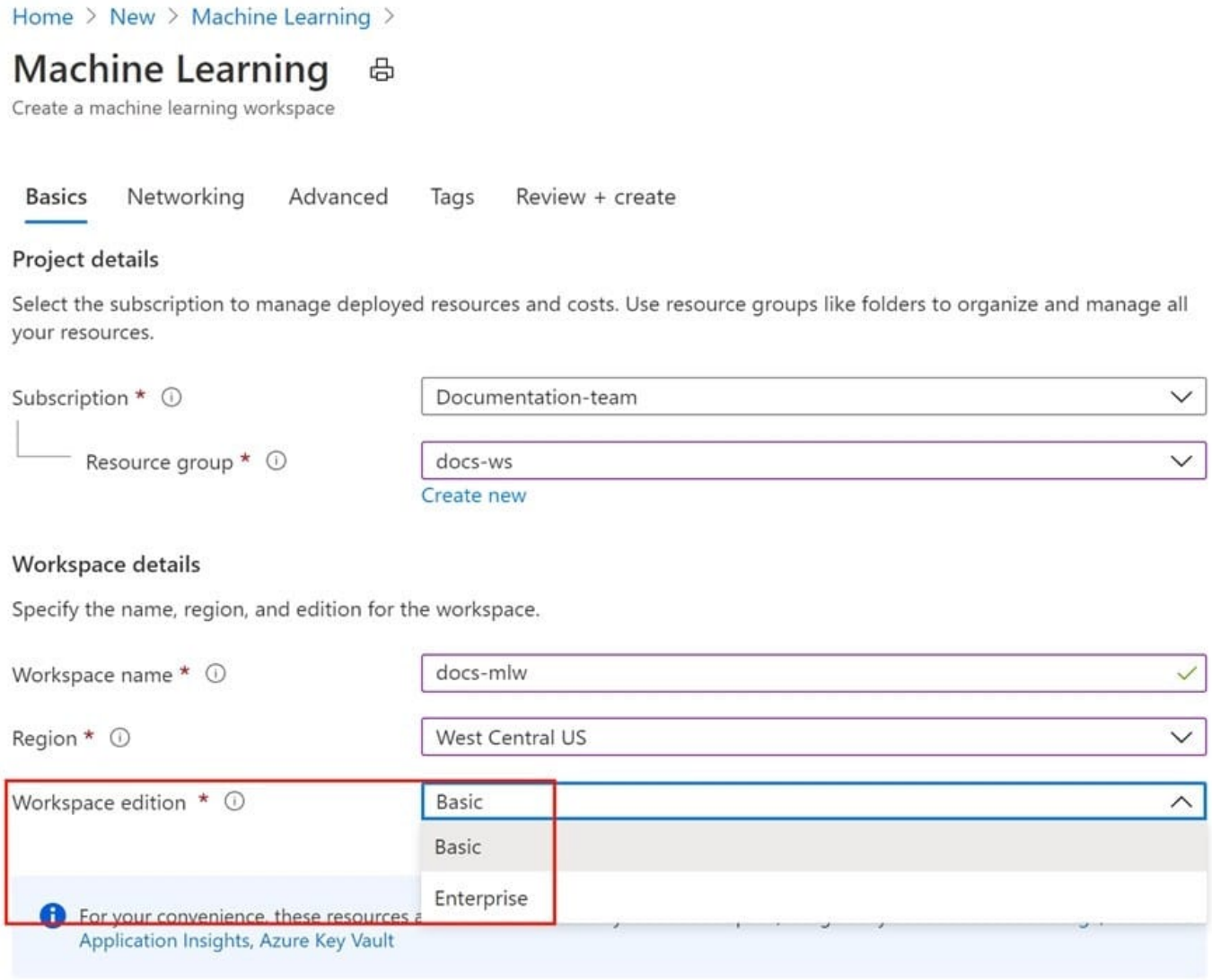
1. Sign in to the Azure portal by using the credentials for your Azure subscription.

2. In the upper-left corner of Azure portal, select + Create a resource.

3. Use the search bar to find Machine Learning.

4. Select Machine Learning.

5. In the Machine Learning pane, select Create to begin.



Reference: https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-workspace-template https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace-cli https://docs.microsoft.com/en-us/azure/machine-learning/how-to-manage-workspace

QUESTION 12

HOTSPOT

You are using the Azure Machine Learning Service to automate hyperparameter exploration of your neural network classification model.

You must define the hyperparameter space to automatically tune hyperparameters using random sampling according to following requirements:

1.

The learning rate must be selected from a normal distribution with a mean value of 10 and a standard deviation of 3.

2.

Batch size must be 16, 32 and 64.

3.

Keep probability must be a value selected from a uniform distribution between the range of 0.05 and 0.1.

You need to use the param_sampling method of the Python API for the Azure Machine Learning Service.

How should you complete the code segment? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

```
from azureml.train.hyperdrive import RandomParameterSampling
param_sampling = RandomParameterSampling( {
"learning_rate" :  [                ▼ ]  ,
                     uniform(10,3)
                     normal(10,3)
                     choice(10,3)
                     Loguniform(10,3)
"batch_size":      [                ▼ ]  ,
                     choice(16,32,64)
                     choice(range(16,64))
                     normal(16,32,64)
                     normal(range(16,64))
"keep_probability" : [                ▼ ]
                     choice(range(0.05, 0.1))
                     uniform(0.05, 0.1)
                     normal(0.05, 0.1)
                     lognormal(0.05, 0.1)
    }
)
```

Correct Answer:

Answer Area

```
from azureml.train.hyperdrive import RandomParameterSampling
param_sampling = RandomParameterSampling( {
"learning_rate" :  [uniform(10,3) | normal(10,3) | choice(10,3) | Loguniform(10,3)] ,
"batch_size":  [choice(16,32,64) | choice(range(16,64)) | normal(16,32,64) | normal(range(16,64))] ,
"keep_probability" :  [choice(range(0.05, 0.1)) | uniform(0.05, 0.1) | normal(0.05, 0.1) | lognormal(0.05, 0.1)]
    }
)
```

In random sampling, hyperparameter values are randomly selected from the defined search space. Random sampling allows the search space to include both discrete and continuous hyperparameters.

Example:

from azureml.train.hyperdrive import RandomParameterSampling

param_sampling = RandomParameterSampling({ "learning_rate": normal(10, 3), "keep_probability": uniform(0.05, 0.1), "batch_size": choice(16, 32, 64) }

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

---

QUESTION 13

You use the Two-Class Neural Network module in Azure Machine Learning Studio to build a binary classification model. You use the Tune Model Hyperparameters module to tune accuracy for the model.

You need to select the hyperparameters that should be tuned using the Tune Model Hyperparameters module.

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

NOTE: Each correct selection is worth one point.

A. Number of hidden nodes

B. Learning Rate

C. The type of the normalizer

D. Number of learning iterations

E. Hidden layer specification

Correct Answer: DE

D: For Number of learning iterations, specify the maximum number of times the algorithm should process the training cases.

E: For Hidden layer specification, select the type of network architecture to create. Between the input and output layers you can insert multiple hidden layers. Most predictive tasks can be accomplished easily with only one or a few hidden

layers.

References:

https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two- class-neural- network

QUESTION 14

DRAG DROP

An organization uses Azure Machine Learning service and wants to expand their use of machine learning.

You have the following compute environments. The organization does not want to create another compute environment.

Environment name	Compute type
nb_server	Compute Instance
aks_cluster	Azure Kubernetes Service
mlc_cluster	Machine Learning Compute

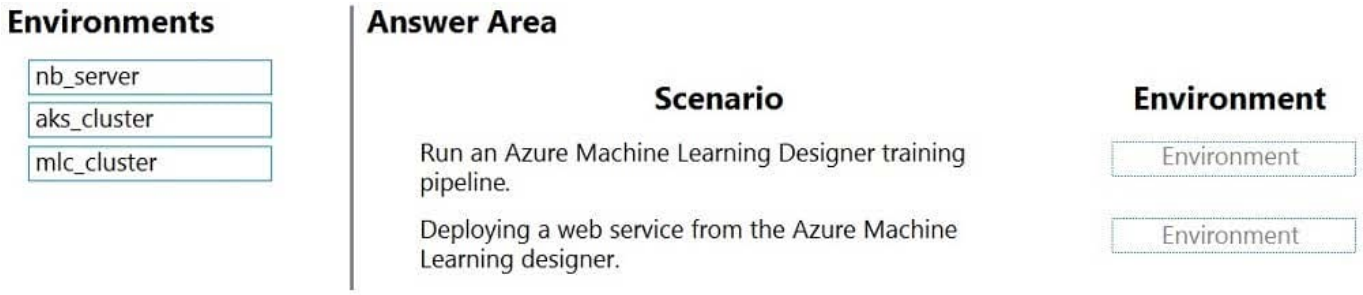
You need to determine which compute environment to use for the following scenarios.

Which compute types should you use? To answer, drag the appropriate compute environments to the correct scenarios. Each compute environment may be used once, more than once, or not at all. You may need to drag the split bar between

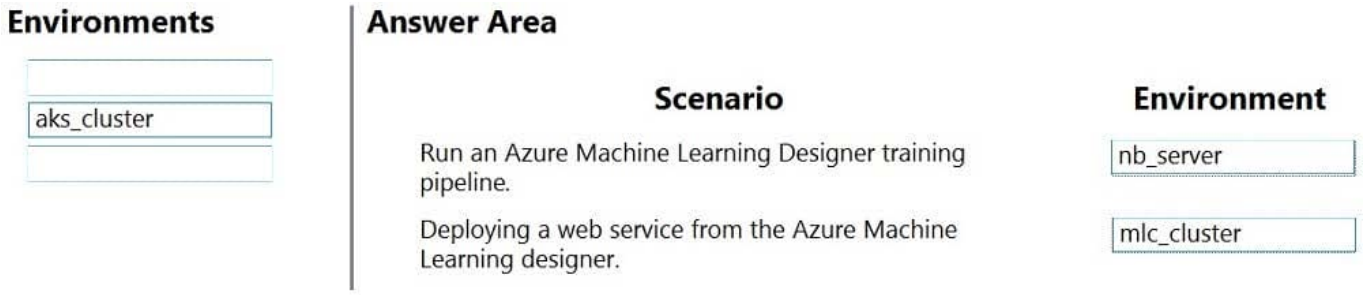
panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Select and Place:



Correct Answer:



Box 1: nb_server

Training targets	Automated ML	ML pipelines	Azure Machine Learning designer
Local computer	yes		
Azure Machine Learning compute cluster	yes & hyperparameter tuning	yes	yes
Azure Machine Learning compute instance	yes & hyperparameter tuning	yes	yes
Remote VM	yes & hyperparameter tuning	yes	
Azure Databricks	yes (SDK local mode only)	yes	
Azure Data Lake Analytics		yes	
Azure HDInsight		yes	
Azure Batch		yes	

Box 2: mlc_cluster With Azure Machine Learning, you can train your model on a variety of resources or environments, collectively referred to as compute targets. A compute target can be a local machine or a cloud resource, such as an Azure Machine Learning Compute, Azure HDInsight or a remote virtual machine.

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

https://docs.microsoft.com/en-us/azure/machine-learning/how-to-set-up-training-targets

QUESTION 15

HOTSPOT

You are running Python code interactively in a Conda environment. The environment includes all required Azure Machine Learning SDK and MLflow packages.

You must use MLflow to log metrics in an Azure Machine Learning experiment named mlflow-experiment.

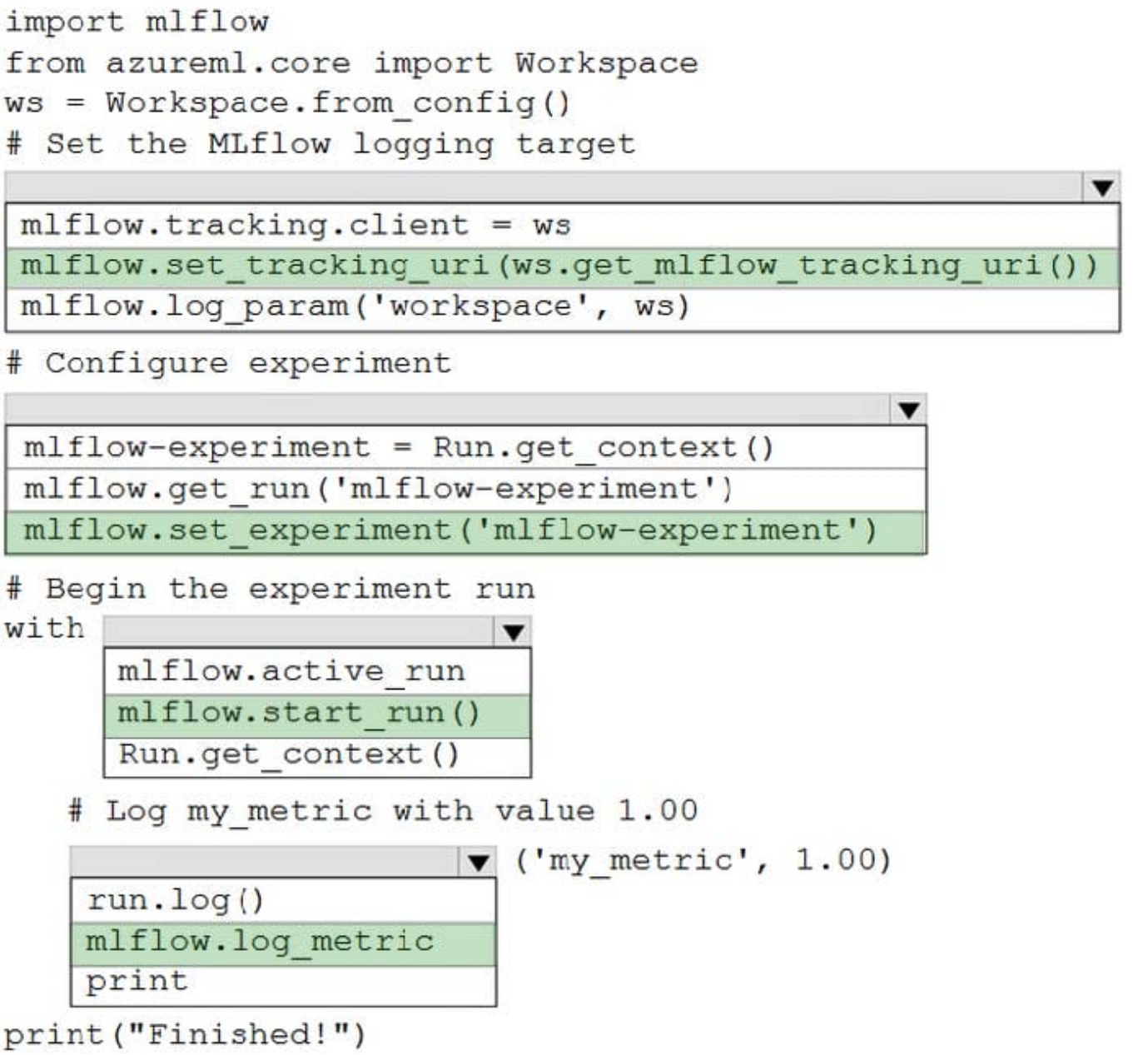
How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

```
import mlflow
from azureml.core import Workspace
ws = Workspace.from_config()
# Set the MLflow logging target
[ mlflow.tracking.client = ws | mlflow.set_tracking_uri(ws.get_mlflow_tracking_uri()) | mlflow.log_param('workspace', ws) ]
# Configure experiment
[ mlflow-experiment = Run.get_context() | mlflow.get_run('mlflow-experiment') | mlflow.set_experiment('mlflow-experiment') ]
# Begin the experiment run
with [ mlflow.active_run | mlflow.start_run() | Run.get_context() ]
    # Log my_metric with value 1.00
    [ run.log() | mlflow.log_metric | print ]('my_metric', 1.00)
print("Finished!")
```

Correct Answer:



Box 1: mlflow.set_tracking_uri(ws.get_mlflow_tracking_uri())

In the following code, the get_mlflow_tracking_uri() method assigns a unique tracking URI address to the workspace, ws, and set_tracking_uri() points the MLflow tracking URI to that address.

mlflow.set_tracking_uri(ws.get_mlflow_tracking_uri())

Box 2: mlflow.set_experiment(experiment_name)

Set the MLflow experiment name with set_experiment() and start your training run with start_run().

Box 3: mlflow.start_run()

Box 4: mlflow.log_metric

Then use log_metric() to activate the MLflow logging API and begin logging your training run metrics.

Reference:

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