

- **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 206

You use the Azure Machine Learning SDK to run a training experiment that trains a classification model and calculates its accuracy metric.

The model will be retrained each month as new data is available.

You must register the model for use in a batch inference pipeline.

You need to register the model and ensure that the models created by subsequent retraining experiments are registered only if their accuracy is higher than the currently registered model.

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

NOTE: Each correct selection is worth one point.

- A. Specify a different name for the model each time you register it.
- B. Register the model with the same name each time regardless of accuracy, and always use the latest version of the model in the batch inferencing pipeline.
- C. Specify the model framework version when registering the model, and only register subsequent models if this value is higher.
- D. Specify a property named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy property value of the currently registered model.
- E. Specify a tag named accuracy with the accuracy metric as a value when registering the model, and only register subsequent models if their accuracy is higher than the accuracy tag value of the currently registered model.

Answer: CE

Explanation:

E: Using tags, you can track useful information such as the name and version of the machine learning library used to train the model. Note that tags must be alphanumeric.

Reference:

<https://notebooks.azure.com/xavierheriat/projects/azureml-getting-started/html/how-to-use-azureml/deployment/register-model-create-image-deploy-service/register-model-create-image-deploy-service.ipynb>

QUESTION 207

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. You must not apply an early termination policy.

- learning_rate: any value between 0.001 and 0.1

- batch_size: 16, 32, or 64

You need to configure the sampling method for the Hyperdrive experiment.

Which two sampling methods can you use? Each correct answer is a complete solution.

NOTE: Each correct selection is worth one point.

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

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

- A. No sampling
- B. Grid sampling
- C. Bayesian sampling
- D. Random sampling

Answer: CD

Explanation:

C: Bayesian sampling is based on the Bayesian optimization algorithm and makes intelligent choices on the hyperparameter values to sample next. It picks the sample based on how the previous samples performed, such that the new sample improves the reported primary metric.

Bayesian sampling does not support any early termination policy Example:

```
from azureml.train.hyperdrive import BayesianParameterSampling from azureml.train.hyperdrive import uniform, choice
param_sampling = BayesianParameterSampling( {
"learning_rate": uniform(0.05, 0.1),
"batch_size": choice(16, 32, 64, 128)
}
)
```

D: 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.

Incorrect Answers:

B: Grid sampling can be used if your hyperparameter space can be defined as a choice among discrete values and if you have sufficient budget to exhaustively search over all values in the defined search space.

Additionally, one can use automated early termination of poorly performing runs, which reduces wastage of resources.

Example, the following space has a total of six samples:

```
from azureml.train.hyperdrive import GridParameterSampling from azureml.train.hyperdrive import choice
param_sampling = GridParameterSampling( {
"num_hidden_layers": choice(1, 2, 3),
"batch_size": choice(16, 32)
}
)
```

Reference:

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

QUESTION 208

You are training machine learning models in Azure Machine Learning. You use Hyperdrive to tune the hyperparameter. In previous model training and tuning runs, many models showed similar performance.

You need to select an early termination policy that meets the following requirements:

- accounts for the performance of all previous runs when evaluating the current run
- avoids comparing the current run with only the best performing run to date

Which two early termination policies should you use? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Median stopping
- B. Bandit
- C. Default
- D. Truncation selection

Answer: AC

Explanation:

The Median Stopping policy computes running averages across all runs and cancels runs whose best performance is worse than the median of the running averages.

If no policy is specified, the hyperparameter tuning service will let all training runs execute to completion.

Incorrect Answers:

B: BanditPolicy defines an early termination policy based on slack criteria, and a frequency and delay interval for evaluation.

The Bandit policy takes the following configuration parameters:

slack_factor: The amount of slack allowed with respect to the best performing training run. This factor specifies the

slack as a ratio.

D: The Truncation selection policy periodically cancels the given percentage of runs that rank the lowest for their performance on the primary metric. The policy strives for fairness in ranking the runs by accounting for improving model performance with training time. When ranking a relatively young run, the policy uses the corresponding (and earlier) performance of older runs for comparison. Therefore, runs aren't terminated for having a lower performance because they have run for less time than other runs.

Reference:

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

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

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

QUESTION 209

You are a data scientist working for a hotel booking website company. You use the Azure Machine Learning service to train a model that identifies fraudulent transactions.

You must deploy the model as an Azure Machine Learning real-time web service using the `Model.deploy` method in the Azure Machine Learning SDK. The deployed web service must return real-time predictions of fraud based on transaction data input.

You need to create the script that is specified as the `entry_script` parameter for the `InferenceConfig` class used to deploy the model.

What should the entry script do?

- A. Register the model with appropriate tags and properties.
- B. Create a Conda environment for the web service compute and install the necessary Python packages.
- C. Load the model and use it to predict labels from input data.
- D. Start a node on the inference cluster where the web service is deployed.
- E. Specify the number of cores and the amount of memory required for the inference compute.

Answer: C

Explanation:

The entry script receives data submitted to a deployed web service and passes it to the model. It then takes the response returned by the model and returns that to the client. The script is specific to your model.

It must understand the data that the model expects and returns.

The two things you need to accomplish in your entry script are:

Loading your model (using a function called `init()`)

Running your model on input data (using a function called `run()`)

Reference: <https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-and-where>

QUESTION 210

You develop and train a machine learning model to predict fraudulent transactions for a hotel booking website.

Traffic to the site varies considerably. The site experiences heavy traffic on Monday and Friday and much lower traffic on other days. Holidays are also high web traffic days.

You need to deploy the model as an Azure Machine Learning real-time web service endpoint on compute that can dynamically scale up and down to support demand.

Which deployment compute option should you use?

- A. attached Azure Databricks cluster
- B. Azure Container Instance (ACI)
- C. Azure Kubernetes Service (AKS) inference cluster
- D. Azure Machine Learning Compute Instance
- E. attached virtual machine in a different region

Answer: D

Explanation:

Azure Machine Learning compute cluster is a managed-compute infrastructure that allows you to easily create a single or multi-node compute. The compute is created within your workspace region as a resource that can be shared with other users in your workspace. The compute scales up automatically when a job is submitted, and can be put in an

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

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

Azure Virtual Network.

Reference:

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-sdk>

QUESTION 211

You use the Azure Machine Learning SDK in a notebook to run an experiment using a script file in an experiment folder.

The experiment fails.

You need to troubleshoot the failed experiment.

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

- A. Use the `get_metrics()` method of the run object to retrieve the experiment run logs.
- B. Use the `get_details_with_logs()` method of the run object to display the experiment run logs.
- C. View the log files for the experiment run in the experiment folder.
- D. View the logs for the experiment run in Azure Machine Learning studio.
- E. Use the `get_output()` method of the run object to retrieve the experiment run logs.

Answer: BD

Explanation:

Use `get_details_with_logs()` to fetch the run details and logs created by the run.

You can monitor Azure Machine Learning runs and view their logs with the Azure Machine Learning studio.

Incorrect Answers:

A: You can view the metrics of a trained model using `run.get_metrics()`.

E: `get_output()` gets the output of the step as `PipelineData`.

Reference:

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

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-monitor-view-training-logs>

QUESTION 212

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.estimator import Estimator
sk_est = Estimator(source_directory='./scripts',
                  compute_target=aml_compute,
                  entry_script='train.py')
```

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

There is a missing line: `conda_packages=['scikit-learn']`, which is needed.

Correct example:

```
sk_est = Estimator(source_directory='./my-sklearn-proj',
                  script_params=script_params,
```

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

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

```
compute_target=compute_target,  
entry_script='train.py',  
conda_packages=['scikit-learn'])
```

Note:

The Estimator class represents a generic estimator to train data using any supplied framework.

This class is designed for use with machine learning frameworks that do not already have an Azure Machine Learning pre-configured estimator. Pre-configured estimators exist for Chainer, PyTorch, TensorFlow, and SKLearn.

Example:

```
from azureml.train.estimator import Estimator  
script_params = {  
# to mount files referenced by mnist dataset  
'--data-folder': ds.as_named_input('mnist').as_mount(),  
'--regularization': 0.8  
}
```

Reference:

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

QUESTION 213

Hotspot Question

You plan to use Hyperdrive to optimize the hyperparameters selected when training a model. You create the following code to define options for the hyperparameter experiment:

```
import azureml.train.hyperdrive.parameter_expressions as pe  
from azureml.train.hyperdrive import GridParameterSampling, HyperDriveConfig  
  
param_sampling = GridParameterSampling({  
    "max_depth" : pe.choice(6, 7, 8, 9),  
    "learning_rate" : pe.choice(0.05, 0.1, 0.15)  
})  
hyperdrive_run_config = HyperDriveConfig(  
    estimator = estimator,  
    hyperparameter_sampling = param_sampling,  
    policy = None,  
    primary_metric_name = "auc",  
    primary_metric_goal = PrimaryMetricGoal.MAXIMIZE,  
    max_total_runs = 50,  
    max_concurrent_runs = 4)
```

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

NOTE: Each correct selection is worth one point.

Answer Area

	Yes	No
There will be 50 runs for this hyperparameter tuning experiment.	<input type="radio"/>	<input type="radio"/>
You can use the policy parameter in the HyperDriveConfig class to specify a security policy.	<input type="radio"/>	<input type="radio"/>
The experiment will create a run for every possible value for the learning rate parameter between 0.05 and 0.15.	<input type="radio"/>	<input type="radio"/>

Answer:

Answer Area

	Yes	No
There will be 50 runs for this hyperparameter tuning experiment.	<input type="radio"/>	<input checked="" type="radio"/>
You can use the policy parameter in the HyperDriveConfig class to specify a security policy.	<input checked="" type="radio"/>	<input type="radio"/>
The experiment will create a run for every possible value for the learning rate parameter between 0.05 and 0.15.	<input type="radio"/>	<input checked="" type="radio"/>

Explanation:

Box 1: No

max_total_runs (50 here)

The maximum total number of runs to create. This is the upper bound; there may be fewer runs when the sample space is smaller than this value.

Box 2: Yes

Policy EarlyTerminationPolicy

The early termination policy to use. If None - the default, no early termination policy will be used.

Box 3: No

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

Reference:

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

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

QUESTION 214

Hotspot Question

You are preparing to build a deep learning convolutional neural network model for image classification. You create a script to train the model using CUDA devices.

You must submit an experiment that runs this script in the Azure Machine Learning workspace.

The following compute resources are available:

- a Microsoft Surface device on which Microsoft Office has been installed. Corporate IT policies prevent the installation of additional software
- a Compute Instance named ds-workstation in the workspace with 2 CPUs and 8 GB of memory
- an Azure Machine Learning compute target named cpu-cluster with eight CPU-based nodes
- an Azure Machine Learning compute target named gpu-cluster with four CPU and GPU-based nodes

You need to specify the compute resources to be used for running the code to submit the experiment, and for running the script in order to minimize model training time.

Which resources should the data scientist use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Resource type	Option
Run code to submit the experiment	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> the Microsoft Surface device the ds-workstation notebook VM the cpu-cluster compute target the gpu-cluster compute target </div> </div>
Run the training script	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> the ds-workstation notebook VM the cpu-compute target the gpu-compute target the Microsoft Surface device </div> </div>

Answer:

Answer Area

Resource type	Option
Run code to submit the experiment	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> the Microsoft Surface device the ds-workstation notebook VM the cpu-cluster compute target the gpu-cluster compute target </div> </div>
Run the training script	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> the ds-workstation notebook VM the cpu-compute target the gpu-compute target the Microsoft Surface device </div> </div>

Explanation:

Box 1: the ds-workstation notebook VM

Box 2: the gpu-compute target

Just as GPUs revolutionized deep learning through unprecedented training and inferencing performance, RAPIDS enables traditional machine learning practitioners to unlock game-changing performance with GPUs. With RAPIDS on Azure Machine Learning service, users can accelerate the entire machine learning pipeline, including data processing, training and inferencing, with GPUs from the NC_v3, NC_v2, ND or ND_v2 families. Users can unlock performance gains of more than 20X (with 4 GPUs), slashing training times from hours to minutes and dramatically reducing time-to-insight.

Reference:

<https://azure.microsoft.com/sv-se/blog/azure-machine-learning-service-now-supports-nvidia-s-rapids/>

QUESTION 215

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

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

Hotspot Question

You are using Azure Machine Learning to train machine learning models. You need to compute target on which to remotely run the training script.

You run the following Python code:

```
from azureml.core.compute import ComputeTarget, AmlCompute
from azureml.core.compute_target import ComputeTargetException
the_cluster_name = "NewCompute"
config = AmlCompute.provisioning_configuration(vm_size= 'STANDARD_D2', max_nodes=3)
the_cluster = ComputeTarget.create(ws, the_cluster_name, config)
```

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

NOTE: Each correct selection is worth one point.

Answer Area

	Yes	No
The compute is created in the same region as the Machine Learning service workspace.	<input type="radio"/>	<input type="radio"/>
The compute resource created by the code is displayed as a compute cluster in Azure Machine Learning studio.	<input type="radio"/>	<input type="radio"/>
The minimum number of nodes will be zero.	<input type="radio"/>	<input type="radio"/>

Answer:

Answer Area

	Yes	No
The compute is created in the same region as the Machine Learning service workspace.	<input checked="" type="radio"/>	<input type="radio"/>
The compute resource created by the code is displayed as a compute cluster in Azure Machine Learning studio.	<input checked="" type="radio"/>	<input type="radio"/>
The minimum number of nodes will be zero.	<input checked="" type="radio"/>	<input type="radio"/>

Explanation:

Box 1: Yes

The compute is created within your workspace region as a resource that can be shared with other users.

Box 2: Yes

It is displayed as a compute cluster.

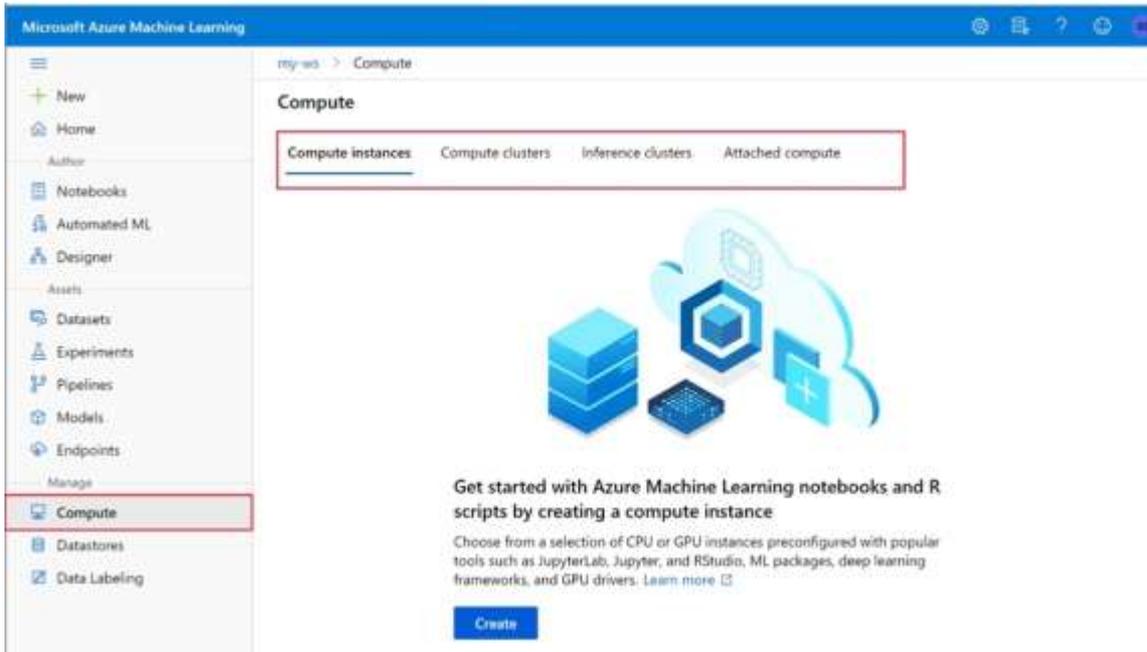
View compute targets

1. To see all compute targets for your workspace, use the following steps:

2. Navigate to Azure Machine Learning studio.

3. Under Manage, select Compute.

4. Select tabs at the top to show each type of compute target.



Box 3: Yes

min_nodes is not specified, so it defaults to 0.

Reference:

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

<https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-studio>

QUESTION 216

Hotspot Question

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.

Answer Area

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: <code>using myDataset_1.to_pandas_dataframe()</code>	<input type="radio"/>	<input type="radio"/>
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.	<input type="radio"/>	<input type="radio"/>
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: <code>myDataset_2.to_pandas_dataframe()</code>	<input type="radio"/>	<input type="radio"/>

Answer:

Answer Area

	Yes	No
The myDataset_1 dataset can be converted into a pandas dataframe by using the following method: <code>using myDataset_1.to_pandas_dataframe()</code>	<input type="radio"/>	<input checked="" type="radio"/>
The myDataset_1.to_path() method returns an array of file paths for all of the TSV files in the dataset.	<input checked="" type="radio"/>	<input type="radio"/>
The myDataset_2 dataset can be converted into a pandas dataframe by using the following method: <code>myDataset_2.to_pandas_dataframe()</code>	<input checked="" type="radio"/>	<input type="radio"/>

Explanation:

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>