Question 1

Domain: Manage Azure resources for machine learning

You are setting up a machine learning environment in Azure. In order to properly control the access of the users to the resources in your workspace, you have created a custom role:

```
# define custom role
{
    "Name": "CustomRole1",
    "IsCustom": true,
    "Description": "My Custom ML Role",
    "Actions": ["*"],
    "NotActions": [
        "Microsoft.MachineLearningServices/workspaces/*/delete",
        "Microsoft.MachineLearningServices/workspaces/write",

"Microsoft.MachineLearningServices/workspaces/computes/*/write",

"Microsoft.MachineLearningServices/workspaces/computes/*/delete",
        "Microsoft.Authorization/*/write"
    ],
    "AssignableScopes":
["/subscriptions/<subscription_id>/resourceGroups/MyRg/providers/Micros
oft.MachineLearningServices/workspaces/MyWsp"
    ]
}
```

With any other privileges granted, users assigned to this role...

- A. can submit any type of model runs
- B. can delete a compute resource
- C. can add, delete, or alter role assignments
- D. can delete the workspace

Explanation:

Correct Answer: A

- Option A is CORRECT because anything except what is explicitly denied by "NotActions" is allowed for the assignees of the role
- Option B is incorrect because deleting compute resources is explicitly denied.
- Option C is incorrect because modifying role assignments are explicitly denied by the role definition.
- Option D is incorrect because deleting the workspace is explicitly denied by the role definition.

Reference:

- https://docs.microsoft.com/en-us/azure/machine-learning/how-to-assign-roles#create-custom-role

Question 2

Domain: Implement responsible machine learning

You need to increase the privacy for your data set and you decide to use the SmartNoise Python library.

How can you ensure the fulfillment of the requirement?

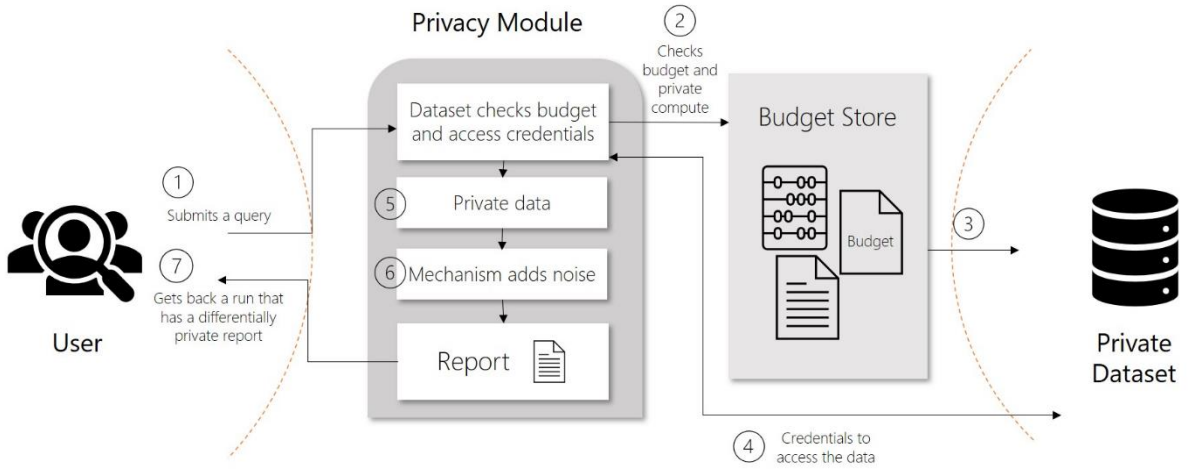
- A. Set epsilon value to -1;
- B. Set the epsilon value to 2
- C. Set delta to -1
- D. Set epsilon value between 0 and 1

Explanation:

Correct Answer: D

- Option A is incorrect because the epsilon value must be a non-negative number.
- Option B is incorrect because the higher the value of the epsilon parameter, the lower the level of privacy (i.e. the level of noise used to distort the raw data). Epsilon should be set below 1 in order to achieve a proper level of differential privacy.
- Option C is incorrect because delta is the probability that the data is not totally private. The higher the epsilon, the higher its value is. In general, the epsilon value is used. Since it is a measure of probability, it must be between 0 and 1.
- Option D is CORRECT because in order to ensure differential privacy, the epsilon value must be set between 0 and 1. This ensures a sufficiently high level of noise to mask the raw data while leaving the statistical characteristics of the data set undistorted.

Diagram:



The above diagram shows the process of adding noise (Step 6) to the raw data in order to decrease the risk of privacy issues.

References:

- https://docs.microsoft.com/en-us/azure/machine-learning/concept-differential-privacy
- https://docs.microsoft.com/en-us/azure/machine-learning/how-to-differential-privacy

Question 3

Domain: Run experiments and train models

While running ML experiments, you want to use MLFlow to track execution of the runs, to monitor the model training process and to store the metrics of the runs in your ML workspace.

Which of the following steps is not necessary while configuring MLFlow for your ML workspace?

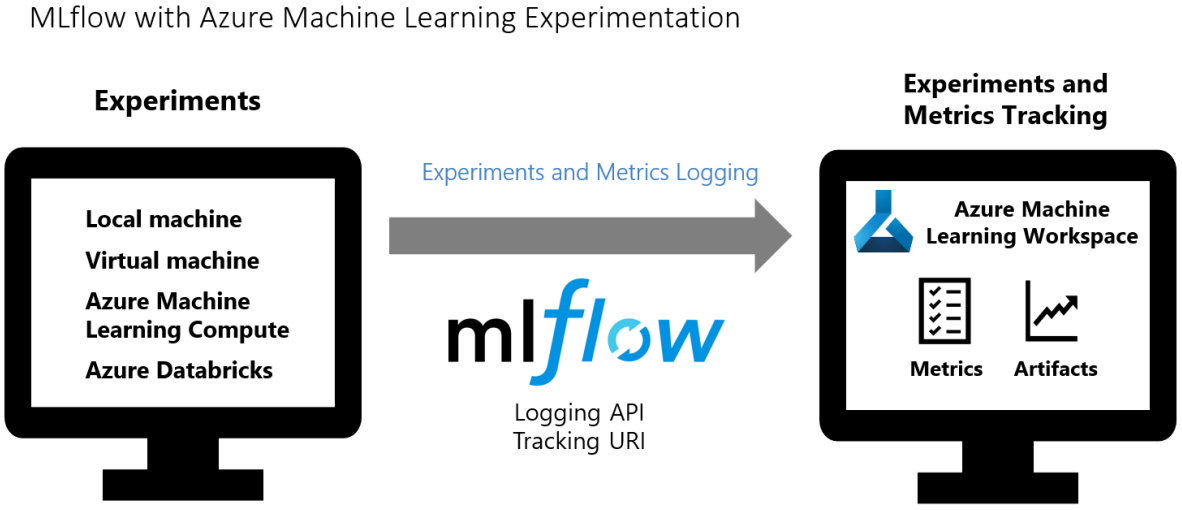
- A. install the azureml-mlflow package
- B. set_tracking_uri
- C. get_mlfow_tracking_uri
- D. Install azureml-core

Explanation:

Correct Answer: D

- Option A is incorrect because the azureml-mlflow package is a prerequisite to using the MLFlow for monitoring your experiments.
- Option B is incorrect because MLFlow's set_tracking_uri is required for MLFlow to point to the URI of an ML workspace.
- Option C is incorrect because the workspace's get_mlflow_tracking_uri is used to achieve a unique tracking URI to the workspace (to be used with the set_tracking_uri() later)
- Option D is CORRECT because the azureml-core package is automatically installed together with the azureml-mlflow package.

Diagram:



Reference:

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

Question 4

Domain: Implement responsible machine learning

You are building a machine learning model to predict the willingness to repay loans among the customers of a large bank. The binary classification model delivers very good performance, but before deploying it to production, you want to make sure that the model is free of age, gender and other sensitive biases, using the Fairlearn unfairness mitigation package.

While evaluating the fairness of the model, you need to balance between...

- A. reduction AND post-processing
- B. model's accuracy AND performance
- C. model's performance AND disparity
- D. disparity AND fairness

Explanation:

Correct Answer: C

- Option A is incorrect because reduction and post-processing are two types of unfairness mitigation algorithms used by the Fairlearn package you have to choose from.
- Option B is incorrect because accuracy is one of the performance metrics of an ML model. Balancing is not applicable between them.
- Option C is CORRECT because unfairness mitigation algorithms always affect the performance of the model. Depending on the "level of unfairness" you want to achieve, certain levels of trade-off between the performance (e.g. accuracy) and the level of disparity achieved.
- Option D is incorrect because "fairness" is the generic term for the un-biased model behavior while disparity is a metric of (un)fairness. No balancing applies here.

Reference:

- https://docs.microsoft.com/en-us/azure/machine-learning/concept-fairness-ml#mitigation-algorithms

Question 5

Domain: Manage Azure resources for machine learning

You have successfully developed your ML model in your local environment and tested it with larger amounts of data on a training compute. Now, you are ready to scale up to productive use and you want to set up an Azure Databricks cluster as deployment environment.

Which of the following options is correct?

- A. You have to create an Azure Databricks cluster in Azure ML and use it as inference compute
- B. You have to create an Azure Databricks Cluster outside Azure ML and attach it to your ML workspace
- C. Include the following code in your ML configuration: databricks_compute = ComputeTarget.create(...)
- D. Databricks clusters cannot be used as inference compute for Azure ML

Explanation:

Correct Answer: B

- Option A is incorrect because Azure Machine Learning cannot create external compute resources like Databricks Clusters.
- Option B is CORRECT because external (unmanaged) compute targets like HDInsight and Databricks must be created externally, and then attached to the ML workspace.
- Option C is incorrect because you cannot create a Databricks compute within your ML configuration. You have to use the ComputeTarget.attach to link to an existing Databricks cluster.
- Option D is incorrect because Databricks is one of the several possible compute targets that can be used with Azure ML.

References:

- https://docs.microsoft.com/en-us/azure/machine-learning/how-to-attach-compute-targets#databricks
- https://docs.microsoft.com/en-us/azure/machine-learning/concept-compute-target
- https://docs.microsoft.com/en-us/azure/machine-learning/how-to-attach-compute-targets

Question 6

Domain: Deploy and operationalize machine learning solutions

You have deployed your ML model and you want to trigger the retraining of it automatically when data drift is detected by an automatic dataset monitor. In this case, an existing Azure Data Factory pipeline must be launched in order to retrain the model. Azure Event Grid, in combination with Azure Logic App should be used.

1. Create a Logic App
2. Select trigger: When an Event Grid resource event occurs
3. Select trigger: When a message is received in a Service Bus Queue
4. Set the Event Type to DatasetDriftDetected
5. Set the Event Type to RunStatusChanged
6. Create an Azure Data Factory pipeline run step

Which of the above actions should be executed to achieve your goal?

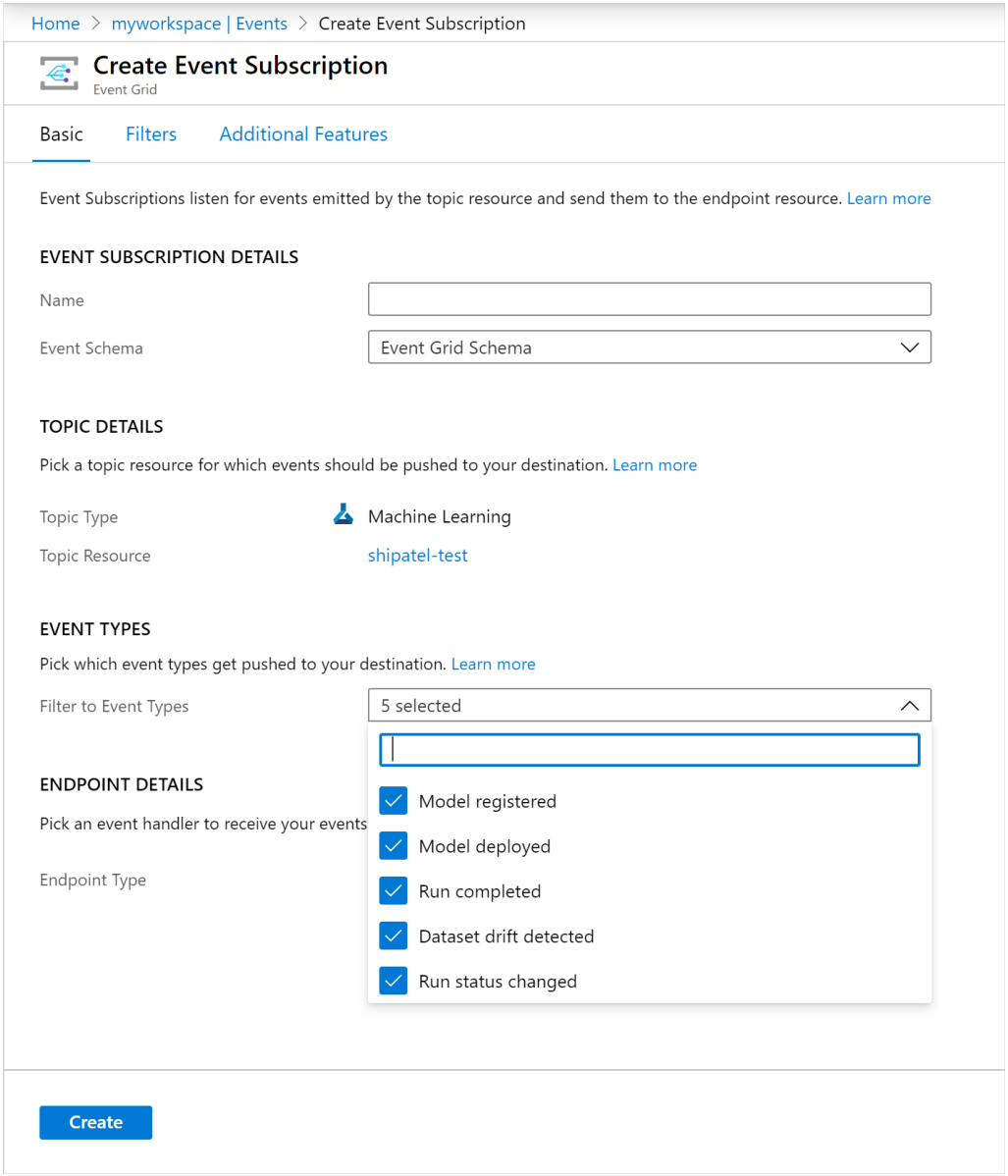
- A. 1, 2, 4, 6
- B. 1, 3, 4, 5
- C. 1, 2, 5, 6
- D. 2, 3, 4, 6

Explanation:

Correct Answer: A

- Option A is CORRECT because a Logic App must be created, triggered by an Event Grid event of type DatesetDriftDetected, the another step to launch the existing AFF pipeline must be added to the Logic App flow.
- Option B is incorrect because the trigger of the Logic App is an Event Grid event, rather than a message in a Service Bus Queue. The event type is incorrect, too.
- Option C is incorrect because the event type DatasetDriftDetected must be set.
- Option D is incorrect because it is an Event Grid event which triggers the execution of the Logic App.

Diagram:



Reference:

- https://docs.microsoft.com/en-us/azure/machine-learning/concept-model-management-and-deployment#notify-automate-and-alert-on-events-in-the-ml-lifecycle
- https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-event-grid

Question 7

Domain: Deploy and operationalize machine learning solutions

While developing your machine learning models and while running experiments, you want to make use of the tools and practices provided by the DevOps-based model management environment. As you develop your models, several versions of the model are created, need to be stored in a central place and to be made available later.

Which of the following features of Azure ML best supports your goal?

- A. Azure ML Environments
- B. Model profiling
- C. Model registration
- D. Model versioning

Explanation:

Correct Answer: C

- Option A is incorrect because Environments describe the pip and Conda dependencies for projects, and are used for training and deployment of models. Environments allow for the reproducibility of the builds.
- Option B is incorrect because model profiling is intended to help you understand the CPU and memory requirements of the service that will be created when deploying the model.
- Option C is CORRECT because it is the Model registration which allows you to store and version your models in your workspace. The model registry provides the organization, tracking and versioning of the trained models. Each time a model with the same name is added to the regisítry, a new version is created.
- Option D is incorrect because while it is actually the versioning that will support your goal, model versioning is an implicit mechanism of the model registration process. It is not a stand-alone feature. Registration should be used.

References:

- https://docs.microsoft.com/en-us/azure/machine-learning/concept-model-management-and-deployment
- https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-profile-model?pivots=py-sdk

Question 8

Domain: Implement responsible machine learning

You have built a ML model to predict the sensitivity of patients to certain drug treatments. You decide to use the Fairlearn package to make your model free of unfairness, i.e. to eliminate any biases and disparities.

Can a properly configured Fairlearn algorithm do it for you automatically, on its own?

- A. Yes
- B. No

Explanation:

Correct Answer: B

- Option A is incorrect because algorithms of the Fairlearn package (or any other existing algorithms) are not capable of mitigating unfairness issues on their own, without human considerations and fine-tuning.
- Option C is CORRECT because mitigating fairness always requires a certain level of human consideration and intervention. Machine algorithms can be very effective in the quantitative part, but they cannot be used to eliminate disparities totally because fairness is a complex phenomenon which always needs qualitative assessment performed by human actors.

Reference:

- https://docs.microsoft.com/en-us/azure/machine-learning/concept-fairness-ml#assess-fairness-in-machine-learning-models

Question 9

Domain: Manage Azure resources for machine learning

You have set up a machine learning workspace in Azure for your team. You want the users to be able to run experiments, therefore they need to be granted access to the ML resources, but prevent them modifying role assignments.

They must be assigned to which built-in role?

- A. Owner
- B. Reader
- C. Contributor
- D. Writer

Explanation:

Correct Answer: C

- Option A is incorrect because the Owner role has full access to the workspace including creating and deleting any assets. This is the highest privilege on the workspace and it should not be granted to a wider group of users.
- Option B is incorrect because assignees to the Reader role cannot create or update any assets within the workspace.
- Option C is CORRECT because it is the Contributor role that gives access to the necessary assets for the users who need to run experiments, while keeping their privileges as limited as possible.
- Option D is incorrect because "Writer" is not a valid built-in role in Azure ML.

Reference:

- https://docs.microsoft.com/en-us/azure/machine-learning/how-to-assign-roles

Question 10

Domain: Run experiments and train models

You are using an Azure Databricks cluster for training your ML model. After successfully training your model built using the scikit-learn framework, you want to register it to the backend tracking server, using the following code:

```
# register model
mlflow.sklearn.log_model(model, artifact_path = "trained_model",
                         registered_model_name = 'my_trained_model')
```

Is this the right formula to register your model with MLFlow backend tracking?

- A. Yes
- B. No

Explanation:

Correct Answer: A

- Option A is CORRECT because it is the correct formula for registering a scikit-learn model with mlflow for tracking. The <sklearn> must be used as “model flavor”.
- Option B is incorrect because the formula actually does the registration of your trained model for being tracked with MLFlow.

References:

- https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow-azure-databricks
- https://mlflow.org/docs/latest/quickstart.html#using-the-tracking-api
- https://mlflow.org/docs/latest/models.html#model-api

Question 11

Domain: Manage Azure resources for machine learning

You are developing your machine learning model which needs to be run on an Azure Databricks cluster. You want to use the following script in order to access your DB cluster from your Azure ML workspace:

```
from azureml.core import Workspace
from azureml.core.compute import ComputeTarget, DatabricksCompute

myws = <.......1......>()

compute_name = 'db_cluster'

db_workspace_name = 'db_workspace'
db_resource_group = 'db_resource_group'
db_access_token = '3747-bxz-xjkh-2293-40...'
db_config =
DatabricksCompute.attach_configuration(resource_group=db_resource_group
,

workspace_name=db_workspace_name,

access_token=db_access_token)

databricks_compute = <......2.....>(myws, compute_name, db_config)
databricks_compute.wait_for_completion(True)
```

Which of the following instructions are missing from the above script?

- A. Workspace.from_config; ComputeTarget.create
- B. ComputeTarget.from_config; Workspace.attach
- C. Workspace.from_config; ComputeTarget.attach
- D. ComputeTarget.create; Workspace.from_config

Explanation:

Correct Answer: C

- Option A is incorrect because in order to use external ML computes like a Databricks cluster, the "attach" method must be used instead of "create".
- Option B is incorrect because the "from_config" method of the Workspace must be used first, then the "attach" method of ComputeTarget is needed.
- Option C is CORRECT because if you want to use your own existing Azure Databricks cluster, you start reading your ML workspace from config, then you need to attach the Databricks workspace to it.
- Option D is incorrect because the two instructions are in the wrong order.

Reference:

- https://docs.microsoft.com/en-us/learn/modules/use-compute-contexts-in-aml/4-creating-compute

Question 12

Domain: Manage Azure resources for machine learning

For your machine learning experiments, you want to use an Azure Databricks cluster as a training compute. You already have your ML environment set up and now you want to configure the Databricks cluster for use. You enter the Azure ML Studio, select the Training clusters and start creating the Databricks cluster.

Is that the right way to reach your goal?

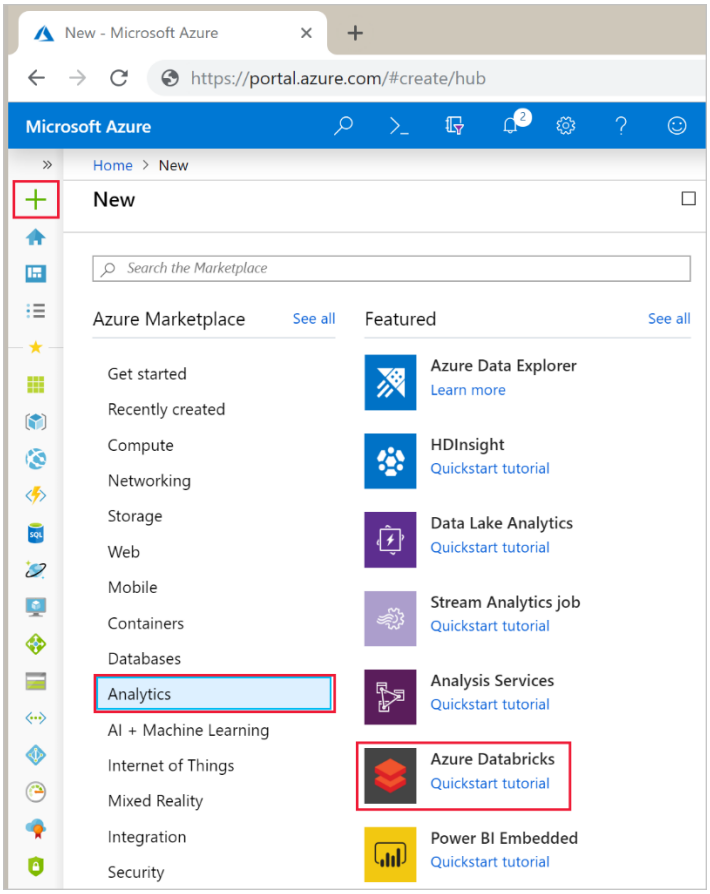
- A. Yes
- B. No

Explanation:

Correct Answer: B

- Option A is incorrect because you cannot create unmanaged compute targets (like Databricks clusters) from Azure ML directly.
- Option B is CORRECT because you cannot create a Databricks cluster from the Azure ML environment. The Azure Databricks workspace must be created first (from Azure Portal, for example), after setting up the DB cluster, it must be then attached to your Azure ML workspace.

Diagram:



References:

- https://docs.microsoft.com/en-us/azure/machine-learning/how-to-attach-compute-targets
- https://docs.microsoft.com/en-us/azure/databricks/scenarios/quickstart-create-databricks-workspace-portal?tabs=azure-portal

Question 13

Domain: Deploy and operationalize machine learning solutions

You want to take the advantages of using the DevOps pipelines provided by Azure. You need to use Data Factory to ingest data and run a notebook on a Databricks cluster, which checks if the data has been ingested correctly and validates the result data file. Steps of your pipeline looks like this:

```
# run pipeline
- job: "test job"
    displayName: "Test job"
    dependsOn: [Deploy_to_Databricks, Deploy_to_ADF]
    pool:
      vmImage: 'ubuntu-latest'
    timeoutInMinutes: 0
    steps:
    - task: <.......task1........>@4
      displayName: DF Pipeline'
      inputs:
        azureSubscription: $(AZURE_RM_CONNECTION)
        ScriptPath:
'$(Build.SourcesDirectory)/adf/temp/My_DFPipeline.ps1'
        ScriptArguments: '-ResourceGroupName $(RESOURCE_GROUP) -
DataFactoryName $(DATA_FACTORY_NAME) -PipelineName $(PIPELINE_NAME)'
        azurePowerShellVersion: LatestVersion
    - task: <.......task2........>@0
      inputs:
        versionSpec: '3.x'
        addToPath: true
        architecture: 'x64'
      displayName: 'Python3.x'

    - task: <.......task3........>@0
      inputs:
        url: '$(DATABRICKS_URL)'
        token: '$(DATABRICKS_TOKEN)'
      displayName: 'Databricks config'

    - task: <.......task4........>@0
      inputs:
        notebookPath: '/Shared/devops-ds/test-data-ingestion'
        existingClusterId: '$(DATABRICKS_CLUSTER_ID)'
        executionParams: '{"bin_file_name":"$(bin_FILE_NAME)"}'
      displayName: 'Ingest data'

    - task: waitexecution@0
      displayName: 'Wait until the testing is done'
```

Match the name of the pipeline steps with the task names in the above script:

- A. Executenotebook; AzurePowerShell; UsePythonVersion; configuredatabricks
- B. AzurePowerShell; configuredatabricks; UsePythonVersion; executenotebook
- C. AzurePowerShell; UsePythonVersion; configuredatabricks; executenotebook
- D. UsePythonVersion; AzurePowerShell; executenotebook; configuredatabricks

Explanation:

Correct Answer: C

- Option A is incorrect because running the notebook mustang be preceded by ingesting data by Data Factory and setting up the environment.
- Option B is incorrect because task2 defines setting the Python version, while task3 defines the Databricks environment.
- Option C is CORRECT because the script first runs a Data Factory pipeline from PowerShell, then sets the Python version and configures Databricks, and finally executes a notebook on a Databricks cluster.
- Option D is incorrect because executenotebook is the last step in the sequence (task4)

Reference:

- https://docs.microsoft.com/en-us/azure/machine-learning/how-to-cicd-data-ingestion#run-the-pipeline-and-check-the-data-ingestion-result

Question 14

Domain: Run experiments and train models

You are using an Azure Databricks cluster for training your ML model. In order to monitor and track the training process of the model, you want to set up MLFlow tracking.

By setting up MLFlow for tracking you can store logs and model artefacts ...

- A. … only in your Azure ML Workspace
- B. … only in your Azure Databricks workspace

- C. … in both of your Azure ML and Azure Databricks workspaces
- D. … in neither of the Azure ML or Databricks workspaces

Explanation:

Correct Answer: C

- Option A is incorrect because MLFlow outputs can also be stored in the Azure Databricks workspace.
- Option B is incorrect because MLFlow outputs can also be directed to your Azure ML workspace.
- Option C is CORRECT because by using MLFlow, the logs and any artefacts of model runs can be stored in both your Azure ML and Azure Databricks workspaces. The two workspaces must be linked together.
- Option D is incorrect because MLFlow logs actually can be stored in both of these types of workspaces.

References:

- https://docs.microsoft.com/en-us/azure/machine-learning/how-to-use-mlflow-azure-databricks
- https://mlflow.org/docs/latest/quickstart.html#using-the-tracking-api
- https://mlflow.org/docs/latest/models.html#model-api

Question 15

Domain: Manage Azure resources for machine learning

While working on your ML experiments, you are setting up compute resources which you want to share with two of your colleagues in your workspace.

For sharing computer resources, which of the following statements are true?

- A. You can share your compute instance for development; you can share your compute cluster for training
- B. You cannot share any of your compute resources
- C. You can share your compute instance for development; you cannot share the compute cluster
- D. You cannot share your compute instance for development; you can share your compute cluster for training

Explanation:

Correct Answer: D

- Option A is incorrect because as a development environment, compute instance cannot be shared with other users in the workspace
- Option B is incorrect because compute resources, except for development compute instances, can be shared among multiple users.
- Option C is incorrect because the compute instance cannot be shared, while the compute cluster can.
- Option D is CORRECT because as a development environment, compute instances cannot be shared with other users in the workspace, while training clusters can be used in shared mode to run jobs launched by several users.

References:

- https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-manage-compute-instance?tabs=python
- https://docs.microsoft.com/en-us/azure/machine-learning/how-to-create-attach-compute-cluster?tabs=python#what-is-a-compute-cluster