Exam DP-100: Designing and Implementing a Data Science Solution on Azure – Skills Measured

This exam will be updated on May 20, 2021. Following the current exam guide, we have included a table that compares the current study guide to the new study guide by category.

Audience Profile

The Azure Data Scientist applies their knowledge of data science and machine learning to implement and run machine learning workloads on Azure; in particular, using Azure Machine Learning Service and Azure Databricks. This entails planning and creating a suitable working environment for data science workloads on Azure, running data experiments and training predictive models, managing and optimizing models, and deploying machine learning models into production.

Skills Measured

NOTE: The bullets that appear below each of the skills measured are intended to illustrate how we are assessing that skill. This list is NOT definitive or exhaustive.

NOTE: Most questions cover features that are General Availability (GA). The exam may contain questions on Preview features if those features are commonly used.

Set up an Azure Machine Learning Workspace (30-35%)

Create an Azure Machine Learning workspace

- create an Azure Machine Learning workspace
- configure workspace settings
- manage a workspace by using Azure Machine Learning studio

Manage data objects in an Azure Machine Learning workspace

- register and maintain datastores
- create and manage datasets

Manage experiment compute contexts

- create a compute instance
- determine appropriate compute specifications for a training workload

- create compute targets for experiments and training

Run Experiments and Train Models (25-30%)

Create models by using Azure Machine Learning Designer

- create a training pipeline by using Azure Machine Learning designer
- ingest data in a designer pipeline
- use designer modules to define a pipeline data flow
- use custom code modules in designer

Run training scripts in an Azure Machine Learning workspace

- create and run an experiment by using the Azure Machine Learning SDK
- configure run settings for a script
- consume data from a dataset in an experiment by using the Azure Machine Learning SDK

Generate metrics from an experiment run

- log metrics from an experiment run
- retrieve and view experiment outputs
- use logs to troubleshoot experiment run errors

Automate the model training process

- create a pipeline by using the SDK
- pass data between steps in a pipeline
- run a pipeline
- monitor pipeline runs

Optimize and Manage Models (20-25%)

Use Automated ML to create optimal models

- use the Automated ML interface in Azure Machine Learning studio
- use Automated ML from the Azure Machine Learning SDK
- select pre-processing options
- determine algorithms to be searched
- define a primary metric
- get data for an Automated ML run
- retrieve the best model

Use Hyperdrive to tune hyperparameters

- select a sampling method
- define the search space
- define the primary metric
- define early termination options
- find the model that has optimal hyperparameter values

Use model explainers to interpret models

- select a model interpreter
- generate feature importance data

Manage models

- register a trained model
- monitor model usage
- monitor data drift

Deploy and Consume Models (20-25%)

Create production compute targets

- consider security for deployed services
- evaluate compute options for deployment

Deploy a model as a service

- configure deployment settings
- consume a deployed service
- troubleshoot deployment container issues

Create a pipeline for batch inferencing

- publish a batch inferencing pipeline
- run a batch inferencing pipeline and obtain outputs

Publish a designer pipeline as a web service

- create a target compute resource
- configure an Inference pipeline
- consume a deployed endpoint

The following table compares the current and new study guides.

Current study guide until May 20, 2021	New study guide as of May 20, 2021
Set up an Azure Machine Learning Workspace (30-35%) **Create an Azure Machine Learning workspace** • create an Azure Machine Learning workspace • configure workspace settings • manage a workspace by using Azure Machine Learning studio **Manage data objects in an Azure Machine Learning workspace** • register and maintain datastores • create and manage datasets **Manage experiment compute contexts** • create a compute instance • determine appropriate compute specifications for a training workload • create compute targets for experiments and training	**Manage Azure resources for machine learning (25-30%)** **Create an Azure Machine Learning workspace** • create an Azure Machine Learning workspace • configure workspace settings • manage a workspace by using Azure Machine Learning studio **Manage data in an Azure Machine Learning workspace** • select Azure storage resources • register and maintain datastores • create and manage datasets **Manage compute for experiments in Azure Machine Learning** • determine the appropriate compute specifications for a training workload • create compute targets for experiments and training • configure Attached Compute resources including Azure Databricks • monitor compute utilization **Implement security and access control in Azure Machine Learning** • determine access requirements and map requirements to built-in roles • create custom roles • manage role membership • manage credentials by using Azure Key Vault **Set up an Azure Machine Learning development environment**

	- create compute instances - share compute instances - access Azure Machine Learning workspaces from other development environments **Set up an Azure Databricks workspace** - create an Azure Databricks workspace - create an Azure Databricks cluster - create and run notebooks in Azure Databricks - link and Azure Databricks workspace to an Azure Machine Learning workspace
Run Experiments and Train Models (25-30%) **Create models by using Azure Machine Learning Designer** - create a training pipeline by using Azure Machine Learning designer - ingest data in a designer pipeline - use designer modules to define a pipeline data flow - use custom code modules in designer **Run training scripts in an Azure Machine Learning workspace** - create and run an experiment by using the Azure Machine Learning SDK - configure run settings for a script - consume data from a dataset in an experiment by using the Azure Machine Learning SDK **Generate metrics from an experiment run** - log metrics from an experiment run - retrieve and view experiment outputs - use logs to troubleshoot experiment run errors **Automate the model training process** - create a pipeline by using the SDK - pass data between steps in a pipeline - run a pipeline	**Run experiments and train models (20-25%)** **Create models by using the Azure Machine Learning designer** - create a training pipeline by using Azure Machine Learning designer - ingest data in a designer pipeline - use designer modules to define a pipeline data flow - use custom code modules in designer **Run model training scripts** - create and run an experiment by using the Azure Machine Learning SDK - configure run settings for a script - consume data from a dataset in an experiment by using the Azure Machine Learning SDK - run a training script on Azure Databricks compute - run code to train a model in an Azure Databricks notebook **Generate metrics from an experiment run** - log metrics from an experiment run - retrieve and view experiment outputs - use logs to troubleshoot experiment run errors - use MLflow to track experiments - track experiments running in Azure Databricks

• monitor pipeline runs	**Use Automated Machine Learning to create optimal models** • use the Automated ML interface in Azure Machine Learning studio • use Automated ML from the Azure Machine Learning SDK • select pre-processing options • select the algorithms to be searched • define a primary metric • get data for an Automated ML run • retrieve the best model **Tune hyperparameters with Azure Machine Learning** • select a sampling method • define the search space • define the primary metric • define early termination options • find the model that has optimal hyperparameter values
Optimize and Manage Models (20-25%) **Use Automated ML to create optimal models** • use the Automated ML interface in Azure Machine Learning studio • use Automated ML from the Azure Machine Learning SDK • select pre-processing options • determine algorithms to be searched • define a primary metric • get data for an Automated ML run • retrieve the best model **Use Hyperdrive to tune hyperparameters** • select a sampling method • define the search space • define the primary metric • define early termination options • find the model that has optimal	**Deploy and operationalize machine learning solutions (35-40%)** **Select compute for model deployment** • consider security for deployed services • evaluate compute options for deployment **Deploy a model as a service** • configure deployment settings • deploy a registered model • deploy a model trained in Azure Databricks to an Azure Machine Learning endpoint • consume a deployed service • troubleshoot deployment container issues **Manage models in Azure Machine Learning** • register a trained model • monitor model usage • monitor data drift

    hyperparameter values

Use model explainers to interpret models

- select a model interpreter
- generate feature importance data

Manage models

- register a trained model
- monitor model usage
- monitor data drift

Create an Azure Machine Learning pipeline for batch inferencing

- configure a ParallelRunStep
- configure compute for a batch inferencing pipeline
- publish a batch inferencing pipeline
- run a batch inferencing pipeline and obtain outputs
- obtain outputs from a ParallelRunStep

Publish an Azure Machine Learning designer pipeline as a web service

- create a target compute resource
- configure an Inference pipeline
- consume a deployed endpoint

Implement pipelines by using the Azure Machine Learning SDK

- create a pipeline
- pass data between steps in a pipeline
- run a pipeline
- monitor pipeline runs

Apply ML Ops practices

- trigger an Azure Machine Learning pipeline from Azure DevOps
- automate model retraining based on new data additions or data changes
- refactor notebooks into scripts
- implement source control for scripts

Deploy and Consume Models (20-25%)

Create production compute targets

- consider security for deployed services
- evaluate compute options for deployment

Deploy a model as a service

- configure deployment settings

Implement Responsible ML (5-10%)

Use model explainers to interpret models

- select a model interpreter
- generate feature importance data

Describe fairness considerations for models

- evaluate model fairness based on prediction

• consume a deployed service • troubleshoot deployment container issues **Create a pipeline for batch inferencing** • publish a batch inferencing pipeline • run a batch inferencing pipeline and obtain outputs **Publish a designer pipeline as a web service** • create a target compute resource • configure an Inference pipeline • consume a deployed endpoint	disparity • mitigate model unfairness **Describe privacy considerations for data** • describe principles of differential privacy • specify acceptable levels of noise in data and the effects on privacy