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

This exam will be updated on December 8, 2020. Following the current exam guide, we have included a version of the exam guide with Track Changes set to "On," showing the changes that will be made to the exam on that date.

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. 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 data stores
- 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
- consume data from a data store in an experiment by using the Azure Machine Learning SDK
- consume data from a dataset in an experiment by using the Azure Machine Learning SDK
- choose an estimator for a training experiment

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 scaling functions and 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 history
- 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 exam guide below shows the changes that will be implemented on December 8, 2020.

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 data store in an experiment by using the Azure Machine Learning SDK~~
- consume data from a dataset in an experiment by using the Azure Machine Learning SDK

- ~~choose an estimator for a training experiment~~

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 ~~scaling functions and~~ 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 ~~history~~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