

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

Introduction

The Designing and Implementing a Data Science Solution on Azure (Exam: DP-100) certification course enables competitive candidates to demonstrate their expertise and understanding of machine learning and data science to current or future employers. In addition to that, the DP-100 certification produces reliable benefits for an aspirant's overall career development by improving their skills in different methods and best practices related to Azure data science and ML services.

Course Outline

MODULE 1: 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 an Azure Databricks workspace to an Azure Machine Learning workspace

MODULE 2: 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
- 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

MODULE 3: Deploy and operationalize machine learning solutions (35–40%)

- Select compute for model deployment
 - consider security for deployed services
 - evaluate compute options for the 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
- 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

MODULE 4: Implement responsible machine learning (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 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

Prerequisites

- Candidates who wish to take up the Designing and Implementing a Data Solution on Azure DP-100 certification exam should have a fundamental knowledge of Microsoft Azure.
- A participant should also be able to write in programming languages such as Python to work with data using various libraries.
- Basic understanding of data science that includes preparing data and train machine learning models using machine learning libraries.

Target Audience

Candidates for the Azure Jo Associate certification should have subject matter expertise applying data science and machine learning to implement and run machine learning workloads on Azure.

Responsibilities for this role include planning and creating a suitable working environment for data science workloads on Azure. You run data experiments and train predictive models. In addition, you manage, optimize, and deploy machine learning models into production.

A candidate for this certification should have knowledge and experience in data science and using Azure Machine Learning and Azure Databricks.

Duration

24 Hours