

# Study Guide

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

### Purpose of this document

This study guide should help you understand what to expect on the exam and includes a summary of the topics the exam might cover and links to additional resources. The information and materials in this document should help you focus your studies as you prepare for the exam.

Useful links	Description
<a href="#">How to earn the certification</a>	Some certifications only require one exam, while others require more. On the details page, you'll find information about what skills are measured and links to registration. Each exam also has its own details page covering exam specifics.
<a href="#">Certification renewal</a>	Once you earn your certification, don't let it expire. When you have an active certification that's expiring within six months, you should renew it—at no cost—by passing a renewal assessment on Microsoft Learn. Remember to renew your certification annually if you want to retain it.
<a href="#">Your Microsoft Learn profile</a>	Connecting your certification profile to Learn brings all your learning activities together. You'll be able to schedule and renew exams, share and print certificates, badges and transcripts, and review your learning statistics inside your Learn profile.
<a href="#">Passing score</a>	All technical exam scores are reported on a scale of 1 to 1,000. A passing score is 700 or greater. As this is a scaled score, it may not equal 70% of the points. A passing score is based on the knowledge and skills needed to demonstrate competence as well as the difficulty of the questions.
<a href="#">Exam sandbox</a>	Are you new to Microsoft certification exams? You can explore the exam environment by visiting our exam sandbox. We created the sandbox as an opportunity for you to experience an exam before you take it. In the sandbox,

Useful links	Description
	you can interact with different question types, such as build list, case studies, and others that you might encounter in the user interface when you take an exam. Additionally, it includes the introductory screens, instructions, and help topics related to the different types of questions that your exam might include. It also includes the non-disclosure agreement that you must accept before you can launch the exam.
<a href="#">Request accommodations</a>	We're committed to ensuring all learners are set up for success. If you use assistive devices, require extra time, or need modification to any part of the exam experience, you can request an accommodation.
<a href="#">Take a practice test</a>	Taking a practice test is a great way to know whether you're ready to take the exam or if you need to study a bit more. Subject-matter experts write the Microsoft Official Practice Tests, which are designed to assess all exam objectives.

## Objective domain: skills the exam measures

The English language version of this exam was updated on May 20, 2021.

Some exams are localized into other languages, and those are updated approximately eight weeks after the English version is updated. Other available languages are listed in the **Schedule Exam** section of the **Exam Details** webpage. If the exam isn't available in your preferred language, you can request an additional 30 minutes to complete the exam.

### Note

The bullets that follow each of the skills measured are intended to illustrate how we are assessing that skill. Related topics may be covered in the exam.

### Note

Most questions cover features that are general availability (GA). The exam may contain questions on Preview features if those features are commonly used.

## Skills measured

- Manage Azure resources for machine learning (25–30%)
- Run experiments and train models (20–25%)
- Deploy and operationalize machine learning solutions (35–40%)
- Implement responsible machine learning (5–10%)

# Functional groups

## 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 dataset

### 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

## 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

## **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

## **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

## **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

# Study Resources

We recommend that you train and get hands-on experience before you take the exam. We offer self-study options and classroom training as well as links to documentation, community sites, and videos.

Study resources	Links to learning and documentation
Get trained	<a href="#">Choose from self-paced learning paths and modules or take an instructor led course</a>
Find documentation	<a href="#">Azure Databricks</a> <a href="#">Azure Machine Learning</a>
Ask a question	<a href="#">Microsoft Q&amp;A   Microsoft Docs</a>
Get community support	<a href="#">AI - Machine Learning - Microsoft Tech Community</a> <a href="#">AI - Machine Learning Blog - Microsoft Tech Community</a>
Follow Microsoft Learn	<a href="#">Microsoft Learn - Microsoft Tech Community</a>
Find a video	<a href="#">Microsoft Learn Shows</a>