

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
How to earn the certification	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.
Certification renewal	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.
Your Microsoft Learn profile	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.
Passing score	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.
Exam sandbox	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.
Request accommodations	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.
Take a practice test	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 will be updated on October 18, 2022. If you're taking this exam's English version before this date, the following Skills Measured is what you should study. If you want to review changes to the future version, scroll to the end of this document.

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	Choose from self-paced learning paths and modules or take an instructor led course
Find documentation	Azure Databricks Azure Machine Learning
Ask a question	Microsoft Q&A Microsoft Docs
Get community support	AI - Machine Learning - Microsoft Tech Community AI - Machine Learning Blog - Microsoft Tech Community
Follow Microsoft Learn	Microsoft Learn - Microsoft Tech Community
Find a video	Microsoft Learn Shows

Future exam skills measured

Our exams are updated periodically to reflect skills that are required to perform a role. The following skills measured list depicts the additions, deletions, and modifications to the exam.

Change log

Key to understanding the table: The topic groups also known as functional groups are in bold typeface followed by the objectives within each group. The table is a comparison between the two versions of the exam skills measured and the third column describes the extent of the changes.

Skill area prior to October 18, 2022	Skill area as of October 18, 2022	Change
Audience profile		Major
Manage Azure resources for machine learning	Design and prepare a machine learning solution	% of exam decreased
	Design a machine learning solution	Added
Create an Azure Machine Learning workspace	Manage an Azure Machine Learning workspace	Major

Manage data in an Azure Machine Learning workspace	Manage data in an Azure Machine Learning workspace	Minor
Manage compute for experiments in Azure Machine Learning	Manage compute for experiments in Azure Machine Learning	Major
Implement security and access control in Azure Machine Learning		Removed
Set up an Azure Machine Learning development environment		Removed
Set up an Azure Databricks workspace		Removed
Run experiments and train models	Explore data and train models	% of exam increased
	Explore data by using data assets and data stores	Added
Create models by using Azure Machine Learning designer	Create models by using the Azure Machine Learning designer	Major
Run model training scripts		Removed
Generate metrics from an experiment run		Removed
Use Automated Machine Learning to create optimal models	Use Automated Machine Learning to explore optimal models	Major
	Use Notebooks for custom model training	Added
Tune hyperparameters with Azure Machine Learning	Tune hyperparameters with Azure Machine Learning	Minor
Deploy and operationalize machine learning solutions	Prepare a model for deployment	% of exam decreased
	Run model training scripts	Added
Select compute for model deployment		Removed
Deploy a model as a service		Removed
Manage models in Azure Machine Learning	Manage models in Azure Machine Learning	Major; Reordered

Create an Azure Machine Learning pipeline for batch inferencing		Removed
Publish an Azure Machine Learning designer pipeline as a web service		Removed
Implement pipelines by using the Azure Machine Learning SDK	Implement training pipelines	Major; Reordered
Apply ML Ops practices		Removed
Implement responsible machine learning		Removed
Use model explainers to interpret models		Removed
Describe fairness considerations for models		Removed
Describe privacy considerations for data		Removed
	Deploy and retrain a model	Added
	Deploy a model	Added
	Apply machine learning operations (MLOps) practices	Added

Audience Profile

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

Responsibilities for this role include designing and creating a suitable working environment for data science workloads; exploring data; training machine learning models; implementing pipelines; running jobs to prepare for production; and managing, deploying, and monitoring scalable machine learning solutions.

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

Functional groups

Design and prepare a machine learning solution (20–25%)

Design a machine learning solution

- Determine the appropriate compute specifications for a training workload
- Describe model deployment requirements
- Select which development approach to use to build or train a model

Manage an Azure Machine Learning workspace

- Create an Azure Machine Learning workspace
- Manage a workspace by using developer tools for workspace interaction
- Set up Git integration for source control

Manage data in an Azure Machine Learning workspace

- Select Azure Storage resources
- Register and maintain datastores
- Create and manage data assets

Manage compute for experiments in Azure Machine Learning

- Create compute targets for experiments and training
- Select an environment for a machine learning use case
- Configure attached compute resources, including Azure Databricks and Azure Synapse Analytics
- Monitor compute utilization

Explore data and train models (35–40%)

Explore data by using data assets and data stores

- Load and transform data
- Analyze data by using Azure Data Explorer
- Use differential privacy

Create models by using the Azure Machine Learning designer

- Create a training pipeline
- Consume data assets from the designer
- Use designer components to define a pipeline data flow
- Use custom code components in designer
- Evaluate the model, including responsible AI guidelines

Use automated machine learning to explore optimal models

- Use automated machine learning for tabular data
- Use automated machine learning for computer vision

- Use automated machine learning for natural language processing (NLP)
- Select and understand training options, including preprocessing and algorithms
- Evaluate an automated machine learning run, including responsible AI guidelines

Use notebooks for custom model training

- Develop code by using a compute instance
- Consume data in a notebook
- Track model training by using MLflow
- Evaluate a model
- Train a model by using Python SDK
- Use the terminal to configure a compute instance

Tune hyperparameters with Azure Machine Learning

- Select a sampling method
- Define the search space
- Define the primary metric
- Define early termination options

Prepare a model for deployment (20–25%)

Run model training scripts

- Configure job run settings for a script
- Configure compute for a job run
- Consume data from a data asset in a job
- Run a script as a job by using Azure Machine Learning
- Use MLflow to log metrics from a job run
- Use logs to troubleshoot job run errors
- Configure an environment for a job run
- Define parameters for a job

Implement training pipelines

- Create a pipeline
- Pass data between steps in a pipeline
- Run and schedule a pipeline
- Monitor pipeline runs
- Create custom components
- Use component-based pipelines

Manage models in Azure Machine Learning

- Describe MLflow model output
- Identify an appropriate framework to package a model

- Assess a model by using responsible AI guidelines

Deploy and retrain a model (10–15%)

Deploy a model

- Configure settings for real-time deployment
- Configure compute for a batch deployment
- Deploy a model to a real-time endpoint
- Deploy a model to a batch endpoint
- Test a real-time deployed service
- Invoke the batch endpoint to start a batch scoring job

Apply machine learning operations (MLOps) practices

- Trigger an Azure Machine Learning pipeline, including from Azure DevOps or GitHub
- Automate model retraining based on new data additions or data changes
- Define event-based retraining triggers