Study guide for 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
Review the skills measured as of March 14, 2023	This list represents the skills measured AFTER the date provided. Study this list if you plan to take the exam AFTER that date.
Review the skills measured prior to March 14, 2023	Study this list of skills if you take your exam PRIOR to the date provided.
Change log	You can go directly to the change log if you want to see the changes that will be made on the date provided.
How to earn the certification	Some certifications only require passing one exam, while others require passing multiple exams.
Certification renewal	Microsoft associate, expert, and specialty certifications expire annually. You can renew by passing a **free** online assessment on Microsoft Learn.
Your Microsoft Learn profile	Connecting your certification profile to Learn allows you to schedule and renew exams and share and print certificates.
Passing score	A score of 700 or greater is required to pass.
Exam sandbox	You can explore the exam environment by visiting our exam sandbox.

Microsoft

Useful links	Description
Request accommodations	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	Are you ready to take the exam or do you need to study a bit more?

Updates to the exam

Our exams are updated periodically to reflect skills that are required to perform a role. We have included two versions of the Skills Measured objectives depending on when you are taking the exam.

We always update the English language version of the exam first. 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 as of March 14, 2023

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.

- Design and prepare a machine learning solution (20–25%)
- Explore data and train models (35–40%)

Microsoft

- Prepare a model for deployment (20–25%)
- Deploy and retrain a model (10–15%)

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 Apache Spark pools
- Monitor compute utilization

Explore data and train models (35–40%)

Explore data by using data assets and data stores

- Access and wrangle data during interactive development
- Wrangle interactive data with Apache Spark

Create models by using the Azure Machine Learning designer

- Create a training pipeline
- Consume data assets from the designer
- 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

Microsoft

- Evaluate an automated machine learning run, including responsible AI guidelines

Use notebooks for custom model training

- Develop code by using a compute instance
- 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

Microsoft

Deploy and retrain a model (10–15%)

Deploy a model

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

Apply machine learning operations (MLOps) practices

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

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
	Azure Synapse Analytics
	MLflow and 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

Microsoft

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 March 14, 2023	Skill area as of March 14, 2023	Change
Design and prepare a machine learning solution	**Design and prepare a machine learning solution**	No % change
Design a machine learning solution	Design a machine learning solution	No change
Manage an Azure Machine Learning workspace	Manage an Azure Machine Learning workspace	No change
Manage data in an Azure Machine Learning workspace	Manage data in an Azure Machine Learning workspace	No change
Manage compute for experiments in Azure Machine Learning	Manage compute for experiments in Azure Machine Learning	Minor
Explore data, and train models	**Explore data and train models**	No % change
Explore data by using data assets and data stores	Explore data by using data assets and data stores	Major
Create models by using the Azure Machine Learning designer	Create models by using the Azure Machine Learning designer	Minor
Use automated machine learning to explore optimal models	Use automated machine learning to explore optimal models	No change
Use notebooks for custom model training	Use notebooks for custom model training	Minor
Tune hyperparameters with Azure Machine Learning	Tune hyperparameters with Azure Machine Learning	No change
Prepare a model for deployment	**Prepare a model for deployment**	No % change
Run model training scripts	Run model training scripts	No change
Implement training pipelines	Implement training pipelines	No change
Manage models in Azure Machine Learning	Manage models in Azure Machine Learning	No change
Deploy and retrain a model	**Deploy and retrain a model**	No % change

Microsoft

Skill area prior to March 14, 2023	Skill area as of March 14, 2023	Change
Deploy a model	Deploy a model	Minor
Apply machine learning operations (MLOps) practices	Apply machine learning operations (MLOps) practices	Minor

Skills measured prior to March 14, 2023

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.

- Design and prepare a machine learning solution (20–25%)
- Explore data, and train models (35–40%)
- Prepare a model for deployment (20–25%)
- Deploy and retrain a model (10–15%)

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

Microsoft

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 SDKv2
- 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

Microsoft

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

Microsoft