Study Guide to Clear Microsoft Certified Azure Data Scientist Associate Exam (DP-100) in 60 Hours

Below is a study guide that helps you clear the Microsoft DP-100: Designing and Implementing a Data Science Solution on Azure in just 60 Hours.

The guide assumes that you have some basic idea on Machine Learning and have at least done some basic End-to-End problems on Classification and Regression.

The first step is to go through the official DP-100 page on Microsoft, https://www.microsoft.com/en-us/learning/exam-dp-100.aspx, to get an understanding of what is to be expected in the exam. From the skills measured section, based on the percentage we can see that majority of the questions will be from Model Selection/Train/Evaluation, EDA/Data Transformation, Feature Engineering and Development Environment/Deployment.

There are 35 questions in the exam and roughly 50% of them are based on generic Machine Learning topics and the type of questions include:

- Case Study with 4-6 Questions.
- Multiple Choice Single Answer
- Multiple Choice Multiple Answers
- Arrange in Correct Order
- Complete the Code

The exam started with the Case Study and once you complete answering all the questions in this section and exit, you cannot go back and review them. So, the ideal approach here is to read through the whole case study once and then go through the part of the section based on the question. The case study could be about a classification problem and a question could be how to fix an overfitting issue.

The other types of questions are self explanatory and you can try the elimination approach to answer the unknown question as there is no negative scoring.

The skills measured include:

- Develop models (40-45%) - This section carries the highest percentage. Most of the questions in the area will be generic and they are programming language and SDK agnostic. So, even if you have worked on scikit-learn you will be able to correctly answer the questions.
- Prepare data for modeling (25-30%) - This section carries the second highest percentage and it will be a mix of generic questions and topics specific to the Microsoft Platform. There could be few tricky questions in this area.
- Perform feature engineering (15-20%) - Though this section carries a small percentage, I felt there are few tricky questions in this area.

- Define and prepare the development environment (15-20%) - This section will contain all the questions related to the Microsoft Azure platform. So, you should have at least a fair idea on the different tools, platforms and services.

Below is a rough guide on how to spend you 60 Hours to clear the DP 100 exam.

1. Complete Free Interactive Training on Microsoft Learn (10 Hours)

Though the complete set of tutorials will take around 30 Hours, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE2PLKZ, you can fast forward few of the tutorials or even skip them as all of them will not appear on the exam.

2. Azure Machine Learning Studio (10 Hours)

This is the starting point to learn more about the Machine Learning platform from Microsoft - Machine Learning Studio, which is a drag-and-drop tool you can use to build, test, and deploy predictive analytics solutions. You need to spend some good quality of time to explore Machine Learning Studio as this is the basic building block and most of the questions specific to Microsoft are based on this application. You can create a free account to explore the Microsoft Azure Machine Learning Studio. Make sure you create few experiments and explore all the options provided by Machine Learning Studio. Click each of the module and go through its properties and also make sure you click the 'Quick Help' on the right bottom of the page to read more about this module.

3. Azure AI Gallery (10 Hours)

Choose the 'Experiments' section on https://gallery.azure.ai/browse to explore the different experiments on Classification, Regression, Clustering, Anomaly Detection, Data Transformation, Train, Test, Evaluate, Statistical Functions, Text Analysis and others. You can start with the content from Microsoft and explore the experiments from other users if required.

4. Machine Learning Studio Module Reference (20 Hours)

Spend a very good amount of time in this section. Read through all the modules and carefully go through the options available for each module, parameters supported, data types on which this module works and any comparisons to other related modules. Each one of the topics on the list is very important. It is difficult to clear the exam without knowing all the topics in this page. So please dedicate a good amount of quality time on this page.

5. General Topics (10 Hours)

Please go through the below topics as there will be few questions on them too. You may just spend some time to understand what this application or service does and in which stage of your work will this be applicable.

- Team Data Science Process - https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/
- Azure Data Science Virtual Machine - https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/
- What is Azure HDInsight - https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-overview
- Machine learning on HDInsight - https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-machine-learning-overview
- What is Apache Spark in Azure HDInsight - https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-overview
- What is Apache Hadoop in Azure HDInsight? - https://docs.microsoft.com/en-us/azure/hdinsight/hadoop/apache-hadoop-introduction
- What is ML Services in Azure HDInsight - https://docs.microsoft.com/en-us/azure/hdinsight/r-server/r-server-overview
- What is Azure Machine Learning - https://docs.microsoft.com/en-gb/azure/machine-learning/service/overview-what-is-azure-ml
- What are compute targets in Azure Machine Learning service - https://docs.microsoft.com/en-gb/azure/machine-learning/service/concept-compute-target
- Deep learning vs. machine learning - https://docs.microsoft.com/en-gb/azure/machine-learning/service/concept-deep-learning-vs-machine-learning
- The Microsoft Cognitive Toolkit Overview - https://docs.microsoft.com/en-us/cognitive-toolkit/
- Intro to Microsoft Cognitive Toolkit - https://www.youtube.com/watch?v=9gDDO5ldT-4&feature=youtu.be
- Azure Notebooks - https://notebooks.azure.com
- Azure Machine Learning Studio vs Azure Machine Learning Services - https://www.codit.eu/blog/azure-machine-learning-studio-vs-services/
- Microsoft Machine Learning for Apache Spark - https://azuremlbuild.blob.core.windows.net/pysparkapi/intro.html
- Apache Zeppelin Notebooks - https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-zeppelin-notebook
- What are the machine learning products at Microsoft? - https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/data-science-and-machine-learning
- Azure HDInsight - https://azure.microsoft.com/en-in/services/hdinsight/
- Azure Storage Documentation - https://docs.microsoft.com/en-us/azure/storage/
- Azure and Power BI - https://docs.microsoft.com/en-us/power-bi/service-azure-and-power-bi
- Azure Kubernetes Service (AKS) - https://azure.microsoft.com/en-in/services/kubernetes-service/
- What is Kubernetes - https://azure.microsoft.com/en-in/topic/what-is-kubernetes/
- Azure Machine Learning SDK for Python - https://docs.microsoft.com/en-us/python/api/overview/azure/ml/intro?view=azure-ml-py
- scikit-learn Tutorials - https://scikit-learn.org/stable/tutorial/index.html