

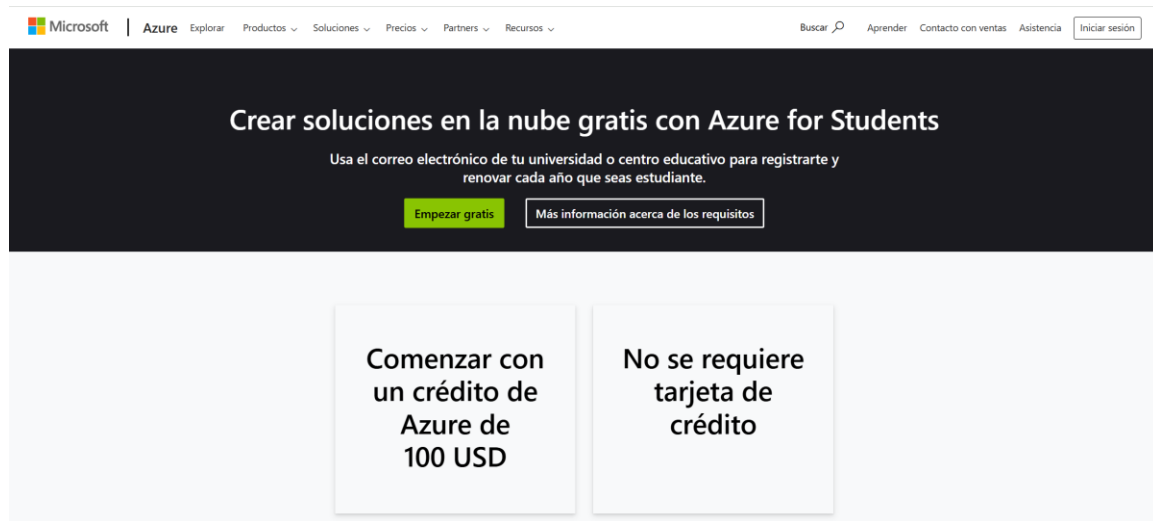


DML-004 - Guía de Laboratorio 1

Esta guía paso a paso te ayudará a crear tu cuenta en Azure como estudiante de la Universidad Católica Boliviana, configurar tu entorno de trabajo en Azure Machine Learning Studio, crear un compute instance, y cargar un dataset desde Kaggle para trabajar con el SDK v2.

PARTE 1 – CREAR CUENTA DE ESTUDIANTE EN AZURE FOR STUDENTS

1. Ir a <https://azure.microsoft.com/en-us/free/students/>



2. Presionar 'Start free' o 'Empieza gratis'.
3. Iniciar sesión con tu correo institucional (terminado en @ucb.edu.bo).
4. Completar el formulario con tus datos reales.
5. Verificar tu cuenta mediante el código enviado a tu correo.
6. Listo: ahora tienes \$100 en créditos para usar en Azure por 1 año.





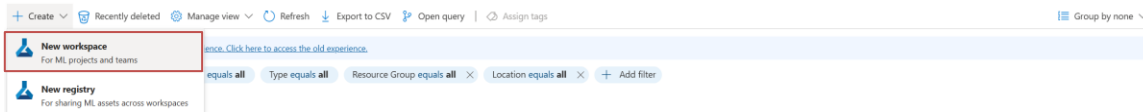
PARTE 2 – CREAR UN WORKSPACE EN AZURE MACHINE LEARNING

1. En el portal de Azure (https://portal.azure.com) buscar 'Machine Learning'.

Azure services



2. Seleccionar 'Create'.



3. Elegir tu suscripción de Azure for Students.
4. En Resource Group, crear uno nuevo llamado rg-mlops-ucb.
5. Workspace Name: ws-mlops-ucb.
6. Region: seleccionar una región disponible (ej. South Central US).
7. Presionar 'Review + Create' y luego 'Create'.

Azure Machine Learning

Create a machine learning workspace

Basics Inbound Access Outbound Access Encryption Identity Tags Review + create

Resource details

Every workspace must be assigned to an Azure subscription, which is where billing happens. You use resource groups like folders to organize and manage resources, including the workspace you're about to create.

[Learn more about Azure resource groups](#)

Subscription *	Azure for Students
Resource group *	rg-mlops-ucb

Workspace details

Configure your basic workspace settings like its storage connection, authentication, container, and more. [Learn more](#)

Name *	ws-mlops-ucb
Region *	South Central US
Storage account *	(new) wsmlopsuc0173583453
Key vault *	(new) wsmlopsuc2008733819
Application insights *	(new) wsmlopsuc0033196058
Container registry	None

8. Esperar a que finalice el despliegue.



PARTE 3 – ENTRAR A AZURE MACHINE LEARNING STUDIO

1. Ir a <https://ml.azure.com>.
2. Seleccionar el workspace creado (ws-mlops-ucb).
3. Verificar que puedas acceder al panel principal.

The screenshot shows the Azure Machine Learning Studio interface for the workspace 'ws-mlops-ucb'. The left sidebar contains navigation links: Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Resource visualizer, Events, Settings, Monitoring, Automation, and Support + troubleshooting. The main area displays the 'Essentials' tab with a table of workspace details:

Property	Value
Resource group	rg-mlops-ucb
Location	South Central US
Subscription	Azure for Students
Storage	wsmlpsucb6379269930
Provisioning State	Succeeded
Studio web URL	https://ml.azure.com?tid=cc28633f-12b8-46cb-bc15-951...
Container Registry	...
Key Vault	wsmlpsucb6089620548
Application Insights	wsmlpsucb0346006038
MLflow tracking URI	azureml://southcentralus.api.azureml.ms/mlflow/v1.0/su...

Below the table, there is a large 'Launch studio' button and a message: 'Work with your models in Azure Machine Learning Studio. The Azure Machine Learning Studio is a web app where you can build, train, test, and deploy ML models. Launch it now to start exploring, or learn more about the Azure Machine Learning studio'.

PARTE 4 – DESCARGAR DATASET DESDE KAGGLE

1. Ir a <https://www.kaggle.com> y registrarte (puede ser con Google o correo UCB).
2. Buscar el dataset 'Heart Disease' <https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset?resource=download>.
3. Descargar el archivo ZIP.
4. Extraer el archivo CSV en tu computadora.

PARTE 5 BIS – CREAR UN COMPUTE INSTANCE

1. En Azure Machine Learning Studio, ir a la sección 'Compute'.

The screenshot shows the 'Compute' section of the Azure Machine Learning Studio interface. The left sidebar has a red box around the 'Compute' link. The main area displays the 'Compute instances' tab with a table of compute instances:

Name	State	Idle shutdown	Applications	Size	Created on	Assn
UCBmlops	Running	1 hour	JupyterLab Jupyter VS Code (Web)	Standard_DS11_v2	Aug 10, 2025 7:52 PM	MAL



2. Seleccionar la pestaña 'Compute instances'.
3. Presionar '+ New'.
4. Configurar:
 - Nombre: compute-ucb
 - Tamaño recomendado: STANDARD_DS11_V2 (o más pequeño para ahorrar créditos).
 - Región: la misma que el workspace.
5. Presionar 'Create' y esperar que esté en estado 'Running'.

Create compute instance

1 Required settings

2 Scheduling optional

3 Security optional

4 Applications optional

5 Tags optional

6 Review

Configure required settings

Select the name and virtual machine size you would like to use for your compute instance

Note that a compute instance can not be shared. It can only be used by a single assigned user. By default, it will be assigned to the creator and you can change this to a different user in the Security step.

Compute name *

Virtual machine type ☒ CPU ☐ GPU

Virtual machine size ☒ Select from recommended options ☐ Select from all options

Name ↑	Category	Workload types	Available quota ⓘ	Cost ⓘ
<input checked="" type="radio"/> Standard_DS11_v2 2 cores, 14GB RAM, 28GB storage	Memory optimized	Development on Notebooks (or other IDE) and light weight testing	4 cores	--
<input type="radio"/> Standard_DS3_v2 4 cores, 14GB RAM, 28GB storage	General purpose	Classical ML model training on small datasets	4 cores	--
<input type="radio"/> Standard_E4ds_v4 4 cores, 32GB RAM, 150GB storage	Memory optimized	Data manipulation and training on medium-sized datasets (1-10GB)	4 cores	\$0.37/hr
<input type="radio"/> Standard_F4s_v2 4 cores, 8GB RAM, 32GB storage	Compute optimiz...	Data manipulation and training on large datasets (>10 GB)	16 cores	\$0.20/hr

[Review + Create](#) [Back](#) [Next](#) [Cancel](#)

6. IMPORTANTE: al terminar, detenerlo para no gastar créditos.

PARTE 6 – SUBIR DATASET COMO DATA ASSET EN AZURE ML

1. En Azure ML Studio, ir a la sección 'Data'.
2. Clic en '+ Create'.

The screenshot shows the Azure ML Studio interface. On the left sidebar, under the 'Assets' section, the 'Data' option is highlighted with a red box. The main panel displays the 'Data' section with tabs for 'Data assets', 'Datastores', 'Dataset monitors', 'Data import', and 'Data connections'. The 'Data assets' tab is active, showing a table of data assets. The table has columns for Name, Version, Data source, Created on, Modified on, Type, Properties, Created by, and Tags. One data asset is listed: 'heart-disease' with version 1, data source 'workspaceblobstore', created on Aug 10, 2025 8:04 PM, modified on Aug 10, 2025 8:04 PM, type 'Folder', and created by 'MAURICIO ALEJANDRO ...'.



3. Nombre: heart-disease
4. Tipo de archivo: File

Create data asset

1 Data type

2 Data source

Set the name and type for your data asset

Name *

heart_disease

Description

Data asset description

Type *

File

Use cases for data types

When should I use File type?

The File type is recommended in most scenarios when you are working with a single data file of any type (including tabular data). This type allows you to specify a file location by URI in a storage location on your local computer, an attached Datastore, blob/ADLS storage, or a publicly available http(s) location. There are many types of supported URIs. In the Azure Machine Learning CLI v2 or Python SDK v2, this data type is called `uri_file`. [Learn more about the uri_file type](#)

When should I use Folder type?

The Folder type has all the same capabilities and use cases as the File type, but is used when specifying a folder location. In the Azure Machine Learning CLI v2 or Python SDK v2, this data type is called `uri_folder`. [Learn more about the uri_folder type](#)

When should I use Table type?

The Table type is most useful for advanced scenarios where you might need to abstract the schema definition for easier sharing. You should use it when you have complex transformations and schema you want to capture in a reusable asset. For simpler tabular data, the File and Folder types are recommended. If you choose the Table type, in the Azure Machine Learning CLI v2 or Python SDK v2, this data type is called `mltable`. [Learn more about the mltable type](#)

Back

Next

Cancel

5. Seleccionar 'From local files'.

Create data asset

3 Data type

4 Data source

5 Destination storage type

6 File or folder selection

7 Review

Choose a source for your data asset

Choose the data source you want to create your asset from. A data source can be from a local storage location on your computer, from an attached datastore, from Azure storage, or from a publicly available web location.

From Azure storage

Create a data asset from registered data storage services including Azure Blob Storage, Azure file share, and Azure Data Lake.

From local files

Create a data asset by uploading files from your local drive.

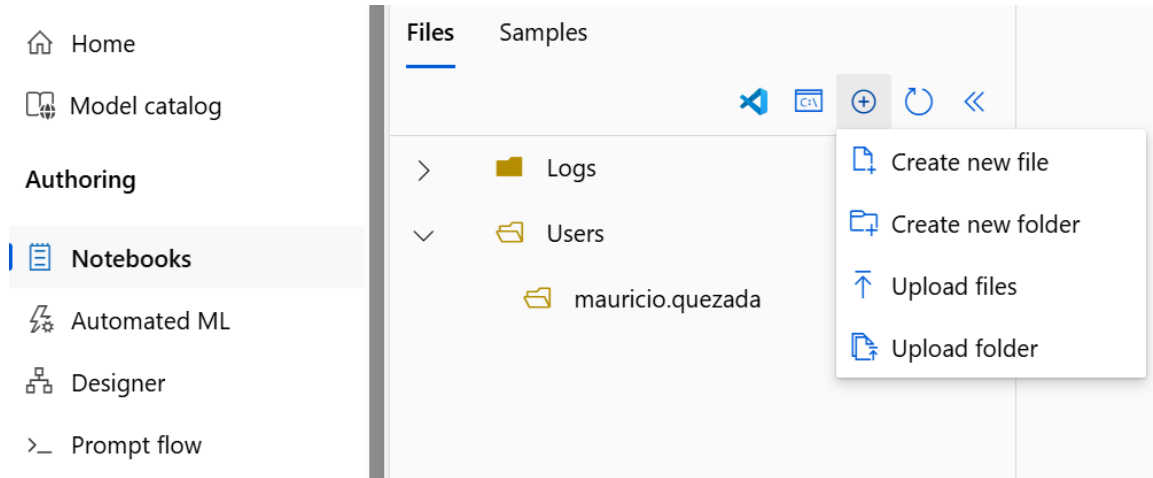
From web files

Create a data asset from a single file located at a public web URL.

6. Cargar el CSV descargado de Kaggle.
7. Guardar.

PARTE 7 – ABRIR NOTEBOOK Y CARGAR DATASET CON SDK V2

1. Crear un nuevo Notebook



2. Iniciar código con pandas

Código de ejemplo:

```
import pandas as pd
from azure.ai.ml import MLClient
from azure.identity import DefaultAzureCredential

ml_client = MLClient.from_config(credential=DefaultAzureCredential())
data_asset = ml_client.data.get('heart-disease', version='1')
df = pd.read_csv(data_asset.path)
df.head()
```

Este código conecta tu notebook con Azure ML, obtiene el dataset 'heart-disease' y lo carga en un DataFrame de Pandas.