



códigofacilito

## — Agenda

- Exploración de las herramientas de desarrollo para la interacción de áreas de trabajo
- Datos y destinos de proceso en Azure Machine Learning

Irving Uribe – ML Engineer / Data  
Tech Lead / AI profesor





## Exploración de las herramientas de desarrollo para la interacción de áreas de trabajo

<https://learn.microsoft.com/es-es/training/modules/explore-developer-tools-for-workspace-interaction/>





# [Repaso] Diferencia entre la consola de Azure y la de Azure AI (ML studio)



La primera, es la consola de administración de la **cuenta**, y la segunda la consola para la gestión de recursos de **Machine Learning**.

The image shows a comparison between the Azure portal and the Azure Machine Learning Studio interface. A red dashed circle highlights the 'Azure Machine Learning' icon in the 'Azure services' section of the Azure portal. A red arrow points from this icon to the 'Azure AI | Machine Learning Studio' interface shown in the foreground.

**Azure portal (Background):**

- Header:** Microsoft Azure, Search resources, services, and docs (G+ /), Copilot, User profile (ayuda@codigofacilito.c...), DEFAULT DIRECTORY.
- Azure services:** Create a resource, Subscriptions, Resource groups, Quotas, **Azure Machine Learning** (highlighted), Storage accounts, Azure OpenAI, Azure AI services, Azure Lab Services, More services.
- Resources:** Recent, Favorite. List of resource names: uriel-facilito-resource-group, handsonautoml5347626068, ayuda-rg, rg-cf. See all.

**Azure AI | Machine Learning Studio (Foreground):**

- Header:** Azure AI | Machine Learning Studio, Search, All workspaces.
- Left sidebar:** Default Directory, Workspaces, Feature stores, Hubs (PREVIEW), Registries, Shared assets, Components, Environments, Models, Data, Admin.
- Main content:**
  - Welcome to the Azure Machine Learning Studio:** Create a new workspace, or open one of your recent workspaces to pick up where you left off.
  - Recent workspaces:** handson\_AutoML (Subscription: Suscripción de Azure 1), ws-cc-ml (Subscription: Suscripción de Azure 1).
  - Learning components:** Build AI solutions with Azure Machine Learning, Introduction to the Azure Machine Learning SDK, Train a machine learning model with Azure ML.
- Right sidebar:** Create workspace, View all workspaces.



## [Repaso] Conexión mediante el SDK de Python



Se puede instalar el sdk mediante el gestor **pip**. Para usar las librerías se solicitarán los **3 insumos de autenticación** `subscription_id`, `resource_group` y `workspace_name`

Python

```
from azure.ai.ml import MLClient
from azure.identity import DefaultAzureCredential

ml_client = MLClient(
    DefaultAzureCredential(), subscription_id, resource_group, workspace
)
```

```
pip install azure-ai-ml
```

*Configuración de cliente*



*Ejecución de Job*

Python

```
from azure.ai.ml import command

# configure job
job = command(
    code="./src",
    command="python train.py",
    environment="AzureML-sklearn-0.24-ubuntu18.04-py37-cpu@latest",
    compute="aml-cluster",
    experiment_name="train-model"
)

# connect to workspace and submit job
returned_job = ml_client.create_or_update(job)
```





## [Repaso] Conexión mediante el Azure CLI



Se puede conectar a los recursos de la cuenta por medio del Command Line Interface (CLI), este permite crear instancias de procesamiento, y clusters

Azure CLI

```
az extension add -n ml -y
```

*Extensión ml*

yml

```
$schema: https://azuremlschemas.azureedge.net/latest/amlCompute.schema.json
name: aml-cluster
type: amlcompute
size: STANDARD_DS3_v2
min_instances: 0
max_instances: 5
```

Azure CLI

Copiar

```
az ml compute create --file compute.yml --resource-group my-resource-group --workspace-name my-workspa
```



# [Duda] ¿Cuál es la diferencia entre una instancia local y una instancia remota?

Mientras que una instancia local (Tu equipo) consume recursos de tu máquina, una instancia remota, necesita de una máquina virtual y contenedor para poder operar.



EXPLORADOR

OCR\_RECOMENDACIONES

- .cache
- .conda
- .ipynb\_checkpoints
- Auditorias
- Datos\_Etiquetados
- documento
- Inferencias
- PDFs2
- 2024-DAI-24110 Infor...
- 2024-GCA-24060 - Inf...
- Alto\_ngramas.xlsx
- Auditoria1.pdf
- Auditoria1.txt
- Auditoria2.pdf
- Auditorias.zip

ESQUEMA

LÍNEA DE TIEMPO Ngramas\_Alt...

El Historial local realizará un seguimiento de los cambios recientes a

	A	B	C
1	Ngrama	Frecuencia	Porcentaje de
2	('proceso', 'dis	2	41.18
3	('demanda', 'p	2	38.24
4	('tecnologicas	2	29.41
5	('soportar', 'pr	2	29.41
6	('proceso', 'pla	2	29.41
7	('integrado', 'c	2	29.41
8	('disposicion',	2	23.53
9	('implementac	2	20.59
10	('final', 'materi	2	20.59
11	('material', 'cla	2	17.65
12	('clasificado',	2	17.65
13	('herramienta'	2	11.76
14	('planeacion',	2	11.76
15	('excedente',	2	11.76
16	('sobrante', 're	2	11.76
17	('superintende	2	11.76

Local

- Default Directory
- Workspaces
- Feature stores
- Hubs PREVIEW
- Registries
- Shared assets
- Components
- Environments**
- Models
- Data
- Admin
- Quota

tensorflow-2.16-cuda12

Version: 1 (latest)

Overview Context Jobs

1

Environment operating system  
Linux

Azure container registry  
mcr.microsoft.com/azureml/curated/tensorflow-2.16-cuda12:1

Asset ID  
azureml://registries/azureml/environments/tensorflow-2.16-cuda12/versions/1

Description

An environment for deep learning with Tensorflow containing the Azure ML SDK and additional python packages.

Tags

GPU : Cuda12 OS : Ubuntu20.04 Preview Python : 3.10  
Tensorflow : 2.16 Training

Remota



## Hacer que los datos estén disponibles en Azure Machine Learning

<https://learn.microsoft.com/es-es/training/modules/make-data-available-azure-machine-learning/>

## Trabajo con destinos de procesos (Compute Targets) en Azure Machine Learning

<https://learn.microsoft.com/es-es/training/modules/work-compute-resources-azure-machine-learning/>





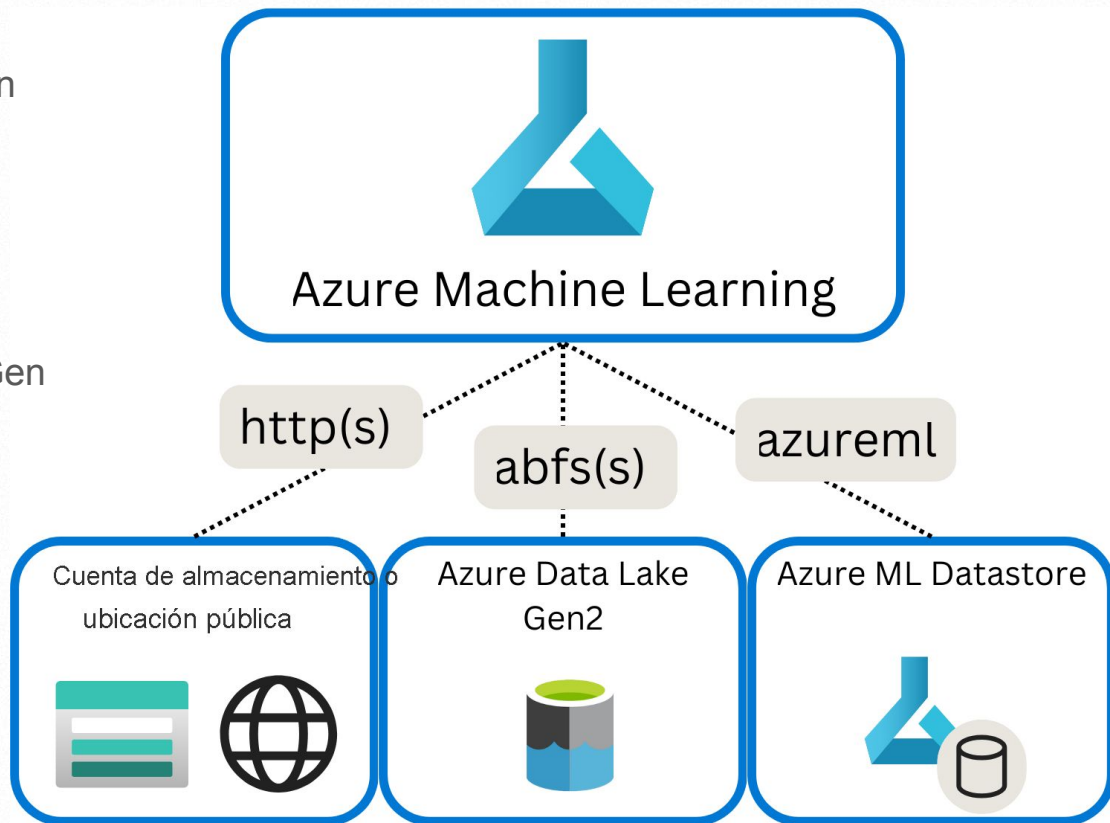


## [Repaso] Azure URI

Una URI hace referencia a la **ubicación de los datos**.



- **http(s)**: Almacenes de datos pública en una ubicación http(s) de Azure Blob Storage o disponible públicamente.
- **abfs(s)**: Almacenes de datos en una instancia de Azure Data Lake Storage Gen 2.
- **azureml**: Para los datos almacenados en una instancia de datos de AzureML, asociados a un Blob Storage.







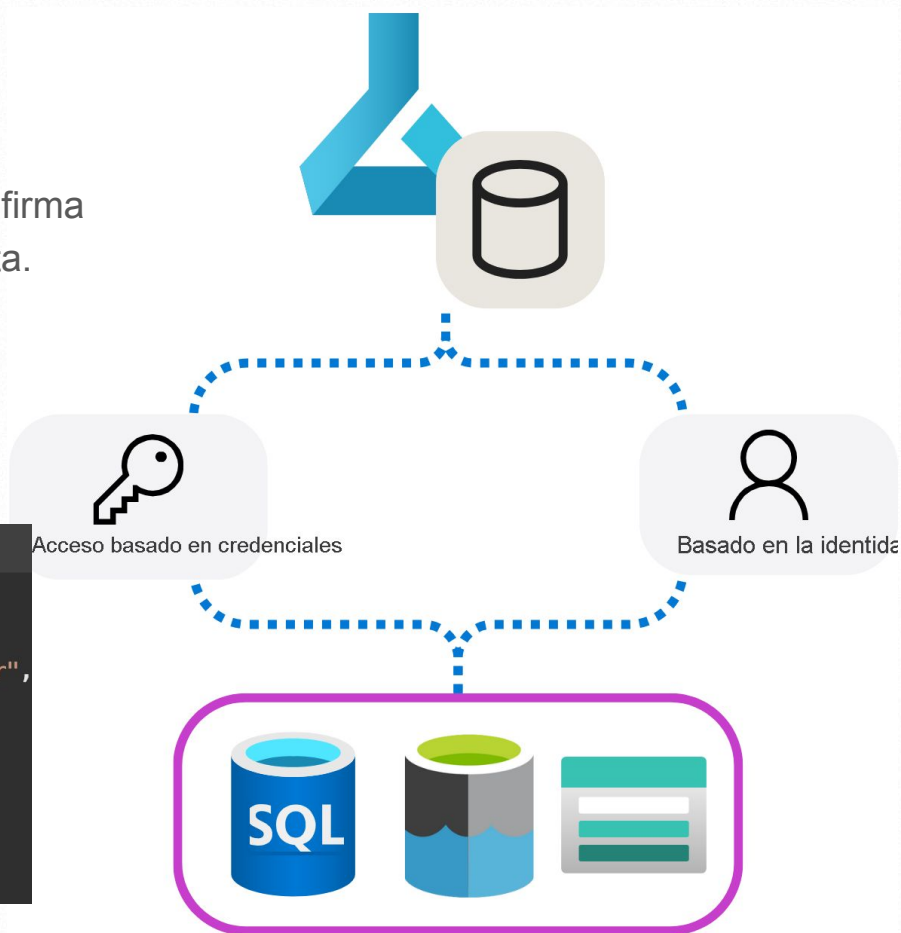
## [Repaso] Métodos de autenticación



- **Basado en credenciales:** Uso de un **Token** de firma de acceso compartido (SAS) o una clave de cuenta.
- **Basado en la identidad:** Use la identidad de Microsoft Entra o la **identidad administrada**.

Python

```
blob_datastore = AzureBlobDatastore(  
    name = "blob_example",  
    description = "Datastore pointing to a blob container",  
    account_name = "mytestblobstore",  
    container_name = "data-container",  
    credentials = AccountKeyConfiguration(  
        account_key="XXXXXXXXXXXXXXXXXXXX"  
    ),  
)  
ml_client.create_or_update(blob_datastore)
```





## [Repaso] Tipos de recursos de datos

Python

```
from azure.ai.ml.entities import Data
from azure.ai.ml.constants import AssetTypes

my_path = '<supported-path>'

my_data = Data(
    path=my_path,
    type=AssetTypes.URI_FILE,
    description="<description>",
    name="<name>",
    version="<version>"
)

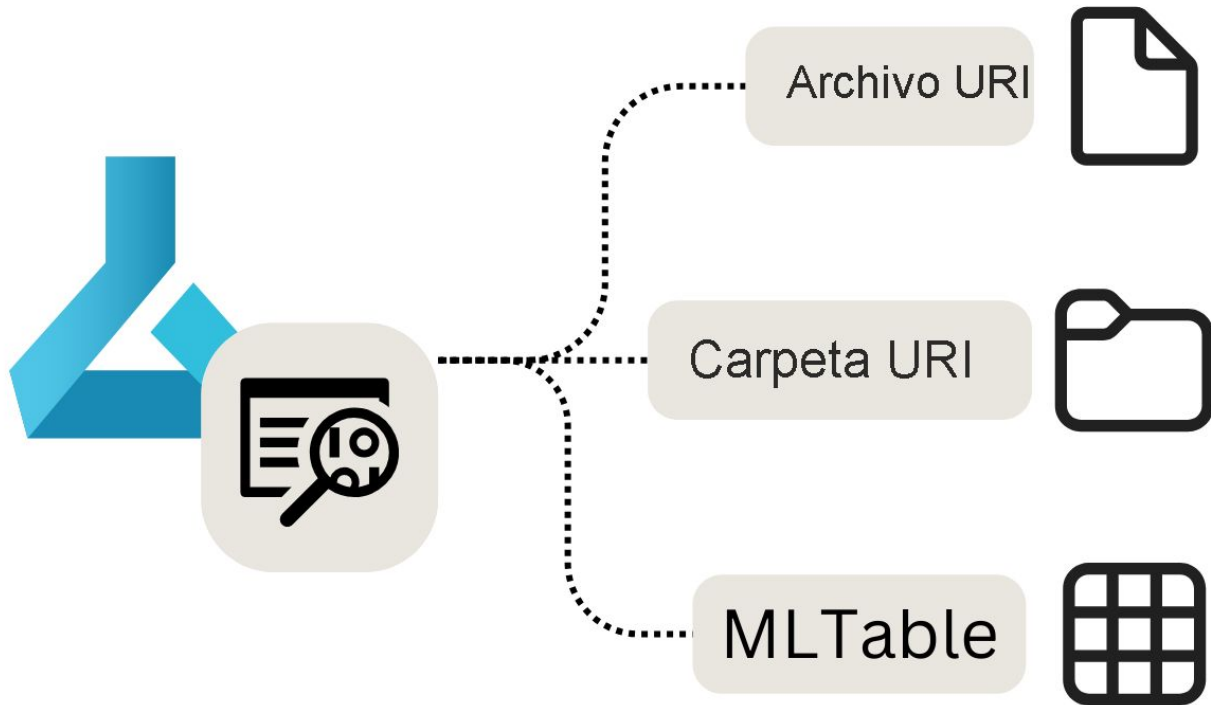
ml_client.data.create_or_update(my_data)
```

Python

```
import argparse
import pandas as pd

parser = argparse.ArgumentParser()
parser.add_argument("--input_data", type=str)
args = parser.parse_args()

df = pd.read_csv(args.input_data)
print(df.head(10))
```



*Creación y lectura de Archivo URI*



# [Repaso] Tipos de recursos de datos



yaml

```
type: mltable

paths:
  - pattern: ./*.txt
transformations:
  - read_delimited:
      delimiter: ','
      encoding: ascii
      header: all_files_same_headers
```

*Esquema para ML Table*

Python

```
from azure.ai.ml.entities import Data
from azure.ai.ml.constants import AssetTypes

my_path = '<path-including-mltable-file>'

my_data = Data(
    path=my_path,
    type=AssetTypes.MLTABLE,
    description="<description>",
    name="<name>",
    version="<version>"
)

ml_client.data.create_or_update(my_data)
```

Python

```
import argparse
import mltable
import pandas

parser = argparse.ArgumentParser()
parser.add_argument("--input_data", type=str)
args = parser.parse_args()

tbl = mltable.load(args.input_data)
df = tbl.to_pandas_dataframe()

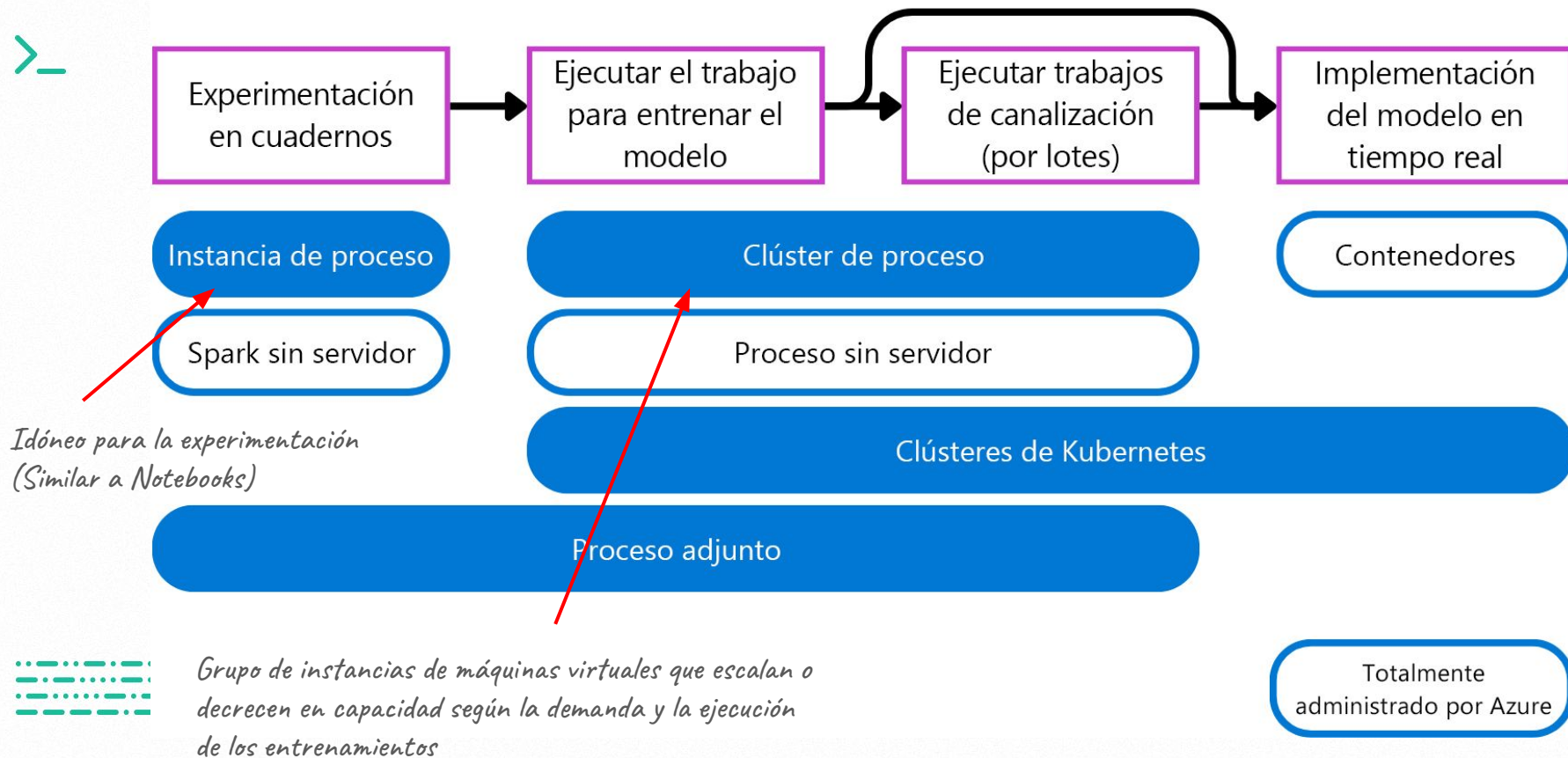
print(df.head(10))
```







## [Repaso] Flujo general del proceso





# [Repaso] Creación de instancias de proceso

Python

```
from azure.ai.ml.entities import ComputeInstance

ci_basic_name = "basic-ci-12345"
ci_basic = ComputeInstance(
    name=ci_basic_name,
    size="STANDARD_DS3_v2"
)
ml_client.begin_create_or_update(ci_basic).result()
```

Python

```
from azure.ai.ml.entities import AmlCompute

cluster_basic = AmlCompute(
    name="cpu-cluster",
    type="amlcompute",
    size="STANDARD_DS3_v2",
    location="westus",
    min_instances=0,
    max_instances=2,
    idle_time_before_scale_down=120,
    tier="low_priority",
)
ml_client.begin_create_or_update(cluster_basic).result()
```

## 1.- Instancia de proceso

Python

```
from azure.ai.ml import command

# configure job
job = command(
    code="./src",
    command="python diabetes-training.py",
    environment="AzureML-sklearn-0.24-ubuntu18.04-py37-cpu@latest",
    compute="cpu-cluster",
    display_name="train-with-cluster",
    experiment_name="diabetes-training"
)

# submit job
returned_job = ml_client.create_or_update(job)
aml_url = returned_job.studio_url
print("Monitor your job at", aml_url)
```

## 3.- Job (Trabajo)

## 2.- Cluster



## [Duda] ¿Para crear un cluster, debo crear primero una instancia de proceso?



Sí, es necesario tener al menos una instancia que puede tener un cluster con 2 o más instancias según las necesidades y la configuración solicitada

Python

```
from azure.ai.ml.entities import ComputeInstance

ci_basic_name = "basic-ci-12345"
ci_basic = ComputeInstance(
    name=ci_basic_name,
    size="STANDARD_DS3_v2"
)
ml_client.begin_create_or_update(ci_basic).result()
```

Python

```
from azure.ai.ml.entities import AmlCompute

cluster_basic = AmlCompute(
    name="cpu-cluster",
    type="amlcompute",
    size="STANDARD_DS3_v2",
    location="westus",
    min_instances=0,
    max_instances=2,
    idle_time_before_scale_down=120,
    tier="low_priority",
)
ml_client.begin_create_or_update(cluster_basic).result()
```







# [Duda] ¿Puedo subir información (DATA) sin usar código?



Sí es posible directamente desde el workspace donde vas a interactuar con el job de entrenamiento del modelo (Se verá más adelante en el curso)

## Create data asset

- 1 Data type
- 2 Data source
- 3 Destination storage type
- 4 File selection
- 5 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 Azure storage, or from a publicly available web location.



From a URI

Create a data asset by providing a web URL or a URI that represents a path to your data.

From local files



From local files

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

Upload files or folder

☐ Overwrite if already exists

### Upload list

Resumen.csv

21.35 KB/21.35 KB



Pandas (SDK v1)

Pandas (SDK v2)



```
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("Tabla", version="1")

df = pd.read_csv(data_asset.path)
df
```

Display as grid



With column headers



Column1	Column2	Column3	Column4	Column5	Column6
Archivo	Página	Alto	Medio	Bajo	null
2022 Au...	10	null	Hallazg...	null	null
2022 Au...	13	null	"Hallazg...	null	null
proyecci...	null	null	null	null	null
2022 Au...	14	null	"Hallazg...	null	null
Capex e...	null	null	null	null	null
2022 Au...	18	null	"Hallazg...	null	null



**[Duda]** ¿Cómo se pueden reutilizar los destinos de proceso en distintos experimentos para reducir costos y tiempo de desarrollo?



**Costos:** Usar Clusters de Computación Reutilizables y escalables



Python

```
from azure.ai.ml.entities import AmlCompute

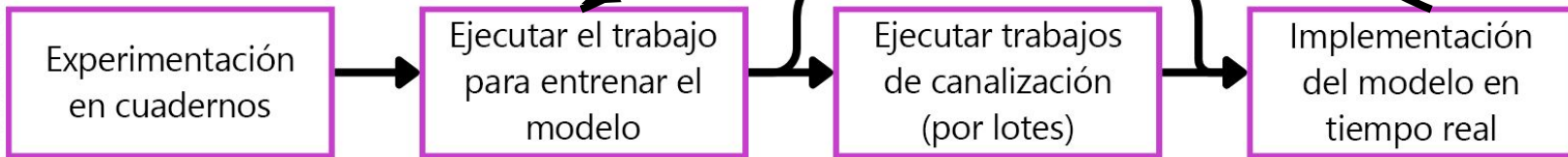
cluster_basic = AmlCompute(
    name="cpu-cluster",
    type="amlcompute",
    size="STANDARD_DS3_v2",
    location="westus",
    min_instances=0,
    max_instances=2,
    idle_time_before_scale_down=120,
    tier="low_priority",
)
ml_client.begin_create_or_update(cluster_basic).result()
```



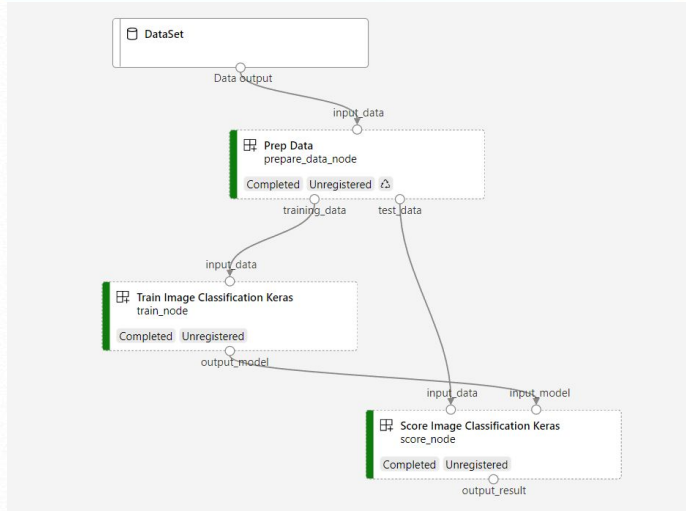
# [Duda] ¿Cómo se pueden reutilizar los destinos de proceso en distintos experimentos para reducir costos y tiempo de desarrollo?



**Tiempo:** Persistencia de Salidas en el Almacenamiento



**Tiempo:** Uso de Pipelines Reusables







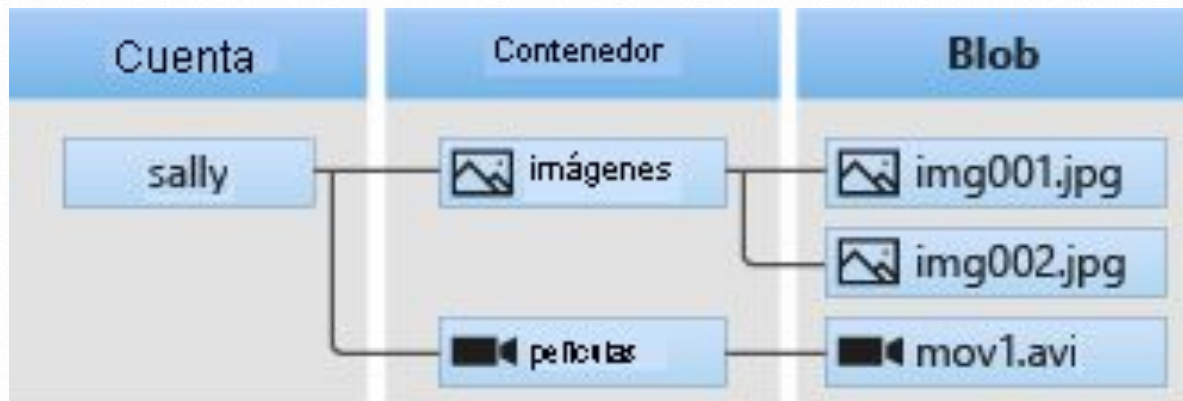
## [Duda] Diferencia entre instancia de BlobStorage y Contenedor



**Blob Storage:** Es el servicio completo de almacenamiento para datos no estructurados en Azure.

**Contenedor:** Es una subunidad dentro de Blob Storage que organiza los blobs. Es necesario crear un contenedor para almacenar blobs.

Por ejemplo: en un proyecto de Machine Learning, puedes tener un Blob Storage llamado `mlprojectstorage`, y dentro de él, contenedores como `datasets`, `models` y `logs`,





# [Duda] Diferencia entre DataStore y Contenedor

Aspecto	DataStore	Contenedor
Definición	Abstracción de acceso a servicios de almacenamiento en Azure Machine Learning.	Unidad organizativa dentro de un Blob Storage.
Función	Facilita el acceso a datos externos en scripts y pipelines de ML.	Agrupar blobs para organizar archivos en el almacenamiento.
Nivel	Abstracción de alto nivel en Azure ML.	Estructura de bajo nivel en el servicio de almacenamiento.
Ejemplo de Uso	Referenciar un Blob Storage desde Azure ML usando un alias y credenciales configuradas.	Almacenar datos dentro de un Blob Storage, agrupados por tipo o propósito.

