



códigofacilito

— Agenda

- Estructura del examen
- Repaso de lecciones (Introducción a Azure y Exploración de recursos)
- Preguntas del formulario
- Dudas adicionales

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





Estructura del examen de certificación DP-100

<https://learn.microsoft.com/es-es/credentials/certifications/resources/study-guides/dp-100>

<https://learn.microsoft.com/en-us/credentials/certifications/azure-data-scientist/?practice-assessment-type=certification>





Nota de conocimientos requeridos



La certificación requiere conocimientos en **Machine Learning**, y uso de las herramientas de **Azure Machine Learning** y **MLflow**

Aptitudes evaluadas a partir del 16 de octubre de 2024

Perfil del público

Como candidato a este examen, debe tener experiencia en la aplicación de ciencia de datos y aprendizaje automático para implementar y ejecutar cargas de trabajo de aprendizaje automático en Azure.

Las responsabilidades de este rol incluyen:

- Diseño y creación de un entorno de trabajo adecuado para cargas de trabajo de ciencia de datos.
- Exploración de datos.
- Entrenamiento de modelos de aprendizaje automático.
- Implementación de canalizaciones.
- Ejecución de trabajos para prepararse para la producción.
- Administración, implementación y supervisión de soluciones de aprendizaje automático.

Como candidato a este examen, debe tener conocimientos y experiencia en ciencia de datos mediante el uso de:

- Azure Machine Learning
- MLflow





Estructura del examen en porcentajes



El examen se divide en **4 secciones** que varían en porcentaje de cobertura en el examen (de manera aleatoria)

Points scored

Implementar y volver a

10,0%

- Implementar un modelo
- Aplicación de los procedimientos en las operaciones de aprendizaje automático (MLOps)

Preparación de un

25,0%

- Ejecución de scripts de entrenamiento del modelo
- Implementación de canalizaciones de entrenamiento
- Administración de modelos en Azure Machine Learning

- Diseño de una solución de aprendizaje automático
- Administración de un área de trabajo de Azure Machine Learning
- Administración de datos en un área de trabajo de Azure Machine Learning

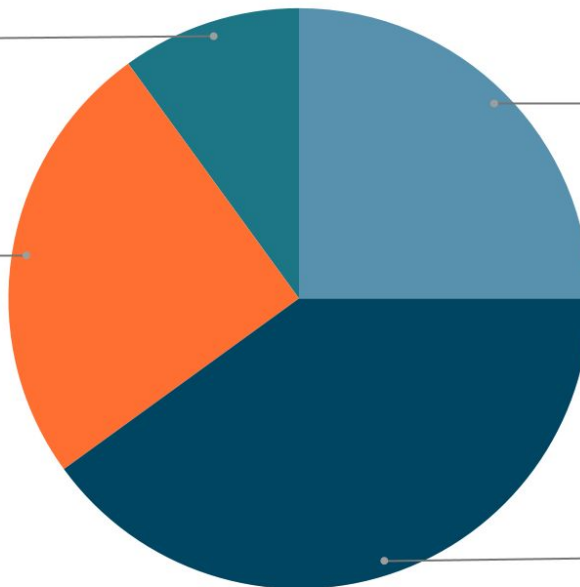
Diseño y preparación

25,0%

- Exploración de datos mediante recursos de datos y almacenes de datos
- Creación de modelos con el diseñador de Azure Machine Learning
- Uso del aprendizaje automático automatizado para explorar modelos óptimos
- Ajuste de hiperparámetros con Azure Machine Learning

Exploración de datos y

40,0%





Repaso de tópicos:

- Introducción a Azure y sus soluciones para Data e IA
- Exploración de recursos del área de trabajo de Azure Machine Learning

<https://learn.microsoft.com/en-us/credentials/certifications/azure-data-scientist/?practice-assessment-type=certification>





Tipos de data



La Data se divide en 3 tipos según su *origen*, y el cómo *pueda ser interpretada* por ciertos aplicativos

Customer				
ID	FirstName	LastName	Email	Address
1	Joe	Jones	joe@litware.com	1 Main St.
2	Samir	Nadoy	samir@northwind.com	123 Elm Pl.

Product		
ID	Name	Price
123	Hammer	2.99
162	Screwdriver	3.49
201	Wrench	4.25

Estructurado



{
...
}



<...>
...
</...>

Semi-estructurado



No estructurado

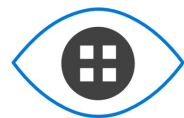
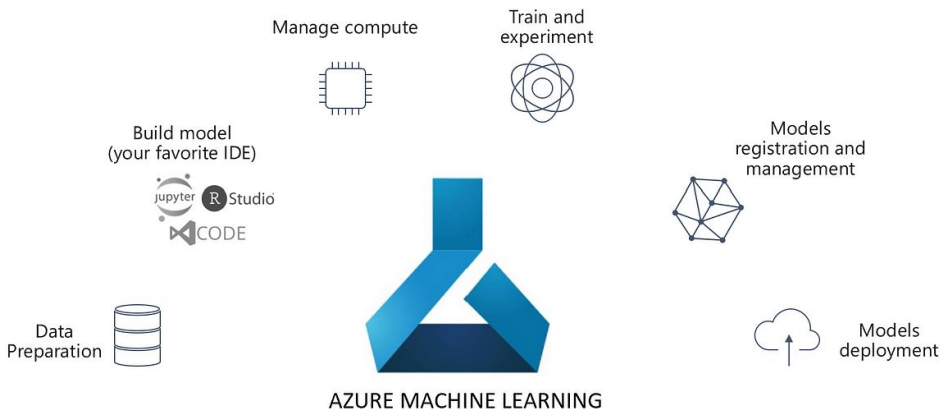




Azure Machine Learning Vs Cognitive Services



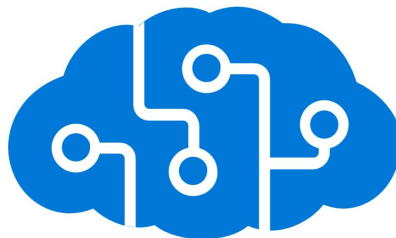
Podemos diferenciar 2 tipos de servicios que involucran IA dentro de Azure, por un lado, **Azure Machine Learning (studio)** y por otro **AI cognitive services**



Vision



Language



Microsoft Azure
Cognitive Services



Speech



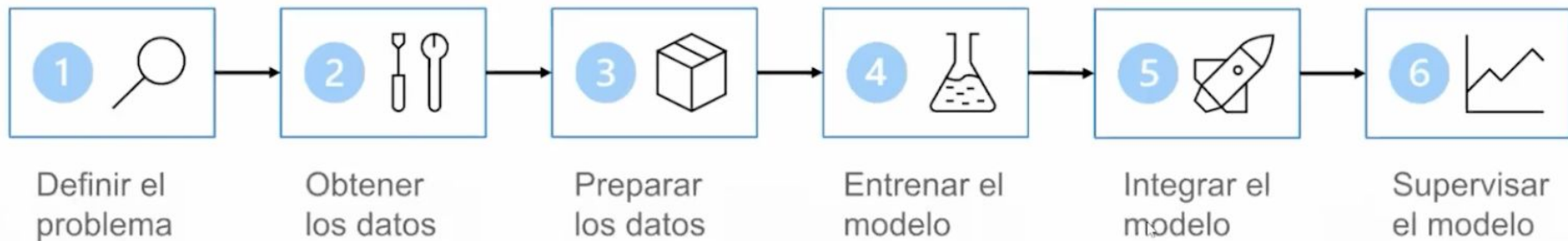
Knowledge



Ciclo de desarrollo de modelos de IA



Los 6 pasos que componen del desarrollo, despliegue y supervisión de un modelo son los siguientes:





Distribución de Azure Machine Learning Services

1.- Azure Portal.

2.- Suscripción de Azure.

3.- Grupo de recursos.

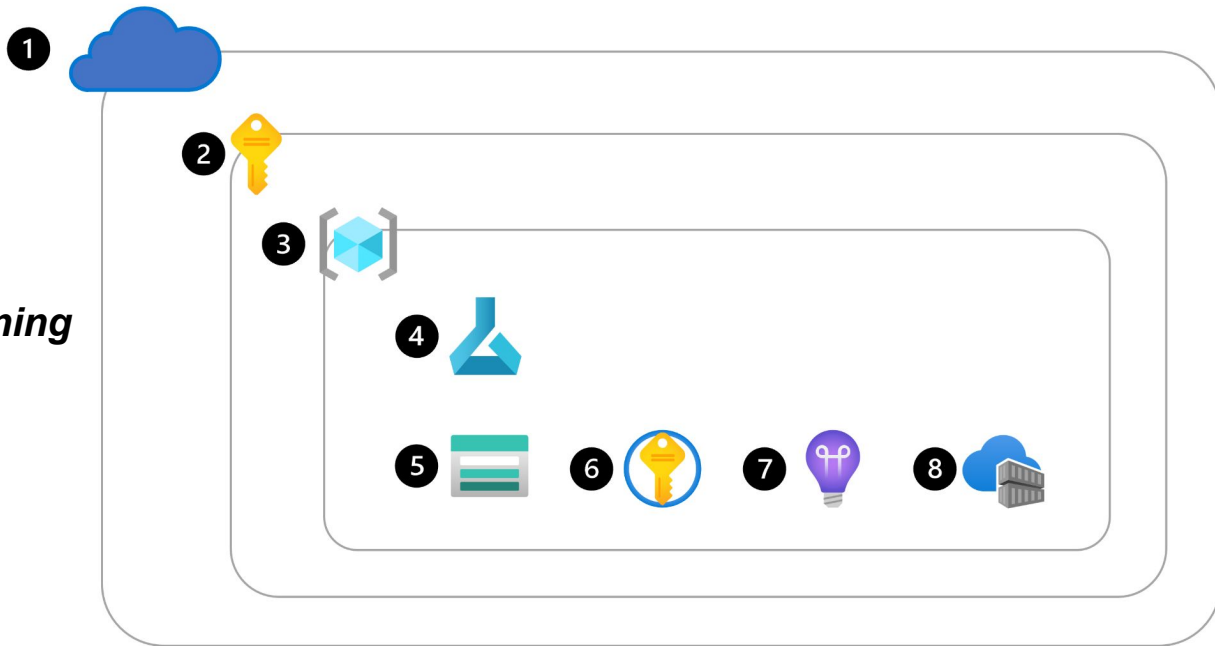
4.- Servicio Azure Machine Learning

5.- Azure Storage.

6.- Azure Key Vault.

7.- Application Insights

8.- Azure Container Registry.





Asesorías





1.- ¿Qué es un "pipeline"? nosotros lo generamos? se genera en automático?



Es una **serie de pasos** que ejecutan procesos encadenados. Un **"pipeline job"** lo puedes crear por medio del **SDK**, el **CLI**, o la consola de azure **"UI"**.

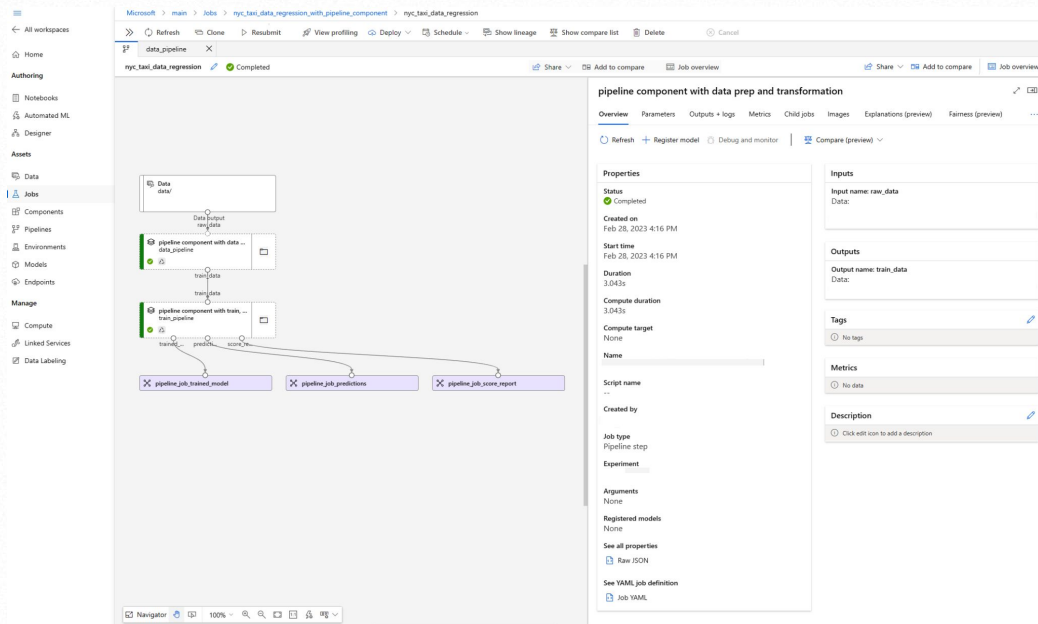
```
Python

# Construct pipeline
@pipeline
def pipeline_with_pipeline_component(
    training_input,
    test_input,
    compute_train_node,
    training_learning_rate=0.1,
    training_learning_rate2=0.01,
):
    # Create two training pipeline component with different learning rate
    # Use anonymous pipeline function for step1
    train_and_evaluate_model1 = train_pipeline_component(
        training_input=training_input,
        test_input=test_input,
        training_learning_rate=training_learning_rate1,
        train_compute=compute_train_node,
    )

    # Use registered pipeline function for step2
    train_and_evaluate_model2 = registered_pipeline_component(
        training_input=training_input,
        test_input=test_input,
        training_learning_rate=training_learning_rate2,
        train_compute=compute_train_node,
    )

    compare2_models = compare2(
        model1=train_and_evaluate_model1.outputs.trained_model,
        eval_result1=train_and_evaluate_model1.outputs.evaluation_report,
        model2=train_and_evaluate_model2.outputs.trained_model,
        eval_result2=train_and_evaluate_model2.outputs.evaluation_report,
    )

    # Return: pipeline outputs
    return {
        "best_model": compare2_models.outputs.best_model,
        "best_result": compare2_models.outputs.best_result,
    }
```



Referencia:

<https://learn.microsoft.com/en-us/azure/machine-learning/how-to-use-pipeline-component?view=azureml-api-2&tabs=ui>



2.- ¿Se puede conectar un espacio de trabajo en Azure Machine Learning con un repositorio de control de versiones como GitHub?





Sí es posible, deberás de enlazar **Azure App Registrations** con tu repositorio de GitHub creando las **secret keys** correspondientes, y dando acceso en la consola de azure, además de configurar en GitHub dichas

Create a new fork

A fork is a copy of a repository. Forking a repository allows you to freely experiment with changes without affecting the original project. [View existing forks.](#)

Owner * Repository name *

 / mlops-v2-gha-demo 

By default, forks are named the same as their upstream repository. You can customize the name to distinguish it further.

Description (optional)

☒ Copy the `main` branch only

Contribute back to Azure/mlops-v2-gha-demo by adding your own branch. [Learn more.](#)

ⓘ You are creating a fork in the microsoft organization (Microsoft Open Source).

Create fork

Wiki

Security

Insights

Settings

Security

Code security and analysis

Deploy keys

Secrets

Actions

Codespaces

Dependabot

Actions secrets / New secret

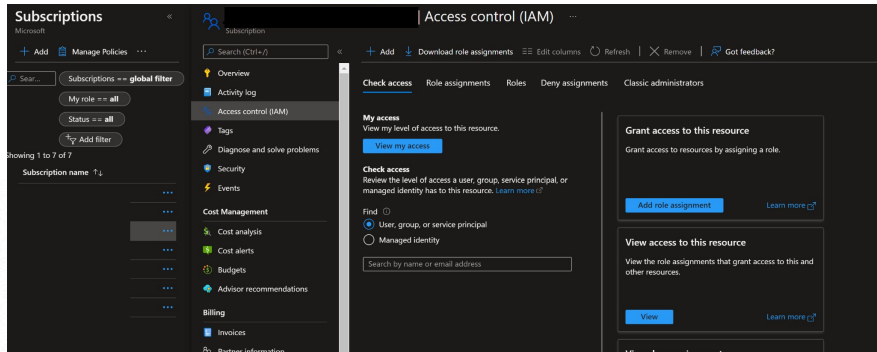
Name *

AZURE_CREDENTIALS

Secret *

```
{
  "clientId": "<service principal client id>",
  "clientSecret": "<service principal client secret>",
  "subscriptionId": "<Azure subscription id>",
  "tenantId": "<Azure tenant id>",
  "activeDirectoryEndpointUrl": "https://login.microsoftonline.com/",
  "resourceManagerEndpointUrl": "https://management.azure.com/",
  "activeDirectoryGraphResourceId": "https://graph.windows.net/",
  "sqlManagementEndpointUrl": "https://management.core.windows.net/8443/"
}
```

Add secret



The screenshot shows the Azure portal interface. On the left, the 'Subscriptions' section is visible with a search bar and filters. The main area displays the 'Access control (IAM)' section for a specific subscription. It includes tabs for 'Check access', 'Role assignments', 'Roles', 'Deny assignments', and 'Classic administrators'. The 'Check access' tab is active, showing a list of role assignments with columns for 'Find', 'User, group, or service principal', and 'Managed identity'. A search bar is provided to find assignments by name or email address. On the right, there are two panels: 'Grant access to this resource' and 'View access to this resource', both with 'Add role assignment' and 'View' buttons respectively.

Referencias:

[Integración de Git en Azure ML](#)
[Configuración de MLOps con GitHub](#)

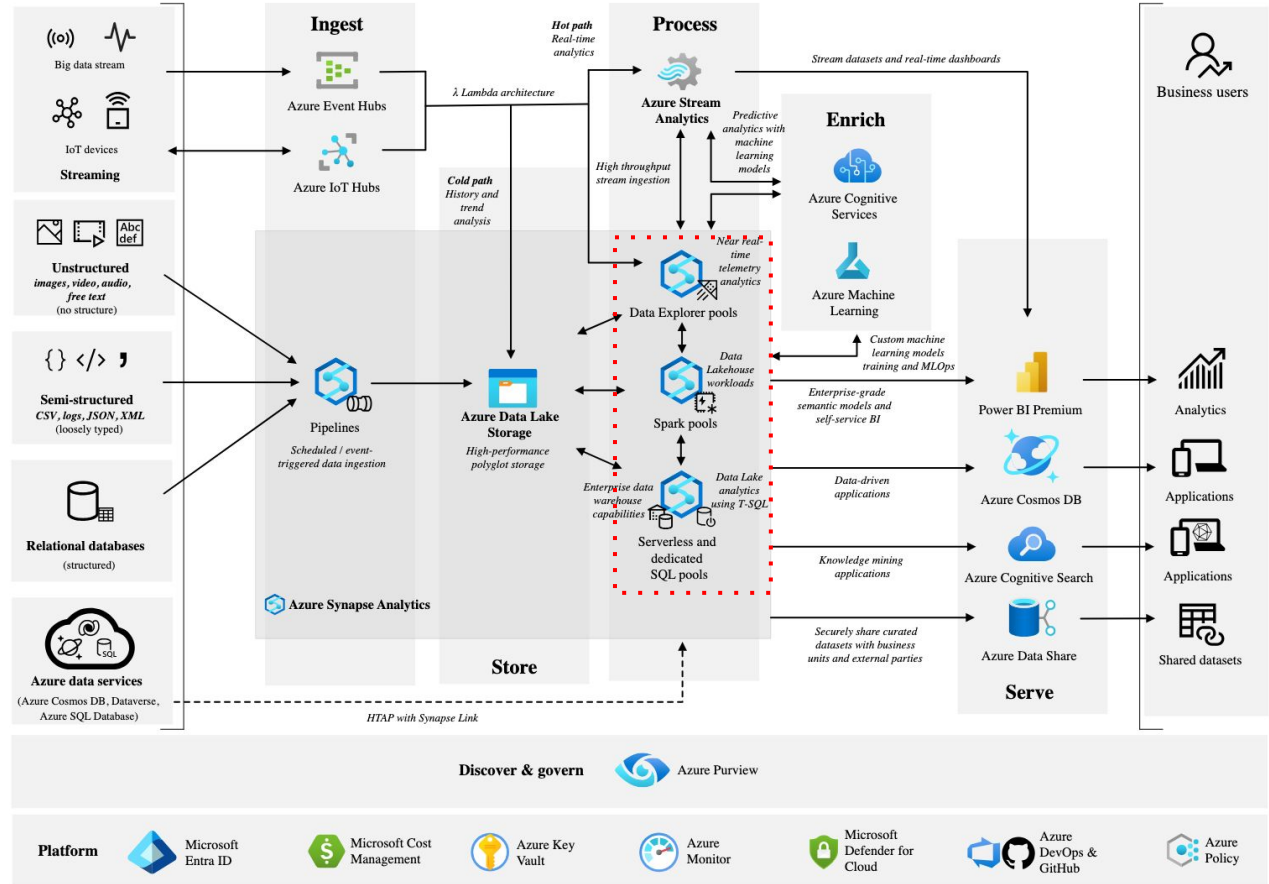


3A.- ¿Podría explicar cómo Azure Synapse Analytics se integra con otros servicios de Azure AI para crear una solución end-to-end de análisis predictivo?

Referencias:

[Integración de Azure Synapse con Azure Machine Learning](#)

[Analytics E2E with azure Synapse](#)





3B.- Específicamente, ¿Cuál sería el flujo de trabajo recomendado cuando queremos combinar datos históricos estructurados con datos no estructurados como texto o imágenes?

Ejemplo de un flujo específico:



- Un proyecto podría empezar con la **ingesta de datos** de ventas históricas (estructurados) y comentarios de clientes (texto no estructurado) en **Azure Synapse Pipelines**, almacenándolos en **Azure Data Lake Storage**.
- Los comentarios se analizarían usando **Azure Cognitive Services** para detectar sentimientos y temas clave.
- Los resultados se **combinarían** con los datos estructurados de ventas y se entrenaría un **modelo en Azure Machine Learning** para predecir comportamientos futuros de los clientes.
- Los resultados finales se **visualizan en Power BI** para decisiones estratégicas.





4A.- El proceso de MLOps en azure simplemente sería observar las métricas de los modelos desarrollados o iría mas alla de crear una interfaz de consumo interactiva de los modelos?



Detalle del proceso de MLOps en Azure:

1. Desarrollo y experimentación:

- **Azure Machine Learning Studio:** Se registran y versionan los experimentos y se comparan las métricas de los modelos.

2. Automatización del pipeline de entrenamiento y despliegue:

- **Pipelines de CI/CD** (Integración Continua y Despliegue Continuo) usando **Azure DevOps** o **GitHub Actions**, que automatizan el proceso de entrenamiento, pruebas y despliegue de modelos.

3. Despliegue en entornos productivos:

- **Azure Kubernetes Service (AKS), Azure Container Instances (ACI) o Azure Functions:** Permite un despliegue flexible y una gestión adecuada del rendimiento.

← All workspaces

Home

Authoring

Notebooks

Automated ML

Designer

Assets

Data

Jobs

Components

Pipelines

Environments

Models

Endpoints

Manage

Compute

Linked Services

Data Labeling



Referencia:

<https://learn.microsoft.com/en-us/azure/machine-learning/concept-model-management-and-deployment?view=azureml-api-2>



- La monitorización de métricas de los modelos en producción es **fundamental**. Azure Machine Learning permite la observación de métricas de rendimiento, como precisión, latencia y uso de recursos.

5.- Gestión de datos y reentrenamiento:

- **Azure Data Factory y Azure Databricks** pueden integrarse para automatizar el flujo de datos y mantener actualizados los modelos.

6.- Interfaces de consumo interactivas:

- Esto incluye el desarrollo de aplicaciones web o APIs que consumen modelos desplegados en servicios de **Azure App Service** o mediante APIs REST servidas por **Azure Machine Learning**.



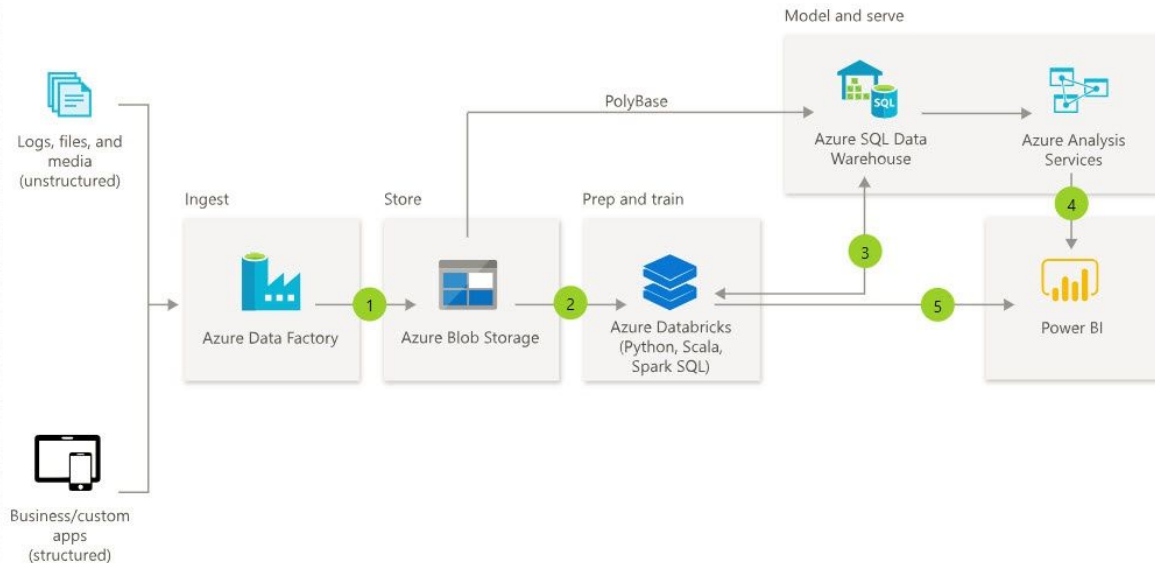
<https://learn.microsoft.com/en-us/azure/machine-learning/concept-model-management-and-deployment?view=azureml-api-2>



5.- ¿Podrían explicar con más detalles azure block store y azure data lake?




La Data ingresada se mantienen dentro de **blobs** en la cuenta de almacenamiento. El servicio que gestiona los blobs es el servicio **Azure Blob Storage**. Data Lake Storage describe las capacidades o "**mejoras**" de este servicio que atienden las **demandas de cargas de trabajo de análisis de big data**.



Referencias:
[Azure Blob Storage](#)
[Azure Data Lake](#)



5.- ¿Podrían explicar con más detalles azure block store y azure data lake?

Característica	Azure Blob Storage	Azure Data Lake Storage
 Definición	Almacenamiento de objetos para datos no estructurados como imágenes, videos y documentos.	Servicio de almacenamiento especializado para análisis de datos a gran escala .
Casos de uso	Almacenamiento de backups, contenido multimedia, archivos de todos los formatos .	Ideal para proyectos de big data e integración con herramientas de análisis .
Niveles de acceso	Ofrece niveles de acceso: caliente y frío optimizando costos según la frecuencia de acceso .	Hereda niveles de acceso de Azure Blob Storage y está optimizado para acceso más rápido y escalable.
Integración con herramientas	Compatible con Azure Media Services y Azure Machine Learning .	Se integra con servicios de big data y análisis como Azure Synapse Analytics y Azure Databricks .
Formato de datos	Almacena cualquier tipo de archivo , sin optimizaciones específicas para big data .	Soporta formatos optimizados para big data, como Apache Parquet , con estructura jerárquica de archivos.



Referencias:
[Azure Blob Storage](#)
[Azure Data Lake](#)



6.- ¿Se puede usar terraform?



Infraestructura como Código (**laC**) desarrollada por HashiCorp. Permite definir y gestionar la infraestructura de TI a través de archivos de configuración en un lenguaje declarativo llamado **HashiCorp Configuration Language (HCL)**, aunque también admite **JSON**.

- **AzureRM**: Máquinas virtuales, cuentas de almacenamiento e interfaces de red.
- **AzureDevops**: Azure DevOps, como agentes, repositorios, proyectos, canalizaciones y consultas.
- **AzureStack**: Azure Stack Hub, como máquinas virtuales, DNS, redes virtuales y almacenamiento.

```
resource "random_pet" "rg_name" {  
  prefix = var.resource_group_name_prefix  
}  
  
resource "azurerm_resource_group" "rg" {  
  location = var.resource_group_location  
  name     = random_pet.rg_name.id  
}
```



Referencia:

<https://learn.microsoft.com/es-es/azure/developer/terraform/overview>