



**MI PRIMER ETL CON
AZURE DATA FACTORY**

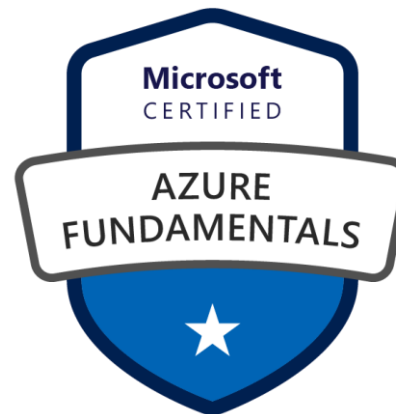
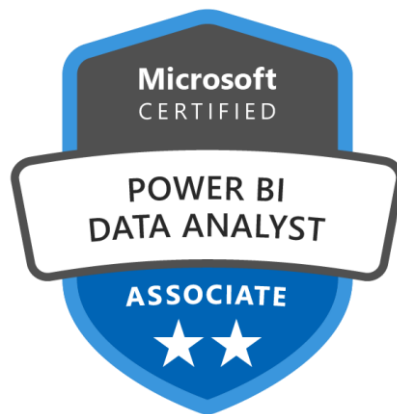


Edgar Quispe

Ingeniero de Sistemas, egresado de la UNMSM , certificado por Microsoft como Power BI Data Analyst Associate, con +5 años de experiencia en herramientas y plataformas como SQL Server, Power BI, SQL Server Integration Services, Python, Power Apps, Power Automate, Azure Data Factory y Databricks, con aplicaciones para empresas en Perú, Latinoamérica y España.



<https://www.linkedin.com/in/edgar-quispe-151051167/>





INGENIERÍA DE DATOS

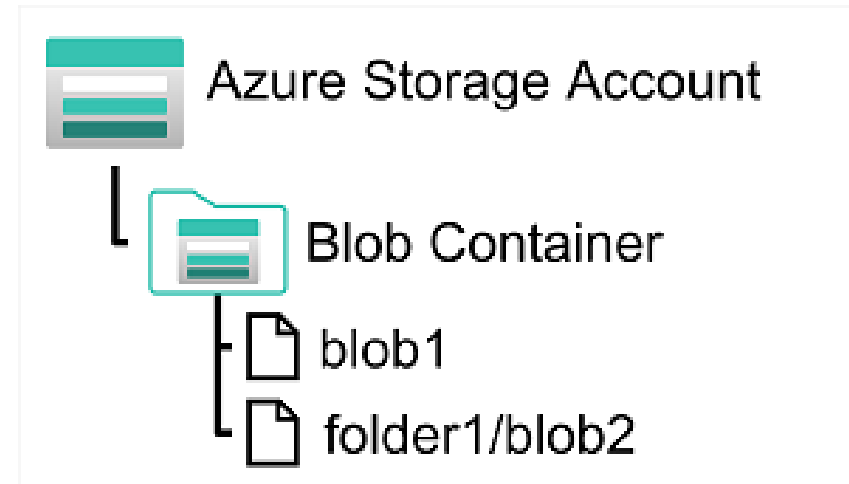
**SERVICIOS
COMPLEMENTARIOS**



Azure Blob Storage es un servicio que le permite almacenar grandes cantidades de datos no estructurados como objetos binarios grandes, o blobs, en la nube.

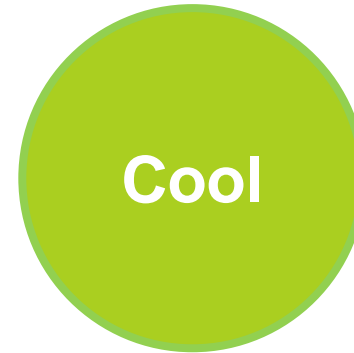
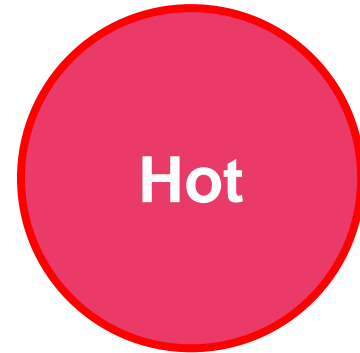
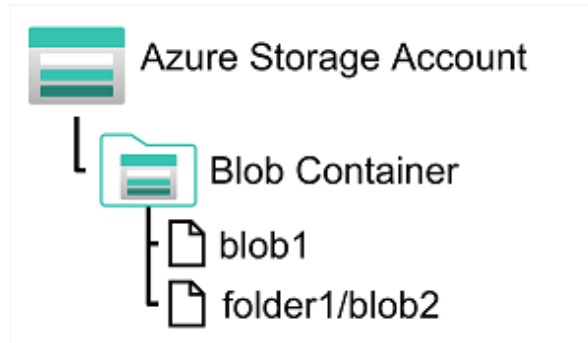
Características

- ✓ Los blobs es una forma eficaz de almacenar archivos en la nube.
- ✓ Los blobs se almacenan en contenedores.





Niveles de Acceso (Tier)



| Característica | Frecuente | Esporádico | Archivo |
|-------------------------|--|--|---|
| Costo de almacenamiento | Alto | Medio | Bajo |
| Costo de acceso (E/S) | Bajo | Alto | Muy Alto |
| Frecuencia de acceso | Frecuente (menor a 30 días) | Poco Frecuente (entre 30 y 90 días) | Muy poco frecuente (mayor a 90 días) |
| Latencia de acceso | Baja (rápido) | Baja (rápido) | Alta (varias horas) |
| Casos de uso | ✓ Archivos (png, mp4) ✓ Datasets para ETL | ✓ Backups ✓ Logs | ✓ Copias de seguridad ✓ Documentos legales |



Azure SQL Database un servicio de base de datos **PaaS** totalmente administrado y altamente escalable que incluye las principales capacidades de base de datos de SQL Server on-premise.

Características

- ✓ Permite escalar automáticamente los recursos de procesamiento y almacenamiento en función de la demanda.
- ✓ Incluye la administración automatizada de backups, actualizaciones de software, parches de seguridad, etc.
- ✓ No tiene habilitado complementos como: SQL Server Agent, Linked Server, etc.
- ✓ Cuenta con los modelos de servicio de **Base de datos Única** y **Grupo Elástico**.





Esta opción le permite configurar y ejecutar rápidamente una sola base de datos de SQL Server.

Características

- ✓ Cuenta con recursos dedicados (como CPU, memoria y almacenamiento).
- ✓ Puedes ajustar los recursos asignados a la base de datos.

Casos de Uso

- ✓ Soluciones donde cada aplicación necesita su propia base de datos.
- ✓ Aplicaciones que requieren alta disponibilidad.





INGENIERÍA DE DATOS

INTRODUCCIÓN A
AZURE DATA FACTORY





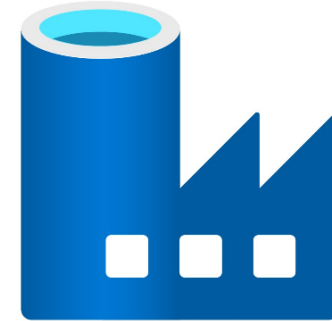
¿Qué es Azure Data Factory?



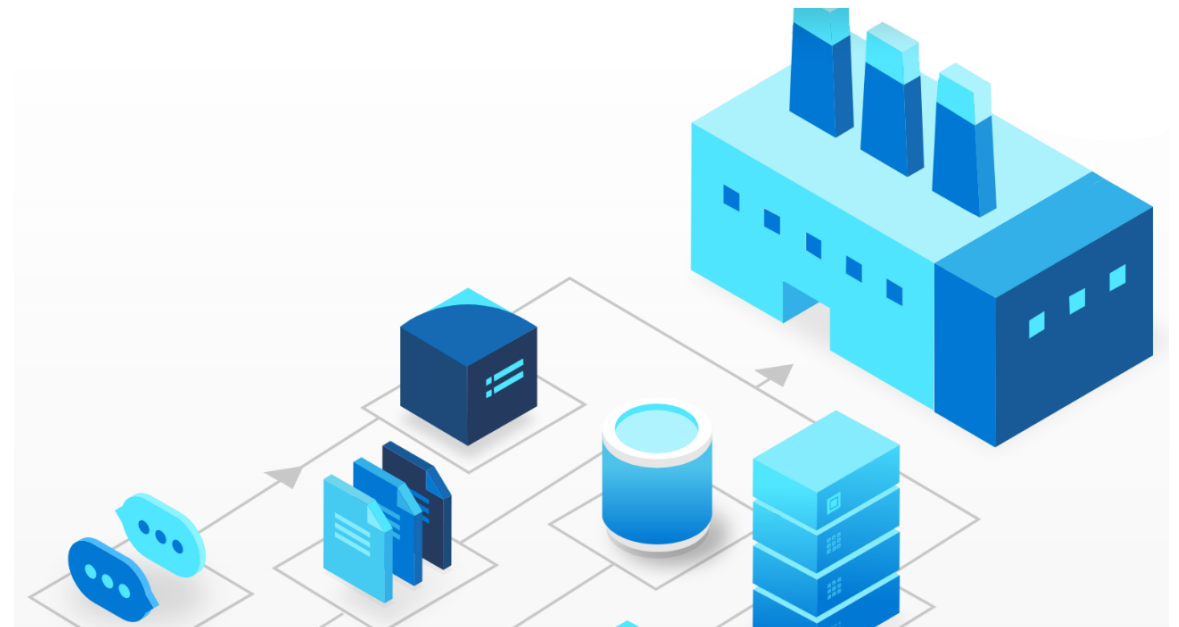
Azure Data Factory (ADF) es un servicio de Azure diseñado para integrar, orquestar y automatizar flujos de datos a gran escala.

ADF te permite

- Conectar múltiples fuentes de datos (locales y en la nube).
- Realizar la limpieza y transformación de datos a un nivel avanzado.
- Cargar datos hacia destinos como data warehouses, Data Lake (ADLS) o Lakehouse (databricks).
- Automatizar la ejecución de los pipelines a través de diferentes eventos.

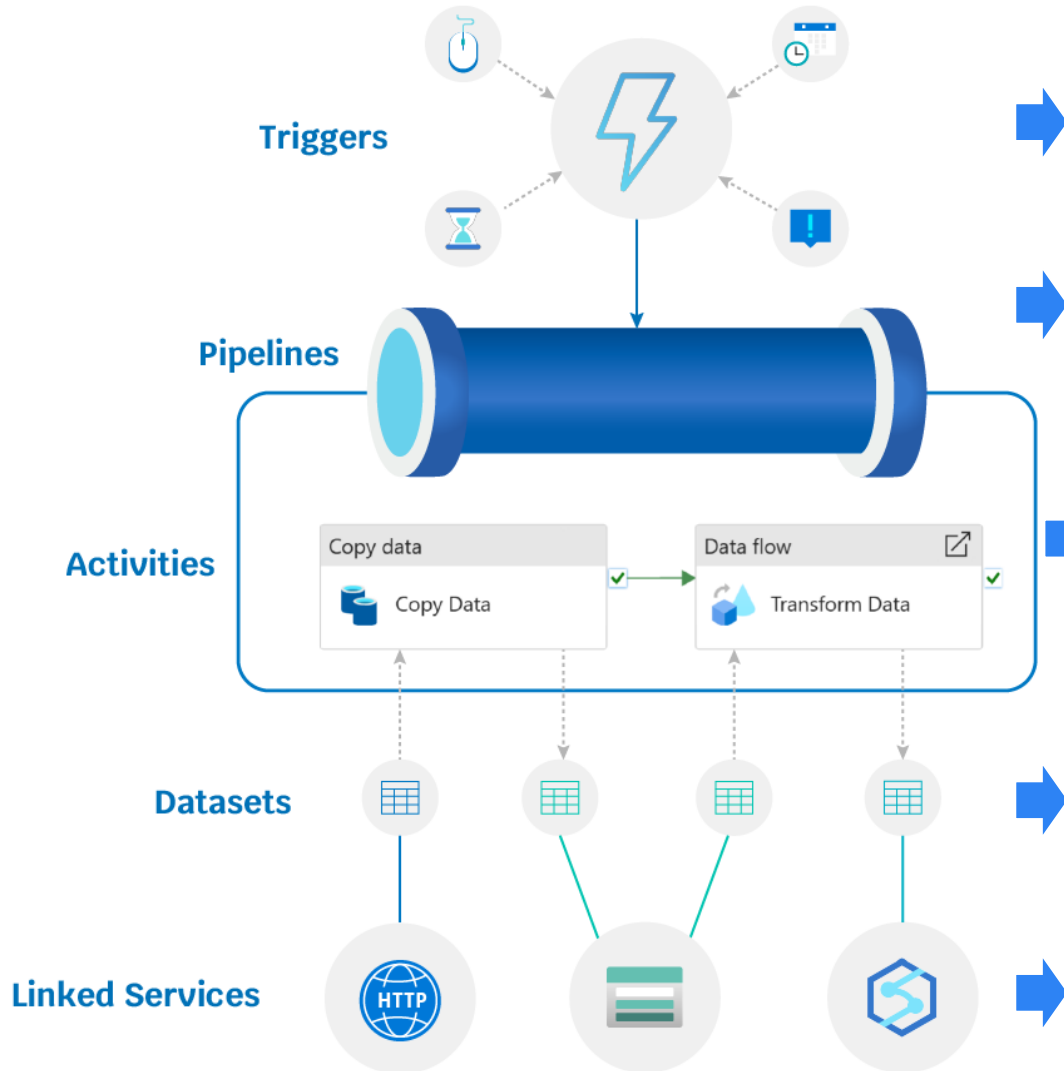


Azure Data Factory





Arquitectura de ADF



Un **trigger** es un mecanismo que inicia la ejecución de un pipeline en respuesta a un evento.



Un **pipeline** es un contenedor lógico que agrupa un conjunto de actividades



Una **actividad** es una tarea específica que realiza una acción dentro de un pipeline.

Copy data: copiar información de un origen a un destino.

Data Flow: limpieza y transformación de datos.



Permite definir la estructura de datos específica del archivo o tabla al que nos queremos conectar.



Permite crear y configurar la conexión a un recurso o servicio, ya sea on-premise o en la nube.



Azure Data Factory Studio



master branch | Validate all | Save all | Publish | Auto Save | Preview experience On

Factory Resources | Filter resources by name | Pipelines (2) | Datasets (4) | Data flows (1) | Power Query (0) | Templates (0)

ETLAtenciones | Activities | Search activities | Move and transform | Copy data | Data flow | Synapse | Azure Data Explorer | Azure Function | Batch Service | Databricks | Data Lake Analytics | General | HDInsight | Iteration & conditionals | Machine Learning | Power Query

Copy data | Copy data | Data flow | Synapse | Azure Data Explorer | Azure Function | Batch Service | Databricks | Data Lake Analytics | General | HDInsight | Iteration & conditionals | Machine Learning | Power Query

Set variable (x) EstablecerFileName | Copy data FileFromFtpToBlob Storage | Data flow DataflowTickets | Copy data CreateCSVFile | Azure Function SendEmail

General | Settings | Parameters | User properties

Data flow * | DataflowTickets | Open | New

Run on (Azure IR) * ⓘ | AutoResolveIntegrationRuntime

Compute size * ⓘ | Small

> Advanced