

# ETL en Azure: Azure Data Factory

Autor: José Luis Cendán Guzmán || Correo:[icendguz@myuax.com](mailto:icendguz@myuax.com) || Fecha: 22/12/2025

## INTRODUCCIÓN

Esta solución aborda el aspecto de ingestión de datos raw a capa Bronze dentro de una arquitectura medallion, implementando un proceso EL (Extract-Load) que extrae datos desde un almacenamiento basado en archivos (Azure Blob Storage) y los carga directamente en un almacén relacional (Azure SQL Database) sin aplicar transformaciones complejas. Los datos se almacenan en su forma cruda/original para ser procesados posteriormente en capas Silver (limpieza y normalización) y Gold (agregaciones y métricas de negocio). En el contexto del pipeline de análisis de datos, esta solución se posiciona en la fase inicial de ingestión, donde Azure Data Factory actúa como herramienta de orquestación para automatizar el movimiento de datos desde el exterior. La implementación requiere una infraestructura cloud compuesta por: suscripción Azure activa, Azure Storage Account con Blob Storage configurado, Azure SQL Database con servidor y base de datos creados, y permisos IAM adecuados. En cuanto a la configuración, tenemos como datos de origen productos\_50.csv (usado en la tarea 1) y esquema de tabla definido en destino. Además, se requieren Linked Services configurados para autenticación entre componentes. El tutorial está bien explicado, pero requiere familiaridad básica en la lógica de este proceso devops, conceptos generales de bases de datos, y capacidad de resolución de problemas técnicos (especialmente en configuración de restricciones de red/firewall y creación de esquemas SQL).

## DESARROLLO

Se crearon los siguientes recursos en Azure: Grupo de recursos para organizar todos los componentes, Cuenta de almacenamiento (Storage Account) con Blob Storage para almacenar el archivo productos\_50.csv, SQL Server con su servidor lógico y SQL Database con la tabla productos creada mediante script SQL, y Data Factory como orquestador del pipeline. Dentro de Data Factory se configuraron los componentes del pipeline: Datasets (CopyProductoCSV para el origen CSV y AzureSqlTable para la tabla productos de destino), y el Pipeline con Copy Activity para realizar el mapeo y transferencia de datos.

La configuración comenzó con la creación del grupo de recursos y la infraestructura base (Storage Account, SQL Server, Data Factory). Se cargó el archivo productos\_50.csv al contenedor de Blob Storage y se importó el script de la tabla productos en SQL Database. Se configuró la seguridad de red para permitir acceso desde servicios Azure y a mi IP. Los Datasets se configuraron especificando formato CSV delimitado por comas con encabezados para el origen (CopyProductoCSV) y la tabla dbo.productos para el destino (AzureSqlTable). El Pipeline se diseñó mediante interfaz drag-and-drop, configurando el mapeo de columnas en la Copy Activity para excluir la columna ID auto-incremental (CSV.nombre → SQL.nombre, CSV.precio → SQL.precio).

La ejecución se realizó en modo debug para validar el flujo, obteniendo como resultado final correcto de 50 registros desde el archivo CSV hacia la tabla SQL Database.

## CONCLUSIONES

Esta práctica permitió consolidar conceptos de adquisición de datos, implementando un proceso EL que representa la fase inicial del ciclo de vida de los datos. Se aprendió la diferencia entre EL y ETL, y cómo la arquitectura medallion organiza datos en capas progresivas, alineándose con principios de Data Lake modernos donde los datos raw se preservan para múltiples usos y reprocesamiento. La relevancia de una buena implementación es crítica en contextos empresariales, ya que una mala adquisición de datos compromete las decisiones del negocio y genera costos excesivos y retrasos. Es fundamental saber elegir el tipo de suscripción y configuración de recursos para evitar gastos innecesarios en cloud. El cumplimiento normativo requiere controles de seguridad desde que se obtiene el dato hasta que se le da su visualización: cifrado, auditoría y segregación. Los principales desafíos encontrados fueron:

- Técnicos: configuración de restricciones de red en SQL Database, gestión de permisos IAM (mínimo privilegio), mapeo de esquemas con columnas auto-incrementales.
- Conceptuales: transición de ETL tradicional (scripts/cron) a orquestación declarativa cloud, representando un cambio hacia DataOps (automatización, monitorización, colaboración).
- Seguridad: implementación de controles desde la obtención del dato hasta su visualización (cifrado, auditoría, segregación).

Mejoras futuras: Data Flows para transformaciones avanzadas, alertas automáticas (Azure Monitor), control de versiones (Git/CI/CD), e integración con Synapse Analytics para Big Data.

# CAPTURAS DE PANTALLAS

## 1. Blog Storage Creado

Todos los servicios > Centro de almacenamiento | Blob Storage >

**Crear una cuenta de almacenamiento**

Datos básicos Avanzado Redes Protección de datos Cifrado Etiquetas Revisar y crear

Azure Storage es un servicio administrado por Microsoft que proporciona almacenamiento en la nube altamente disponible, seguro, duradero, escalable y eficiente. Azure Storage incluye Azure Blob (objetos), Azure Data Lake Storage Gen2, Azure Files, Azure Queues y Azure Tables. El costo de una cuenta de Storage depende del uso y de las opciones que elija a continuación. Más información sobre las cuentas de almacenamiento junto con otros recursos.

**Detalles del proyecto**

Seleccione la suscripción en la que se creará la nueva cuenta de almacenamiento. Baja un grupo de recursos nuevo o uno ya existente para organizar y administrar la cuenta de almacenamiento junto con otros recursos.

Suscripción \* MÁSTER UNIVERSITARIO EN ANÁLISIS DE DATOS MASIVOS

Grupo de recursos \* dev-rg-devops-jose-cendan-guzman

**Detalles de la instancia**

Nombre de la cuenta de almacenamiento strluiscendan2025

Región \* (Europe) West Europe

Tipo de almacenamiento preferido Azure Blob Storage o Azure Data Lake Storage Gen2

Rendimiento \* Estándar: Opción recomendada para la mayoría de los escenarios (cuenta de usuario v2)

Redundancia \* Almacenamiento con redundancia local (LRS)

**Autenticación**

Usar la autenticación solo de Microsoft Entra

Inicio de sesión del administrador del servidor \* luiscendan

Contraseña \*

Confirmar contraseña \*

**Anterior Siguiente Revisar y crear**

## 2. SQL Server Creado

Todos los servicios > Azure SQL | SQL databases > Crear base de datos SQL >

**Crear un servidor de SQL Database**

**Detalles del servidor**

Especifique la configuración necesaria para este servidor, incluida la inclusión de un nombre y una ubicación. Este servidor se creará en la misma suscripción y grupo de recursos que la base de datos.

Nombre del servidor \* sqlserver-luiscendan .database.windows.net

Ubicación \* (Europe) Spain Central

**Autenticación**

Usar la autenticación ahora en Microsoft Entra ID: [Más información](#)

Select your preferred authentication methods for accessing this server. We recommend using only Microsoft Entra authentication. [Más información](#)

Método de autenticación Uso de la autenticación de SQL y Microsoft Entra

Inicio de sesión del administrador del servidor \* luiscendan

Contraseña \*

Confirmar contraseña \*

## 3. Tabla Creada y 4.Carga en CSV

bd-luiscendan (luiscendan@myvus.com)

Consultas 1

```
1: USE [bd_luiscendan]
2: GO
3: IF OBJECT_ID('dbo.products', 'U') IS NOT NULL
4: BEGIN
5:     DROP TABLE products;
6: END;
7: GO
8: CREATE TABLE products (
9:     id INT IDENTITY(1,1) PRIMARY KEY,
10:    name NVARCHAR(100),
11:    price DECIMAL(18,2)
12: );
13: GO
14: -- verificar que la tabla se crea correctamente
15: SELECT TABLE_NAME, COLUMN_NAME, DATA_TYPE
16: FROM INFORMATION_SCHEMA.COLUMNS
17: WHERE TABLE_NAME = 'products'
18: ORDER BY ORDINAL_POSITION;
19: GO
```

Resultados Información

TABLE\_NAME

products

products

**Cargar blob**

1 archivo seleccionados: productos.csv

Arrastrar y colocar archivos aquí o [Buscar archivos](#)

Selección un contenedor existente

productos

Crear nuevo

Subscribir los archivos si ya existen

**Cargar**

[Enviar comentarios](#)

## 5.Data Factory Creado

**Se completó la implementación**

Nombre de implementación : MicrosoftDataFactory-2025121719... Hora de inicio : 17/12/2025, 19:26:39

Suscripción : MÁSTER UNIVERSITARIO EN ANÁLISIS DE DATOS MASIVOS

Grupo de recursos : dev-rg-devops-jose-cendan-guzman

**Detalles de implementación**

Recurso	Tipo	Estado	Detalles de la operación
df-analysis-2025	Factoría de datos (V2)	OK	<a href="#">Detalles de la operación</a>

**Pasos siguientes**

[Ir al recurso](#)

## 6. SQL Database Creado

**Se completó la implementación**

Nombre de implementación : MicrosoftSQLDatabaseNewDatabaseNewServer\_caca449c62d45bb4abf

Suscripción : MÁSTER UNIVERSITARIO EN ANÁLISIS DE DATOS MASIVOS

Grupo de recursos : dev-rg-devops-jose-cendan-guzman

**Detalles de implementación**

Recurso

- sqlserver-luiscendan/bd-luiscendan
- sqlserver-luiscendan/Default
- sqlserver-luiscendan

## 8.Dataset Origen (CSV)

Conexión Esquema Parámetros

Servicio vinculado : Azureblobstorage1

Ruta de acceso del archivo : productos / Dirección : productos\_30.csv

Tipo de compresión : Sin compresión

Delimitador de columnas : Comma (,)

Delimitador de filas : Default (CRLF o VTFN)

Codificación : Valor predeterminado(UTF-8)

Carácter de comillas : Double quote (")

Carácter de escape : Backslash (\)

Primera fila como encabezado :

Valor nulo :

para aprovechar las características y mejoras más recientes:

**Publicar todo**

products AzureSqlTable1

Conexión Esquema Parámetros

Servicio vinculado : AzureSqlDatabase1

Tabla : dbo.products

Prueba de conexión

## 9. Dataset Destino (SQL)

## 7. Objetos creados

- bd-luiscendan (sqlserver-luiscendan/bd-luiscendan)
- sqlserver-luiscendan
- strluiscendan2025
- df-analysis-2025
- dev-rg-devops-jose-cendan-guzman

## 10.Configurando Copy Activity

Nuevo servicio vinculado

Almacenamiento de blobs de Azure

Nombre : AzureblobStorage1

Descripción :

Conectar mediante Integration Runtime \*  AzureResolveIntegrationRuntime

Tipo de autenticación : Clave de cuenta

Cadena de conexión : Azure Key Vault

Método de selección de cuenta : From Azure subscription  Enter manually

Suscripción de Azure : MÁSTER UNIVERSITARIO EN ANÁLISIS DE DATOS MASIVOS (33b67b58-295c-435c-8a86-9a...

Nombre de cuenta de almacenamiento : strluiscendan2025

Propiedades de conexión adicionales

Nuevo

Prueba de conexión :  Al servicio vinculado  A la ruta de acceso de archivo

Anotaciones

+ Nuevo

Parámetros

Conexión correcta Prueba de conexión

## 11. Pipeline exitoso

Publicar todo

products AzureSqlTable1 CopyProductsToCSV

Actividades

Copy data! (Copy data1)

Parámetros Variables Configuración Salida

Id de ejecución de canalización : ffa902-5d4f-4302-a656-ba70ef5219

Estado de la canalización : Correcto

All status :

Mostrando elementos del 1 al 1 de un total de 1

Nombre de actividad : Estado de ... Tipo d... Inicio de la ejecu... Duraci... Entorno de ejecución de integración

Copy data! : Correcto Copiar datos 12/10/2025, 6:33:34 PM 14s Autofsolveintegrationruntime (Spain Central)

## 12..Datos en SQL Database

SELECT \* FROM dbo.products

Resultados Mensajes

ID	nombre	precio
1	Mandarina Golden kg	1.92
2	Platano kg	1.53
3	Naranja kg	1.53
4	Tomate kg	2.39