

ETL en Azure: Azure Data Factory

Autor: José Luis Cendán Guzmán || **Correo:** jcendguz@myuax.com || **Fecha:** 22/12/2025

INTRODUCCIÓN

Esta solución aborda el aspecto de ingesta 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 ingesta, 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

Los todos los >

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 redundante. 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 de Azure >

Detalles del proyecto

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

Suscripción *

Grupo de recursos *

Máster Universitario en Análisis de Datos Masivos

dev-rg-devops-jose-cendan-guzman

Clear nuevo

Detalles de la instancia

Nombre de la cuenta de almacenamiento *

Región *

Grupo de recursos *

Tipos de almacenamiento preferido

Rendimiento *

Redundancia *

trufiscendan2025

(Europe) West Europe

Implementación en una zona extendida de Azure

Azure Blob Storage o Azure Data Lake Storage Gen2

Esto nos ayuda a ofrecer información relevante. No limita su almacenamiento a este tipo de recursos. [Obtén más información](#)

☒ Estándar: Opción recomendada para la mayoría de los escenarios (cuenta de uso general v2)

☐ Premium: Se recomienda para escenarios que requieren una latencia baja

Almacenamiento con redundancia local (LRS)

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

Microsoft

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 * | <input type="text" value="sqlserver-luisccandn"/> | ✓ |
| Ubicación * | <input type="text" value="(Europe) Spain Central"/> | database.windows.net |

Autenticación

Azure Active Directory ahora es Microsoft Entra ID. [Más información](#) ⓘ

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


Microsoft Entra [Más información](#) ⓘ

| | |
|-------------------------|---|
| Método de autenticación | <input type="radio"/> Usar la autenticación solo de Microsoft Entra <input type="radio"/> Uso de la autenticación de SQL y Microsoft Entra <input checked="" type="radio"/> Uso de la autenticación de SQL. |
|-------------------------|---|

| | | |
|---|---|---|
| Inicio de sesión del administrador del servidor * | <input type="text" value="luisccandn"/> | ✓ |
| Contraseña * | <input type="password" value="*****"/> | ✓ |
| Confirmar contraseña * | <input type="password" value="*****"/> | ✓ |

The screenshot displays a web application interface. At the top, there's a navigation bar with a user profile icon, a search bar, and links for 'Inicio', 'Cargar', 'Ver', 'Eliminar', 'Actualizar', and 'Compartir'. Below this, the main content area is titled 'Cargar blob' and features a file selection button. A list of files is shown, including 'productos_50.csv' and 'Bunacar archivos'. The right sidebar contains a code editor with a SQL query for inserting data into a 'productos' table.

5.Data Factory Creado





Se completó la implementación

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

Suscripción : Máster Universitario en Análisis de ... Id. de correlación : c7ae629b-d390-48fc-9748-ac3a11e9342a

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

Detalles de implementación

| Recurso | Tipo | Estado | Detalles de la operación |
|---|--|--------|--|
|  df-analisis-2025 |  Factoría de datos (V2) | OK | Detalles de la operación |

Pasos siguientes

[Ir al recurso](#)

6. SQL Database Creado

Se completó la implementación

Nombre de implementación: Microsoft.SqlDatabase.newDatabaseNewServer_caca446c026458b0bbaf4

Suscripción: Máster Universitario en Análisis de Datos Masivos

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

Detalles de implementación

| Recurso |
|----------------------------------|
| sqlserver-luisendan/bd-luisendan |
| sqlserver-luisendan/Default |
| sqlserver-luisendan |

8.Dataset Origen (CSV)

Comandos

Esquema

Parámetros

Servicio utilizado *

AzureFileStorage1

Prueba de conexión

Editar

Nuevo

Problemas

Ruta al archivo del archivo *

Actualmente

Directorio

productivo_50_52

Tipo de compresión

Sin compresión

Delimitador de columnas

Comma (,)

Delimitador de filas

Default (r\n, or r\r\n)

Codificación

Valor predefinido(UTF-8)

Carácter de comillas

Double quote (")

Carácter de escape






Backslash (\)

Primera fila como encabezado

☒

Valor nulo

7. Objetos creados


| | | | |
|--------------------------|---|-----|--------------------------|
| <input type="checkbox"/> |  bd-luisendan (sqlserver-luisendan/bd-luisendan) | ... | Base de datos SQL |
| <input type="checkbox"/> |  sqlserver-luisendan | ... | SQL Server |
| <input type="checkbox"/> |  strluisendan2025 | ... | Cuenta de almacenamiento |
| <input type="checkbox"/> |  df-analisis-2025 | ... | Factores de datos (V2) |
| <input type="checkbox"/> |  dev-rig-devops-jose-cendan-guzman | ... | Grupo de recursos |

9. Dataset

Destino (SQL)

10. Configurando Copy Activity


Nuevo servicio vinculado


 Almacenamiento de blobs de Azure [Más información](#)


Nombre *

AzureBlobStorage1

Descripción


Conectar mediante Integration Runtime * 

 AutoResolveIntegrationRuntime




Tipo de autenticación


Clave de cuenta





[Centro de conexión](#)

[Azure Key Vault](#)


Método de selección de cuenta 

 From Azure subscription

 Enter manually


Subscripción de Azure 


Máster Universitario en Análisis de Datos Masivos (33b67b58-255c-435c-8a8e-9aa...




Nombre de cuenta de almacenamiento *


stfhuiscandar2025







Propiedades de conexión adicionales


 Nuevo


Prueba de conexión 


 Al servicio vinculado


 O la ruta de acceso de archivo


Anotaciones

 Nuevo

 Parámetros









Crear

Cancelar

 Conexión correcta

 Prueba de conexión

11. Pipeline exitoso

The screenshot shows the Azure Data Explorer web interface. At the top, there's a navigation bar with 'Productos' and a search bar. Below it, the 'Actividades' (Activities) section is active, displaying a list of operations. A modal dialog titled 'Copiar datos' (Copy data) is open, showing a green checkmark and the text 'Copy data1'. The main panel shows a table with columns: 'Id de ejecución de consulta' (Query execution ID), 'Nombre de actividad' (Activity name), 'Estado de...' (Status), 'Tipo de...' (Type), 'Inicio de la ejecución...' (Execution start), 'Duración...' (Duration), 'Extremo de ejecución de integración...' (Integration execution end), and 'Ph...'. The first row shows a successful copy operation with ID 'fa7903a-5d8f-4302-a956-3a70e6a521f9' and status 'Completado' (Completed). The bottom status bar indicates 'AutoscaleIntegrationsRuntime (Spain Central)'.

12..Datos en SQL Database