

**Diseño de una plataforma de integración de datos con transformaciones automatizadas  
(ETL inteligente) utilizando Azure Data Factory, Databricks y Power BI**

Paola Fernanda Lopez Recalde

Asesor

Lina Rocío Rivadeneira

Universidad Nacional Abierta y a Distancia UNAD  
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI  
Especialización en Ciencia de Datos y Analítica

2025

## Resumen

Las empresas que no logran una integración de datos eficiente experimentan una disminución del 30 % en su productividad operativa, según un informe de Deloitte sobre el impacto de la transformación digital en las empresas. Esta falta de integración puede colocarlas en desventaja frente a competidores que tienen acceso a datos en tiempo real para apoyar decisiones estratégicas. En sectores como el retail y la logística, donde la capacidad de responder rápidamente a los cambios del mercado es crucial, esta desventaja puede traducirse en pérdida de ingresos y menor capacidad de innovación.

La intervención manual para ajustar procesos de ETL puede ser lenta, más aún cuando ocurren cambios inesperados en los datos de origen, originando que los datos no estén disponibles en el tiempo que se requieren para el análisis, afectando la capacidad de la empresa para responder rápidamente a las tendencias y necesidades del mercado.

Con cada ajuste manual, aumenta la posibilidad de cometer errores, especialmente cuando las reglas de negocio o los esquemas de datos son difíciles, estos pueden generar datos incorrectos que afectan la calidad del análisis y la confianza en los datos por parte de los usuarios.

Con este trabajo de grado se diseñó una solución basada en los servicios de Azure, integrando Azure Data Factory para la orquestación de procesos ELT, Azure Databricks para el desarrollo de transformaciones inteligentes de datos y Power BI para la construcción de tableros de control interactivos. El objetivo fue disponibilizar todos los procesos de ingesta de información, generación de reportes, dashboards e indicadores clave, con el fin de optimizar los procesos y tecnologías actualmente utilizadas por la empresa.

Asimismo, se propició una transición operativa gradual hacia la nueva plataforma, acompañada de un proceso de transferencia de conocimiento que permitió desarrollar capacidades en el personal para el uso, explotación y aprovechamiento de esta solución de Business Intelligence (BI) moderna y escalable.

Usar Power BI integrado con Databricks combina lo mejor de dos mundos: el procesamiento avanzado y la escalabilidad de Databricks con la facilidad de uso y la potencia de visualización de Power BI, haciendo que los datos sean más accesibles, accionables y relevantes para todos los niveles de la organización.

**Palabras clave:** Integración de datos, ETL inteligente, azure databricks, azure data factory, power BI, transformación digital, business intelligence, automatización de procesos, plataforma tecnológica

## Abstract

Companies that fail to achieve efficient data integration experience a 30% decrease in operational productivity, according to a Deloitte report on the impact of digital transformation on businesses. This lack of integration can place them at a disadvantage compared to competitors who have access to real-time data to support strategic decisions. In sectors such as retail and logistics, where the ability to quickly respond to market changes is crucial, this disadvantage can translate into loss of revenue and reduced innovation capacity.

Manual intervention to adjust ETL processes can be slow, especially when unexpected changes occur in source data, resulting in data not being available when needed for analysis. This affects the company's ability to respond quickly to market trends and demands.

With each manual adjustment, the likelihood of making mistakes increases, especially when business rules or data schemas are complex. These mistakes can lead to inaccurate data, which in turn affects the quality of analysis and user trust in the data.

This thesis project involved the design of a solution based on Azure services, integrating Azure Data Factory for the orchestration of ELT processes, Azure Databricks for the development of intelligent data transformations, and Power BI for the creation of interactive dashboards. The objective was to make all information ingestion processes, report generation, dashboards, and key performance indicators available in order to optimize the current processes and technologies used by the company.

Additionally, a gradual operational transition to the new platform was promoted, along with a knowledge transfer process that enabled the development of staff capabilities for the use, exploitation, and effective adoption of this modern and scalable Business Intelligence (BI) solution.

Using Power BI integrated with Databricks combines the best of both worlds: the advanced processing and scalability of Databricks with the ease of use and powerful visualization capabilities of Power BI, making data more accessible, actionable, and relevant at all levels of the organization.

**Keywords:** Data integration, intelligent ETL, azure databricks, azure data factory, power BI, digital transformation, business intelligence, process automation, technology platform

## Tabla de Contenido

Introducción .....	10
Justificación .....	11
Objetivos.....	14
Objetivo General.....	14
Objetivos Específicos.....	14
Marco Conceptual y Teórico .....	15
Metodología .....	18
Análisis y Diagnóstico Inicial.....	18
Diseño de la Solución .....	18
Desarrollo de Flujos ETL.....	19
Configuración del Entorno Databricks .....	19
Creación de Tableros en Power BI .....	19
Monitoreo y Evaluación.....	20
Conclusiones .....	23
Recomendaciones .....	25
Referencias Bibliograficas .....	26
Apéndices.....	28

## Lista de Tablas

<b>Tabla 1</b> <i>Cronograma de Actividades</i> .....	19
<b>Tabla 2</b> <i>Recursos Necesarios</i> .....	20
<b>Tabla 3</b> <i>Resultado de Productos Esperados</i> .....	21

## Lista de Apéndices

<b>Apéndice A</b> <i>Análisis y Diagnóstico Inicial- Inventario_ETL_Analisis_FaseA.xlsx</i> .....	28
<b>Apéndice B</b> <i>Diseño de la Solución, Diagrama de Arquitectura de la Tecnología Propuesta</i> .....	29
<b>Apéndice C</b> <i>Diseño Lógico del Flujo de Datos y Transformación</i> .....	30
<b>Apéndice D</b> <i>Especificación de los Procesos ETL/ELT Automatizados</i> .....	31
<b>Apéndice E</b> <i>Herramientas y Componentes Seleccionados</i> .....	32
<b>Apéndice F</b> <i>Reglas de Negocio Aplicadas al Tratamiento de los Datos</i> .....	33
<b>Apéndice G</b> <i>Desarrollo de Flujos ETL, Canalizaciones (Pipelines)</i> .....	34
<b>Apéndice H</b> <i>Flujo de Datos</i> .....	35
<b>Apéndice I</b> <i>Origen</i> .....	36
<b>Apéndice J</b> <i>Destino</i> .....	37
<b>Apéndice K</b> <i>Almacenamiento</i> .....	38
<b>Apéndice L</b> <i>Visualización del Archivo CSV Exportado desde Azure Data Factory</i> .....	39
<b>Apéndice M</b> <i>Configuración del Entorno Databricks, Configuración de Clústeres en Azure Databricks</i> .....	40
<b>Apéndice N</b> <i>Notebooks Desarrollados en Scala para Capa Silver</i> .....	41
<b>Apéndice O</b> <i>Notebooks Desarrollados en Scala para Capa Gold</i> .....	42
<b>Apéndice P</b> <i>Código Generado en Lenguaje Scala para Capa Silver</i> .....	43
<b>Apéndice Q</b> <i>Código Generado en Lenguaje Scala para Capa Gold</i> .....	44
<b>Apéndice R</b> <i>Tablero Desarrollado en Power BI – Avance de Contratos</i> .....	45
<b>Apéndice S</b> <i>Conexión DirectQuery desde Power BI a Azure Databricks</i> .....	46
<b>Apéndice T</b> <i>Vista Previa de Datos Conectados desde Azure Databricks en Power</i> .....	47
<b>Apéndice U</b> <i>Capturas del Tablero Publicado</i> .....	48

<b>Apéndice V</b> <i>Métricas y KPIs Implementados</i> .....	49
<b>Apéndice W</b> <i>Monitoreo y Evaluación</i> .....	50
<b>Apéndice X</b> <i>Manual de Usuario y Guía Operativa del Entorno Databricks y Power BI</i> .....	51

## Introducción

En la era del big data y la transformación digital, las organizaciones enfrentan el reto de procesar e integrar grandes volúmenes de datos provenientes de múltiples fuentes, en tiempo real y con altos niveles de confiabilidad. No hacerlo puede generar consecuencias significativas: pérdida de productividad, errores operativos, aumento de costos, e incluso pérdida de competitividad. Según estudios recientes de firmas como Deloitte y McKinsey, más del 70% de las empresas aún presentan dificultades importantes en sus procesos de integración de datos, especialmente aquellas que operan con sistemas ETL tradicionales (Deloitte, 2023; McKinsey, 2022).

Este proyecto surge precisamente ante esa necesidad: mejorar la eficiencia, la calidad y la escalabilidad de los procesos de datos en una empresa especializada en distribución, comercialización y dispensación de medicamentos y que diariamente procesa información desde varias fuentes. En la actualidad, estos procesos son mayoritariamente manuales, poco flexibles y propensos a errores, lo que impacta directamente la toma de decisiones estratégicas, el cumplimiento normativo y la capacidad de adaptación de la empresa frente a un entorno cambiante.

Como respuesta a esta problemática, se diseñó una solución inteligente basada en los servicios de Azure para la automatización del proceso ELT y la visualización y análisis estratégico de datos. Esta propuesta permite no solo garantizar la disponibilidad de la información en tiempo real, sino también reducir el esfuerzo operativo, mejorar la calidad de los datos, y facilitar la toma de decisiones a todos los niveles de la organización.

## Justificación

Actualmente, las organizaciones enfrentan un entorno cada vez más dinámico y competitivo donde la capacidad de procesar e integrar datos de manera eficiente es crucial para mantener su ventaja competitiva. Los datos de origen provienen de diversos sistemas que, con frecuencia, actualizan sus formatos, fuentes, esquemas o estructuras sin previo aviso. Esto genera la necesidad de ajustes manuales constantes para garantizar la integración adecuada de los datos, lo cual no solo es costoso y propenso a errores, sino que también ralentiza los procesos de toma de decisiones.

Resolver este problema es esencial porque permite a las organizaciones obtener varios beneficios concretos:

**Mejora en la Toma de Decisiones:** Al automatizar los procesos ETL y garantizar datos confiables y actualizados en tiempo real, las empresas pueden basar sus decisiones en información precisa, lo que les permite responder de manera más ágil a los cambios del mercado y a las necesidades de sus clientes.

**Reducción de Costos:** La automatización reduce la necesidad de intervención manual, minimiza los errores y disminuye los tiempos de parada en los sistemas. Esto conlleva una disminución significativa en los costos operativos y en los recursos dedicados a la corrección de errores.

**Cumplimiento Normativo:** En sectores altamente regulados, como el de salud y financiero, la capacidad de consolidar y reportar datos de manera eficiente y confiable es fundamental para evitar sanciones e incumplimientos normativos.

**Escalabilidad y Adaptabilidad:** Un sistema ETL inteligente basado en la nube permite adaptarse dinámicamente a cambios en las fuentes de datos, formatos y esquemas sin

interrupciones. Esto es crucial para organizaciones que operan en múltiples mercados con distintas regulaciones y características.

**Impacto Social y Ambiental Positivo:** Además de los beneficios operativos, esta solución puede contribuir a objetivos más amplios de sostenibilidad y equidad social. Al optimizar los flujos de datos, se reduce el consumo de recursos computacionales y energéticos asociados con procesos manuales ineficientes, disminuyendo así la huella de carbono de las operaciones digitales. Por otro lado, al proporcionar datos confiables y accesibles, las organizaciones pueden tomar decisiones más responsables, como optimizar cadenas de suministro para reducir desperdicios o implementar estrategias más equitativas en sus mercados, beneficiando tanto a la sociedad como al medio ambiente.

La implementación de una solución basada en Azure Data Factory, Databricks y Power BI es altamente viable por las siguientes razones:

**Entorno Unificado y Escalable:** Azure Data Factory permite la orquestación de flujos de datos automatizados y seguros, facilitando la integración entre múltiples fuentes de información. En combinación, Databricks proporciona una plataforma robusta que integra el poder de Apache Spark con herramientas avanzadas de procesamiento distribuido y transformación de datos. Esto permite manejar grandes volúmenes de datos en tiempo real, adaptarse a cambios en las fuentes y construir flujos de trabajo optimizados y eficientes.

**Integración con Power BI:** Power BI complementa la solución al ofrecer herramientas de visualización avanzadas que permiten a las organizaciones transformar datos complejos en información clara y accionable. Esta integración facilita la creación de tableros dinámicos y reportes en tiempo real que apoyan la toma de decisiones estratégicas.

**Alineación con Objetivos de Ciencia de Datos:** Este proyecto está alineado con los objetivos de la especialización en Ciencia de Datos al integrar técnicas avanzadas de procesamiento y análisis de datos con herramientas de visualización y toma de decisiones. Además, fomenta el uso de tecnologías modernas para resolver problemas reales, como la automatización y la escalabilidad en los flujos de datos.

**Impacto Estratégico:** Adoptar una solución de ETL inteligente no solo resuelve los problemas actuales de integración y transformación de datos, sino que también posiciona a la organización como líder en el uso de tecnologías avanzadas. Esto les permite:

Aprovechar al máximo sus datos significa usar la información que ya tienen las empresas para tomar decisiones más inteligentes y rápidas, al hacerlo, pueden destacarse frente a la competencia en un mundo cada vez más digital, además fomenta una forma de trabajo donde las decisiones importantes se basan en hechos y números, lo que ayuda a las empresas a mantenerse fuertes y preparadas para los retos del futuro.

Con base en lo anterior, esta solución no solo es relevante, sino que es un paso estratégico hacia la transformación digital, cumpliendo con las necesidades actuales y futuras del sector. Además, al contribuir al uso eficiente de recursos y a decisiones empresariales responsables, esta iniciativa refuerza el compromiso de las organizaciones con la sostenibilidad y el impacto positivo en la sociedad.

## **Objetivos**

### **Objetivo General**

Diseñar una solución inteligente y automatizada de integración y análisis de datos, basada en Azure Data Factory, Azure Databricks y Power BI, que optimice los procesos tradicionales de ETL, garantizando la disponibilidad de datos en tiempo real y apoyando la toma de decisiones estratégicas de la empresa.

### **Objetivos Específicos**

Analizar los procesos ETL actuales y las necesidades de la organización para identificar puntos de mejora y diseñar una transición gradual hacia un sistema inteligente y automatizado.

Desarrollar flujos de trabajo ETL automatizados utilizando Azure Data Factory como orquestador y Azure Databricks para las transformaciones, garantizando la adaptabilidad a cambios en las fuentes de datos y la calidad de los datos procesados.

Integrar Power BI con Databricks para construir tableros de control interactivos que faciliten el análisis en tiempo real y la visualización efectiva de métricas clave.

Evaluar el impacto de la solución implementada mediante métricas de desempeño como la reducción de tiempos de procesamiento, mejora en la calidad de los datos y satisfacción de los usuarios.

## Marco Conceptual y Teórico

El presente proyecto se fundamenta en conceptos clave que respaldan la transición de sistemas ETL tradicionales hacia plataformas inteligentes apoyadas en tecnologías modernas como Azure Data Factory, Azure Databricks y Power BI. Esta transformación resulta esencial para responder a las crecientes demandas de escalabilidad, automatización y análisis en tiempo real, características indispensables en la actual era de la ciencia de datos y la integración en la nube.

ETL Tradicional vs. ETL Inteligente: Los sistemas tradicionales de ETL (Extract, Transform, Load) han sido ampliamente utilizados para extraer datos de diversas fuentes, transformarlos según las necesidades del negocio y cargarlos en un almacén de datos. No obstante, presentan limitaciones notables como rigidez, bajo rendimiento y escasa flexibilidad (Kimball & Ross, 2013).

En contraste, los sistemas ETL inteligentes aprovechan tecnologías como Apache Spark y Databricks, que permiten automatizar flujos de trabajo y escalar el procesamiento de datos con ventajas como:

Automatización mediante reglas dinámicas y flujos programables.

Procesamiento distribuido para manejar grandes volúmenes de datos.

Integración en tiempo real con capacidades de respuesta ante cambios instantáneos (Databricks, 2023; Microsoft Azure, 2023).

ELT (Extract, Load, Transform) : El enfoque ELT representa una evolución del modelo tradicional ETL. En este esquema, los datos se cargan primero en el Data Lake o Data Warehouse y se transforman posteriormente, aprovechando las capacidades de procesamiento

distribuido de la nube. Esto permite mayor rendimiento, escalabilidad y rapidez en la integración (Databricks, 2023; Azure Data Factory, 2023).

Databricks, en particular, automatiza las transformaciones en Data Lakes e integra herramientas analíticas para acelerar la entrega de valor a las áreas de negocio.

**Data Lakes:** Un Data Lake es un repositorio centralizado que permite almacenar datos estructurados, semiestructurados y no estructurados en su formato original. A diferencia de los Data Warehouses, los Data Lakes ofrecen mayor flexibilidad y costos más bajos de almacenamiento, lo que permite utilizar los datos sin necesidad de transformaciones previas (AWS, 2022).

Databricks, por ejemplo, integra datos desde Data Lakes utilizando Apache Spark para el procesamiento distribuido eficiente (Spark, 2022).

**Computación en la Nube:** La computación en la nube ha transformado la forma en que las organizaciones gestionan sus datos, permitiendo:

Escalabilidad bajo demanda.

Reducción de costos mediante el modelo de pago por uso.

Accesibilidad desde cualquier ubicación, en tiempo real.

Plataformas como Azure y Databricks aprovechan esta tecnología para facilitar la ingesta, transformación y análisis de grandes volúmenes de datos sin necesidad de infraestructura física local (Microsoft Azure, 2023; IDC, 2021).

**Metodología Kimball y Ross:** La metodología de Kimball y Ross (2013) sigue siendo un enfoque clásico y vigente en el diseño de Data Warehouses. Sus principios clave incluyen:

Integración de datos desde múltiples fuentes.

Optimización de rendimiento mediante esquemas dimensionales.

Aseguramiento de la calidad de los datos para decisiones confiables.

Este enfoque es complementario a herramientas modernas como Databricks, que permiten aplicar dichos principios en entornos altamente dinámicos.

Databricks: Databricks es una plataforma de análisis de datos en la nube que combina Apache Spark con funcionalidades avanzadas de integración, transformación y análisis. Ofrece:

Procesamiento distribuido eficiente.

Automatización de flujos de trabajo.

Capacidades de análisis en tiempo real.

Esta plataforma facilita la creación de soluciones modernas de ETL inteligente integradas con herramientas de visualización como Power BI, permitiendo una toma de decisiones más ágil y basada en datos actualizados (Databricks, 2023; Spark, 2022).

## **Metodología**

Este proyecto se desarrolló mediante una metodología de tipo proyectiva-aplicada, orientada a la solución de una problemática específica mediante el diseño e implementación de una solución tecnológica en entornos reales como el de esta investigación. La estructura metodológica se basó en fases secuenciales y progresivas, las cuales permitieron abordar cada etapa del proyecto de forma organizada y controlada. A continuación, se describen las fases y actividades desarrolladas a lo largo de los 6 meses de ejecución:

### **Análisis y Diagnóstico Inicial**

Durante el primer mes se llevó a cabo el levantamiento de información y el diagnóstico de los procesos ETL actuales de la organización. Esta fase incluyó reuniones con usuarios clave e ingenieros de datos para elaborar un documento que contiene el resumen del proyecto, el plan de migración, el inventario de objetos, la equivalencia por tipo de datos, la revisión de la documentación técnica y el análisis de los principales cuellos de botella. Este trabajo permitió establecer una línea base y comprender las limitaciones del sistema actual.

### **Diseño de la Solución**

En el segundo mes se llevó a cabo el diseño conceptual y técnico de la solución propuesta. Durante esta etapa, se definieron la arquitectura de datos, los componentes del flujo ELT y las herramientas a utilizar: Azure Data Factory como orquestador de procesos, Azure Databricks para la transformación y normalización de los datos, y Power BI para la visualización y análisis de la información.

Asimismo, se establecieron los requerimientos funcionales y no funcionales, y se definieron las capas de extracción Bronze, Silver y Gold, siguiendo las buenas prácticas en arquitectura de datos moderna. Esta etapa fue fundamental para garantizar la alineación entre los

objetivos del proyecto y la solución tecnológica planteada, asegurando escalabilidad, trazabilidad y eficiencia en el procesamiento de datos.

### **Desarrollo de Flujos ETL**

En el tercer mes se inició el desarrollo de los flujos ETL automatizados utilizando Azure Data Factory, el servicio de integración y transformación de datos en la nube de Azure. Se diseñaron flujos modulares y reutilizables, orientados a la extracción, transformación y carga de datos desde Dynamics 365 (D365) mediante el consumo de APIs. Esta fase también incluyó la ejecución de pruebas de calidad y validación de datos, asegurando la integridad y confiabilidad de la información procesada.

### **Configuración del Entorno Databricks**

Durante el cuarto mes se realizó la configuración inicial de los servicios en la nube mediante Azure Databricks, lo que incluyó la creación de clústeres, la definición de rutas de almacenamiento en Data Lake, la conexión con las fuentes de datos y la ejecución de pruebas de conectividad.

Adicionalmente, se desarrollaron notebooks en Scala y Python para llevar a cabo la transformación y normalización de los datos, los cuales fueron almacenados en las capas Silver y Gold, conforme al diseño de arquitectura previamente establecido. Esta fase fue clave para habilitar una infraestructura escalable y eficiente para el procesamiento avanzado de datos.

### **Creación de Tableros en Power BI**

En el quinto mes se desarrollaron los tableros de control (dashboards) con Power BI, conectados directamente al Data Lake a través de Databricks. Se diseñaron visualizaciones dinámicas, KPIs estratégicos e indicadores operativos que permiten a los usuarios tomar decisiones informadas a partir de datos actualizados en tiempo real.

## Monitoreo y Evaluación

Durante el sexto mes se llevó a cabo el monitoreo de los triggers en Azure Data Factory, así como la validación de los datos procesados y finales. Además, se verificó el desempeño de los tableros desarrollados en Power BI conectados a Databricks, evaluando su tiempo de respuesta, exactitud de la información presentada y consistencia con las fuentes de datos. Esta etapa fue clave para asegurar la confiabilidad operativa y la calidad de los entregables de la solución implementada.

**Tabla 1**

### *Cronograma de Actividades*

Actividad	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6
Análisis y diagnóstico inicial	X					
Diseño de la solución		X				
Desarrollo de flujos ETL			X			
Configuración del entorno Databricks				X		
Creación de tableros en Power BI					X	
Monitoreo y evaluación						X

*Nota.* Esta tabla refleja una planificación eficiente y estructurada, acorde con los objetivos planteados y con una adecuada distribución del tiempo para cada fase crítica del proyecto.

*Fuente.* Autor.

**Tabla 2***Recursos Necesarios*

Recurso	Descripción
1. Equipo Humano	Un equipo conformado por un analista de datos, un desarrollador ETL y un especialista en Power BI.
2. Equipos y Software	Licencias para Databricks, Power BI Pro, y servicios de nube (Azure).
3. Viajes y Salidas de Campo	Reuniones con clientes o stakeholders (en caso de ser necesario).
4. Materiales y suministros	Manuales de usuario, material de capacitación (presentaciones, guías técnicas).
5. Bibliografía	Libros, artículos y referencias relacionadas con ETL, BI y tecnologías en la nube.

*Nota.* Esta tabla detalla los recursos necesarios para la ejecución del proyecto, clasificados en cinco categorías, estos recursos son fundamentales para garantizar la correcta implementación de los procesos ETL automatizados y la generación de tableros analíticos con Power BI. *Fuente.*

Autor.

**Tabla 3***Resultado de Productos Esperados*

Resultado/Producto esperado	Indicador	Beneficiario
Automatización de los procesos ETL actuales mediante Databricks	Porcentaje de flujos de datos automatizados y reducción del tiempo promedio de procesamiento.	Área de TI y análisis de la organización.
Tableros interactivos y dinámicos en Power BI para la visualización en tiempo real de métricas clave	Número de tableros creados y validados por los usuarios finales.	Equipo directivo y analistas de negocio.
Mejora en la calidad y accesibilidad de los datos procesados	Porcentaje de reducción en errores de datos y disponibilidad de datos en tiempo real.	Usuarios finales de datos e información
Informe final con los resultados obtenidos y recomendaciones para la optimización continua del sistema implementado	Documento entregado que resuma el impacto y las áreas de mejora futuras.	Dirección de la organización.

*Nota.* Esta tabla presenta los principales resultados esperados del proyecto, junto con sus respectivos indicadores de cumplimiento y los beneficiarios directos dentro de la organización.

*Fuente.* Autor.

## Conclusiones

Implementación exitosa de una solución integral de datos. La plataforma desarrollada logró integrar satisfactoriamente los servicios de Azure Data Factory, Azure Databricks y Power BI, permitiendo automatizar los procesos de extracción, transformación y carga de datos (ETL) desde fuentes como Dynamics 365. Esta integración permitió construir una solución robusta, escalable y adaptable a entornos reales de producción.

Transformaciones eficientes mediante notebooks en Databricks. El uso de notebooks en Scala y Python dentro de Azure Databricks fue clave para llevar a cabo transformaciones complejas y normalización de los datos, respetando los principios de calidad, trazabilidad y eficiencia. Los datos transformados fueron almacenados en las capas Silver y Gold, siguiendo una arquitectura modular y basada en buenas prácticas.

Automatización y orquestación de flujos ELT con Azure Data Factory. La creación de flujos modulares y reutilizables en Azure Data Factory permitió orquestar las tareas de integración de forma automatizada y supervisada. El monitoreo de los triggers y la validación de cada etapa del proceso garantizó la confiabilidad y continuidad operativa del sistema.

Visualización y toma de decisiones basadas en datos en tiempo real. La implementación de tableros dinámicos en Power BI, conectados directamente al Data Lake a través de Databricks, permitió a los usuarios acceder a indicadores estratégicos y operativos actualizados en tiempo real, fortaleciendo la toma de decisiones basada en datos (data-driven, decisión making).

Metodología aplicada adecuada para entornos empresariales reales. La metodología proyectiva-aplicada, basada en fases secuenciales, permitió estructurar el proyecto de forma

ordenada y efectiva. Cada etapa respondió a necesidades específicas, desde el diagnóstico inicial hasta la validación final, alineándose con los objetivos técnicos y organizacionales.

Escalabilidad y sostenibilidad de la solución. La arquitectura propuesta no solo resuelve la problemática identificada, sino que también sienta las bases para su evolución y expansión futura. La solución puede adaptarse fácilmente a nuevas fuentes de datos, transformaciones adicionales o requerimientos analíticos, garantizando sostenibilidad tecnológica en el tiempo.

## Recomendaciones

Continuar con la capacitación del personal técnico: Es importante ofrecer entrenamientos continuos sobre el uso de Databricks y Power BI, asegurando que el equipo interno pueda mantener, escalar y adaptar la solución de forma autónoma.

Expandir la automatización a otros procesos clave: Se recomienda replicar este modelo en otras áreas de la empresa que aún utilizan procesos manuales o semiautomáticos, con el fin de lograr una estandarización integral del flujo de datos.

Implementar un sistema de monitoreo continuo: Desarrollar alertas y mecanismos de monitoreo que permitan detectar errores en los pipelines de datos en tiempo real, mejorando la respuesta ante incidencias.

Documentar y versionar los notebooks y flujos: Mantener una documentación técnica actualizada y utilizar control de versiones para los notebooks ayudará a facilitar el mantenimiento y colaboración futura en el entorno Databricks.

Evaluar la incorporación de modelos predictivos: Una vez estabilizado el entorno de análisis, se sugiere explorar la integración de modelos de ciencia de datos y aprendizaje automático (machine learning) para anticipar comportamientos del negocio y optimizar decisiones estratégicas.

## Referencias Bibliográficas

- AWS Insights. (2023). *Migración a la nube: Estrategias para la integración de datos escalables*. AWS Whitepaper. <https://aws.amazon.com>
- Blog Databricks. (2022). Desafíos comunes en la implementación de ETL. Databricks Engineering Blog. <https://databricks.com>
- Conecta Magazine. (2023). ETL y ELT: Principales diferencias y aplicaciones. Conecta Insights. <https://conecta-magazine.com>
- Data Integration Weekly. (2024). Transformación de datos en tiempo real en los modernos pipelines ETL. <https://dataintegrationweekly.com>
- Databricks. (2024). Guía de Servicios Databricks. Databricks Documentation. <https://databricks.com>
- Deloitte. (2023). *Informe sobre transformación digital y productividad*. Deloitte Insights.
- Deloitte Insights. (2023). Tendencias de integración de datos en la nube. Deloitte Tech Reports. <https://deloitte.com>
- Forrester Research. (2023). Recursos de integración de API, SaaS, SOA. Forrester Reports. <https://forrester.com>
- Gartner. (2022). Arquitectura de integración: Un modelo empresarial moderno. Gartner Insights. <https://gartner.com>
- Gupta, A. (2021). *La arquitectura de tejido de datos moderniza la integración*. Data Architecture Journal.
- Guzmán, C. (2021). *Automated ETL Platforms for Big Data*. Springer. <https://link.springer.com>
- IBM Research. (2023). Uso de plataformas iPaaS para la integración de datos. IBM Whitepaper. <https://ibm.com>

- Jace, A. (2020). *77 bloques para construir una transformación digital: El modelo de capacidad digital*. Wiley Press.
- Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling* (3rd ed.). Wiley.
- McKinsey & Company. (2022). *Tendencias de integración de datos en empresas modernas*. McKinsey Reports.
- Microsoft. (2023). Azure: Integraciones de plataformas para la transformación digital. Microsoft Azure Docs. <https://azure.microsoft.com>
- Microsoft. (2024). ¿Qué es la transformación de datos?. Microsoft Learn. <https://learn.microsoft.com>
- Power BI Blog. (2023). Análisis en tiempo real para la toma de decisiones estratégicas: Casos de éxito con Power BI y Databricks. <https://powerbi.microsoft.com>
- Smith, J., & Taylor, R. (2024). *Optimización de procesos ETL para almacenamiento de datos*. O'Reilly Media.
- Taylor, M. (2022). Beneficios de ETL en la gestión empresarial. *Harvard Business Review*. <https://hbr.org>
- Taylor & Francis. (2023). *Integración y procesamiento de Big Data*. Taylor & Francis Publishing.
- White, J., & Miller, R. (2023). *Data engineering: Transformaciones automatizadas para la era digital*. Pearson.

## Apéndices

### Apéndice A

*Análisis y Diagnóstico Inicial- Inventario\_ETL\_Analisis\_FaseA.xlsx*



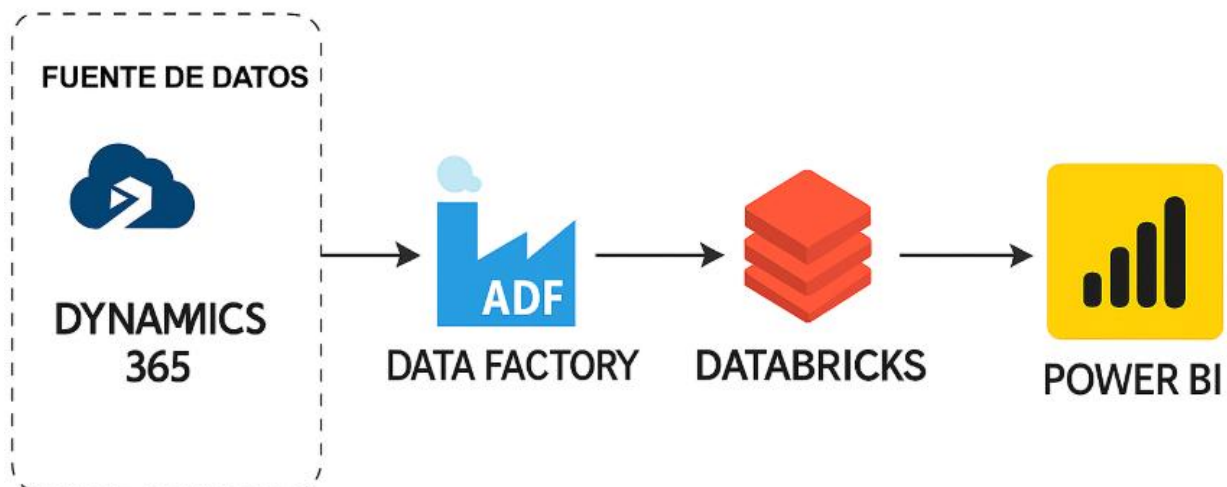
Inventario\_ETL\_Analisi  
s\_FaseA.xlsx.xlsx

*Nota.* Esta evidencia corresponde al archivo Excel que contiene el inventario general y analítico de los objetos de base de datos. Incluye el mapeo de datos origen–destino, la identificación de riesgos y el alcance del proceso de migración.

## Apéndice B

*Diseño de la Solución, Diagrama de Arquitectura de la Tecnología Propuesta*

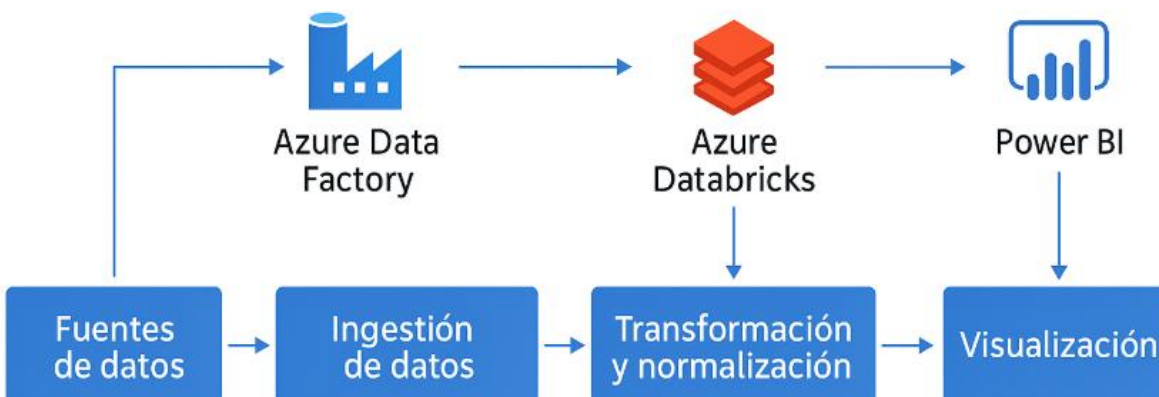
### DIAGRAMA DE ARQUITECTURA TECNOLÓGICA PROPUESTA



*Nota.* Representación gráfica de los componentes de la solución, flujos de datos y herramientas seleccionadas.

## Apéndice C

### *Diseño Lógico del Flujo de Datos y Transformación*



*Nota.* Esquema lógico que describe el recorrido de los datos desde su origen hasta su visualización final.

**Apéndice D***Especificación de los Procesos ETL/ELT Automatizados*

<b>ETAPA</b>	<b>EXTRACCIÓN</b>	<b>TRANSFORMACIÓN</b>	<b>CARGA</b>
<b>BRONCE</b>	Extracción		Carga
<b>SILVER</b>		Transformación	
<b>GOLD</b>			

*Nota.* Diagrama o tabla que muestra las etapas de extracción, carga y transformaciones definidas en el proyecto.

## Apéndice E

### *Herramientas y Componentes Seleccionados*

Tecnología	Rol / uso en el proyecto
Azure Data Factory (ADF)	Orquestación de <i>pipelines</i> de datos e integración desde múltiples fuentes.
Azure Databricks	Procesamiento distribuido de grandes volúmenes con <i>notebooks</i> en PySpark y SQL.
Power BI	Visualización y análisis: creación de <i>dashboards</i> y reportes estratégicos.
Data Lake Storage (ADLS Gen2)	Almacenamiento en la nube por capas (Bronce, Silver, Gold) para datos crudos, refinados y curados.
Azure Active Directory	Autenticación y gestión de usuarios, permisos y accesos (incluido acceso a Databricks).
Microsoft Excel	Documentación de inventarios, cronogramas y generación de entregables operativos.

*Nota.* La tabla muestra las principales herramientas utilizadas en el desarrollo de la solución tecnológica propuesta.

## Apéndice F

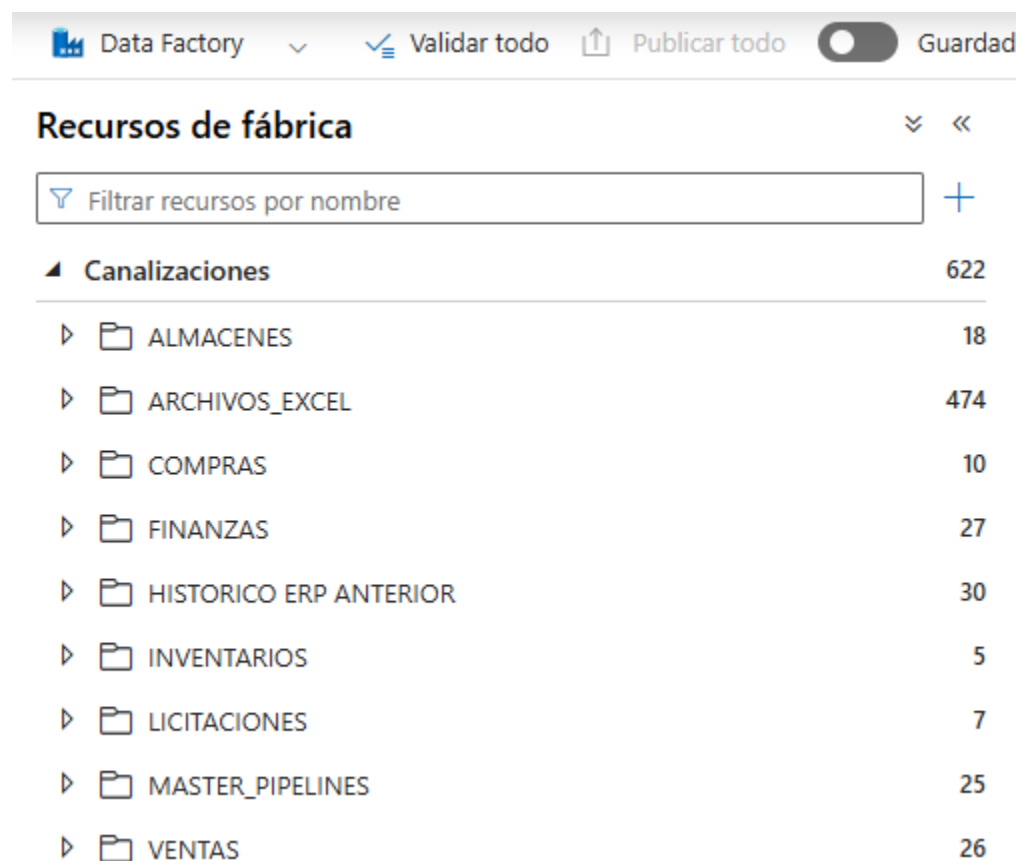
### Reglas de Negocio Aplicadas al Tratamiento de los Datos

Regla del negocio aplicada	¿Qué se hace?	Propósito / resultado
Normalización de nombres de columnas	Se convierten todos los nombres a minúsculas.	Coherencia de esquema y evitar errores por <i>case sensitive</i> .
Limpieza de valores vacíos en cadenas	Los "" en columnas String se transforman a NULL.	Facilitar filtros y evitar fallos en agregaciones/joins.
Selección y renombramiento de columnas	Se eligen las columnas requeridas y se renombran.	Estandarizar nombres y alinear con el modelo destino.
Unificación de fuentes (JOIN)	Se combinan tablas: Guías de viaje, Contratos, Pedidos remisionados, Remisiones facturadas; luego se vuelve a seleccionar/renombrar el resultado.	Obtener un dataset integrado y consistente.
Eliminación de duplicados	<i>Dedup</i> usando claves como ID documento, cliente, fecha.	Evitar registros repetidos en la salida final.
Conversión de tipos de datos	Estandarización por columna (p. ej., fecha yyyy-MM-dd, importe Decimal(16,6)).	Compatibilidad con sistemas de carga y exactitud de tipos.
Generación de campo de fecha de carga	Se añade fecha_carga con la fecha de ejecución.	Trazabilidad para auditoría y seguimiento.
Aplicación de filtros condicionales	Se excluyen registros anulados, incompletos o con fechas fuera de rango.	Asegurar calidad del conjunto final.

*Nota.* La tabla describe, de forma resumida, las reglas de negocio aplicadas en la preparación e integración del dataset: qué acción se ejecuta (normalización, limpieza, selección/renombrado, uniones, de duplicación, conversiones de tipo, generación de fecha\_carga y filtros) y con qué propósito, para garantizar la coherencia, trazabilidad y calidad de los datos en la salida final.

## Apéndice G

### *Desarrollo de Flujos ETL, Canalizaciones (Pipelines)*



The screenshot shows the 'Recursos de fábrica' (Factory Resources) page in Azure Data Factory. At the top, there is a navigation bar with 'Data Factory', 'Validar todo', 'Publicar todo', and a 'Guardad' toggle. Below the title, there is a search box labeled 'Filtrar recursos por nombre'. The main content is a list of folders representing functional areas, each with a right-pointing arrow and a count of pipelines. The folders are: ALMACENES (18), ARCHIVOS\_EXCEL (474), COMPRAS (10), FINANZAS (27), HISTORICO ERP ANTERIOR (30), INVENTARIOS (5), LICITACIONES (7), MASTER\_PIPELINES (25), and VENTAS (26).

Área Funcional	Cantidad de Canalizaciones
ALMACENES	18
ARCHIVOS_EXCEL	474
COMPRAS	10
FINANZAS	27
HISTORICO ERP ANTERIOR	30
INVENTARIOS	5
LICITACIONES	7
MASTER_PIPELINES	25
VENTAS	26

*Nota.* Captura de pantalla del entorno de desarrollo de Azure Data Factory, donde se muestra la organización jerárquica de las canalizaciones (pipelines) agrupadas por áreas funcionales. Se observa un total de 622 canalizaciones distribuidas en carpetas como ALMACENES, ARCHIVOS\_EXCEL, FINANZAS, VENTAS, entre otras, lo cual facilita la gestión modular, reutilizable y ordenada del flujo de integración de datos.

## Apéndice H

### Flujo de Datos




*Nota.* Vista simplificada de un pipeline en Azure Data Factory, donde se realiza la actividad de copia de datos desde la fuente Ext\_Ptycsa\_Consulta\_List\_Remision hacia el flujo de transformación denominado cifras\_control.

## Apéndice I

### Origen

---

 REST  
EXT\_Ptycsa\_Proveedores\_inp

---

Conexión    Parámetros

---

Servicio vinculado \*    PHARMATYCSA\_BC\_API    Prueba de conexión    Editar    + Nuevo    Más información

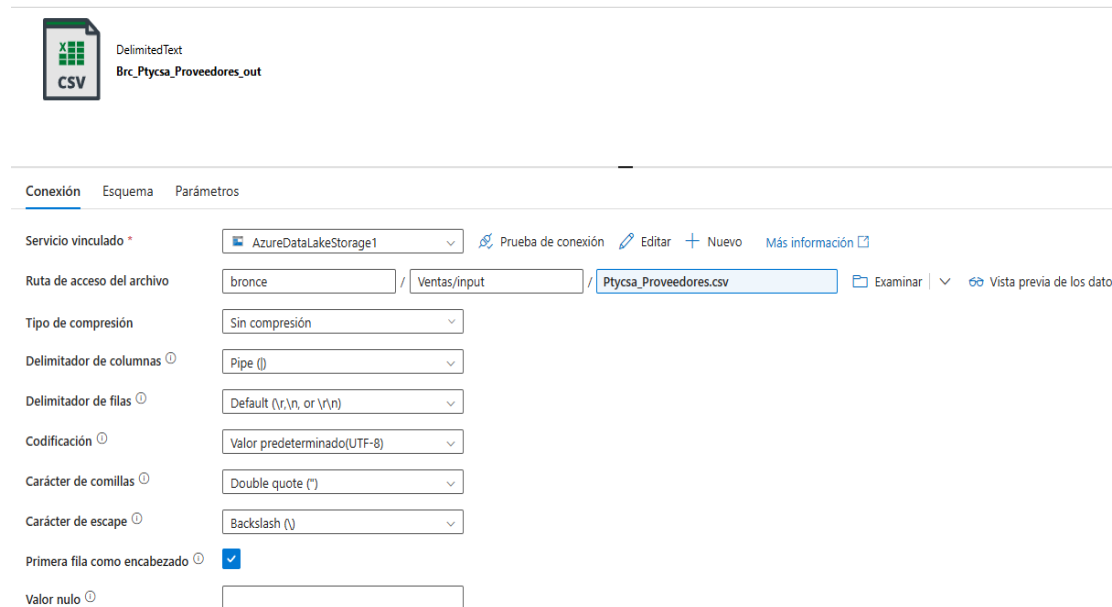
URL base    https://api.businesscentral.dynamics.com...

Dirección URL relativa ⓘ    Company('PHARMA%20TYCSA%20S....'    Vista previa de los datos

*Nota.* Captura muestra la configuración de un conector REST en Azure Data Factory, que establece la conexión con la API de Dynamics 365 Business Central utilizando el servicio vinculado PHARMATYCSA\_BC\_API. Se observa la URL base del endpoint y la dirección relativa configurada para acceder a los datos de la compañía.

## Apéndice J

### Destino



The screenshot shows the configuration interface for a sink in Azure Data Factory. At the top left, there is a file icon labeled "DelimitedText" and "Brc\_Ptycsa\_Proveedores\_out". Below this, the configuration is organized into three tabs: "Conexión", "Esquema", and "Parámetros". The "Conexión" tab is active, showing the following settings:

- Servicio vinculado \***: AzureDataLakeStorage1 (with options for Prueba de conexión, Editar, + Nuevo, and Más información).
- Ruta de acceso del archivo**: bronze / Ventas/input / Ptycsa\_Proveedores.csv (with an Examinar button and a link to Vista previa de los datos).
- Tipo de compresión**: Sin compresión.
- Delimitador de columnas**: Pipe (|).
- Delimitador de filas**: Default (\r, \n, or \r\n).
- Codificación**: Valor predeterminado(UTF-8).
- Carácter de comillas**: Double quote (").
- Carácter de escape**: Backslash (\).
- Primera fila como encabezado**: .
- Valor nulo**: (empty text box).

*Nota.* Interfaz de configuración de un sink en Azure Data Factory, donde los datos extraídos y transformados se escriben como archivo CSV en el contenedor bronze/Ventas/input/ de Azure Data Lake Storage, con delimitador de columnas tipo pipe (|) y codificación UTF-8.

## Apéndice K

### Almacenamiento

Inicio > prodstorageeastusiga

prodstorageeastusiga | Explorador de almacenamiento

Contenedores de blobs > bronze

Método de autenticación: Clave de acceso (Cambiar a la cuenta de usuario de Microsoft Entra)

Buscar blobs por prefijo (distingue mayúsculas de minúsculas)

Mostrando todos los elementos 9

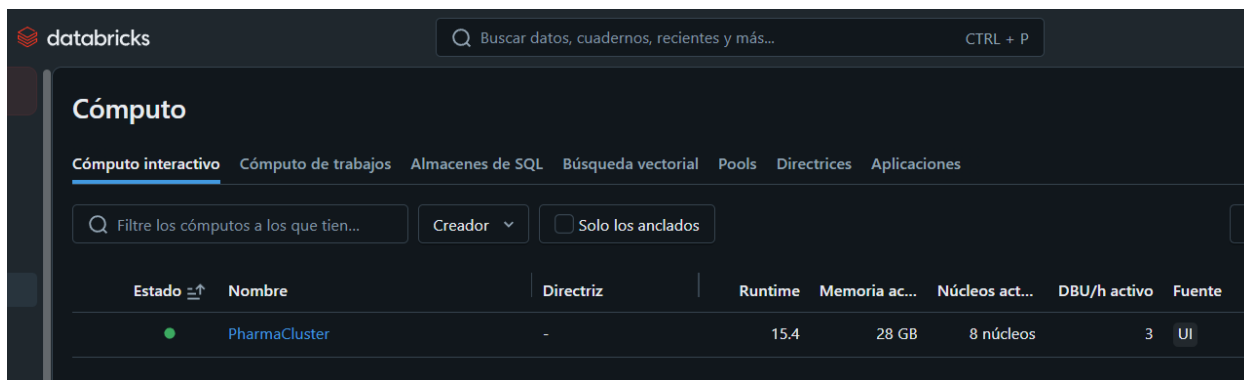
Nombre	Última modificación	Nivel de acceso	Tipo de blob	Tamaño
ARCHIVOS_EXCEL	10/3/2025, 1:34:41 p. m.			
Almacen	25/9/2024, 6:31:54 p. m.			
Compras	10/9/2024, 3:42:55 p. m.			
ERP_ANTERIOR	20/2/2025, 4:34:04 p. m.			
Finanzas	10/9/2024, 3:44:19 p. m.			
Inventarios	10/9/2024, 3:46:09 p. m.			
Licitaciones	10/9/2024, 3:47:53 p. m.			
Ventas	10/9/2024, 3:49:16 p. m.			
Ptycsa_InfoPedidoCompra.csv	9/9/2024, 4:16:13 p. m.	Esporádico (inferido)	Blob en bloques	663 I

*Nota.* Imagen muestra el explorador del contenedor bronze dentro de Azure Data Lake Storage, que muestra carpetas organizadas por dominio funcional (como Ventas, Finanzas, Compras, etc.) y archivos almacenados como Ptycsa\_InfoPedidoCompra.csv, resultado de los procesos de ingesta definidos en los pipelines de Azure Data Factory.



## Apéndice M

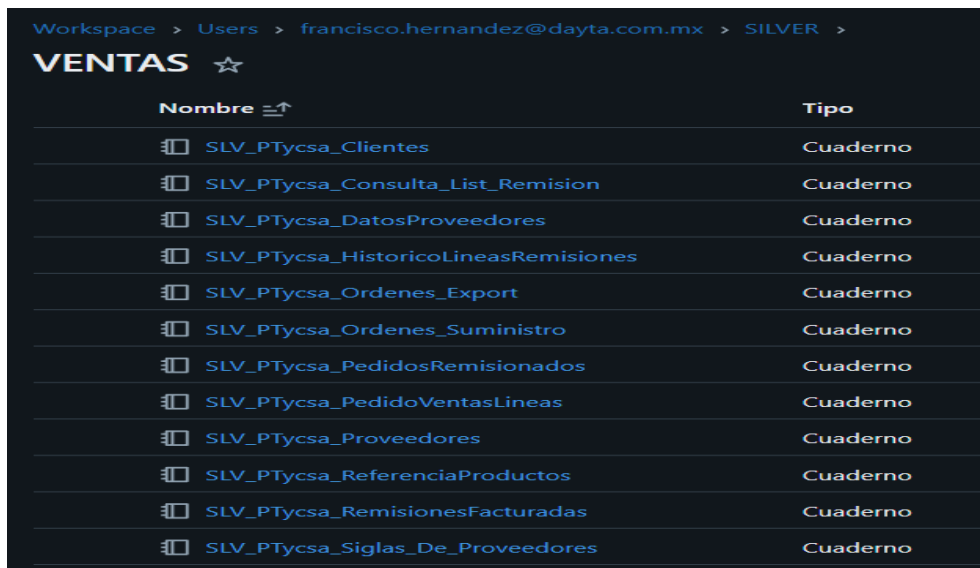
### *Configuración del Entorno Databricks, Configuración de Clústeres en Azure Databricks*



*Nota.* Interfaz del módulo de cómputo interactivo en Azure Databricks, donde se muestra el clúster PharmaCluster configurado con 8 núcleos, 28 GB de memoria, y versión de runtime 15.4. Este clúster es utilizado para ejecutar notebooks de transformación de datos en las capas Silver y Gold.

## Apéndice N

### *Notebooks Desarrollados en Scala para Capa Silver*



The screenshot shows the Azure Databricks interface. The breadcrumb navigation is 'Workspace > Users > francisco.hernandez@dayta.com.mx > SILVER > VENTAS'. Below the breadcrumb, there is a star icon and the word 'VENTAS'. A table lists the notebooks with columns 'Nombre' and 'Tipo'. Each notebook name starts with 'SLV\_PTydsa\_' followed by a specific entity name. All notebooks are of type 'Cuaderno'.

Nombre	Tipo
SLV_PTydsa_Clientes	Cuaderno
SLV_PTydsa_Consulta_List_Remision	Cuaderno
SLV_PTydsa_DatosProveedores	Cuaderno
SLV_PTydsa_HistoricoLineasRemisiones	Cuaderno
SLV_PTydsa_Ordenes_Export	Cuaderno
SLV_PTydsa_Ordenes_Suministro	Cuaderno
SLV_PTydsa_PedidosRemisionados	Cuaderno
SLV_PTydsa_PedidoVentasLineas	Cuaderno
SLV_PTydsa_Proveedores	Cuaderno
SLV_PTydsa_ReferenciaProductos	Cuaderno
SLV_PTydsa_RemisionesFacturadas	Cuaderno
SLV_PTydsa_Siglas_De_Proveedores	Cuaderno

*Nota.* Listado de cuadernos (notebooks) en el entorno de trabajo de Azure Databricks, específicamente dentro del directorio SILVER > VENTAS. Cada notebook realiza transformaciones específicas sobre entidades como clientes, órdenes, remisiones y proveedores. Estos cuadernos forman parte del procesamiento intermedio antes de consolidar los datos en la capa silver.

## Apéndice O

### *Notebooks Desarrollados en Scala para Capa Gold*



The screenshot shows the Azure Databricks interface. At the top, the breadcrumb navigation reads: 'Workspace > Users > francisco.hernandez@dayta.com.mx > GOLD >'. Below this, the directory name 'VENTAS' is displayed with a star icon. A table lists the notebooks in this directory, with columns for 'Nombre' and 'Tipo'.

Nombre 	Tipo
 GLD_PTyca_Concentrado_Consulta_List_Remision	Cuaderno
 GLD_PTyca_HistoricoLineasRemisiones	Cuaderno
 GLD_PTyca_PedidosRemisionados_Clientes	Cuaderno
 GLD_PTyca_Pedidosremisionados_Contratos_Proveed...	Cuaderno
 GLD_PTyca_PedidosRemisionados_OrdenesSuministro	Cuaderno
 GLD_PTyca_Pedidosremisionados_Refproduc_Proveed...	Cuaderno
 GLD_Ventas_Consulta_List_Remisions	Cuaderno
 GLD_Ventas_GL	Cuaderno
 GLD_Ventas_PedidosRemisionados	Cuaderno

*Nota.* Listado de cuadernos (notebooks) en el entorno de trabajo de Azure Databricks, específicamente dentro del directorio SILVER > VENTAS. Cada notebook realiza transformaciones específicas sobre entidades como clientes, órdenes, remisiones y proveedores. Estos cuadernos forman parte del procesamiento intermedio antes de consolidar los datos en la capa Gold.

## Apéndice P

### Código Generado en Lenguaje Scala para Capa Silver

```

1 // Import necessary libraries
2 import java.sql.{Connection, DriverManager, ResultSet}
3 import java.time.LocalDate
4 import java.util.Properties
5 import org.apache.spark.sql.{DataFrame, SparkSession}
6 import spark.implicits._
7 var path_to_bronze = s"abfss://bronze@prodstorageeastusigsa.dfs.core.windows.net/Ventas/input/Ptycsa_Proveedores.csv"
8 println(s"path_file_name ${path_to_bronze}")
9 // Read the CSV file into DataFrame
10 val df = spark.read
11   .format("csv")
12   .option("delimiter", "|")
13   .option("header", "true")
14   .option("inferSchema", "true")
15   .load(path_to_bronze)
16 //normalizar datos
17 import java.text.Normalizer
18 def cleanString(input: String): String = {
19   val normalized = Normalizer.normalize(input, Normalizer.Form.NFD)
20   val cleaned = normalized.replaceAll("[^\\p{Alnum}]", "")
21   cleaned
22 }
23 // guardar datos
24 var stage = "silver"
25 var source = "ventas"
26 var table = "ptycsa_proveedores"
27 var schema = "default"
28 var silverPath = s"${stage}.${schema}.${source}.${table}"
29 df4.write.option("overwriteSchema", "true").mode("overwrite").format("delta").saveAsTable(silverPath)
30 // cifras control
31 val count = df4.count()
32 val fechaActual = java.sql.Timestamp.valueOf(LocalDate.now())
33 val tabla = "silver.ventas_ptycsa_proveedores"
34 val observacion = "Carga exitosa desde pipeline"
35 val logDf = Seq((fechaActual, tabla, count, observacion))
36   .toDF("fecha_registro", "tabla_origen", "cantidad_registros", "observacion")
37 logDf.write
38   .mode("append")

```

*Nota.* Fragmento de código Scala utilizado en un notebook de Azure Databricks que lee un archivo CSV desde la capa Bronze, aplica limpieza y normalización de texto mediante expresiones regulares, y guarda los resultados en formato Delta en la capa Silver, también registra información de control como fecha, tabla origen y cantidad de registros procesados.

## Apéndice Q

### *Código Generado en Lenguaje Scala para Capa Gold*

```
// Import necessary libraries
import java.sql.{Connection, DriverManager, ResultSet}
import java.time.LocalDate
import java.util.Properties
import org.apache.spark.sql.{DataFrame, SparkSession}
import spark.implicits._
val sabanaQuery = ""
with
pedrem as (
  select
    contract_no_1 as contract_no_pedrem ,
    item_reference_no_1 as item_reference_no_pedrem ,
    selltocustomerno as selltocustomerno_pedrem ,
    licitation_no_apm_2 as licitation_no_apm_2_pedrem ,
    pedrem.posting_date as posting_date_pedrem ,
    pedrem.no,
    pedrem.selltocustomername,
    pedrem.status,
    pedrem.document_date,
    pedrem.orderdate,
    pedrem.duedate,
    pedrem.status_pharmatycsa,
    pedrem.requesteddeliverydate,
    pedrem.description,
    pedrem.quantity,
    pedrem.unit_price,
    pedrem.amount,
    pedrem.amount_including_vat
    from silver.default.ventas_ptycsa_pedidosremisionados as pedrem
),
contratos as (
  select
    cn.contract_no_1 as contrato_no_contratos, -- unir pedrem
    cn.billtocustomerno as billtocustomerno_contratos , -- unir pedrem
    cn.item_reference_no_1 as item_reference_no_contratos , -- unir pedrem
    cn.licitation_no_apm_2 as licitacion_contratos, -- unir pedrem
)
var stage = "gold"
var source = "ventas"
var table = "ptycsa_pedidosremisionados_contratos_proveedores"
var schema = "default"
var pedidosRemisionadosRefProdProvGoldPath = s"${stage}.${schema}.${source}.${table}"
sabanaDF.write.option("overwriteSchema", "true").mode("overwrite").format("delta").saveAsTable(pedidosRemisionadosRefProdProvGoldPath)
```

*Nota.* Fragmento de código Scala utilizado en un notebook de Azure Databricks que lee datos de la tabla en Silver, aplica limpieza y normalización de texto mediante expresiones regulares, y guarda los resultados en formato Delta en la capa Gold, también registra información de control como fecha, tabla origen y cantidad de registros procesados.

## Apéndice R

### Tablero Desarrollado en Power BI – Avance de Contratos



*Nota.* Figura “Avance de Contratos – Pharmacy CSA”. Tablero en Power BI conectado a capas procesadas del Data Lake vía Azure Databricks. Resume KPI (264 contratos totales, 171 formalizados, 187 M entregado, 294 M facturado, % avance requerido 4,321.52 %, atendido 10.78 %, facturado 16.03 %) y muestra: anillos por estado del pedido, serie temporal de facturación y unidades, mapa con concentración en México, barras comparativas por institución (> y < 100 000 unidades) y dispersión con línea de tendencia. Incluye filtros por contrato, institución y estado; actualización automática desde la capa Gold.

## Apéndice S

### Conexión DirectQuery desde Power BI a Azure Databricks

**Azure Databricks**

Nombre de la base de datos (obligatorio)

Base de datos (opcional)

Detección automática de proxy (opcional)

Consulta nativa (Requiere: Catálogo predeterminado... (opcional)

Modo Conectividad de datos  Importar  DirectQuery

**Aceptar** **Cancelar**

Nota. Captura de pantalla del entorno de Power BI donde se configura la conexión a la base de datos Azure Databricks, utilizando el modo DirectQuery.

## Apéndice T

### Vista Previa de Datos Conectados desde Azure Databricks en Power

Es posible que esta vista previa tenga 27 días. Actualizar

Value.NativeQuery(Databricks.Catalogs("adb-2417435740458809.9.azure.databricks.net", "/sql/1.0/warehouses/1e5189e1983c3472",

order_no_clr	item_reference_no_clr	bill_to_customer_no_clr	1.2 cantidad_atendidas	1.2 importe_atendidas	posting_date	
1	PV3PL-AU-004301	010.000.2617.00	3PL-AU-0064	40	2900	02/05
2	PV3PL-BI-001545	010.000.1051.01	3PL-BI-0001	214	14265.24	02/05
3	PV3PL-AU-004236	010.000.2618.00	3PL-AU-0064	60	5400	26/04
4	PV3PL-AU-004209	010.000.4273.00	3PL-AU-0001	3	915	25/04
5	PV3PL-AU-004221	010.000.5309.01	3PL-AU-0064	5000	65000	25/04
6	PV3PL-AU-004224	010.000.5735.01	3PL-AU-0064	50	6000	25/04
7	PV3PL-AU-004218	010.000.2650.00	3PL-AU-0064	10	0.0001	25/04
8	PV3PL-AU-004229	010.000.5490.00	3PL-AU-0064	120	16800	24/04
9	PV3PL-AU-004192	010.000.3047.00	3PL-AU-0001	3838	86355	23/04
10	PV3PL-AU-004174	010.000.3047.00	3PL-AU-0001	76	1710	23/04
11	PV3PL-AU-004149	010.000.4110.00	3PL-AU-0001	25	823.5	23/04
12	PV3PL-B-001794	010.000.5544.00	3PL-B-0005	2404	587104.88	23/04
13	PV3PL-AU-004137	010.000.4110.00	3PL-AU-0001	10906	359243.64	23/04
14	PV3PL-AU-004183	010.000.1937.00	3PL-AU-0064	360	6480	22/04
15	PV3PL-AU-004178	010.000.1937.00	3PL-AU-0064	3000	54000	22/04
16	PV3PL-RE-000096	010.000.6125.00	3PL-R-0005	13	290225	21/04
17	PV3PL-AU-004055	010.000.5490.00	3PL-AU-0008	2	70	21/04
18	PV3PL-AU-004099	010.000.2650.00	3PL-AU-0001	276	7728	16/04
19	PVD-009149	010.000.3662.00	DIS-0068	85	104550	16/04
20	PV3PL-AU-004085	010.000.2650.00	3PL-AU-0001	322	9016	16/04
21	PV3PL-AU-004072	010.000.2650.00	3PL-AU-0001	127	3556	16/04
22	PV3PL-EX-000211	010.000.4217.01	3PL-EX-0032	105	13650	14/04
23	PV3PL-EX-000205	010.000.4217.01	3PL-EX-0032	57	7410	14/04
24						

Configuración de la consulta

PROPIEDADES

Nombre  
ptycsa\_consulta\_list\_remision

Todas las propiedades

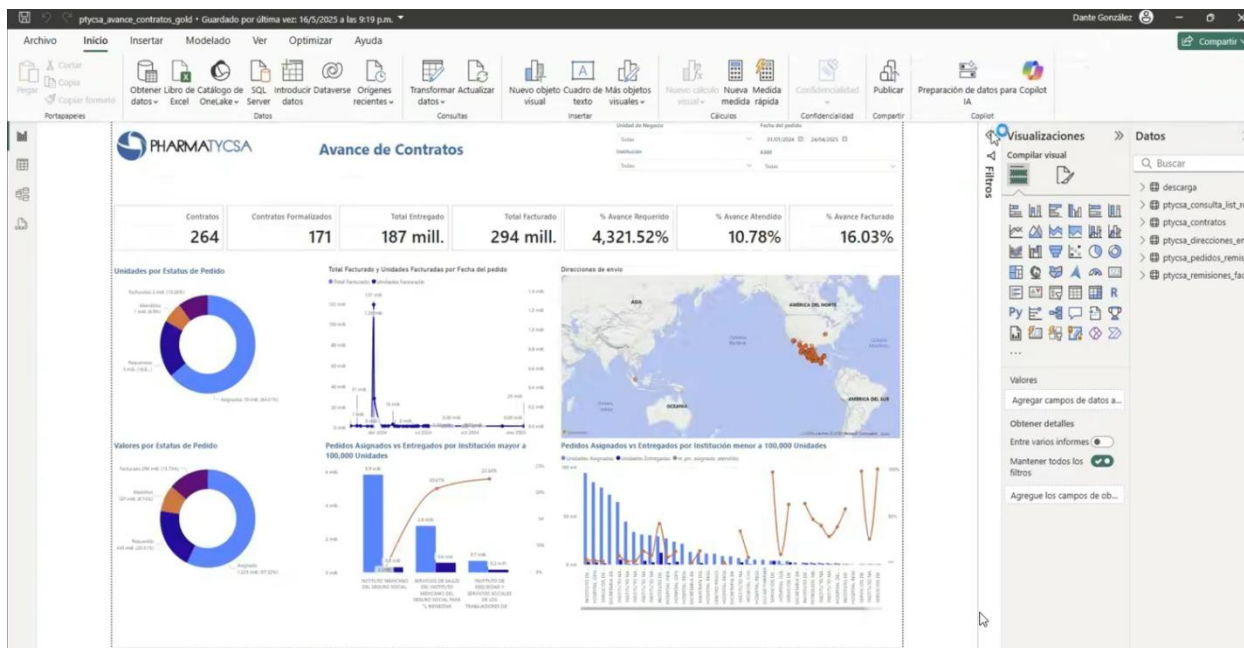
PASOS APLICADOS

Origen

*Nota.* Visualización de la consulta `ptycsa_consulta_list_remision` cargada desde Azure Databricks mediante DirectQuery en Power BI. Se muestran columnas como número de orden, cantidad e importe atendidos, listas para ser usadas en el modelo de datos y los tableros de análisis.

## Apéndice U

### Capturas del Tablero Publicado



*Nota.* Vista final de los dashboards construidos para el análisis.

## Apéndice V

### Métricas y KPIs Implementados

Indicadores Estratégicos (KPIs)	
KPI	Descripción
Total, de Contratos	Número total de contratos cargados desde Dynamics 365.
Contratos Formalizados	Total, de contratos con estatus final de formalización.
Total, Entregado (Millones)	Monto total entregado en unidades monetarias (calculado por suma de entregas).
Total, Facturado (Millones)	Monto total facturado acumulado por cliente o pedido.
% Avance Requerido	Porcentaje calculado sobre el total de lo esperado a entregar.
% Avance Atendido	Porcentaje de avance basado en las unidades entregadas frente a lo requerido.
% Avance Facturado	Porcentaje del valor facturado respecto al total requerido.
Indicadores Operativos	
Indicador	Descripción
Unidades por Estado de Pedido	Distribución de unidades en estados como completado, entregado, pendiente.
Valores por Estado de Pedido	Análisis financiero por estado de los pedidos.
Entregas por Institución (>100k y <100k)	Separación de análisis para instituciones de gran y pequeña escala.
Total, Facturado por Fecha de Pedido	Línea de tendencia temporal para la facturación histórica.
Mapa de Entregas por Región/País	Distribución geográfica de los pedidos cumplidos.
Órdenes por Cliente o Entidad Compradora	Número de pedidos asociados por cliente.

*Nota.* Tabla muestra los indicadores clave de desempeño (KPIs) y métricas desarrolladas en el tablero *Avance de Contratos* de Power BI, alimentado desde Azure Databricks.

## **Apéndice W**

### *Monitoreo y Evaluación*



Anexo%20F1.docx

*Nota.* Informe final contiene el detalle en cuanto al resumen general del proyecto, así como también los resultados, recomendaciones y conclusiones.

## Apéndice X

*Manual de Usuario y Guía Operativa del Entorno Databricks y Power BI*



Anexo%20F2.docx

*Nota.* Documento que describe paso a paso el uso básico de la solución y las instrucciones operativas para usuarios y analistas.