

Richard Guanoluisa
Bases de Datos II
Pontificia Universidad Católica
del Ecuador



Panorama de Bases de Datos

Bases de Datos

Las bases de datos son el núcleo del ecosistema de información. Permiten almacenar, organizar y acceder a datos estructurados y no estructurados, sirviendo como la fuente central para la toma de decisiones y el análisis empresarial.

Roles

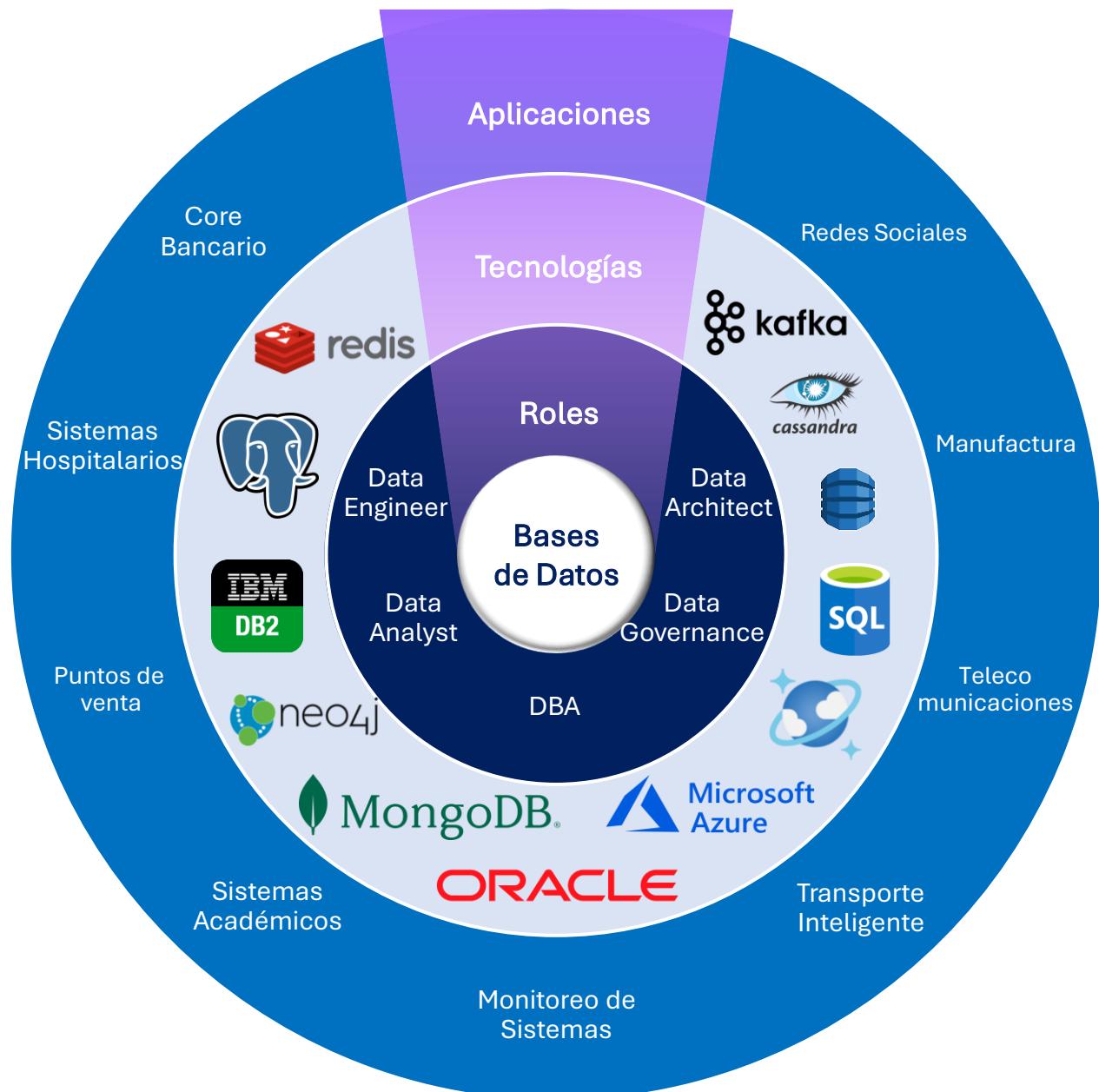
Desde ingenieros que diseñan y gestionan pipelines, hasta analistas y científicos que extraen valor e inteligencia, cada rol contribuye al ciclo completo de gestión y aprovechamiento de la información.

Tecnologías

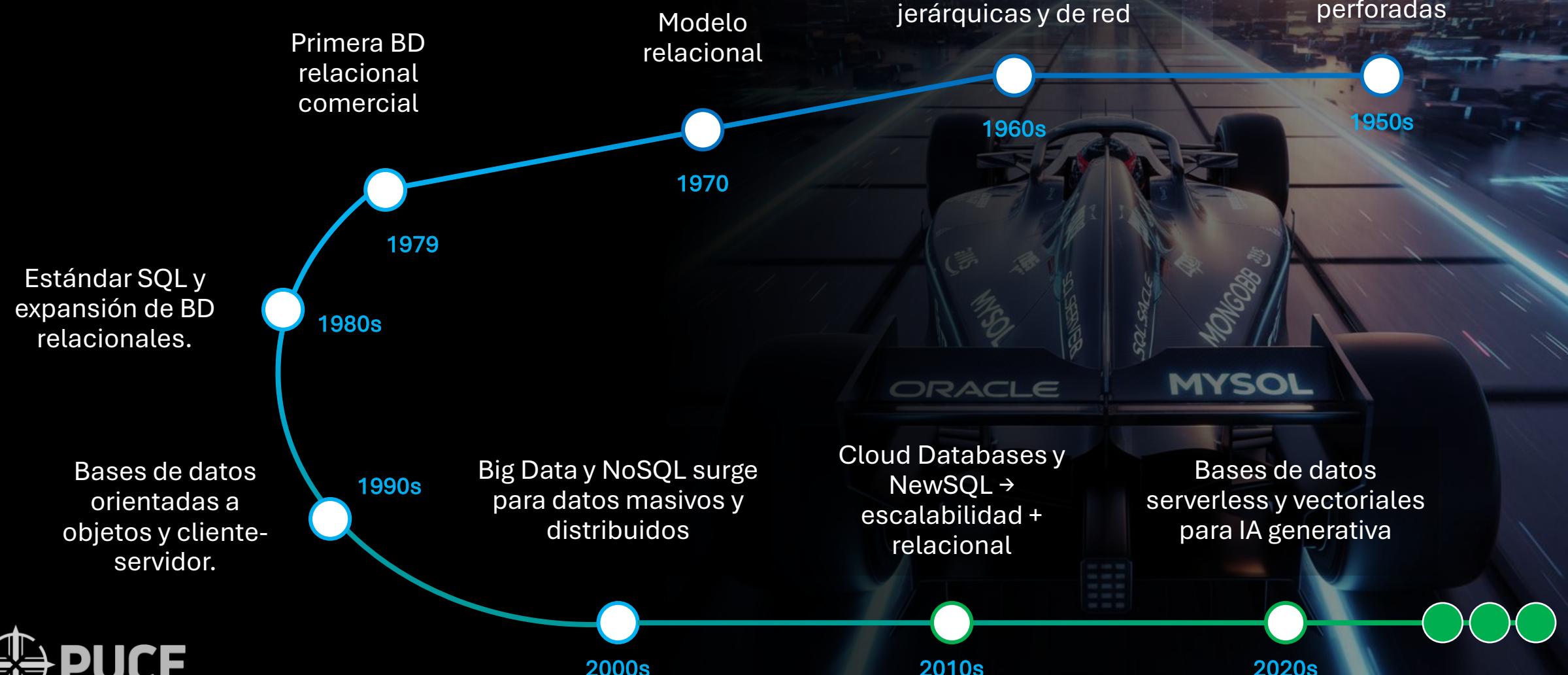
Permiten procesar, transformar y movilizar los datos. Incluyen herramientas y plataformas que facilitan la extracción, integración, análisis y automatización de flujos de información dentro de la organización.

Aplicaciones

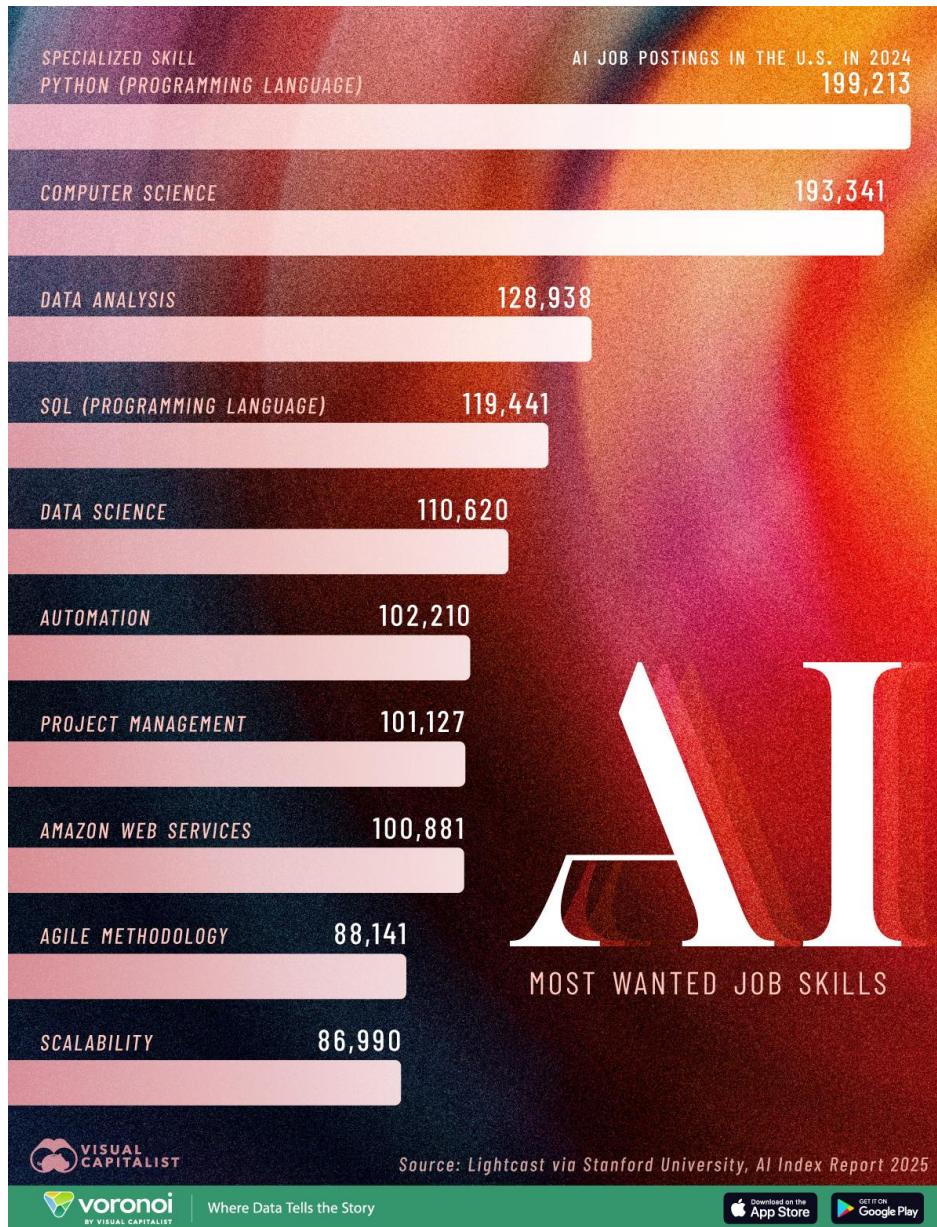
Capa donde los datos se convierten en valor tangible. A través de sistemas y soluciones específicas, impulsan procesos clave en industrias como banca, salud, educación, retail y manufactura, optimizando decisiones y resultados operativos.



Transformación de la Gestión de Datos



¿Dónde puedo aprender más?



Ejemplo 1

Escenario

Comportamiento Transaccional

Las organizaciones necesitan comprender cómo se comportan sus clientes y cómo se realizan las transacciones para tomar decisiones acertadas. Este caso integra información de diferentes fuentes y permite analizar patrones de consumo, identificar tendencias y apoyar decisiones estratégicas basadas en datos.

Objetivo General:

Proveer insights sobre el comportamiento transaccional para apoyar decisiones de negocio.

Objetivos Específicos:

- Identificar patrones de consumo y tendencias en los clientes.
- Visualizar información clave que facilite la toma de decisiones.
- Entender cómo los datos resuelven necesidades reales del negocio.

Orígenes de Datos

- Archivo .csv con Tabla de Transacciones en S3
- Archivo .csv con Tabla de Catálogo de Agencias en S3
- Archivo .csv con Tabla de Catálogo de Código de Transacción en S3
- Archivo Imagen con Tabla de Catálogo de Canales en Azure Blob Storage

Ejemplo 1

Estrategia de Solución

Comportamiento Transaccional

- Archivo .csv con Tabla de Transacciones en **S3**
- Archivo .csv con Tabla de Catálogo de Agencias en **S3**
- Archivo .csv con Tabla de Catálogo de Código de Transacción en **S3**



- Archivo Imagen con Tabla de Catálogo de Canales en **Azure Blob Storage**



Ejemplo 1

Estrategia de Solución

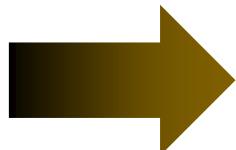
Comportamiento Transaccional

- Archivo .csv con Tabla de Transacciones en **S3**
- Archivo .csv con Tabla de Catálogo de Agencias en **S3**
- Archivo .csv con Tabla de Catálogo de Código de Transacción en **S3**

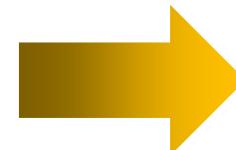


amazon
S3

- Archivo Imagen con Tabla de Catálogo de Canales en **Azure Blob Storage**



Información Procesada y
Estructurada



Información Lista
para Analizar

Ejemplo 1

Estrategia de Solución

Comportamiento Transaccional

Arquitectura Medallón

Bronce

- Archivo .csv con Tabla de Transacciones en **S3**
- Archivo .csv con Tabla de Catálogo de Agencias en **S3**
- Archivo .csv con Tabla de Catálogo de Código de Transacción en **S3**



- Archivo Imagen con Tabla de Catálogo de Canales en **Azure Blob Storage**



Plata

Información Procesada y Estructurada



Oro

Información Lista para Analizar



Ejemplo 1

Definición de Roles

Comportamiento Transaccional

Esquema

- Archivo .csv con Tabla de Transacciones en **S3**
- Archivo .csv con Tabla de Catálogo de Agencias en **S3**
- Archivo .csv con Tabla de Catálogo de Código de Transacción en **S3**



- Archivo Imagen con Tabla de Catálogo de Canales en **Azure Blob Storage**



Se utiliza Microsoft Fabric como entorno cloud para la ingestión, transformación, almacenamiento y consumo analítico de datos.



Información Procesada y Estructurada



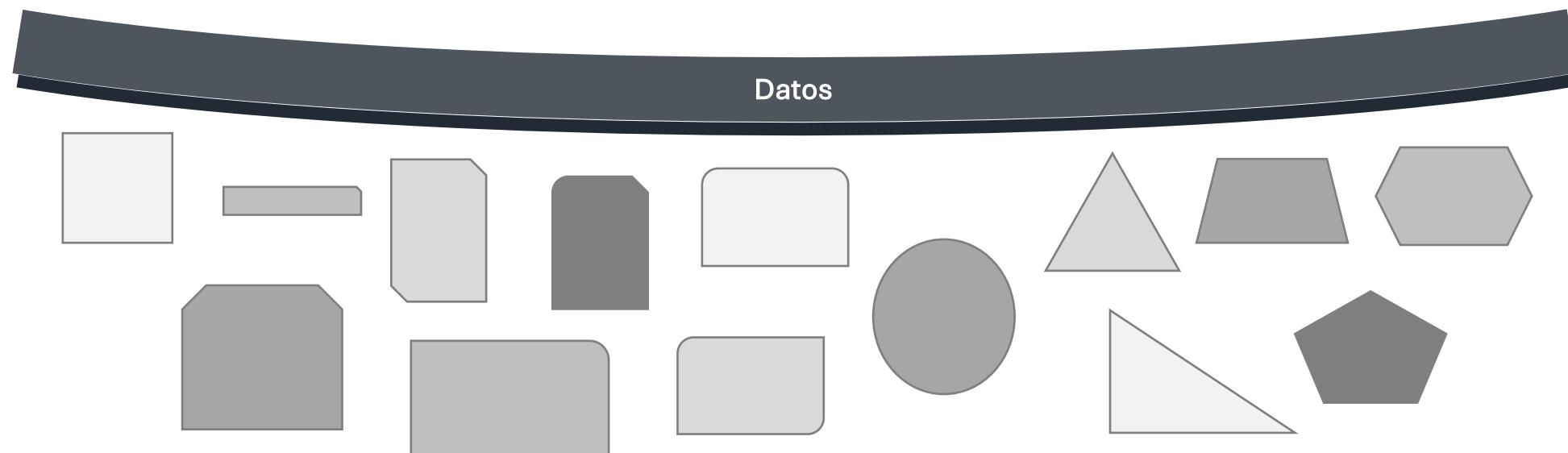
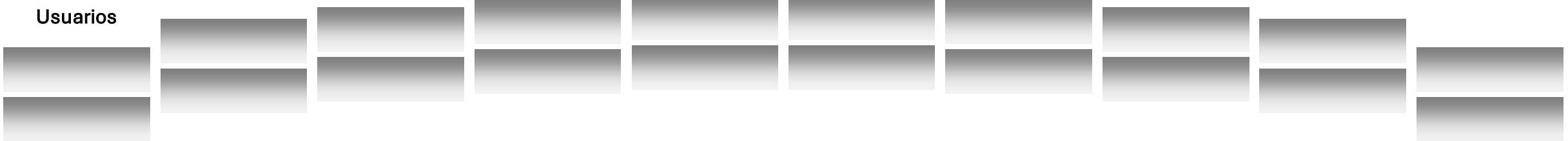
Información Lista para Analizar



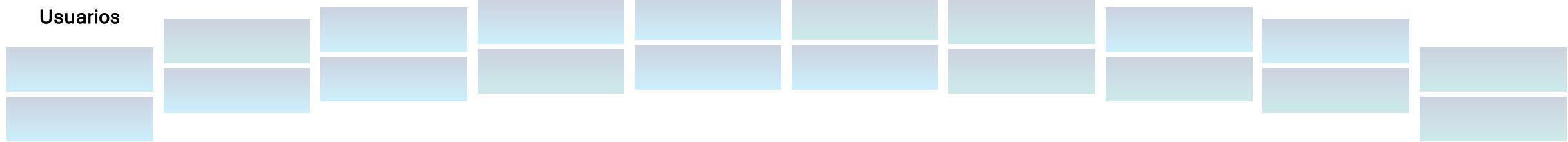
Consumo de Información



Definición de Roles



Definición de Roles



Valor de los Datos

Ingeniero de IA

Ingeniero de Datos

DBA

Analista de Datos

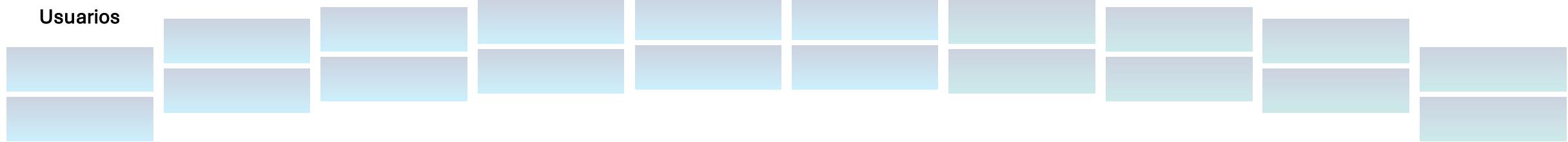
Científico de Datos

Técnicos

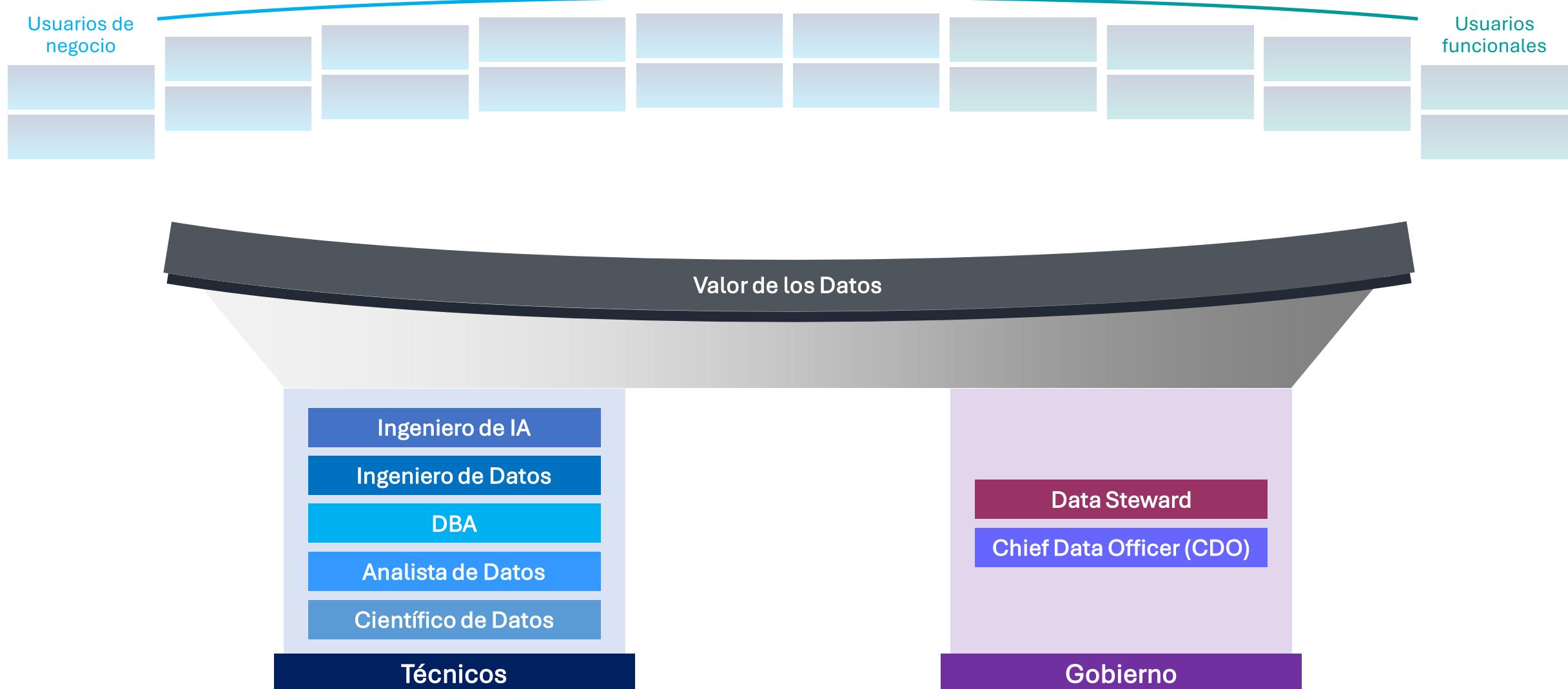
Encargados de construir, administrar y mantener la infraestructura, los procesos y las soluciones tecnológicas que permiten capturar, almacenar, transformar, analizar y aprovechar los datos.



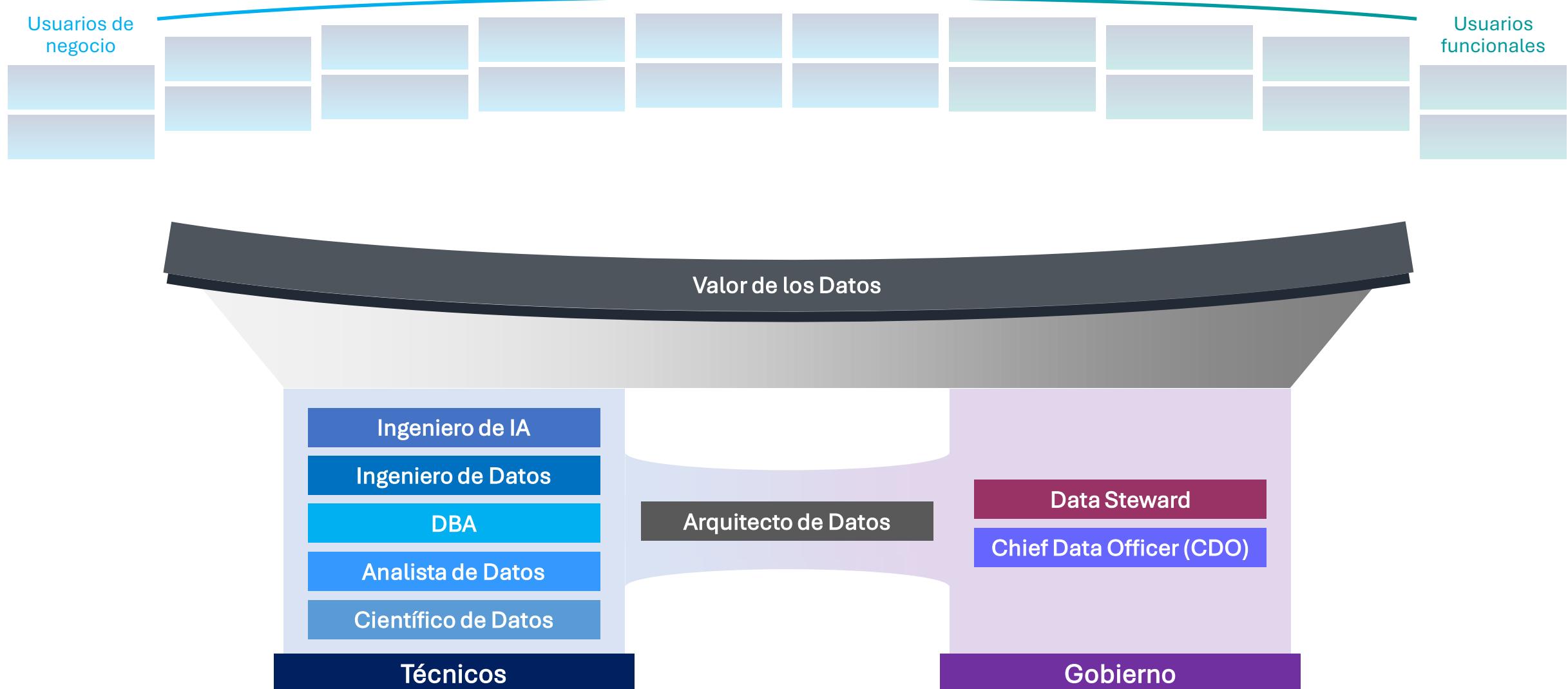
Definición de Roles



Definición de Roles



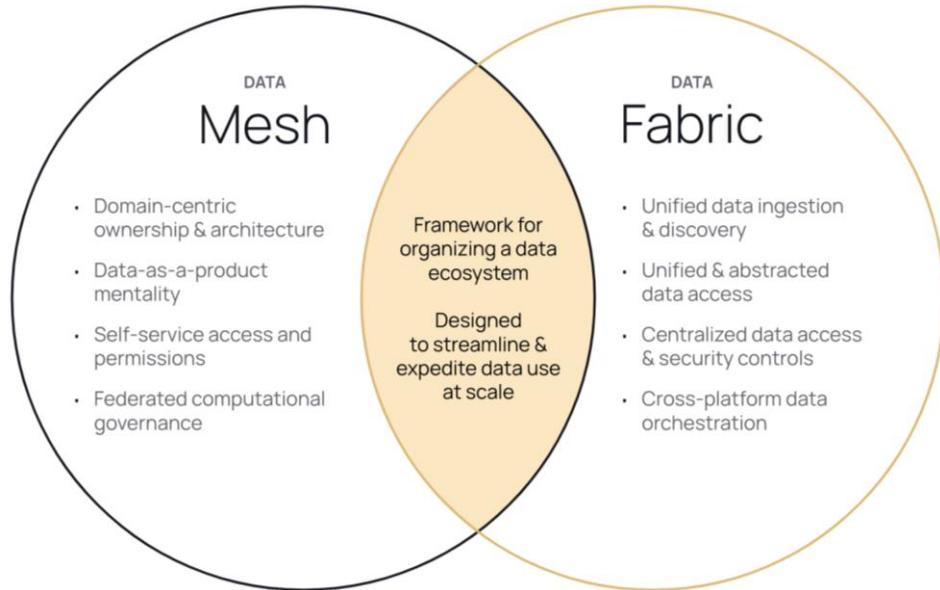
Definición de Roles



Definición de Roles

	Rol	Objetivo	Actividades	Herramientas
Técnicos	Ingeniero de IA	Implementar y desplegar modelos de inteligencia artificial en entornos productivos.	Entrenamiento y optimización de modelos, integración con APIs, automatización de inferencias.	 TensorFlow 
	Ingeniero de Datos	Diseña, construye y mantiene pipelines de datos	Construcción de pipelines ETL/ELT, integración de fuentes, optimización de rendimiento, orquestación de flujos.	  python™
	DBA	Asegurar el correcto funcionamiento, disponibilidad y seguridad de las bases de datos.	Instalación, configuración, respaldo, tuning, monitoreo, control de accesos.	  mongoDB
	Analista de Datos	Transformar datos en información útil para la toma de decisiones.	Limpieza y exploración de datos, creación de dashboards, elaboración de reportes.	  Power BI
	Científico de Datos	Extraer conocimiento y generar modelos predictivos basados en datos.	Análisis estadístico, minería de datos, machine learning, validación de modelos.	  python™
	Arquitecto de Datos	Diseñar la arquitectura integral de datos alineando tecnología y estrategia de gobierno.	Modelado de datos, diseño de flujos, definición de estándares, soporte a ingenieros y CDO.	  Microsoft Azure
Gobierno	Data Steward	Garantizar la calidad, integridad y coherencia de los datos dentro de la organización.	Validación de datos, definición de estándares, monitoreo de calidad, documentación de metadatos.	 Power BI
	Chief Data Officer (CDO)	Liderar la estrategia de datos y fomentar su uso ético y alineado con los objetivos del negocio.	Definir políticas de datos, gobierno, compliance, impulsar cultura del dato.	

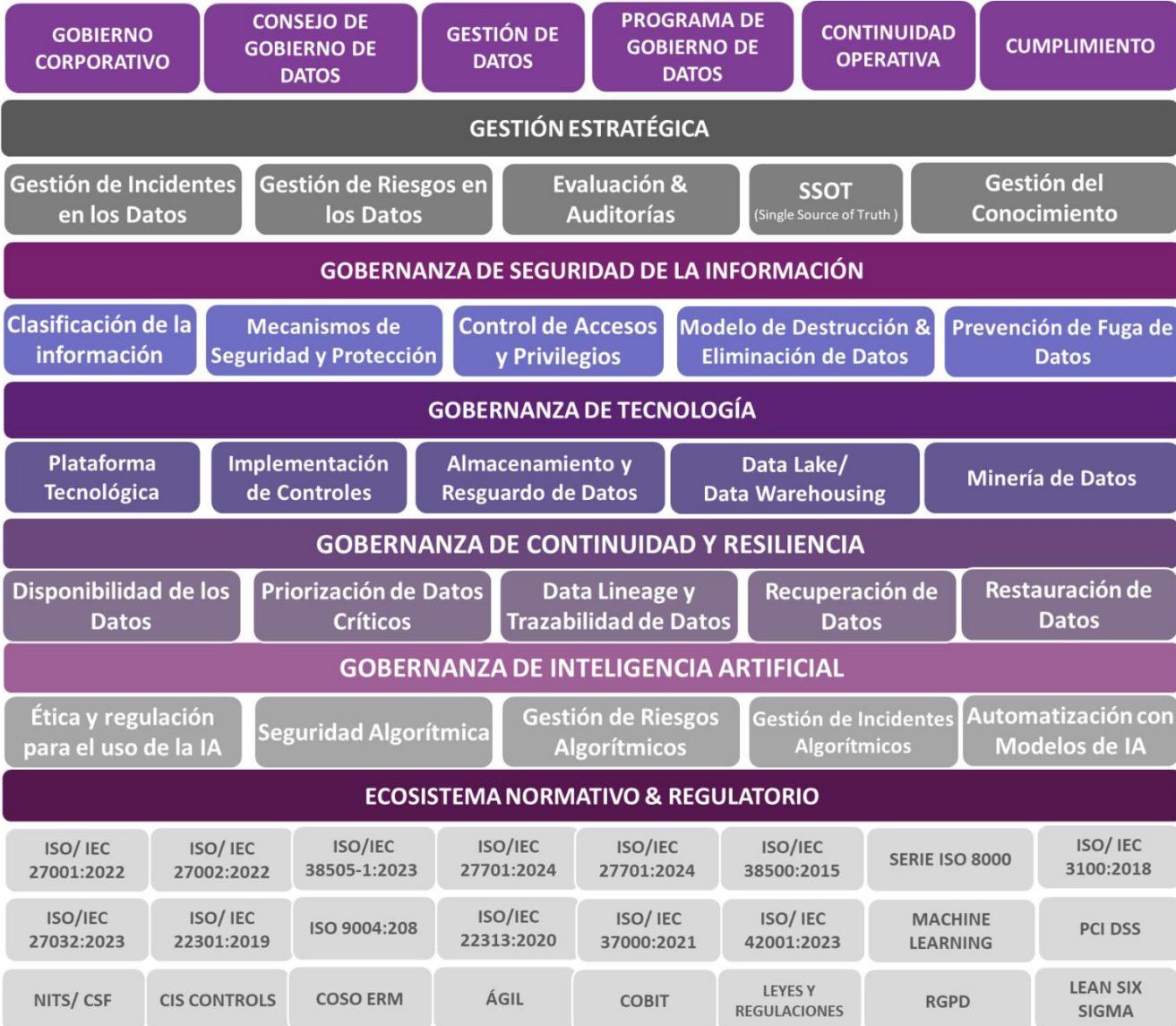
Componentes & Conceptos



El **Data Mesh** distribuye la responsabilidad de los datos entre los equipos de negocio, mientras que el **Data Fabric** centraliza la gestión y el control. Ambos necesitan un Gobierno de Datos sólido: el Mesh lo aplica de forma federada, y el Fabric lo automatiza de forma centralizada.

En conjunto, permiten que los datos sean confiables, accesibles y gobernados a escala.

GOBIERNO DE DATOS



Ejemplo 1

Definición de Roles

Comportamiento Transaccional

Esquema

- Archivo .csv con Tabla de Transacciones en **S3**
- Archivo .csv con Tabla de Catálogo de Agencias en **S3**
- Archivo .csv con Tabla de Catálogo de Código de Transacción en **S3**



- Archivo Imagen con Tabla de Catálogo de Canales en **Azure Blob Storage**



Se utiliza Microsoft Fabric como entorno cloud para la ingestión, transformación, almacenamiento y consumo analítico de datos.



Información Procesada y Estructurada



Información Lista para Analizar



Consumo de Información



Ingeniero de IA

Aplica modelos de inteligencia artificial para extraer valor de los datos

Ingeniero de Datos

Diseña, construye y mantiene pipelines de datos

DBA

Encargado de la gestión, seguridad y optimización de bases de datos.

Analista de Datos

Interpreta datos para generar insights, dashboards y reportes de negocio.

Componentes & Conceptos



Amazon S3 es un servicio de almacenamiento de objetos ofrecido por Amazon Web Services (AWS). Está diseñado para almacenar y recuperar cualquier tipo de dato (archivos, imágenes, logs, backups, datasets, etc.) de forma segura, escalable y altamente disponible a través de Internet.

S3 no almacena archivos en carpetas físicas, sino como objetos dentro de buckets, lo que le da gran flexibilidad y escalabilidad.

Componentes & Conceptos



Google Colab es un entorno de ejecución en la nube basado en Jupyter Notebook, que permite escribir, ejecutar y compartir código Python directamente desde el navegador, sin necesidad de instalar software localmente.

<https://colab.google/>



Componentes & Conceptos



Apache Spark es un motor de procesamiento distribuido de datos en memoria, diseñado para ejecutar tareas de análisis, transformación y machine learning sobre grandes volúmenes de información de forma rápida y escalable.

Funciona bajo el modelo cluster computing, dividiendo las operaciones en múltiples nodos (o procesos paralelos), lo que le permite procesar desde gigabytes hasta terabytes de datos mucho más eficientemente que frameworks tradicionales como Hadoop MapReduce.

Spark SQL

Spark Streaming / Structured Streaming

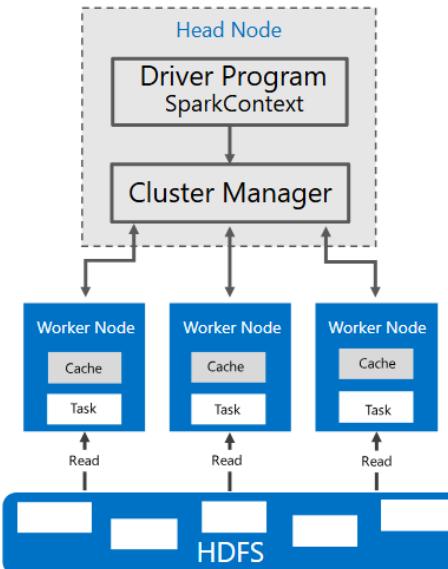
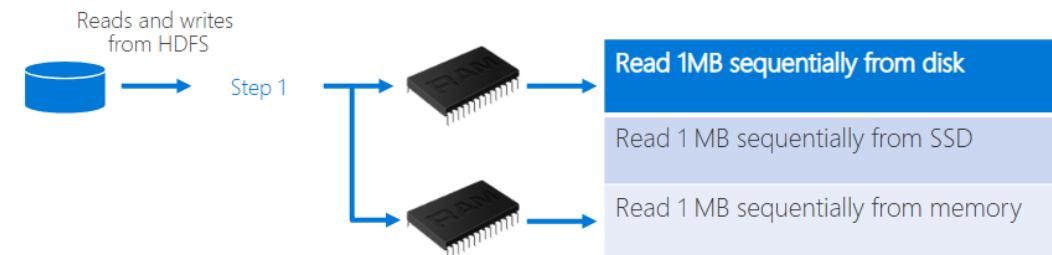
GraphX

SparkR / PySpark

In traditional
MapReduce



In Spark



Componentes & Conceptos



[Guía de inicio rápido de la API de Gemini](#) | [Gemini API](#) | [Google AI for Developers](#)

Conoce los modelos

[Comienza a crear con Gemini](#)

2.5 Pro ✨

Nuestro modelo de razonamiento más potente, con funciones para el razonamiento complejo y mucho más

2.5 Flash ✨

Nuestro modelo más equilibrado, con una ventana de contexto de 1 millón de tokens y mucho más

2.5 Flash-Lite ✨

Nuestro modelo multimodal más rápido y rentable, con un gran rendimiento para tareas de alta frecuencia

Veo 3 🎥

Nuestro modelo de generación de videos de vanguardia, con audio nativo

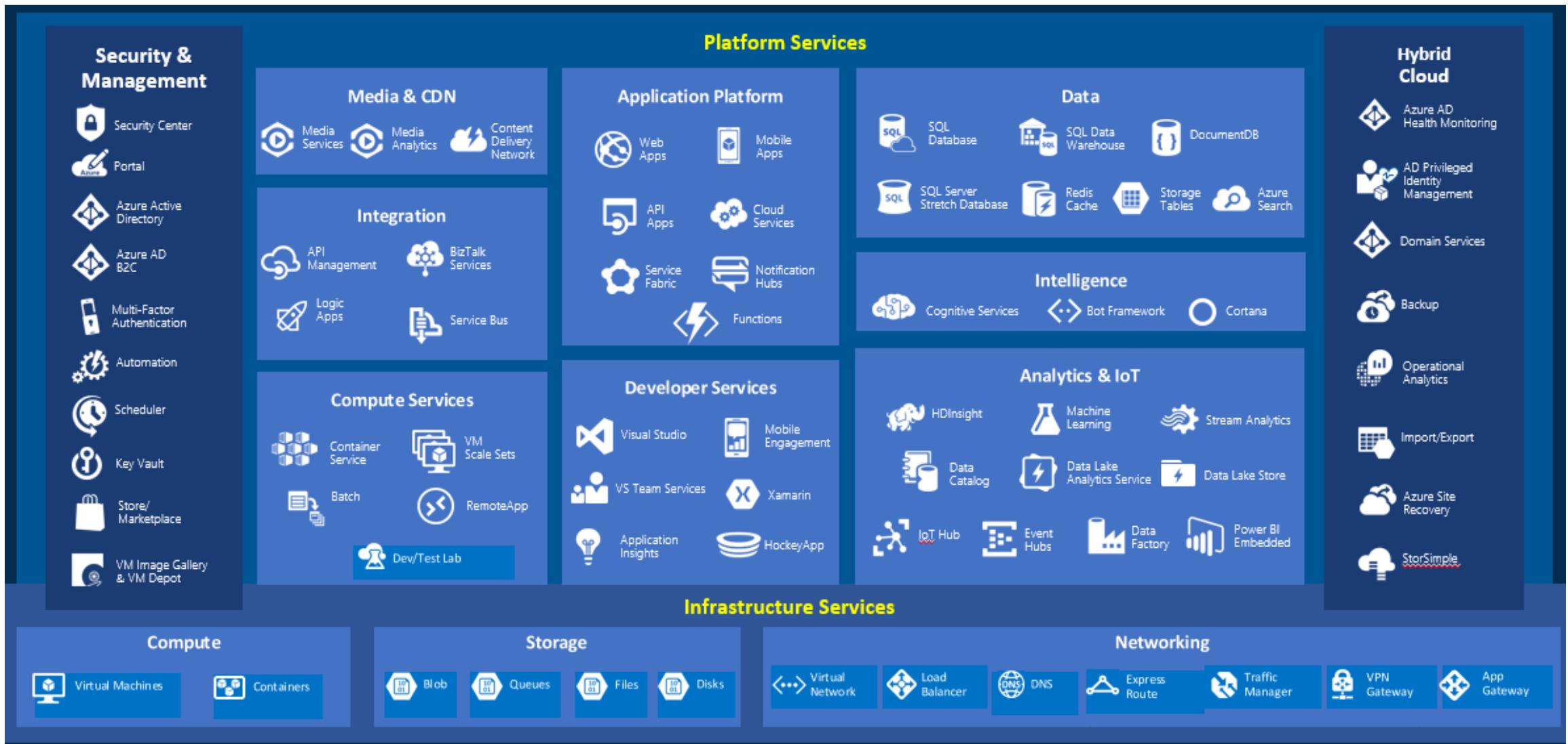
Gemini 2.5 Flash Image 🖼

(Nano Banana), nuestro modelo de generación de imágenes altamente eficaz y preciso

Gemini Embeddings []

Nuestro primer modelo de embedding de Gemini, diseñado para flujos de trabajo de RAG de producción

Componentes & Conceptos



Componentes & Conceptos



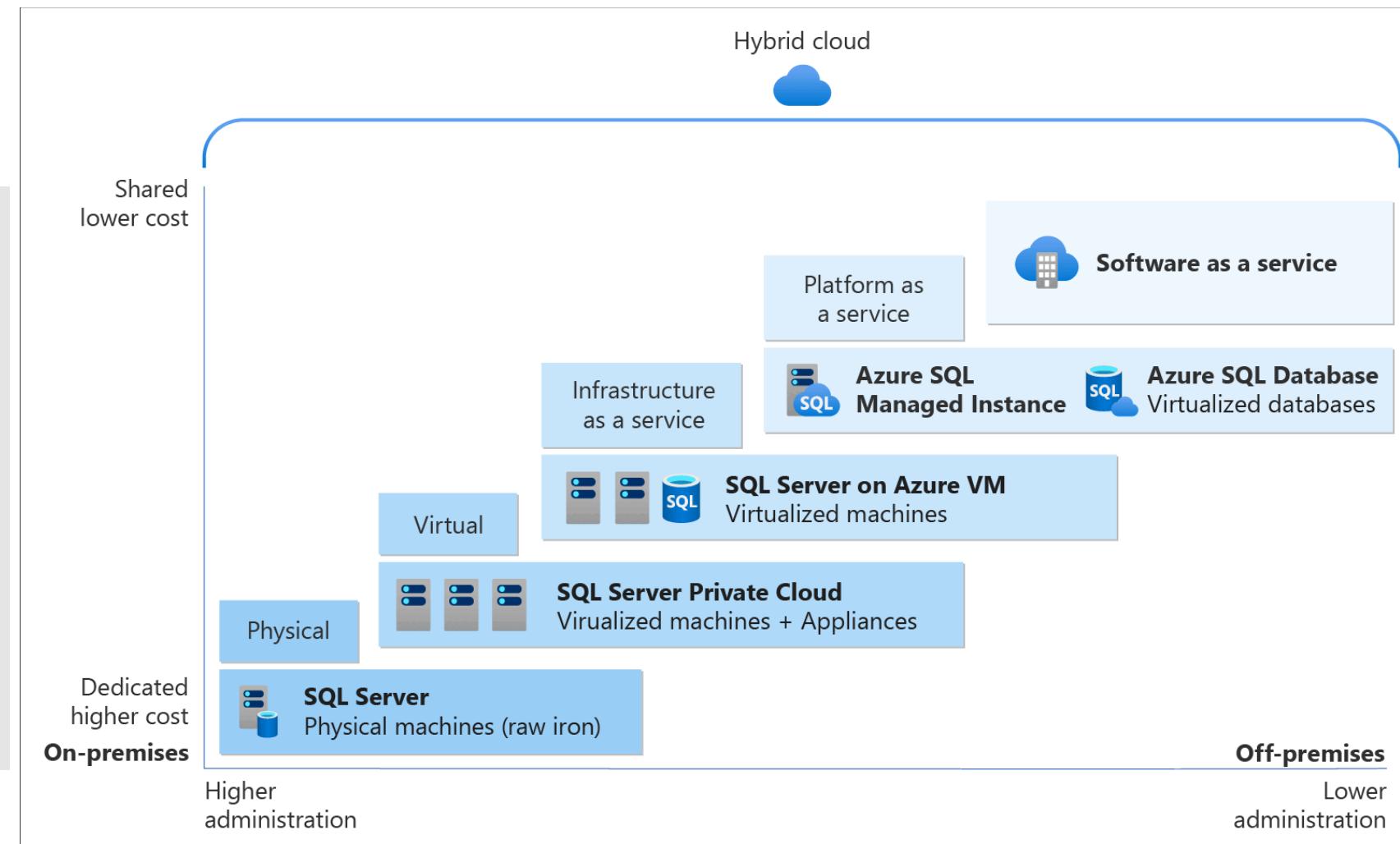
- Azure, la plataforma de informática en la nube de Microsoft, ofrece más de 200 productos y servicios en una red global de centros de datos.
- Con más de 400 centros de datos altamente seguros en más de 70 regiones, Azure ofrece la superficie de nube más amplia, lo que proporciona a las empresas más regiones entre las que elegir que cualquier otro proveedor de nube.
- El 95 % de las empresas de la lista Fortune 500 confían en Azure para ayudarles a dirigir sus negocios y satisfacer sus demandas a escala empresarial.
- ...

Azure Arc Incorpore los servicios y la administración de Azure a cualquier infraestructura	Azure Cosmos DB Base de datos NoSQL rápida con API abiertas para cualquier escala	Azure Kubernetes Service (AKS) Simplifique la implementación, la administración y las operaciones de Kubernetes	Azure Operator Insights Análisis de datos de red de varios orígenes	Azure Quantum (versión preliminar) Experimente el impacto actual de Quantum en Azure
Azure SQL Familia de SQL moderna para la migración y la modernización de aplicaciones	Azure Virtual Desktop La mejor experiencia de escritorio virtual que se ofrece en Azure	Funciones de Azure Procese eventos con código sin servidor	Fundición de IA de Azure Cree aplicaciones listas para el mercado para su organización con inteligencia artificial	Máquinas virtuales Aprovisionamiento de máquinas virtuales para Ubuntu, Red Hat, Windows y mucho más
Microsoft Copilot en Azure Simplificación de las operaciones y la administración de la nube al perímetro con un compañero de inteligencia artificial	Microsoft Entra ID Sincronice los directorios locales y habilite el inicio de sesión único	Servicio de Aplicaciones Cree aplicaciones eficaces en la nube con rapidez para la web y móviles		

Componentes & Conceptos



Base de datos Azure SQL está disponible como una base de datos única con su propio conjunto de recursos administrados a través de un servidor lógico y como una base de datos agrupada en un grupo elástico con un conjunto compartido de recursos administrados a través de un servidor lógico. En general, los grupos elásticos están diseñados para un patrón de aplicación de software como servicio (SaaS) típico, con una base de datos por cliente o inquilino. Con los grupos, administra el rendimiento colectivo y las bases de datos se escalan automáticamente.

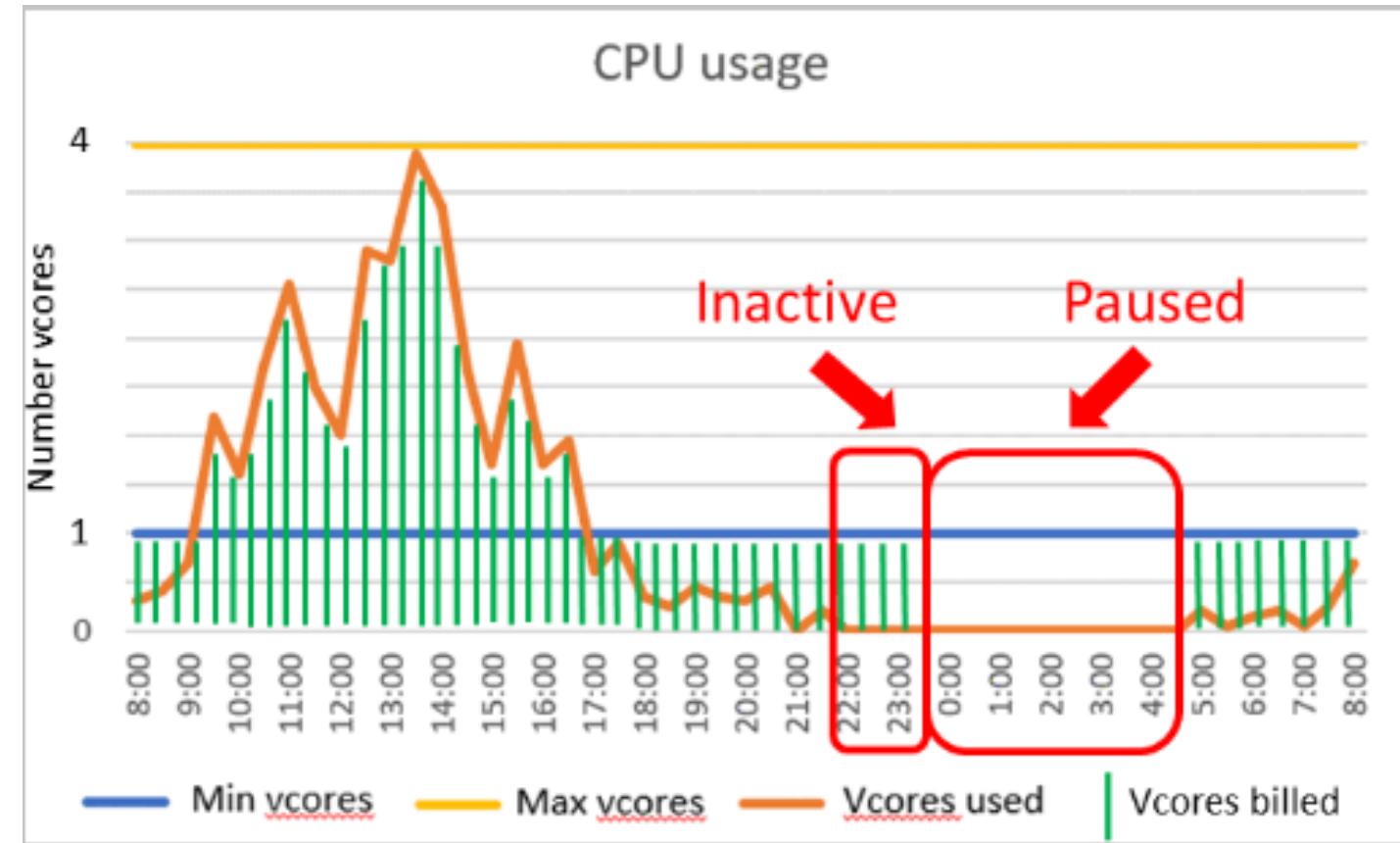


Componentes & Conceptos



Nivel de computación sin servidor para Azure SQL Database

El nivel de proceso sin servidor para las bases de datos únicas de Azure SQL Database escala automáticamente el proceso en función de la demanda de carga de trabajo y se factura según la cantidad de proceso usado por segundo. El nivel de proceso sin servidor también detiene automáticamente las bases de datos durante períodos inactivos cuando solo se factura el almacenamiento y reanuda automáticamente las bases de datos cuando finaliza la actividad. El nivel de proceso sin servidor está disponible en el nivel de servicio De uso general y en el nivel de servicio Hiperescala.



[Nivel de computación sin servidor - Azure SQL Database | Microsoft Learn](#)

Componentes & Conceptos



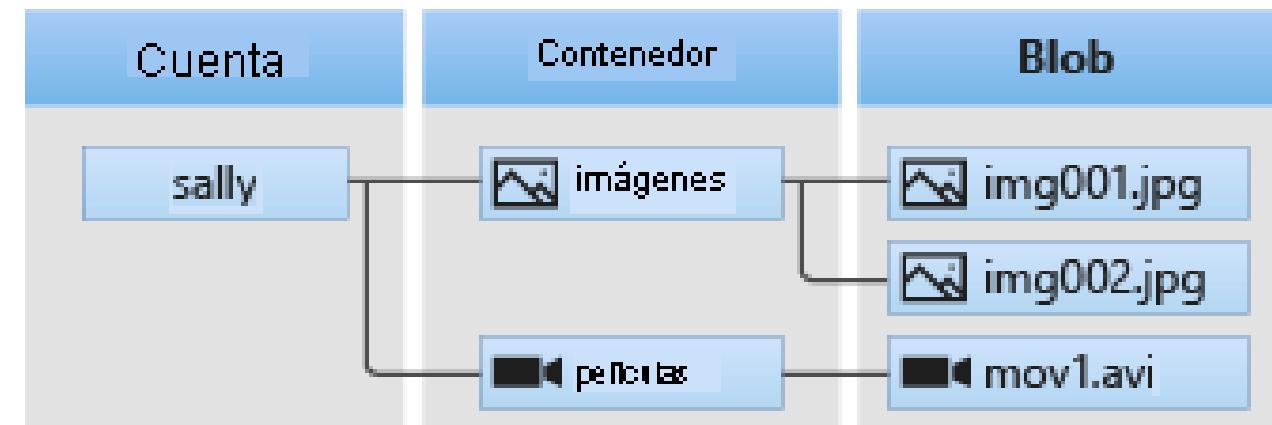
Microsoft Azure
Blob Storage



Azure Blob Storage es la solución de almacenamiento de objetos de Microsoft para la nube. Blob Storage está optimizado para el almacenamiento de cantidades masivas de datos no estructurados. Los datos no estructurados son datos que no se ciñen a ningún un modelo de datos o definición concretos, como texto o datos binarios.

Diseñado para:

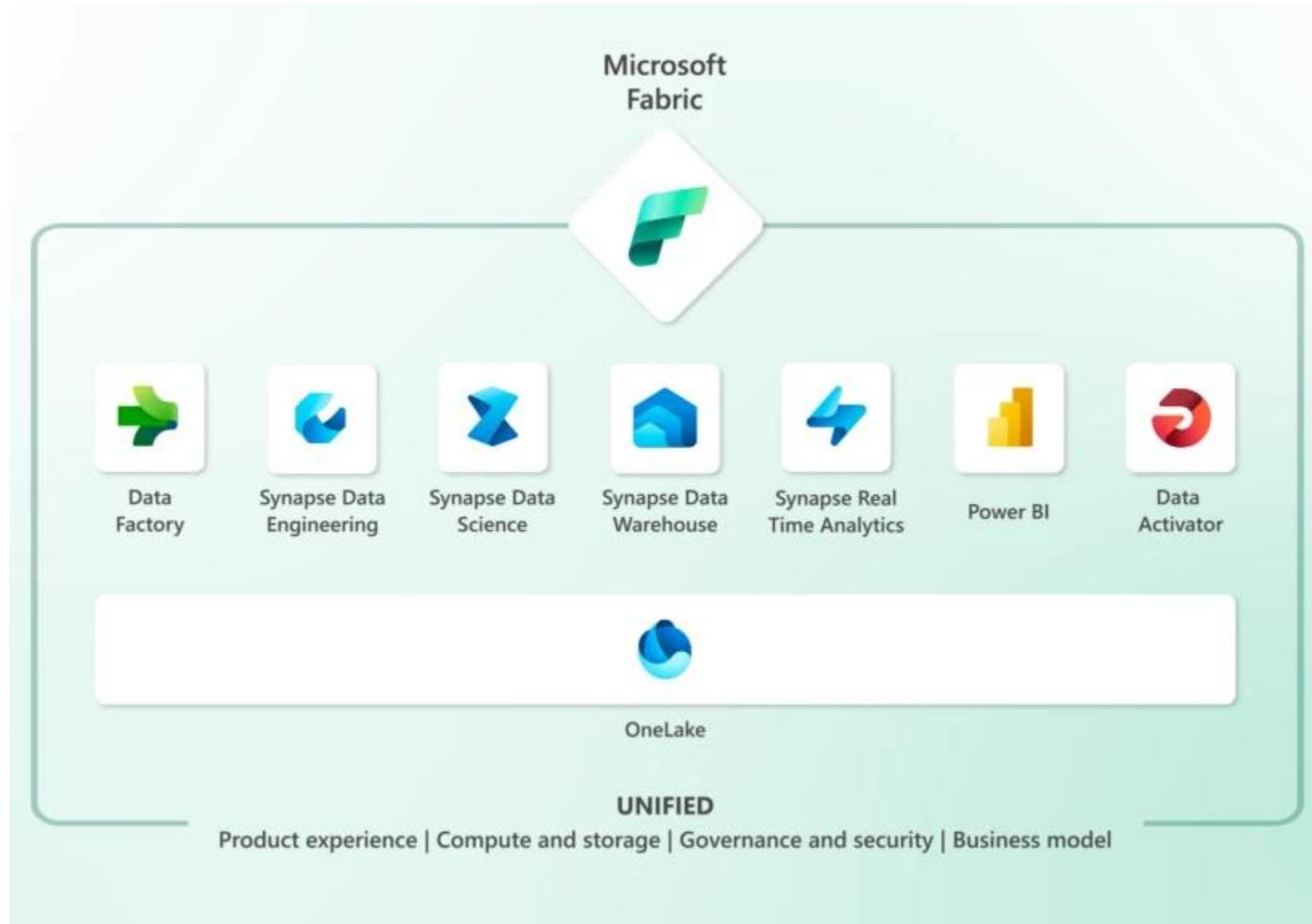
- Visualización de imágenes o documentos directamente en un explorador.
- Almacenamiento de archivos para acceso distribuido.
- Streaming de audio y vídeo.
- Escribir en archivos de registro.
- Almacenamiento de datos para copia de seguridad y restauración, recuperación ante desastres y archivado.
- Almacenamiento de datos para el análisis en local o en un servicio hospedado de Azure.



Componentes & Conceptos



Microsoft Fabric es una solución de análisis de un extremo a otro con capacidades de servicio completo como, por ejemplo, movimiento de datos, lagos de datos, ingeniería de datos, integración de datos, ciencia de datos, análisis en tiempo real, supervisión de datos e inteligencia empresarial, todo ello con el respaldo de una plataforma compartida que proporciona una seguridad, un gobierno y un cumplimiento de datos de primera línea.

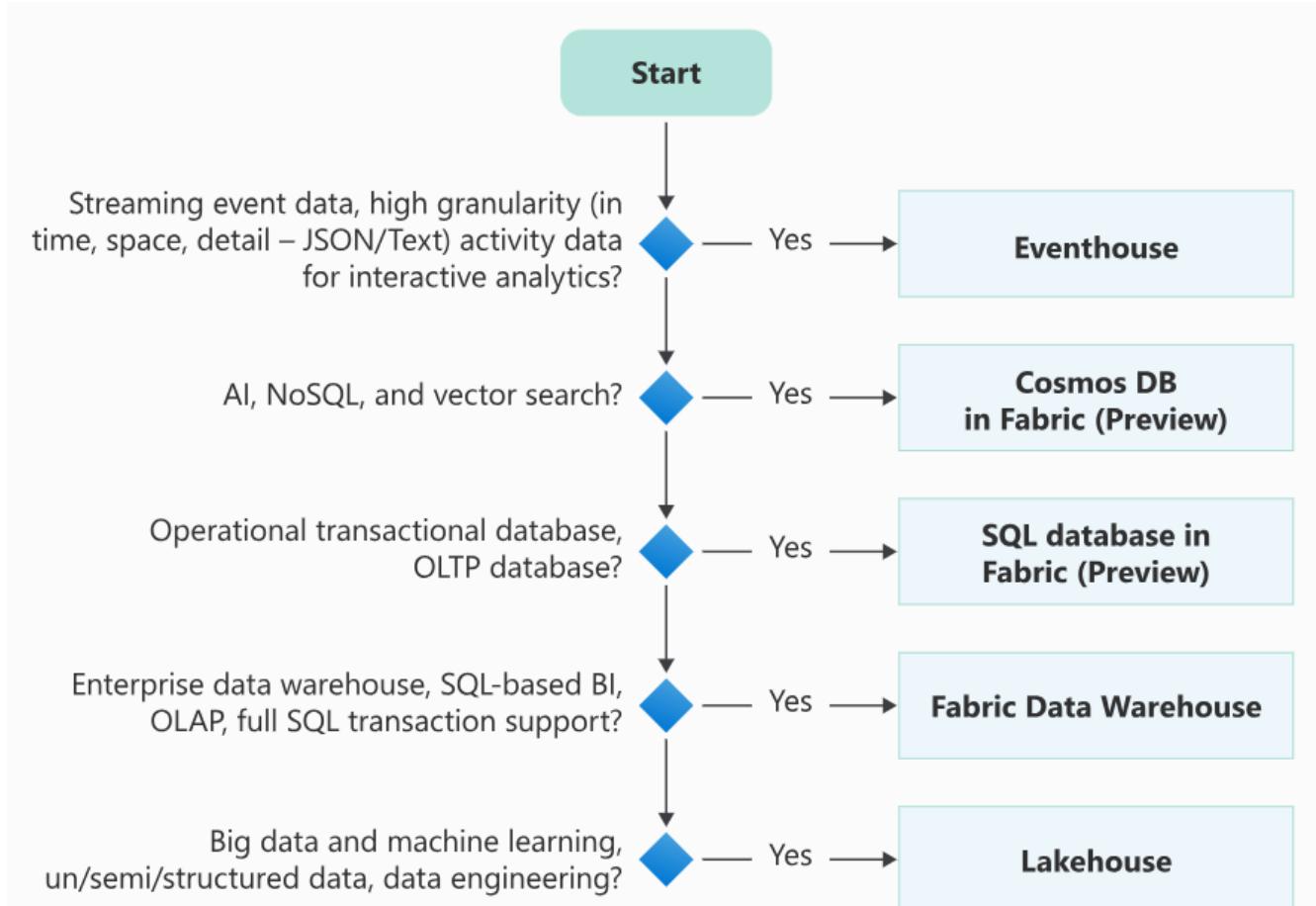


Componentes & Conceptos



Microsoft Fabric

Guía de decisiones de Microsoft Fabric:
selección de un almacén de datos



OneLake

Unified storage foundation
All data available for all workloads

Componentes & Conceptos



Microsoft Fabric Lakehouse es una plataforma de arquitectura de datos para almacenar, administrar y analizar datos estructurados y no estructurados en una sola ubicación. Es una solución flexible y escalable que permite a las organizaciones controlar grandes volúmenes de datos mediante diversas herramientas y marcos para procesar y analizar esos datos. Se integra con otras herramientas de análisis y administración de datos para proporcionar una solución completa para la ingeniería y el análisis de datos. Una instancia de Lakehouse combina la escalabilidad de un lago de datos con el rendimiento y la estructura de un almacenamiento de datos, lo que proporciona una plataforma unificada para el almacenamiento de datos, la administración y el análisis.

A screenshot of the Microsoft Fabric Lakehouse interface. The left sidebar shows navigation options like Home, Create, Browse, Data hub, Monitoring hub, Workspaces, LHRedesign_AvWS, ContosoDailySales, and Power BI. The main area has tabs for Home, Get data, New Power BI dataset, and Open notebook. A message at the top says: "A SQL endpoint for SQL querying and a default dataset for reporting were created and will be updated with any tables added to the lakehouse. You can access the SQL endpoint using the dropdown." Below this is the "Lakehouse explorer" section, which lists "ContosoDailySales" with its "Tables" (Customer, inventory, product, sales, Transactions) and "Unidentified" files (CustomerFeedbackAudio). To the right is a "Customer" dataset preview table with columns: Index, UserId, FirstName, LastName, Sex, Email, Phone, DateOfBirth, and JobTitle. The table shows 11 rows of sample data.

Componentes & Conceptos



Es una plataforma de Business Intelligence (BI) desarrollada por Microsoft que permite conectar, transformar, modelar y visualizar datos de múltiples fuentes para crear informes interactivos y paneles analíticos.

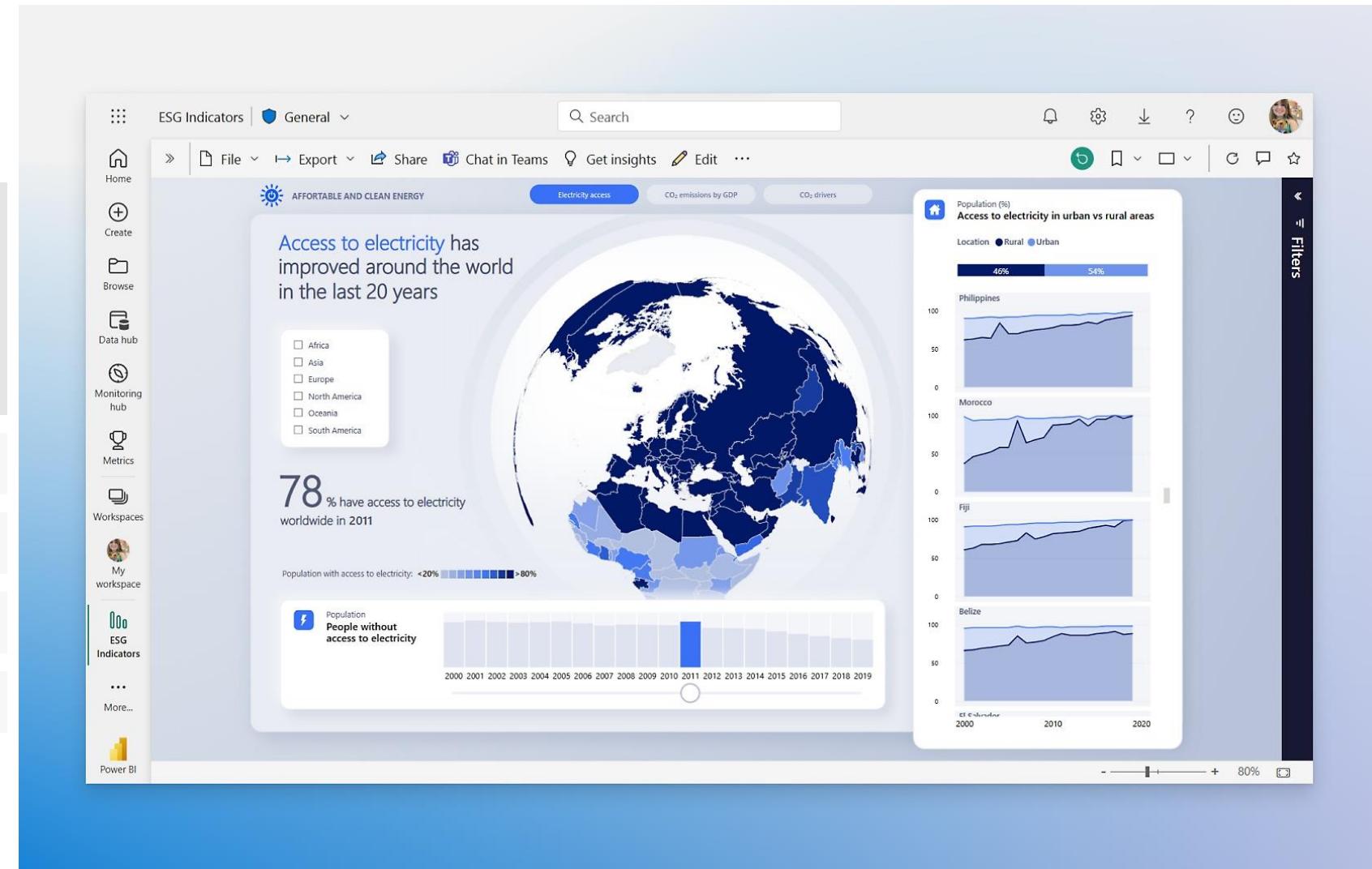
Power BI Desktop

Power BI Service (Cloud)

Power BI Data Gateway

Power BI Mobile

https://www.microsoft.com/es-es/power-platform/products/power-bi#tabs-pill-bar-ocb9d418_tab0

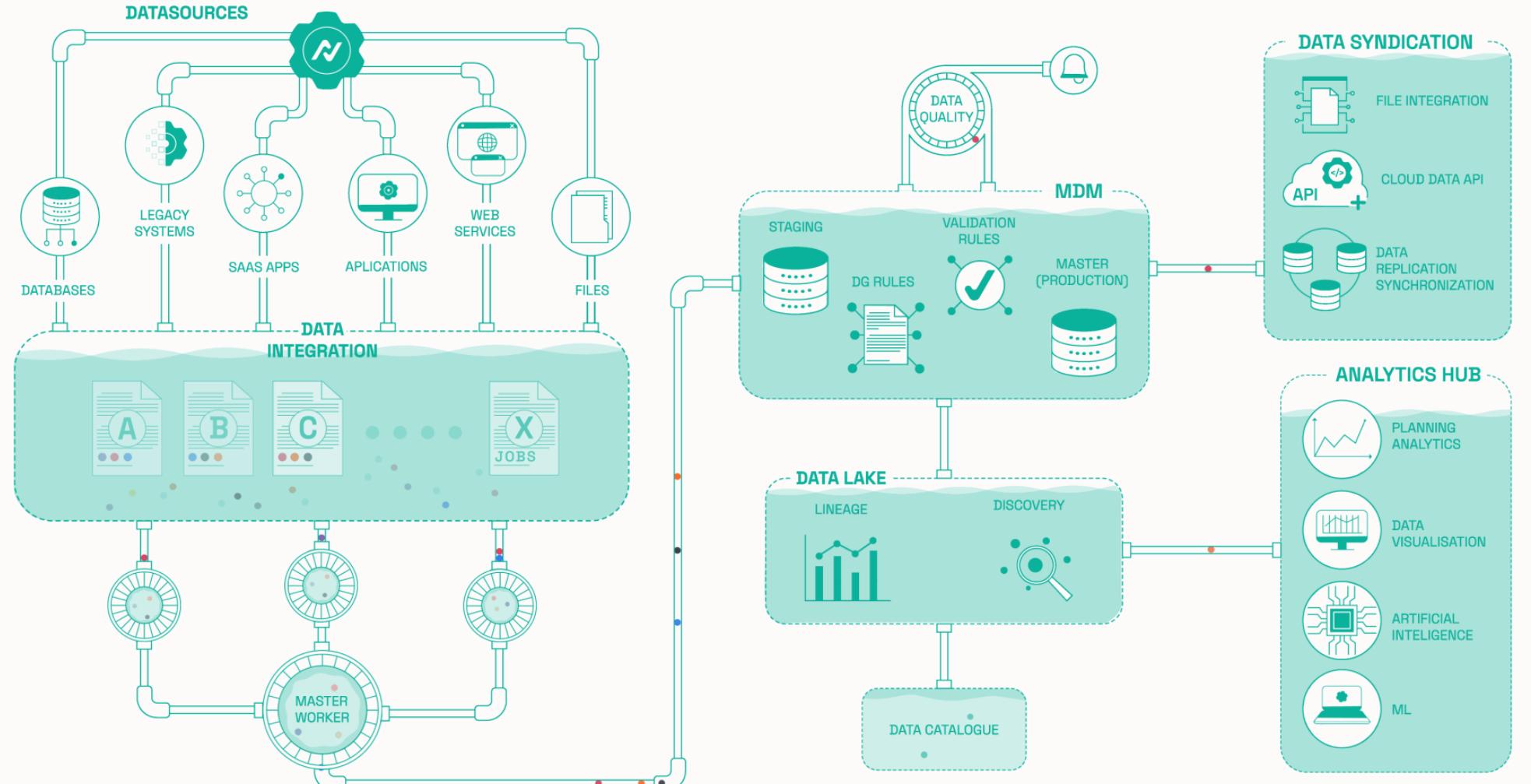


Componentes & Conceptos

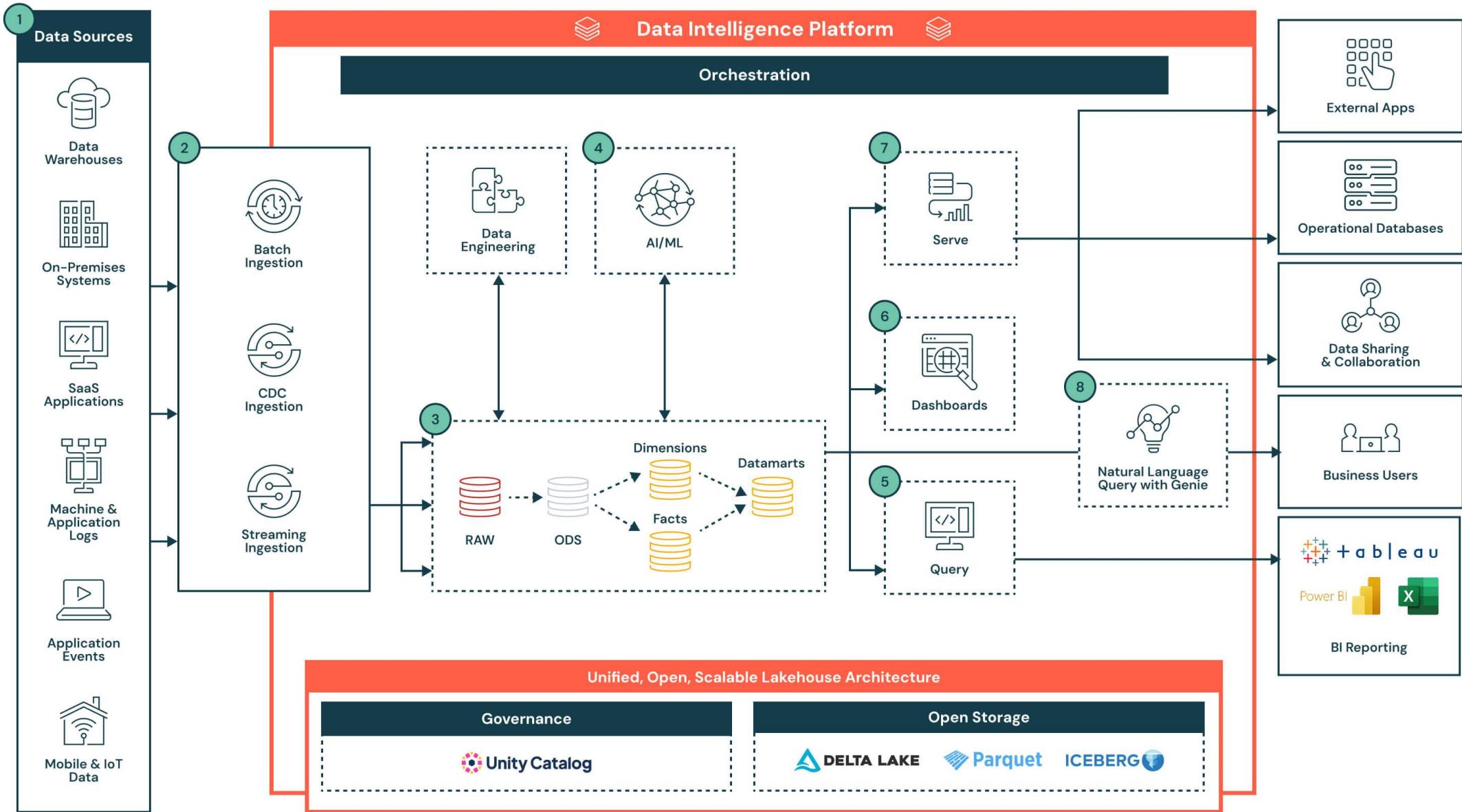


Pregunta	Bronce	Plata	Oro
¿Qué ocurre en esta capa?	Ingesta de datos sin procesar	Limpieza y validación de datos	Modelado y agregación dimensionales
¿Quién es el usuario previsto?	<ul style="list-style-type: none">•Ingenieros de datos•Operaciones de datos•Equipos de cumplimiento y auditoría	<ul style="list-style-type: none">•Ingenieros de datos•Analistas de datos (use la capa Silver para un conjunto de datos más refinado que conserva la información detallada necesaria para el análisis detallado)•Científicos de datos (crear modelos y realizar análisis avanzados)	<ul style="list-style-type: none">•Analistas de negocios y desarrolladores de BI•Científicos de datos e ingenieros de aprendizaje automático (ML)•Ejecutivos y responsables de la toma de decisiones•Equipos operativos

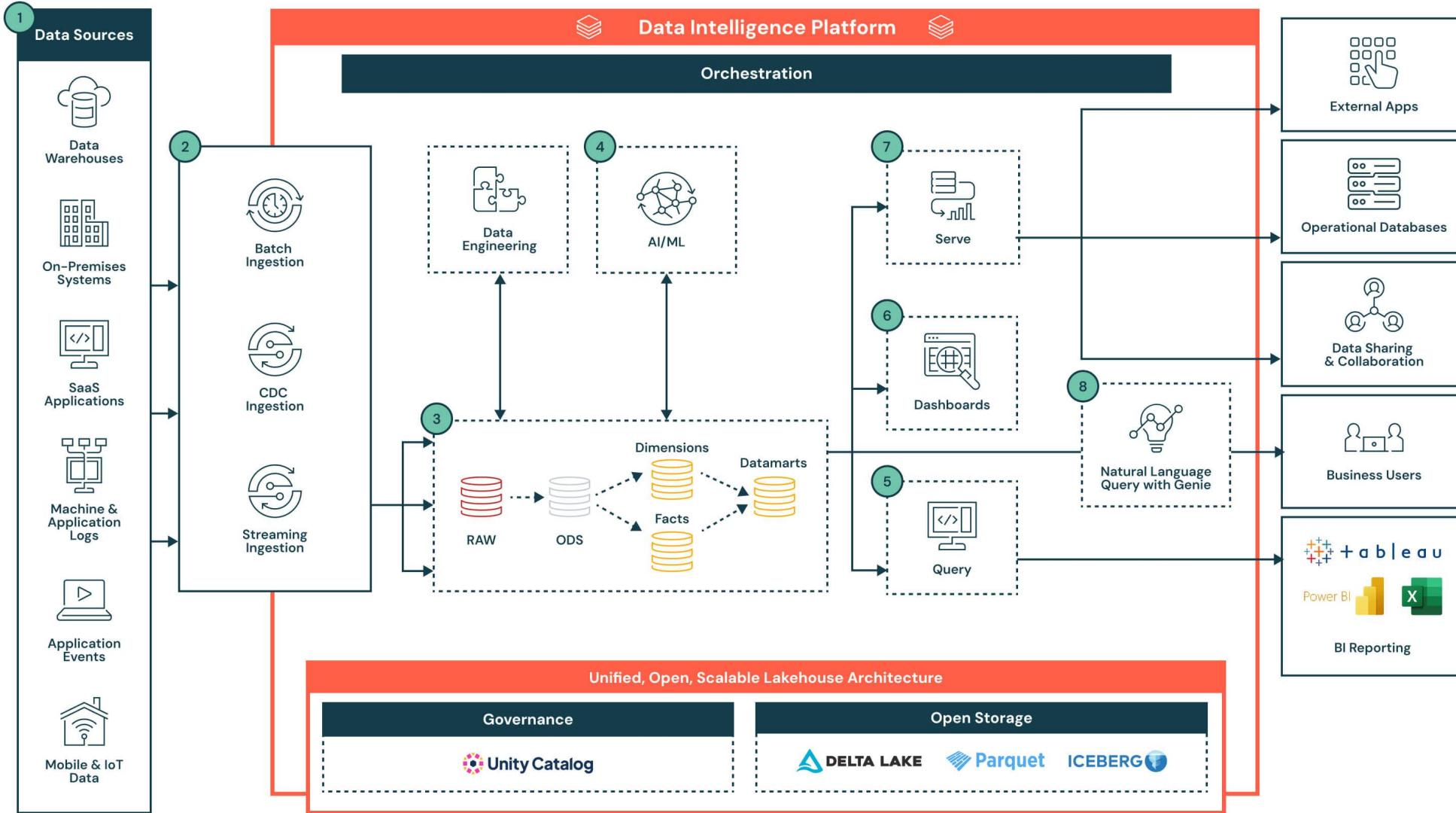
Compreensión de una Solución



Compreensión de una Solución



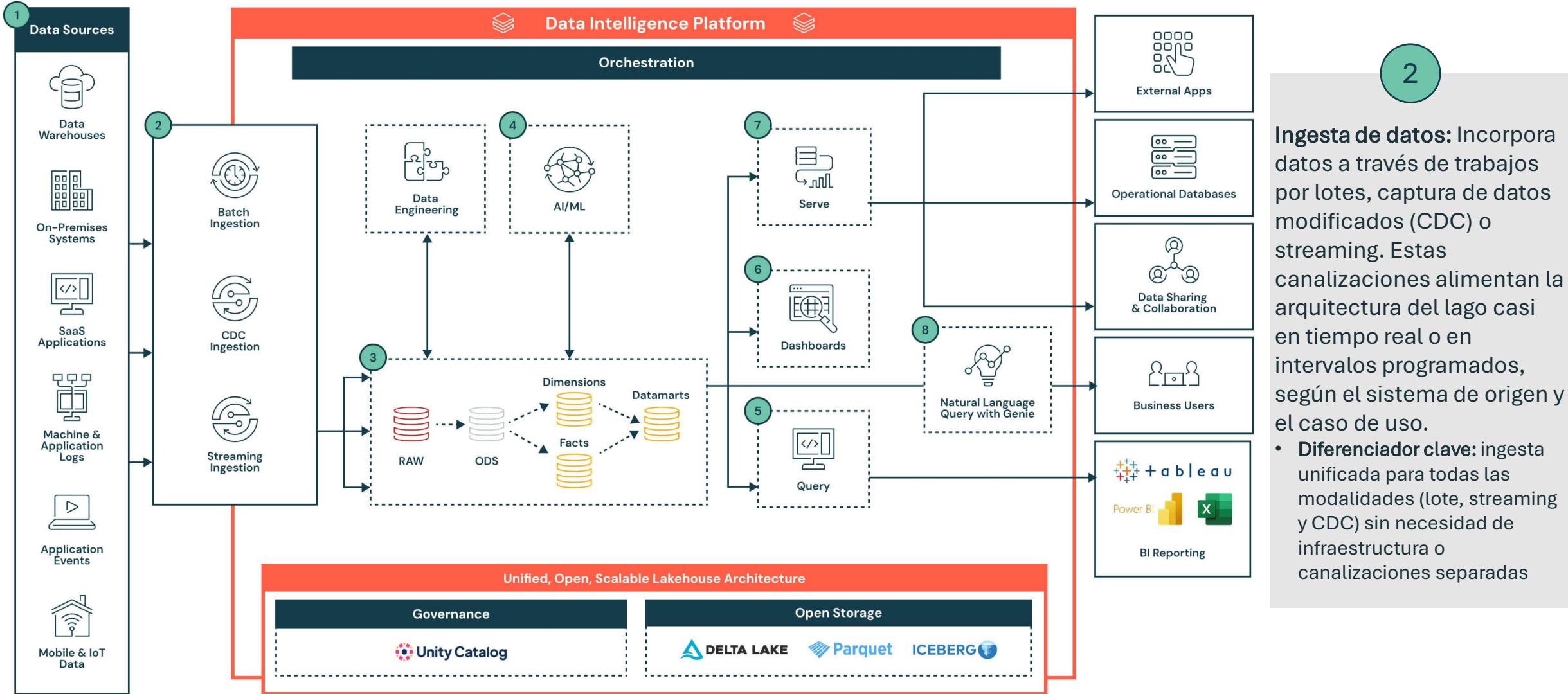
Compreensión de una Solución



1

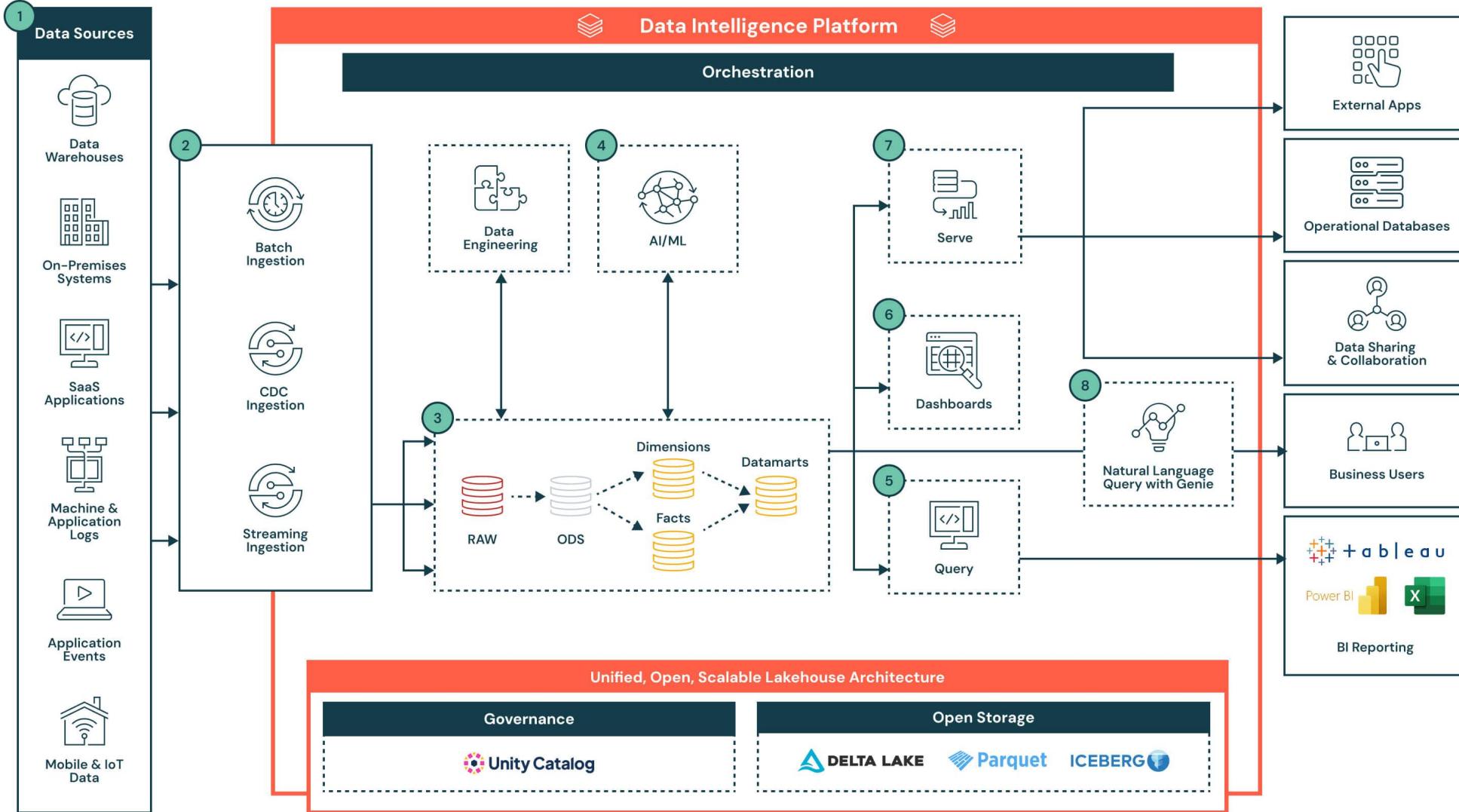
Fuentes de datos: los datos se almacenan en una amplia variedad de sistemas, incluidas aplicaciones empresariales (por ejemplo, SAP, Salesforce), bases de datos, dispositivos IoT, registros de aplicaciones y API externas. Estas fuentes pueden producir datos estructurados, semiestructurados o no estructurados.

Compreensión de una Solución



Compreensión de una Solución

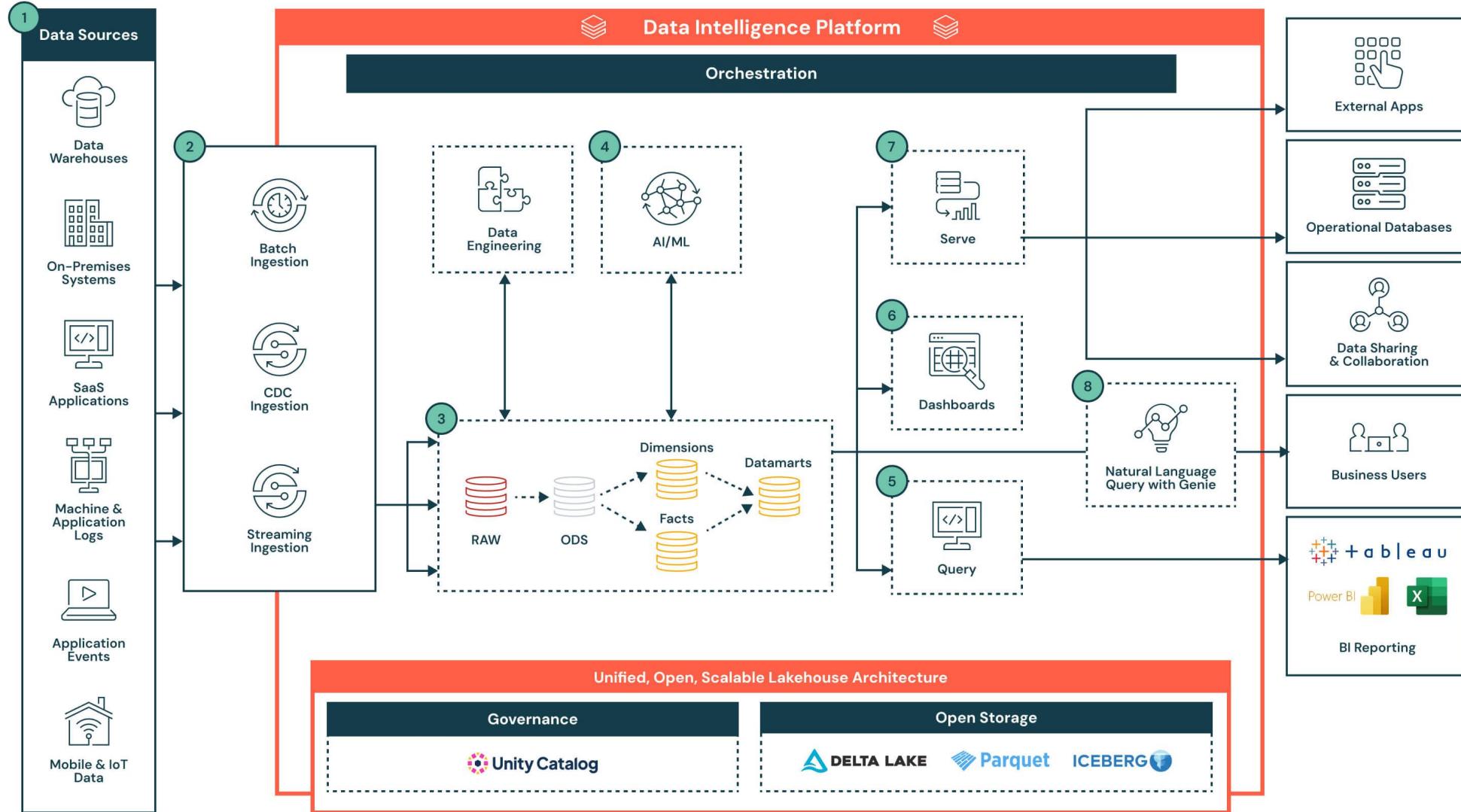
3



Zona sin procesar a zona Bronce: datos ingeridos desde sistemas de origen externos donde las estructuras de esta capa corresponden a las estructuras de tabla del sistema de origen "tal cual", sin transformación ni actualizaciones de los datos

- Zona Bronce a zona Plata:** Estandarizar y limpiar los datos entrantes
- Zona Plata a Zona Oro:** Aplicar lógica de negocios para crear modelos reutilizables
- Hechos y dimensiones → data marts:** agregue y seleccione datos para análisis posteriores
- Diferenciador clave:** canalizaciones declarativas de nivel de producción con linaje, observabilidad y evolución de esquemas integrados

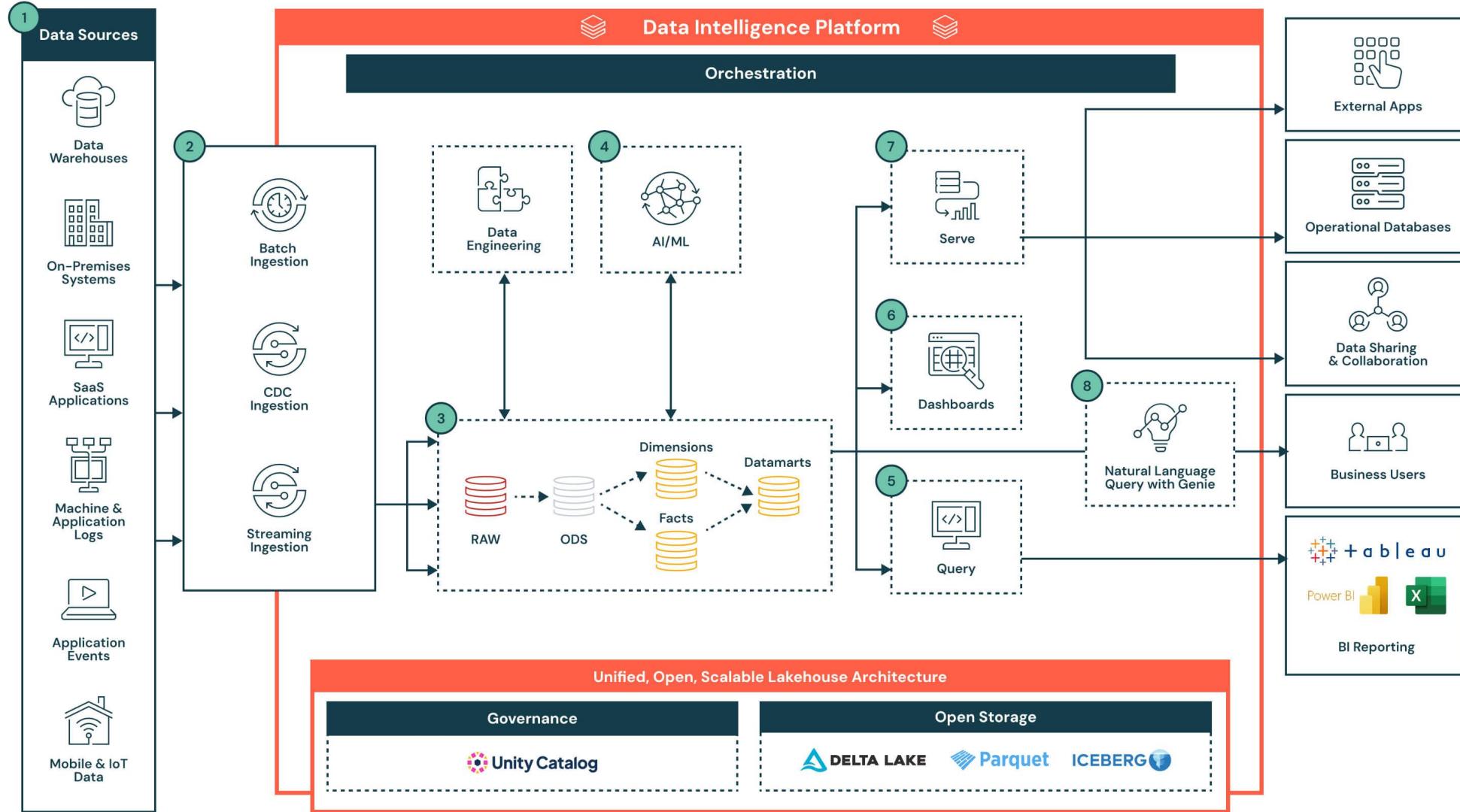
Compreensión de una Solución



Datos seleccionados para casos de uso de IA: los datos seleccionados de los data marts se pueden usar para entrenar o aplicar modelos de aprendizaje automático. Estos modelos admiten casos de uso como la previsión de la demanda, la detección de anomalías y la puntuación de clientes.

- Los resultados del modelo se almacenan junto con los datos tradicionales del almacén para facilitar el acceso a través de SQL o paneles
- Los resultados se pueden actualizar en un cronograma o calificar en tiempo real, según los requisitos
- **Diferenciador clave:** Cargas de trabajo de análisis e IA colocadas en la misma plataforma, sin necesidad de movimiento de datos. Las salidas del modelo se tratan como activos gobernados nativos y consultables.

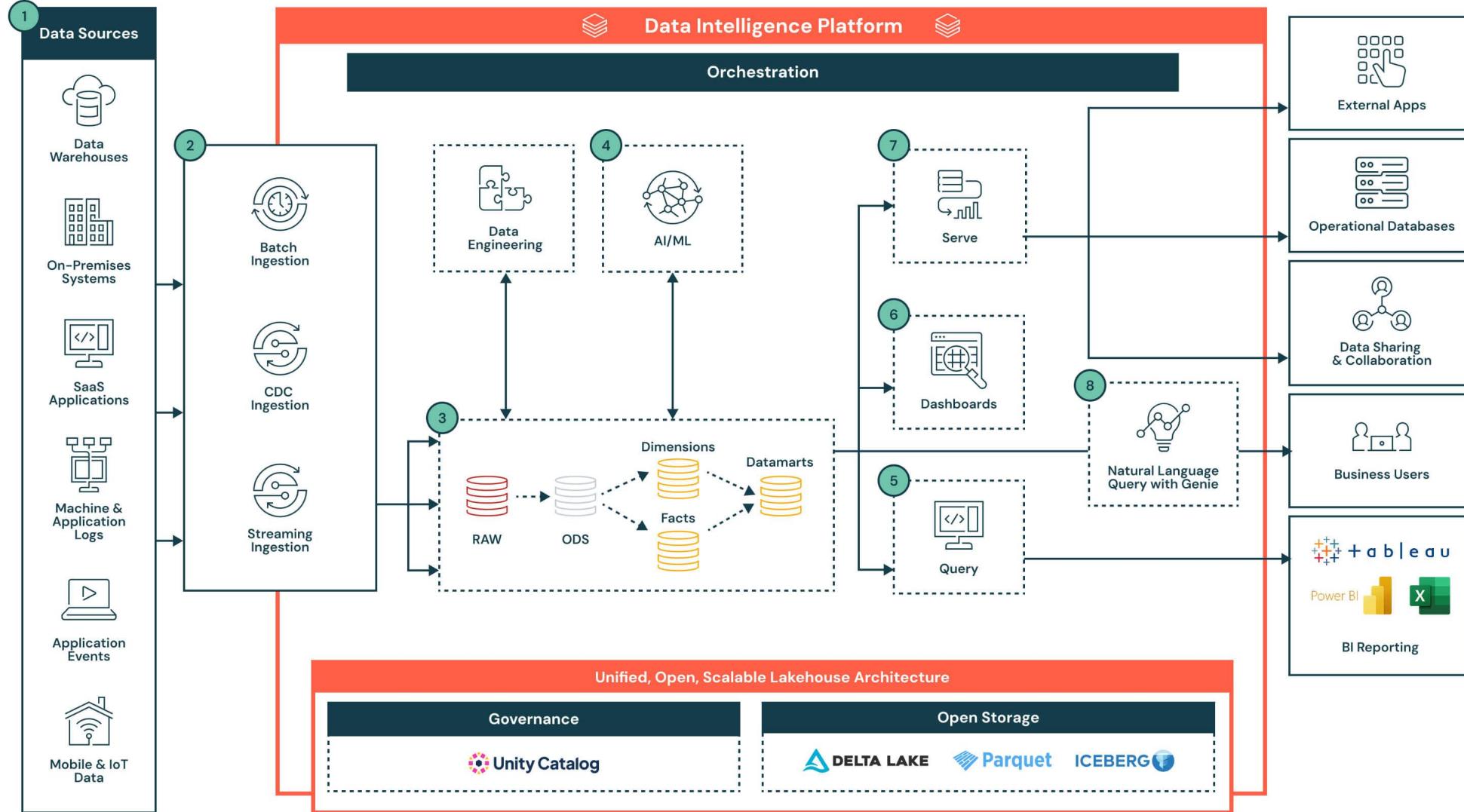
Compreensión de una Solución



Herramientas de informes de BI de alimentación de consultas: Databricks SQL admite consultas de alta simultaneidad y baja latencia a través de procesos sin servidor y se conecta fácilmente a herramientas de BI populares.

- Editor de consultas integrado e historial de consultas
- Las consultas devuelven resultados gobernados y actualizados de data marts o salidas de modelos enriquecidos
- **Diferenciador clave:** Databricks SQL permite que las herramientas de BI consulten datos directamente, sin replicación, lo que reduce la complejidad, evita costos de licencia adicionales y reduce el TCO general. Combinado con la computación sin servidor y la optimización inteligente, ofrece un rendimiento de nivel de almacén con un ajuste mínimo.

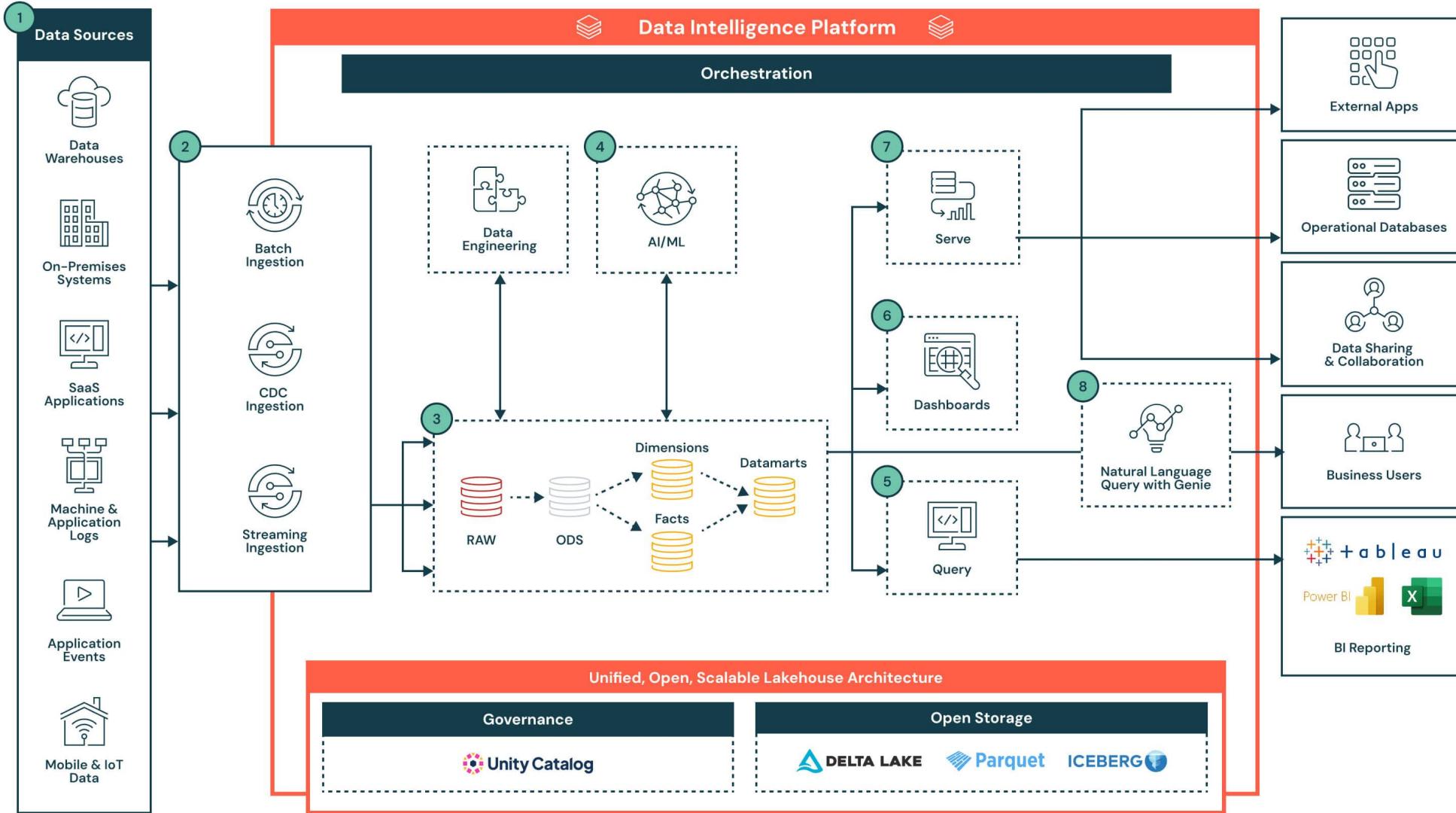
Compreensión de una Solución



Paneles: se pueden crear directamente en Databricks o en herramientas de BI externas como Power BI o Tableau. Los usuarios pueden describir objetos visuales en lenguaje natural y Databricks Assistant generará los gráficos correspondientes, que luego se pueden refinar mediante una interfaz de apuntar y hacer clic.

- Crear visualizaciones con entrada de lenguaje natural
- Modifique y explore paneles de forma interactiva con filtros y desgloses
- Publicar y compartir paneles de forma segura en toda la organización, incluso con usuarios fuera del área de trabajo de Databricks
- **Diferenciador clave:** Ofrece una experiencia de bajo código y asistida por IA para crear y explorar paneles en datos gobernados en tiempo real

Compreensión de una Solución

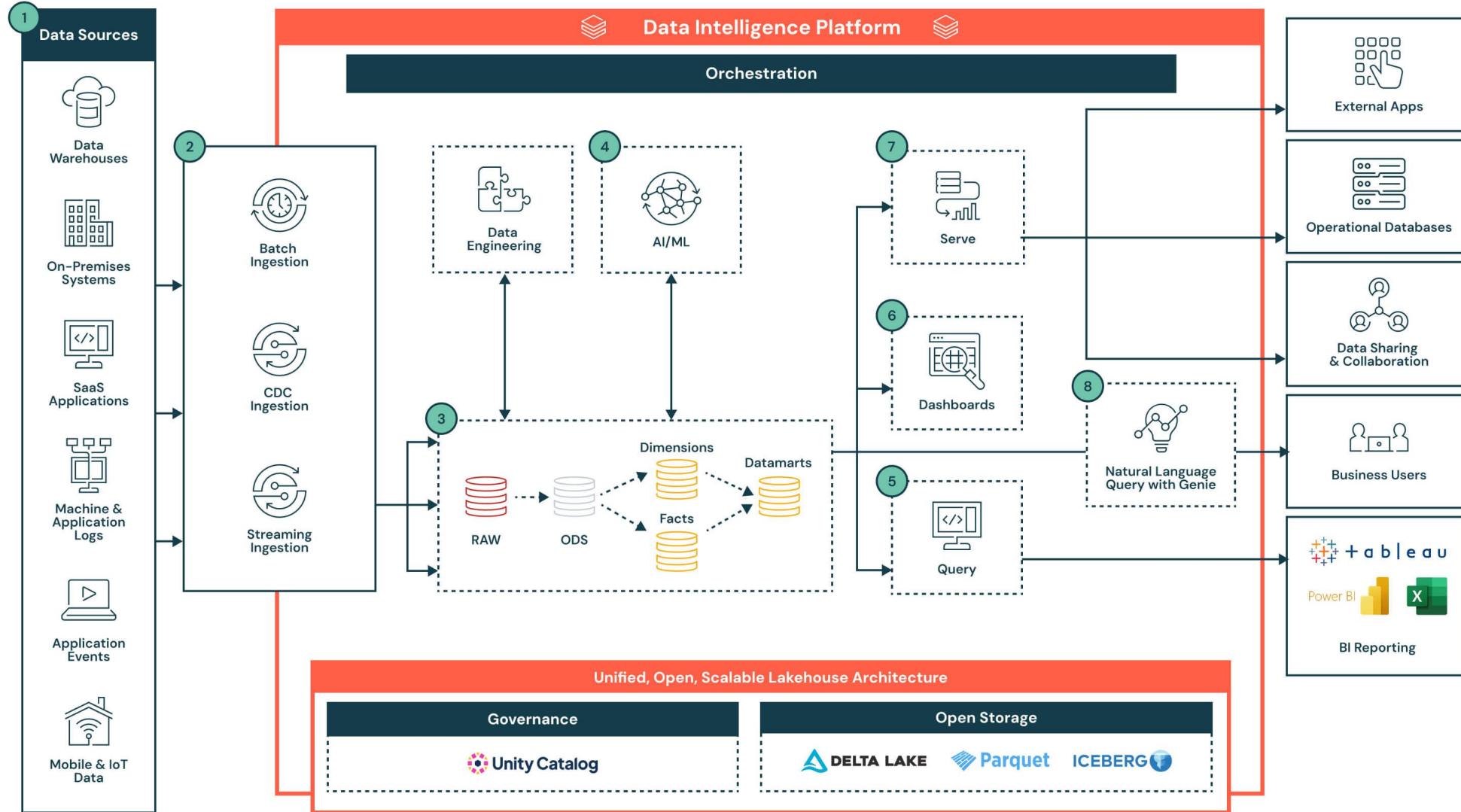


7

Entrega de datos seleccionados: Una vez refinados, los datos se pueden servir más allá de los paneles:

- Compartido con aplicaciones posteriores o bases de datos operativas para la toma de decisiones transaccionales
- Se utiliza en cuadernos colaborativos para el análisis
- Distribuido a través de Delta Sharing a socios, equipos o consumidores externos con gobernanza unificada

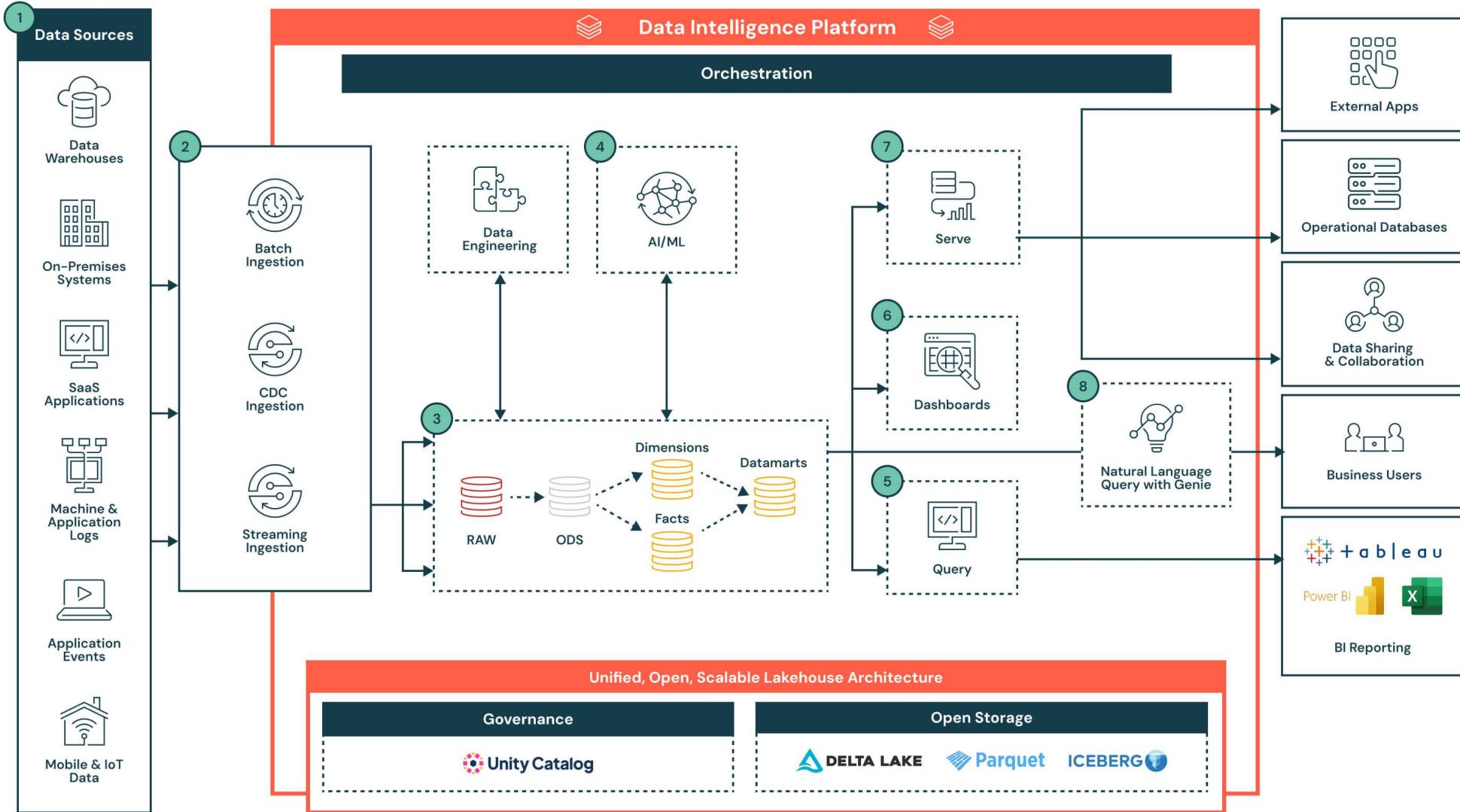
Compreensión de una Solución



Consulta en lenguaje natural (NLQ): los usuarios empresariales pueden acceder a los datos gobernados mediante lenguaje natural. Esta experiencia conversacional, impulsada por IA generativa, permite a los equipos ir más allá de los paneles estáticos y obtener información de autoservicio en tiempo real. NLQ traduce la intención del usuario en SQL aprovechando la semántica y los metadatos de la organización de Unity Catalog.

- Admite preguntas ad hoc, interactivas y en tiempo real que no están preintegradas en los paneles
- Se adapta de forma inteligente a la evolución de la terminología y el contexto empresarial a lo largo del tiempo
- Aprovecha la gobernanza de datos y los controles de acceso existentes a través de Unity Catalog

Compreensión de una Solución



9

Capacidades de la plataforma: Gobernanza, rendimiento, orquestación y almacenamiento abierto: La arquitectura está respaldada por un conjunto de capacidades nativas de la plataforma que admiten seguridad, optimización, automatización e interoperabilidad en todo el ciclo de vida de los datos.

Capacidades clave:

- Gobernanza
- Rendimiento
- Orquestación
- Almacenamiento abierto
- Supervisión y auditabilidad
- Diferenciador clave

¿Dónde puedo aprender más?



Ingeniero de IA

Exam



Ingeniero de Datos

Exam



DBA

Exam



Científico de Datos

Exam



Ingeniero de Datos

Exam



Analista de Datos

Exam

Ejemplo 1

Definición de Roles

Comportamiento Transaccional

Esquema

- Archivo .csv con Tabla de Transacciones en **S3**
- Archivo .csv con Tabla de Catálogo de Agencias en **S3**
- Archivo .csv con Tabla de Catálogo de Código de Transacción en **S3**



- Archivo Imagen con Tabla de Catálogo de Canales en **Azure Blob Storage**



Se utiliza Microsoft Fabric como entorno cloud para la ingestión, transformación, almacenamiento y consumo analítico de datos.



Información Procesada y Estructurada



Información Lista para Analizar



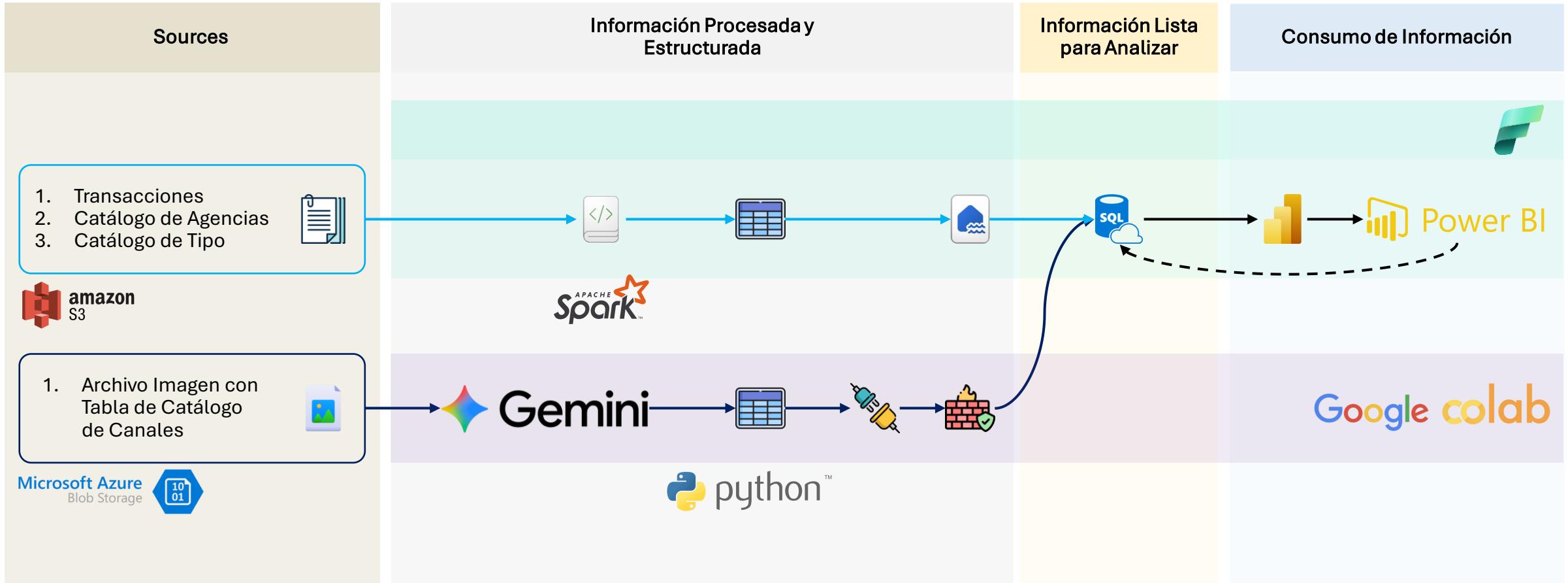
Consumo de Información



Ejemplo 1 – I

Comportamiento Transaccional

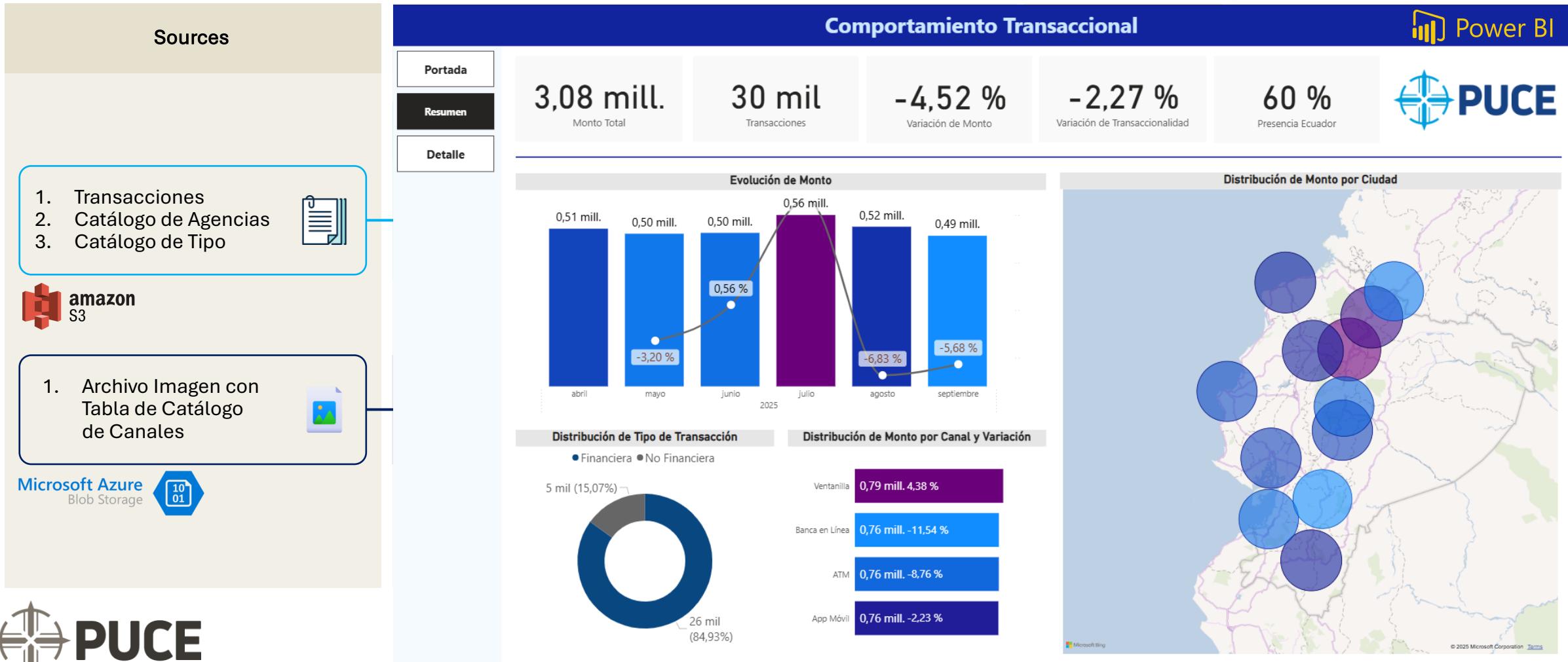
Diagrama de alto nivel



Ejemplo 1 – I

Comportamiento Transaccional

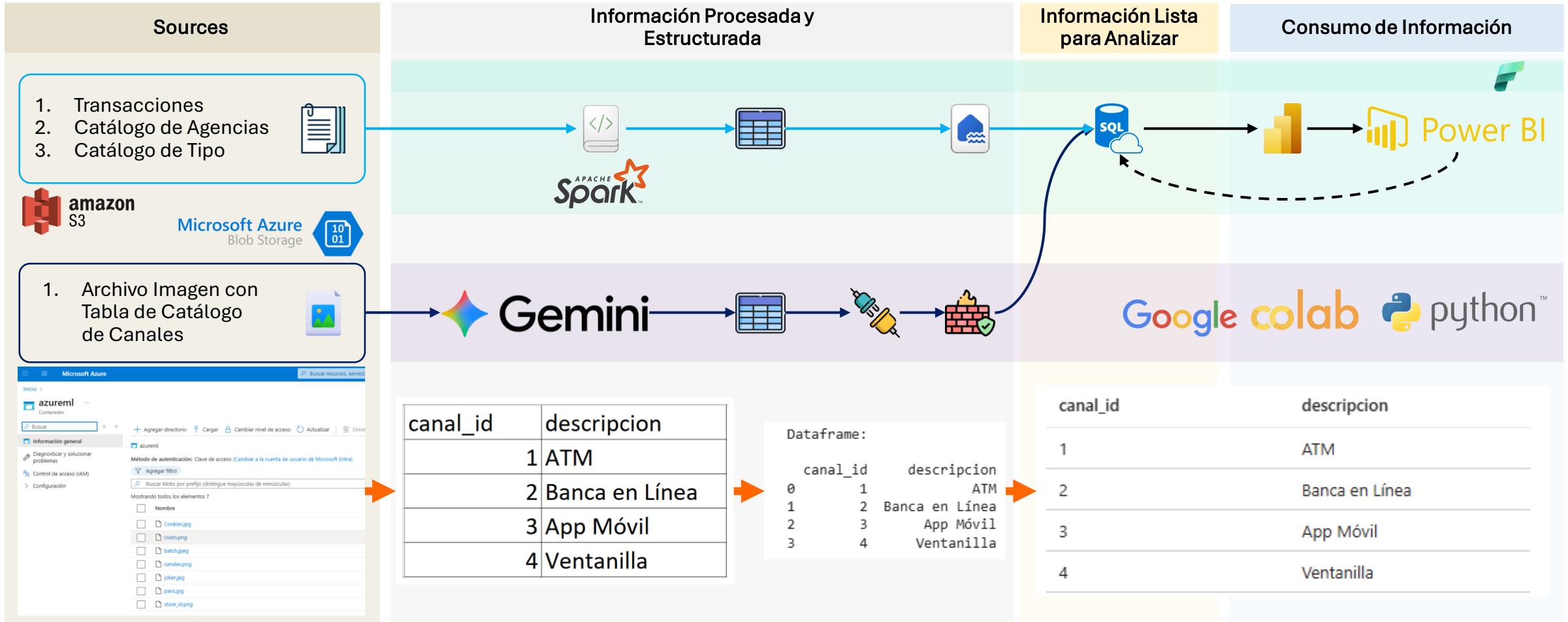
Diagrama de alto nivel



Ejemplo 1 – I

Comportamiento Transaccional

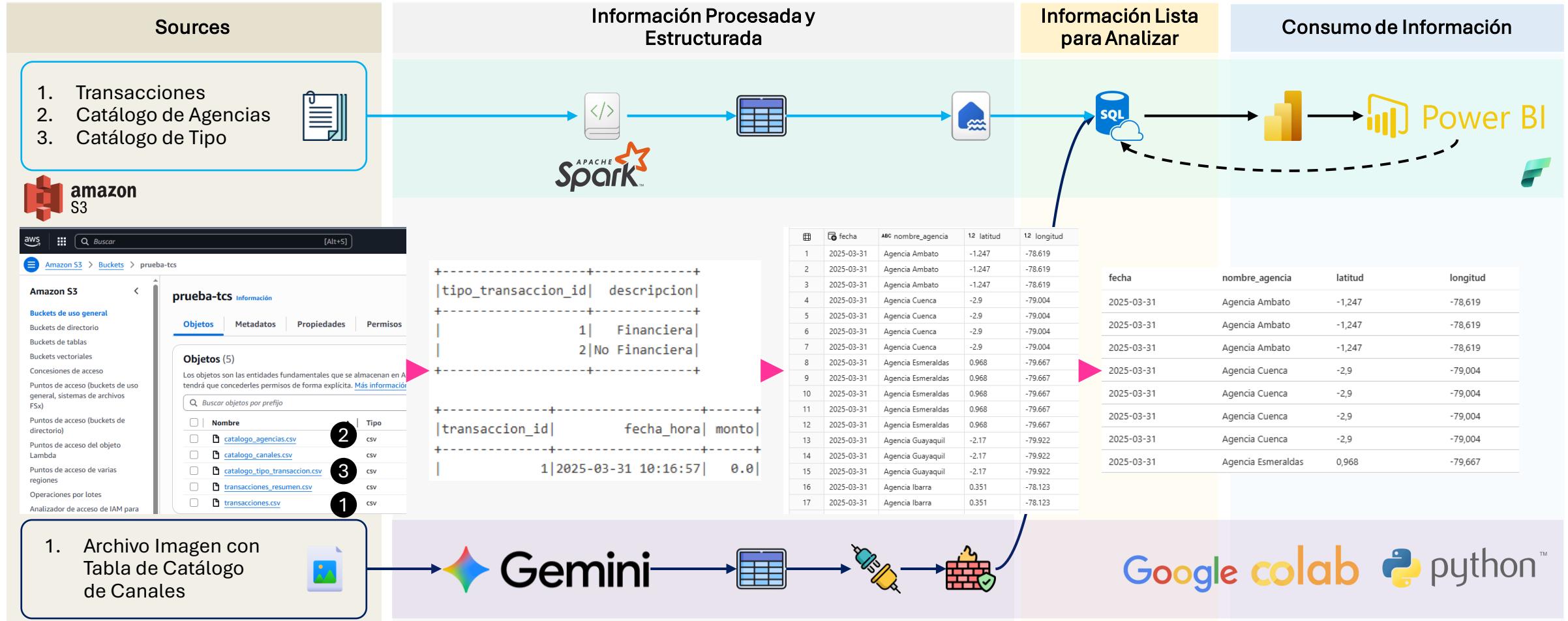
Diagrama de alto nivel



Ejemplo 1 – I

Comportamiento Transaccional

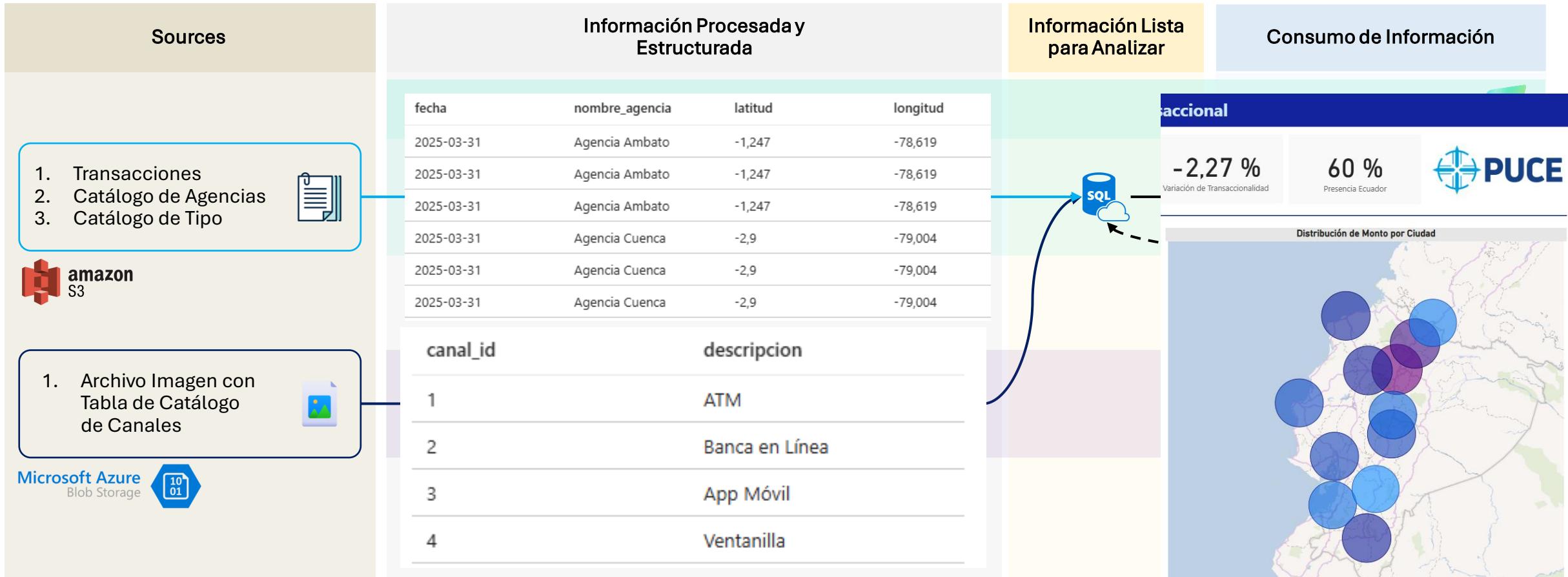
Diagrama de alto nivel



Ejemplo 1 – I

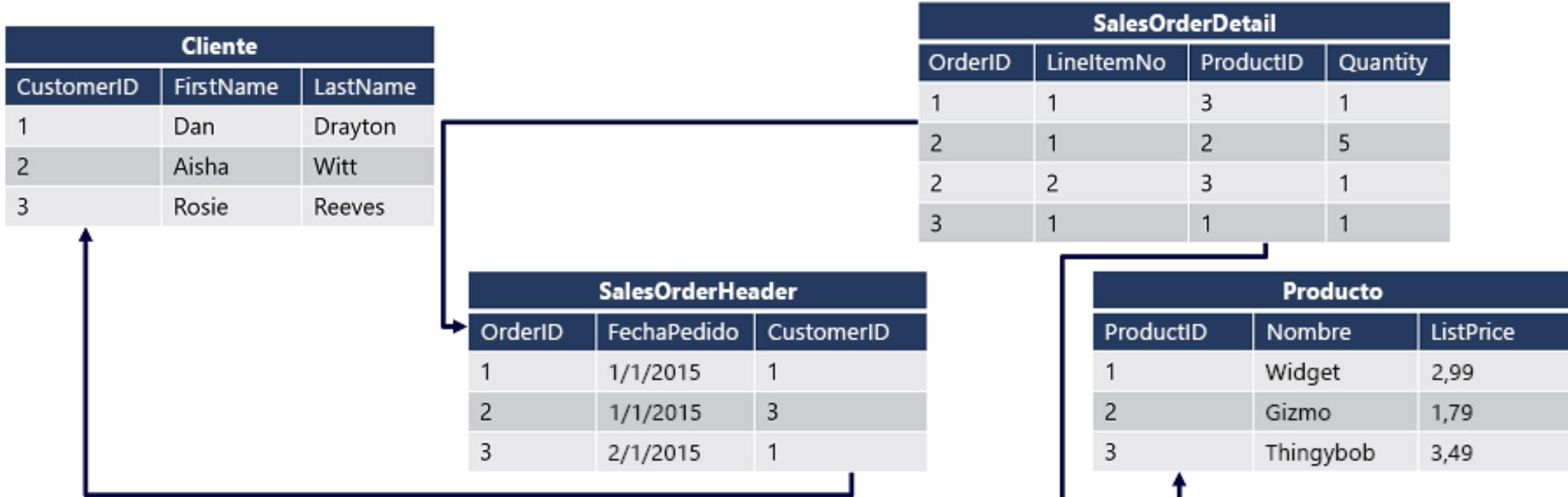
Comportamiento Transaccional

Diagrama de alto nivel



SQL Fundamentos

Datos relacionales

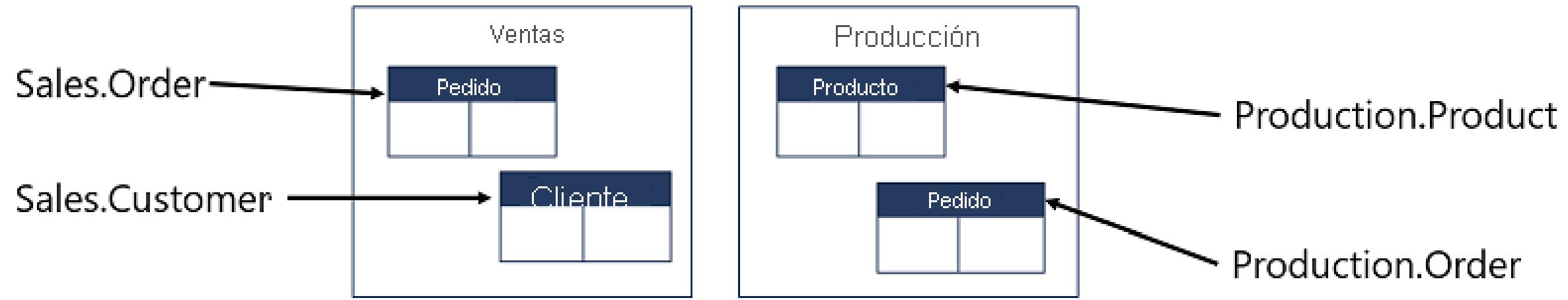


Una **base de datos relacional** es aquella en la que los datos se han organizado en varias tablas (denominadas técnicamente relaciones), cada una de las cuales representa un tipo determinado de **entidad** (por ejemplo, un cliente, un producto o un pedido de ventas). Los **atributos** de estas entidades (por ejemplo, el nombre de un cliente, el precio de un producto o la fecha de pedido de un encargo de ventas) se definen como columnas, o atributos, de la tabla y cada fila de la tabla representa una instancia del tipo de entidad (por ejemplo, un cliente, producto o pedido de ventas específico).

Las **tablas de la base de datos** están relacionadas entre sí mediante columnas de **clave** que identifican de forma única la entidad determinada representada. Se define una **clave principal** para cada tabla y una referencia a esta clave se define como una **clave externa** en cualquier tabla relacionada.

SQL Fundamentos

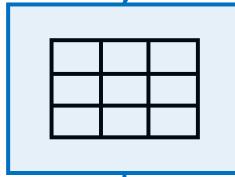
Trabajo con esquemas



El nombre completo de un objeto incluye el **nombre de una instancia del servidor de bases de datos** en la que se almacena la base de datos, el **nombre de la base de datos**, el **nombre del esquema** y el **nombre de la tabla**. Por ejemplo: **Server1.StoreDB.Sales.Order**.

SQL Fundamentos

Instrucciones SQL



Data Definition Language

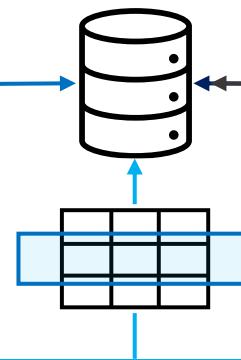
DDL

Conjunto de instrucciones SQL que **controla la definición y el ciclo de vida de los objetos** de base de datos, como tablas, vistas y procedimientos. DDL incluye instrucciones como **CREATE, ALTER y DROP**.

```
CREATE TABLE Estudiantes (
    id INT PRIMARY KEY IDENTITY(1,1),
    nombre VARCHAR(50),
    apellido VARCHAR(50),
    carrera VARCHAR(100),
    promedio DECIMAL(4,2)
);
```

```
ALTER TABLE Estudiantes
ADD correo VARCHAR(100);
```

```
DROP TABLE Estudiantes;
```



Lenguaje de manipulación

DML

Conjunto de instrucciones SQL que se centra en **consultar y modificar datos**. Las instrucciones DML incluyen **SELECT**, el punto central de este entrenamiento, y las instrucciones de modificación como **INSERT, UPDATE y DELETE**.

```
SELECT nombre, promedio
FROM Estudiantes
WHERE carrera = 'Ingeniería en Software';
```

```
INSERT INTO Estudiantes (nombre, apellido,
carrera, promedio)
VALUES ('Ana', 'Torres', 'Ingeniería en Software',
9.3);
```

```
UPDATE Estudiantes
SET promedio = 9.5
WHERE nombre = 'Ana';
```

```
DELETE FROM Estudiantes
WHERE promedio < 7;
```

Data Control Language

DCL

Conjunto de instrucciones SQL que se usan para **administrar permisos de seguridad para usuarios y objetos**. DCL incluye instrucciones como **GRANT, REVOKE y DENY**.

```
GRANT SELECT, INSERT ON Estudiantes TO
usuario_analista;
```

```
REVOKE INSERT ON Estudiantes FROM
usuario_analista;
```

```
DENY DELETE ON Estudiantes TO
usuario_analista;
```

SQL Fundamentos

Instrucciones SQL

Data Query Language

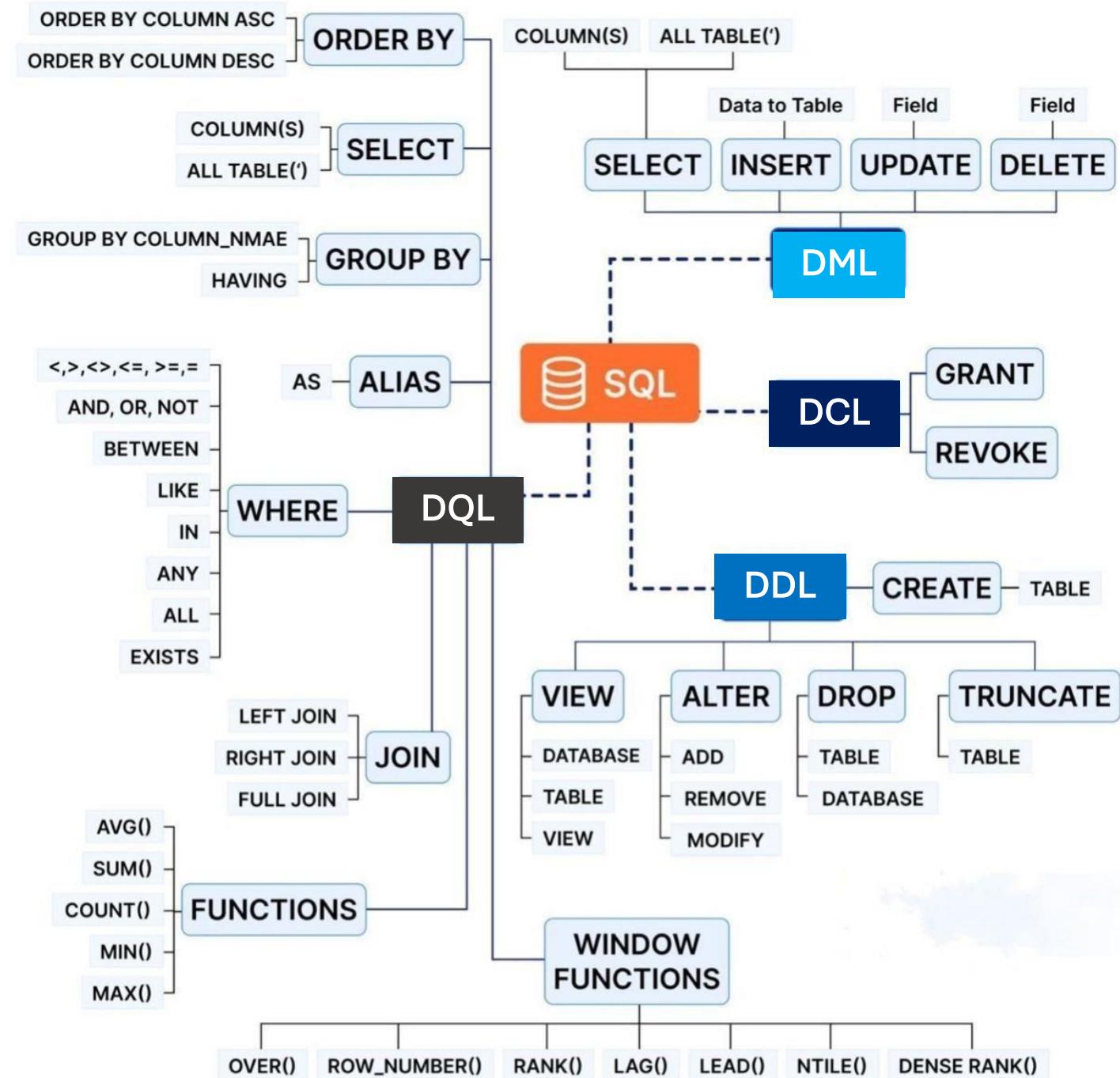
DQL

Conjunto de instrucciones SQL que se usan para **consultar y recuperar información** almacenada en las bases de datos. DQL incluye principalmente la instrucción SELECT, junto con cláusulas para filtrar, ordenar y agrupar datos.

```
SELECT nombre, promedio  
FROM Estudiantes  
WHERE promedio > 8.5;
```

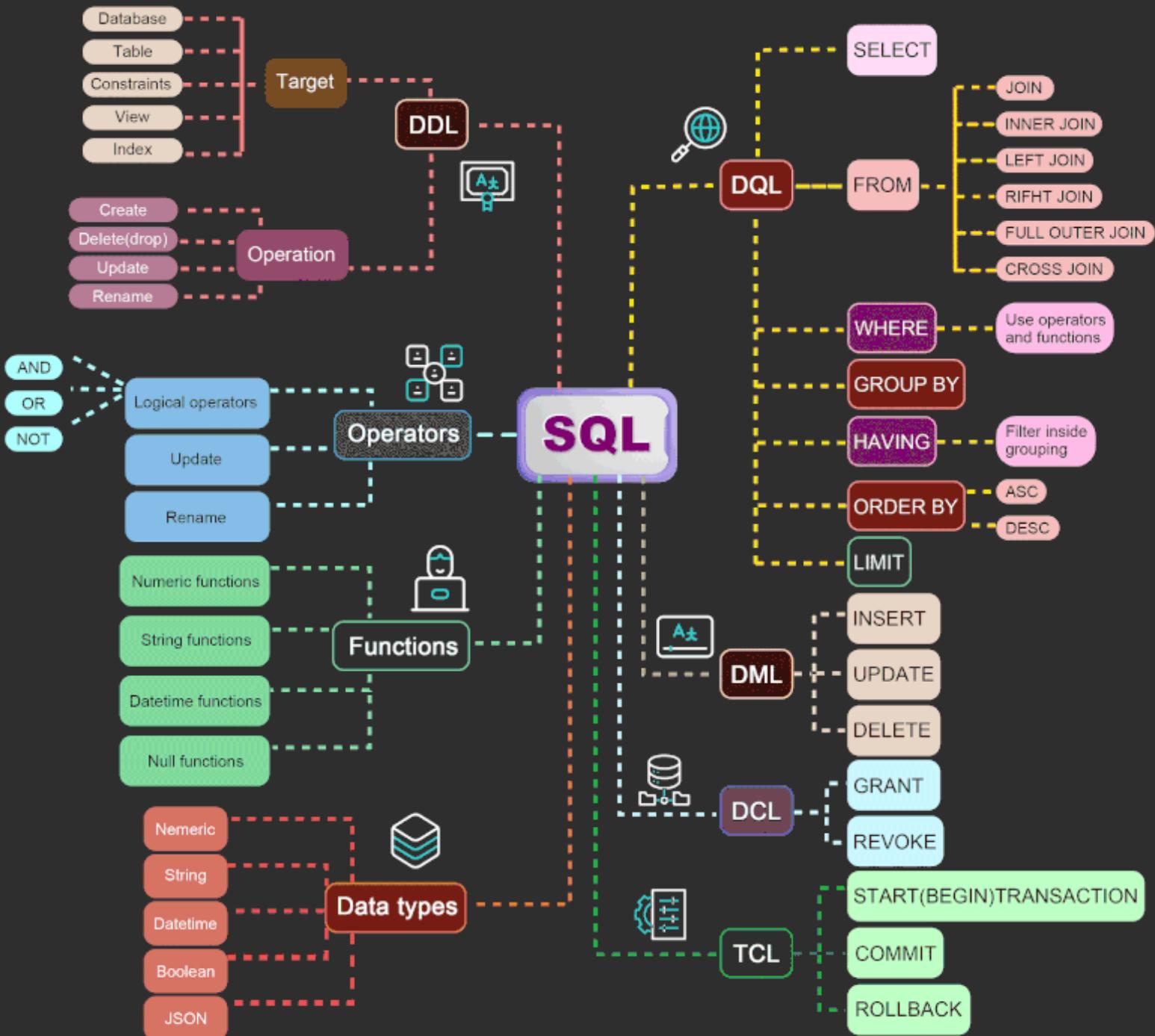
```
SELECT carrera, COUNT(*) AS total_estudiantes  
FROM Estudiantes  
GROUP BY carrera;
```

```
SELECT nombre, promedio  
FROM Estudiantes  
ORDER BY promedio DESC;
```



SQL Fundamentos

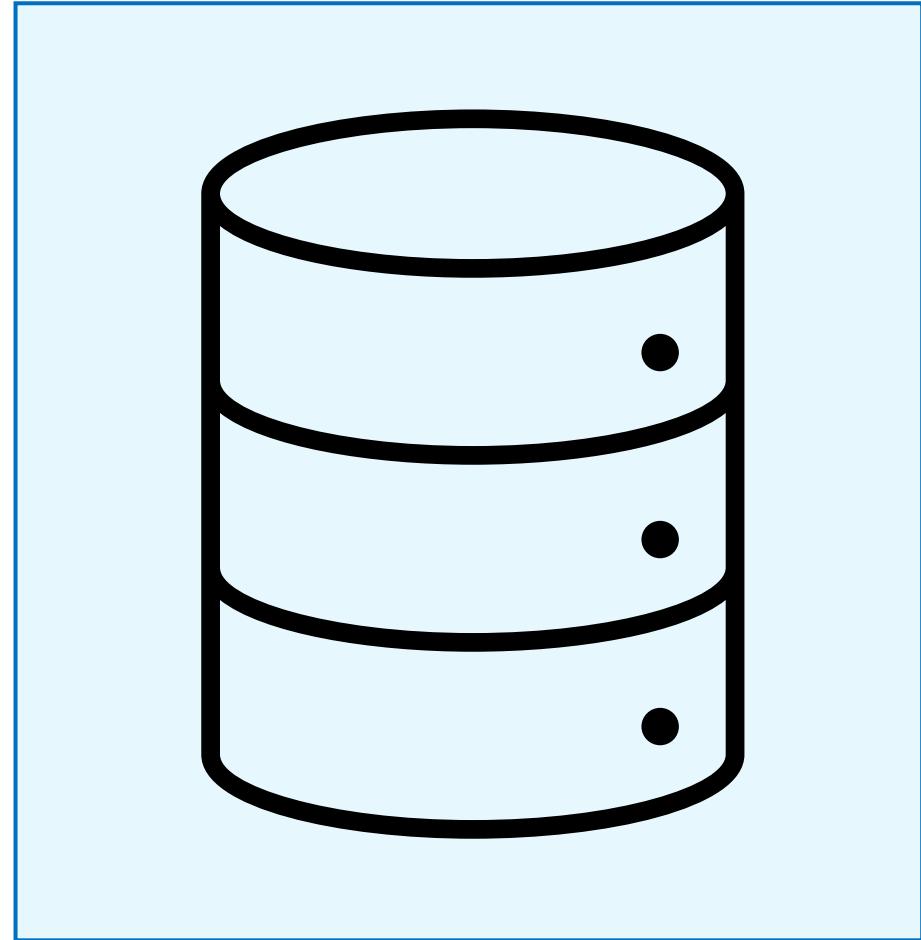
Instrucciones SQL



SQL Fundamentos

Ejecución

```
SELECT OrderDate, COUNT(OrderID) AS Orders  
FROM Sales.SalesOrder  
WHERE Status = 'Shipped'  
GROUP BY OrderDate  
HAVING COUNT(OrderID) > 1  
ORDER BY OrderDate DESC;
```

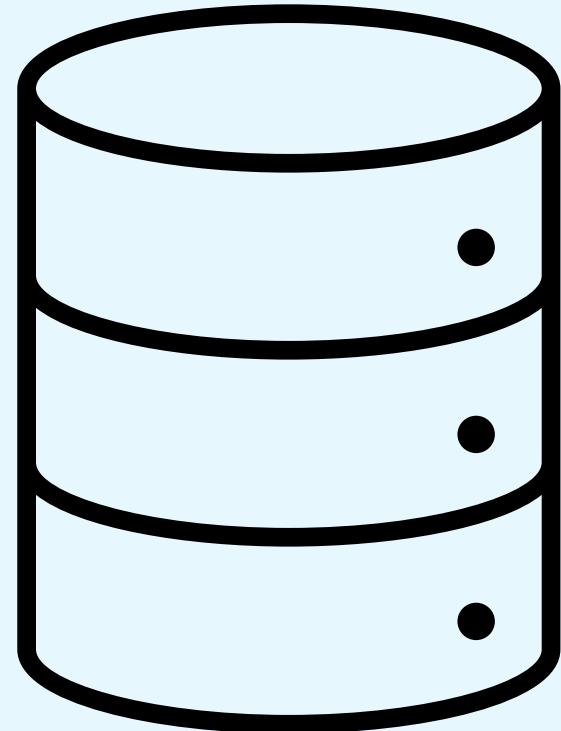


SQL Fundamentos

Ejecución

```
SELECT OrderDate, COUNT(OrderID) AS Orders  
FROM Sales.SalesOrder  
WHERE Status = 'Shipped'  
GROUP BY OrderDate  
HAVING COUNT(OrderID) > 1  
ORDER BY OrderDate DESC;
```

El orden en el que se escribe una instrucción SELECT no es el orden en que el motor de base de datos de SQL Server la evalúa y procesa.

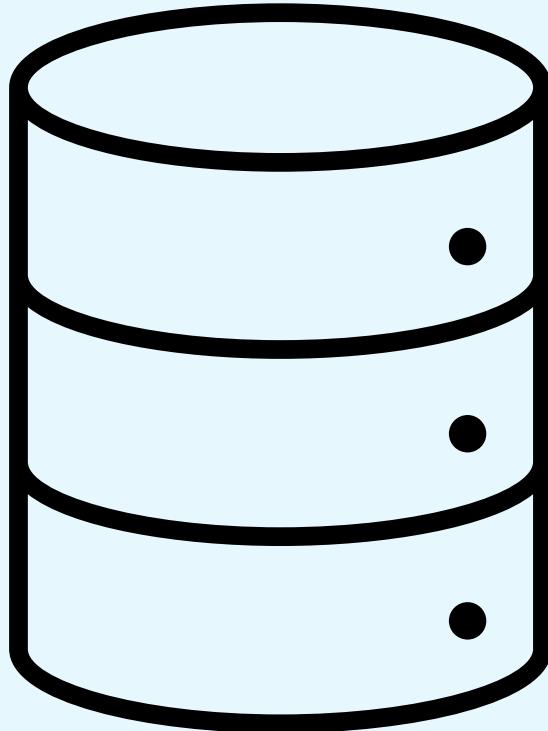


SQL Fundamentos

Ejecución

```
SELECT OrderDate, COUNT(OrderID) AS Orders  
FROM Sales.SalesOrder  
WHERE Status = 'Shipped'  
GROUP BY OrderDate  
HAVING COUNT(OrderID) > 1  
ORDER BY OrderDate DESC;
```

1. La cláusula **FROM** se evalúa primero para proporcionar las filas de origen para el resto de la instrucción. Se crea una tabla virtual y se pasa al paso siguiente.
2. La cláusula **WHERE** es la siguiente en evaluarse, filtrando las filas de la tabla de origen que coinciden con un predicado. La tabla virtual filtrada se pasa al paso siguiente.
3. **GROUP BY** es la siguiente, que organiza las filas de la tabla virtual según los valores únicos que se encuentran en la lista GROUP BY. Se crea una nueva tabla virtual que contiene la lista de grupos y se pasa al paso siguiente. A partir de este punto del flujo de operaciones, otros elementos solo pueden hacer referencia a las columnas de la lista GROUP BY o a las funciones de agregado.
4. Despues se evalúa la cláusula **HAVING**, que filtra grupos completos en función de su predicado. La tabla virtual creada en el paso 3 se filtra y se pasa al paso siguiente.
5. Por último, se ejecuta la cláusula **SELECT**, que determina qué columnas aparecerán en los resultados de la consulta. Dado que la cláusula SELECT se evalúa después de los otros pasos, los alias de columna (en nuestro ejemplo, Orders) creados allí no se pueden usar en las cláusulas GROUP BY o HAVING.
6. La cláusula **ORDER BY** es la última en ejecutarse, que ordena las filas según lo determinado por su lista de columnas.



```
FROM Sales.SalesOrder  
WHERE Status = 'Shipped'  
GROUP BY OrderDate  
HAVING COUNT(OrderID) > 1  
SELECT OrderDate, COUNT(OrderID) AS Orders  
ORDER BY OrderDate DESC;
```

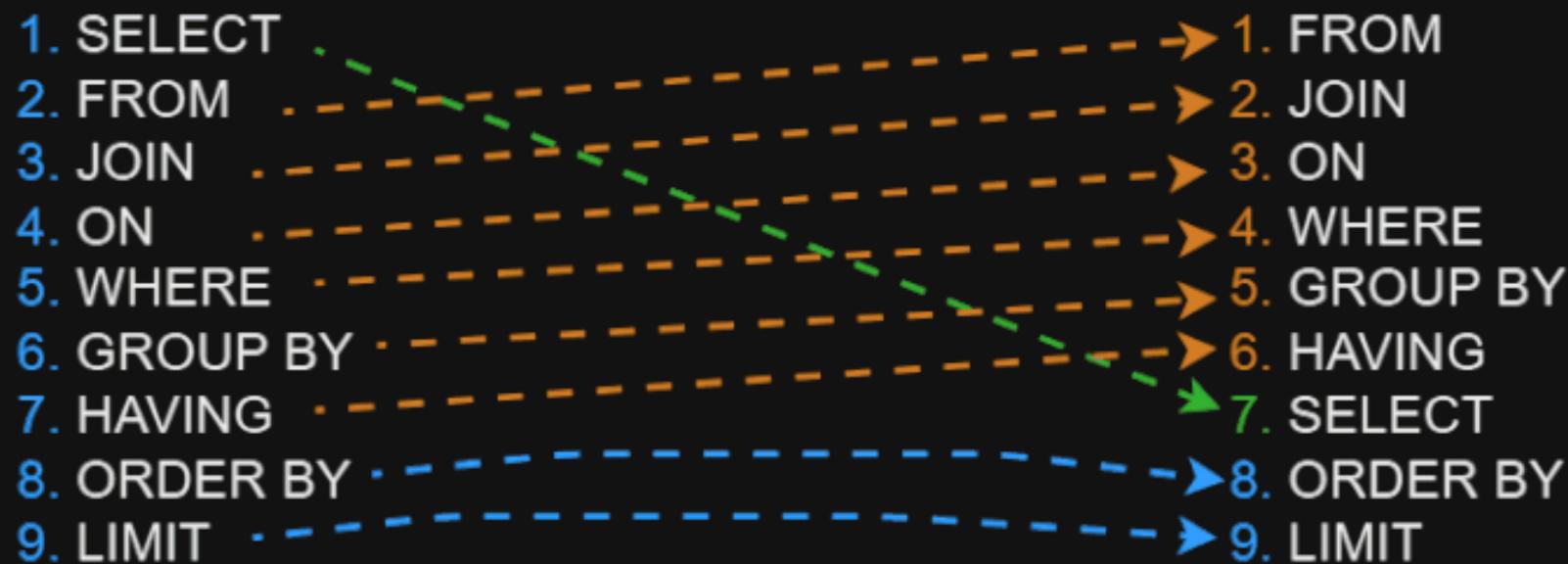
SQL Fundamentos

Ejecución

Coding order



Execution order



Piratas de Datos

y el océano infinito



Piratas de Datos

y el océano infinito

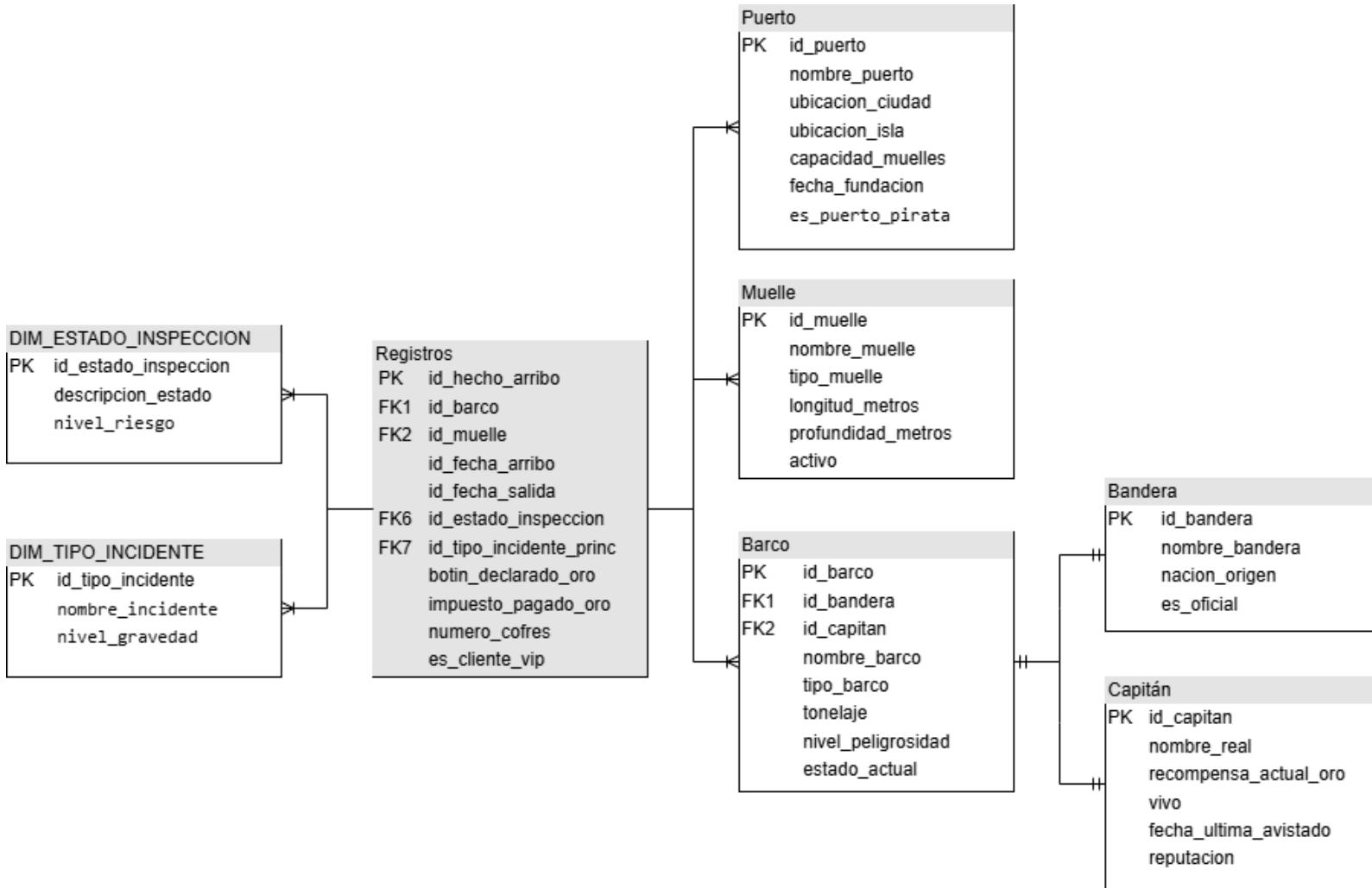
En el creciente negocio de la piratería marítima, la información se ha convertido en el tesoro más valioso. Cada barco que arriba al puerto deja un rastro de registros: manifiestos de carga, tripulaciones, inspecciones y bolines que deben ser controlados y analizados con precisión. Sin embargo, el aumento constante de datos ha generado un caos en los registros y la dirección del puerto necesita una solución que transforme la información dispersa en conocimiento estratégico.



Piratas de Datos

y el océano infinito

Tabla	Descripción
HECHO_ARRIBO_PIRATA	Registra cada arribo con medidas de negocio
DIM_BARCO	Información técnica y jerárquica de barcos
DIM_CAPITAN	Líder de embarcaciones
DIM_BANDERA	Origen o facción de barcos
DIM_MUELLE	Infraestructura portuaria
DIM_PUERTO	Ciudad o isla base del puerto
DIM_ESTADO_INSPECCION	Resultado del control portuario
DIM_TIPO INCIDENTE	Catálogo de incidentes



SQL Fundamentos

Conversión de tipo de datos	Limitación	Combinaciones	Agregado	Generación
CAST y TRY_CAST	TOP	INNER JOIN	SUM	IDENTITY
CONVERT y TRY_CONVERT	WITH TIES	OUTER JOIN	MEDIA	SEQUENCE
PARSE y TRY_PARSE	PERCENT	CROSS JOIN	MÍN	Actualización
STR	OFFSET-FETCH	Subconsultas	MÁX	UPDATE
Valores NULL	Filtrado con predicados	EXISTS	COUNT	Eliminación
ISNULL	WHERE	Funciones escalares	Resumen	DELETE
COALESCE	ES NULO / NO ES NULO	Funciones lógicas	GROUP BY	TRUNCATE TABLE
NULLIF	IN	IIF	Filtrado	
Ordenación	LIKE	CHOOSE	HAVING	
ORDER BY	Duplicados	Clasificación	Insertado	
	DISTINCT	OVER	INSERT	

SQL Fundamentos

Conversión de tipo de datos				
CAST y TRY_CAST	Convierte un valor a un tipo de datos específico.	CAST(expresión AS tipo_dato)	Igual que CAST, pero devuelve NULL si la conversión falla.	TRY_CAST(expresión AS tipo_dato)
CONVERT y TRY_CONVERT	Convierte valores con formato (útil para fechas y estilos).	CONVERT(tipo_dato, expresión [, estilo])	Igual que CONVERT, pero devuelve NULL si falla.	CONVERT y TRY_CONVERT
PARSE y TRY_PARSE	Convierte texto a número o fecha según configuración regional.	PARSE(expresión AS tipo_dato USING 'cultura')	Igual que PARSE, pero devuelve NULL si falla.	TRY_PARSE(expresión AS tipo_dato USING 'cultura')
STR	Convierte valores numéricos a texto.	STR(valor [, longitud [, decimales]])		
Valores NULL				
ISNULL	Reemplaza NULL por un valor definido.	ISNULL(expresión, valor_reemplazo)		
COALESCE	Devuelve el primer valor NO NULL de una lista.	COALESCE(valor1, valor2, valor3, ...)		
NULLIF	Devuelve NULL si dos expresiones son iguales.	NULLIF(exp1, exp2)		
Ordenación				
ORDER BY	Ordena el resultado de una consulta.	ORDER BY columna ASC		

SQL Fundamentos

Limitación	
TOP	Limita el número de filas devueltas.
WITH TIES	Incluye empates del último valor al usar ORDER BY.
PERCENT	Devuelve un porcentaje de filas.
OFFSET-FETCH	Paginación de resultados (omite y luego toma filas).
Filtrado con predicados	
WHERE	Filtra filas según una condición.
ES NULO / NO ES NULO	Evalúa si un campo tiene valor nulo o no.
IN	Compara si un valor pertenece a una lista o subconsulta.
LIKE	Filtra texto usando patrones (%,_).
Duplicados	
DISTINCT	Elimina filas duplicadas en el resultado.

`SELECT TOP (n) * FROM tabla;`

`SELECT TOP (n) WITH TIES * FROM tabla ORDER BY columna;`

`SELECT TOP (10) PERCENT * FROM tabla;`

`ORDER BY columna OFFSET n ROWS FETCH NEXT m ROWS ONLY;`

`WHERE condición`

`WHERE columna IS NULL`

`WHERE columna IN (valor1, valor2, ...)`

`WHERE columna LIKE 'A%'`

`SELECT DISTINCT columna FROM tabla;`

SQL Fundamentos

Combinaciones

INNER JOIN

Devuelve filas que tienen coincidencia en ambas tablas.

```
SELECT * FROM A INNER JOIN B  
ON A.id = B.id;
```



```
1 SELECT *  
2 FROM A  
3 INNER JOIN B ON A.key =  
B.key;
```

INNER JOIN

Retrieves rows where there is a match between tables A and B based on a common key.

OUTER JOIN

Devuelve todas las filas (LEFT, RIGHT o FULL).

```
SELECT * FROM A LEFT JOIN B ON  
A.id = B.id;
```



```
1 SELECT *  
2 FROM A  
3 FULL JOIN B ON A.key =  
B.key;
```

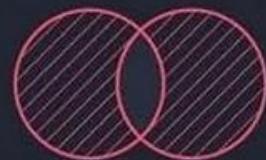
FULL JOIN

Retrieves rows where there is a match in either table A, table B, or both, including all data from both tables.

CROSS JOIN

Combina todas las filas de ambas tablas (producto cartesiano).

```
SELECT * FROM A CROSS JOIN B;
```



```
1 SELECT *  
2 FROM A  
3 FULL JOIN B ON A.key = B.key;  
4 WHERE A.key IS NULL OR B.key  
IS NULL;
```

FULL JOIN (WITH NULL CHECK)

Retrieves rows where there is no match between tables A and B, capturing only unmatched rows from both tables.

Subconsultas

EXISTS

Comprueba si una subconsulta devuelve resultados.

```
WHERE EXISTS (SELECT * FROM  
tabla WHERE condición)
```



```
1 SELECT *  
2 FROM A  
3 LEFT JOIN B ON A.key =  
B.key;
```

LEFT JOIN

Returns all rows from table A, along with matched rows from table B. Rows in A without a match in B will still appear, with NULL values for B's columns.

Funciones escalares

Funciones lógicas

IIF

Condición rápida: devuelve valor1 si es verdadero, valor2 si no.

```
IIF(condición, valor_si_verdadero,  
valor_si_falso)
```



```
1 SELECT *  
2 FROM A  
3 LEFT JOIN B ON A.key =  
B.key;  
4 WHERE B.key IS NULL;
```

LEFT JOIN (WITH NULL CHECK)

Selects rows from table A that do not have a corresponding match in table B, filtering out only unmatched rows.

Clasificación

OVER

E Define particiones o ventanas para funciones agregadas.

```
SUM(columna) OVER(PARTITION  
BY columna)
```



```
1 SELECT *  
2 FROM A  
3 RIGHT JOIN B ON A.key =  
B.key;
```

RIGHT JOIN

Returns all rows from table B, along with matching rows from table A. Rows in B without a match in A will still appear, with NULL values for A's columns.



```
1 SELECT *  
2 FROM A  
3 RIGHT JOIN B ON A.key =  
B.key;  
4 WHERE A.key IS NULL;
```

RIGHT JOIN (WITH NULL CHECK)

Selects rows from table B that lack a corresponding match in table A, isolating only unmatched rows.

SQL Fundamentos

Agregado		
SUM	Suma valores numéricos.	SUM(columna)
MEDIA	Calcula el promedio.	AVG(columna)
MÍN	Devuelve el valor mínimo.	MIN(columna)
MÁX	Devuelve el valor máximo.	MAX(columna)
COUNT	Cuenta filas o valores.	COUNT(*) o COUNT(columna)
Resumen		
GROUP BY	Agrupa resultados para aplicar funciones agregadas.	GROUP BY columna
Filtrado		
HAVING	Filtira grupos después de GROUP BY.	HAVING condición
Insertado		
INSERT	Aggrega nuevos registros.	INSERT INTO tabla (col1, col2) VALUES (val1, val2);

SQL Fundamentos

Generación

IDENTITY

Genera valores automáticos consecutivos (clave primaria).

IDENTITY(inicio, incremento)

SEQUENCE

Genera secuencias personalizadas reutilizables.

CREATE SEQUENCE nombre AS tipo
START WITH x INCREMENT BY y;

Actualización

UPDATE

Modifica registros existentes.

UPDATE tabla SET columna = valor
WHERE condición;

Eliminación

DELETE

Elimina registros.

DELETE FROM tabla WHERE
condición;

TRUNCATE TABLE

Borra todos los registros de una tabla (más rápido que DELETE).

TRUNCATE TABLE tabla;

Richard Guanoluisa
Bases de Datos II
Pontificia Universidad Católica
del Ecuador