



SNOWFLAKE WEBINAR

Iván Legorreta

ÍNDICE

- Batch vs Streaming
- SnowPark
- Dynamic Tables
- Streamlit
- Snowflake Cortex

- Data Warehouses
- ¿Qué es Snowflake?
- ¿Por qué Snowflake es diferente?
- Métodos de Ingesta

DATA WAREHOUSES

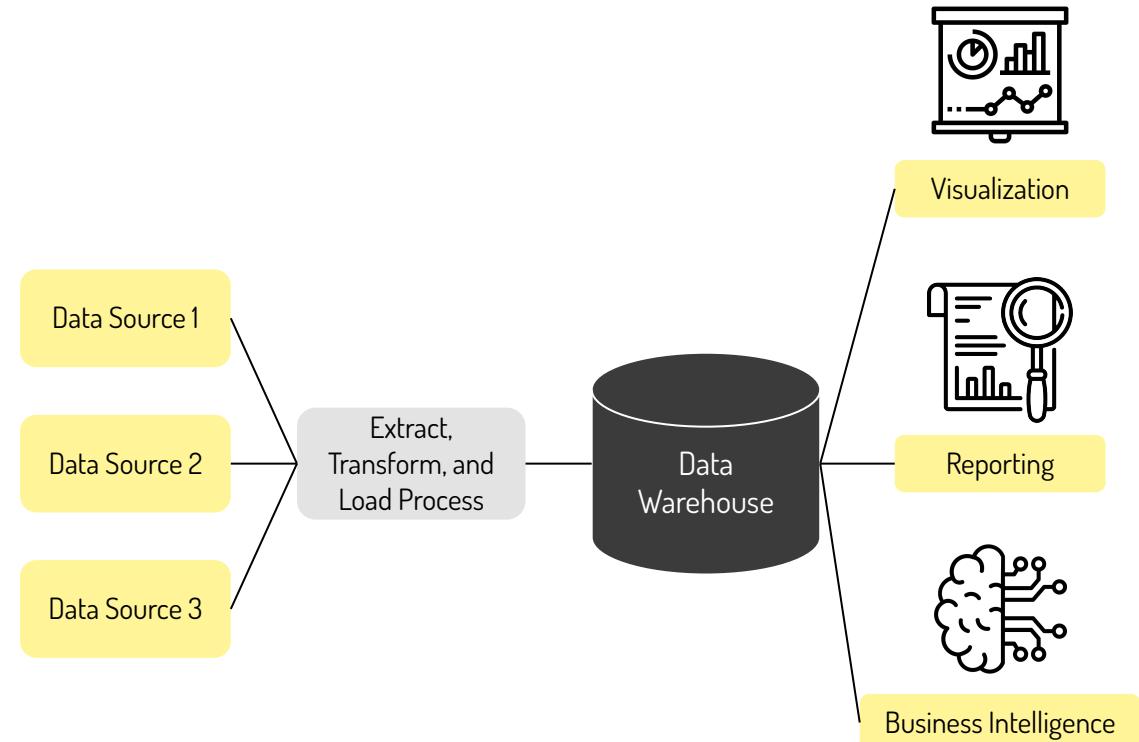
DATA WAREHOUSE

Infraestructura de almacenamiento de datos utilizada para almacenar, organizar y analizar grandes volúmenes de datos estructurados.

Datos pueden provenir de diversas fuentes dentro de una organización, como:

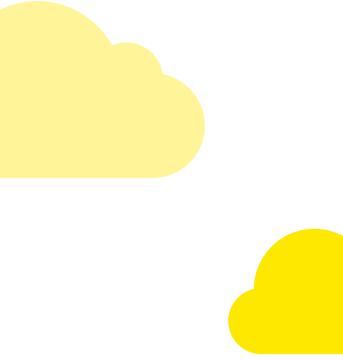
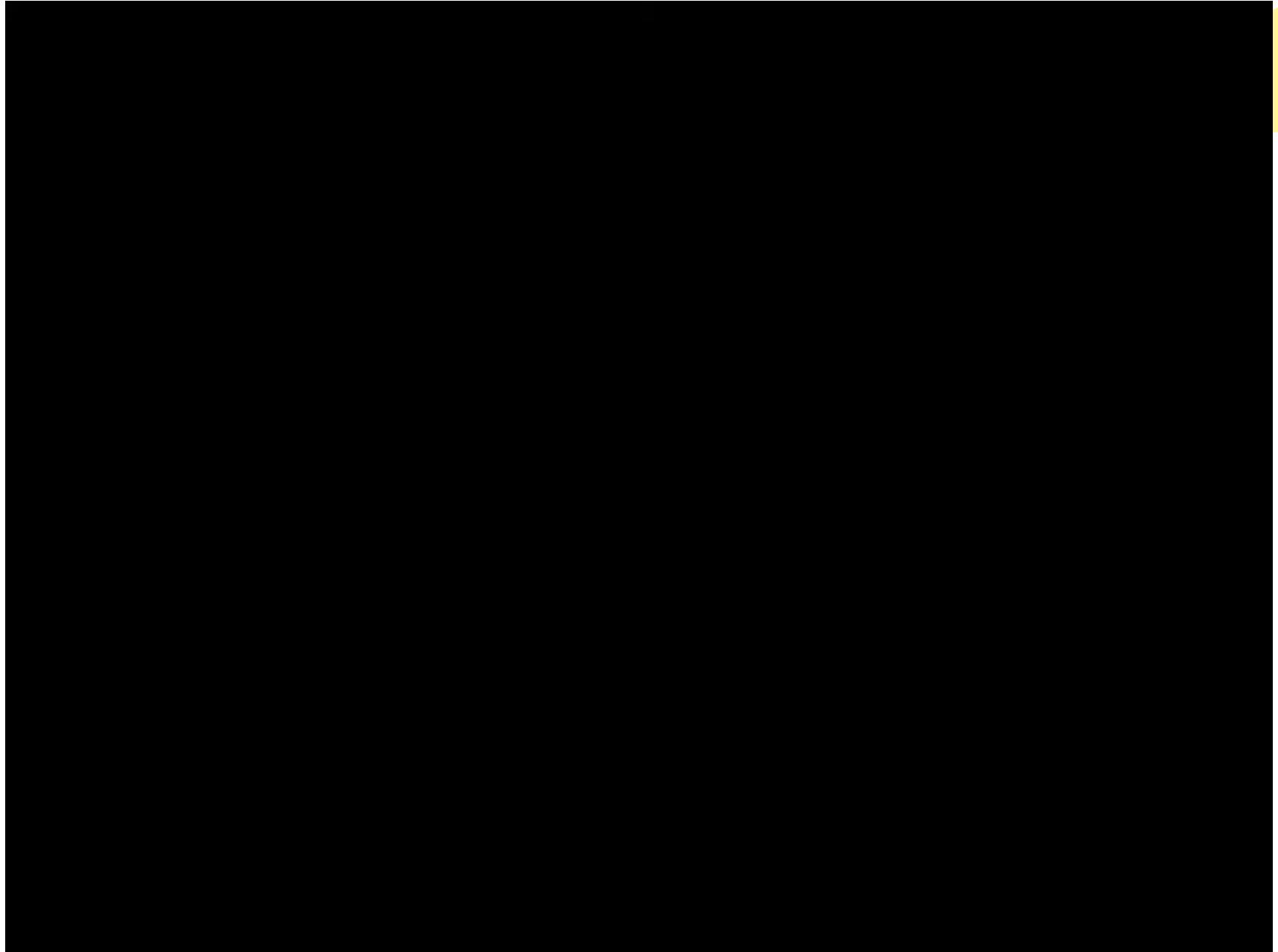
- Bases de datos transaccionales
- Sistemas de gestión de relaciones con los clientes (CRM)
- Sistemas de gestión de recursos empresariales (ERP)
- Servicios Web
- Archivos planos

- Reducen silos de información.
- Optimizados para consultas analíticas



¿QUÉ ES SNOWFLAKE?

All rights reserved. XAL DIGITAL Confidential and Trademark

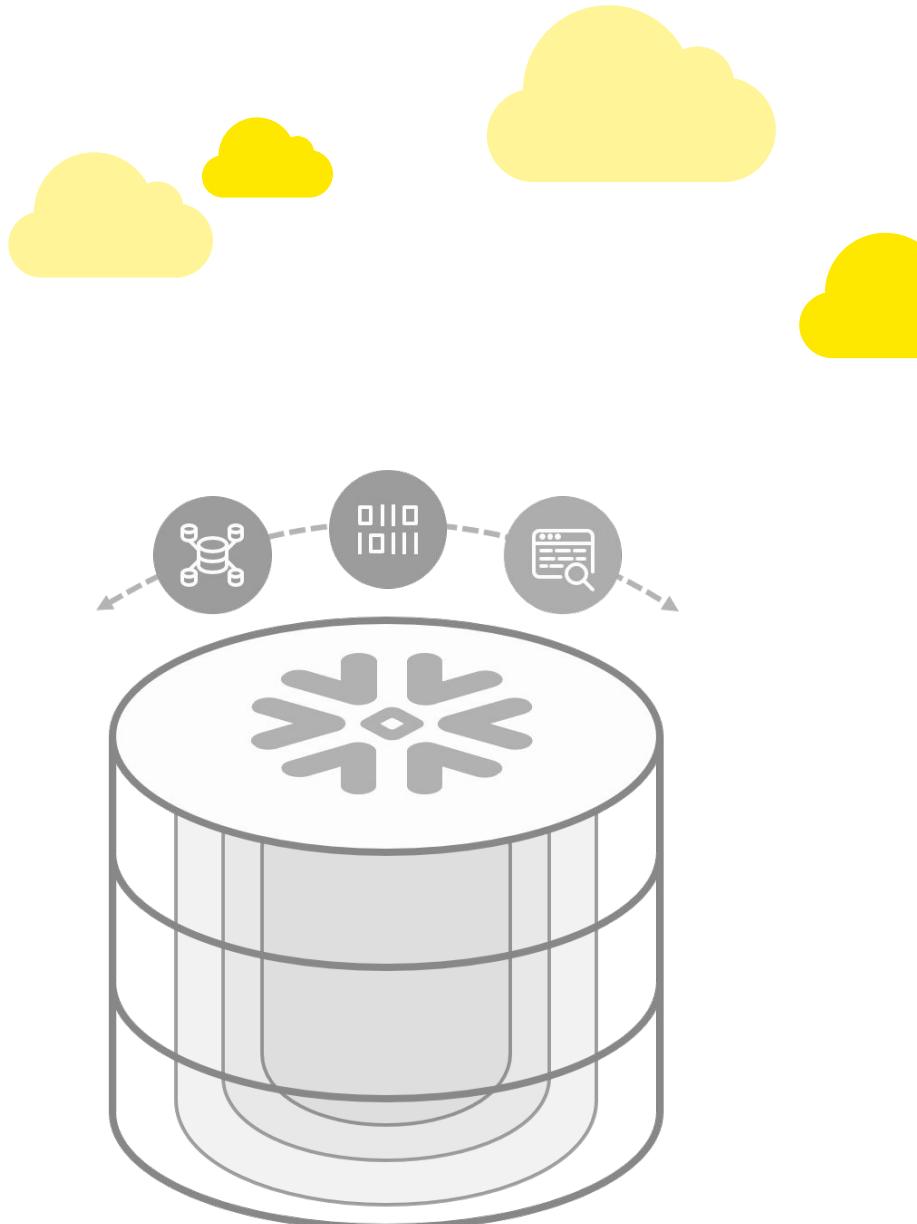


XD

¿QUÉ ES SNOWFLAKE?

Data Warehouse construido desde y para nube

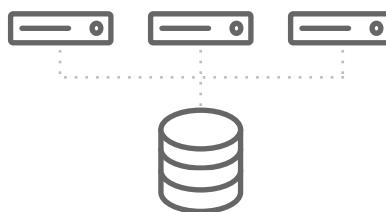
- Snowflake es un Data Warehouse aprovisionado como Software-as-a-Service (SaaS). Snowflake ofrece un Data Warehouse que es más rápido, fácil de usar y más flexible que otros Data Warehouses tradicionales.
- El Data Warehouse de Snowflake no está construido sobre bases de datos existentes ni sobre otras plataformas de Big Data como Hadoop.
- El Data Warehouse de Snowflake utiliza un nuevo motor de base de datos SQL con una arquitectura única diseñada para la nube.
- Altamente disponible y completamente administrado.
- Infraestructura aprovisionada desde nubes públicas (AWS, Azure y GCP).



UNA NUEVA ARQUITECTURA PARA DATA WAREHOUSING

Data multi-cluster compartida en la nube

Arquitecturas Tradicionales



Shared-disk

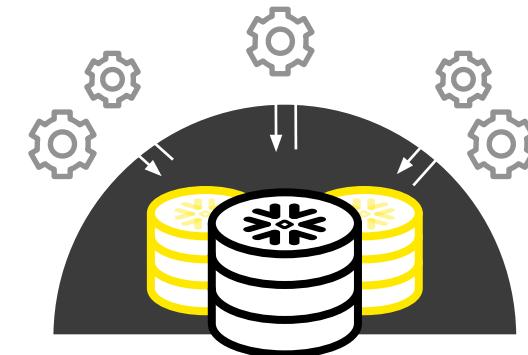
Almacenamiento Compartido
Cluster único



Shared-nothing

Descentralizado,
almacenamiento local
Cluster único

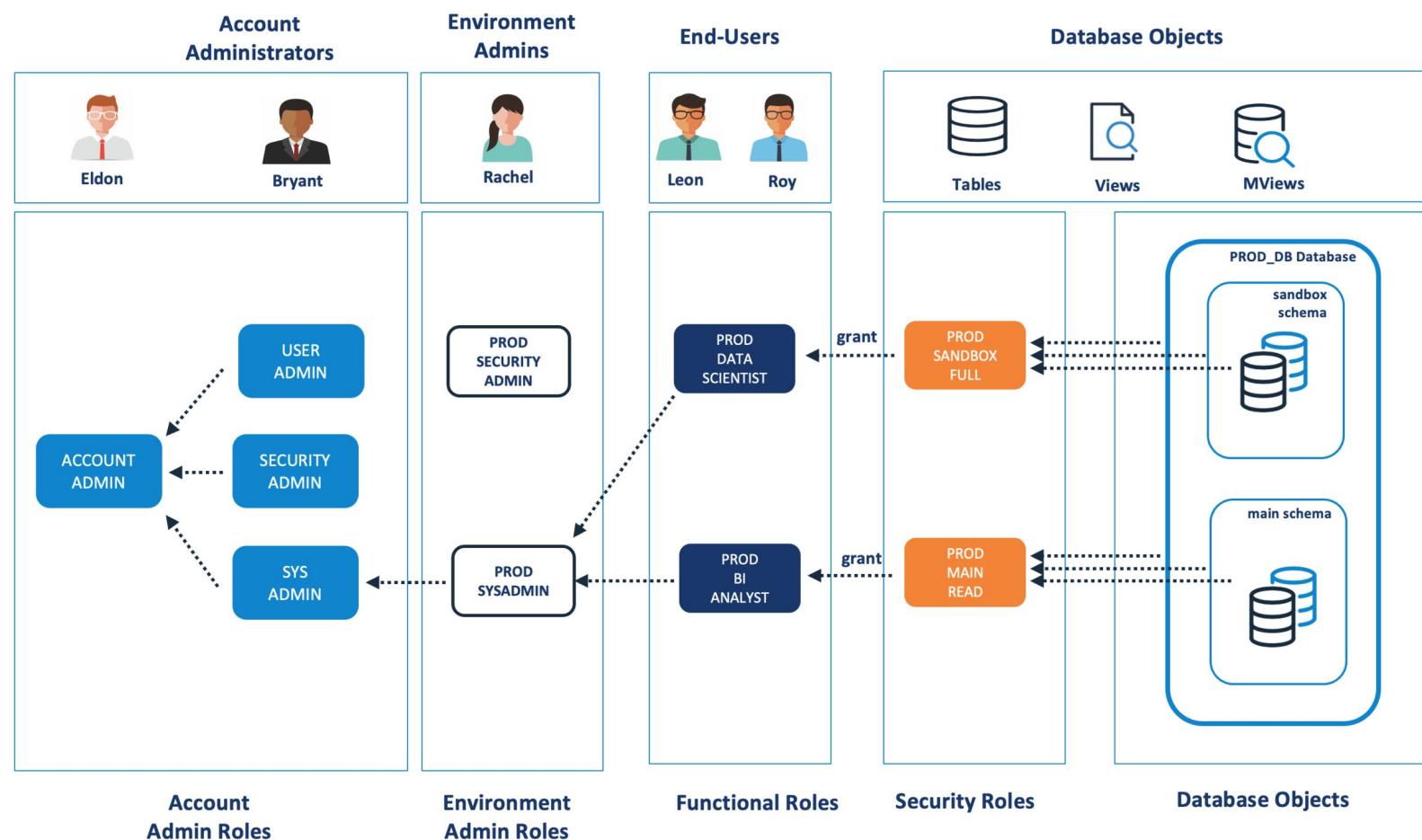
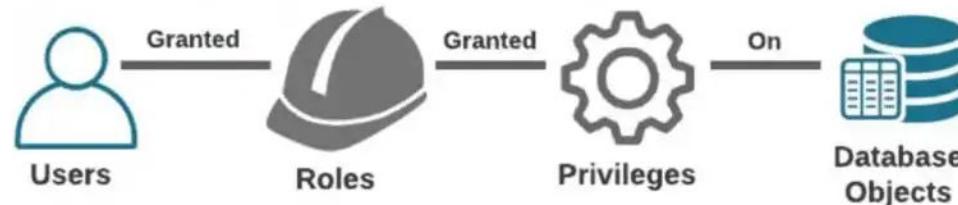
Snowflake



Multi-cluster, shared data

Almacenamiento centralizado -
scale-out
Múltiples clusters de procesamiento
independientes (Warehouses
Virtuales)

CONTROL DE ACCESO BASADO EN ROLES (RBAC)



¿POR QUÉ SNOWFLAKE ES DIFERENTE?

¿POR QUÉ SNOWFLAKE ES DIFERENTE?



Repository central
de datos



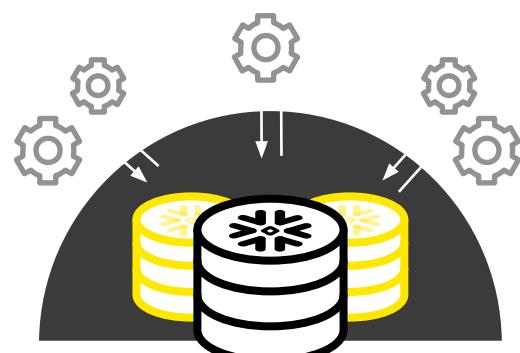
Escalamiento y
conurrencia



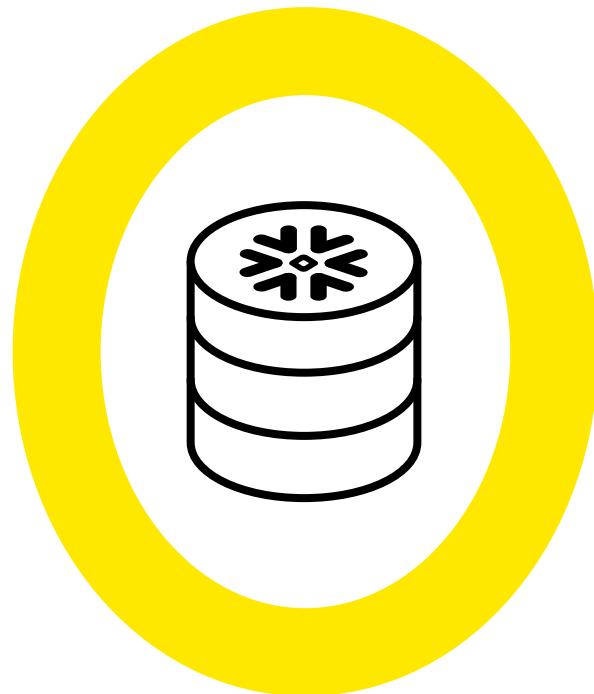
Cero
administración



Datos compartidos en
vivo y al instante



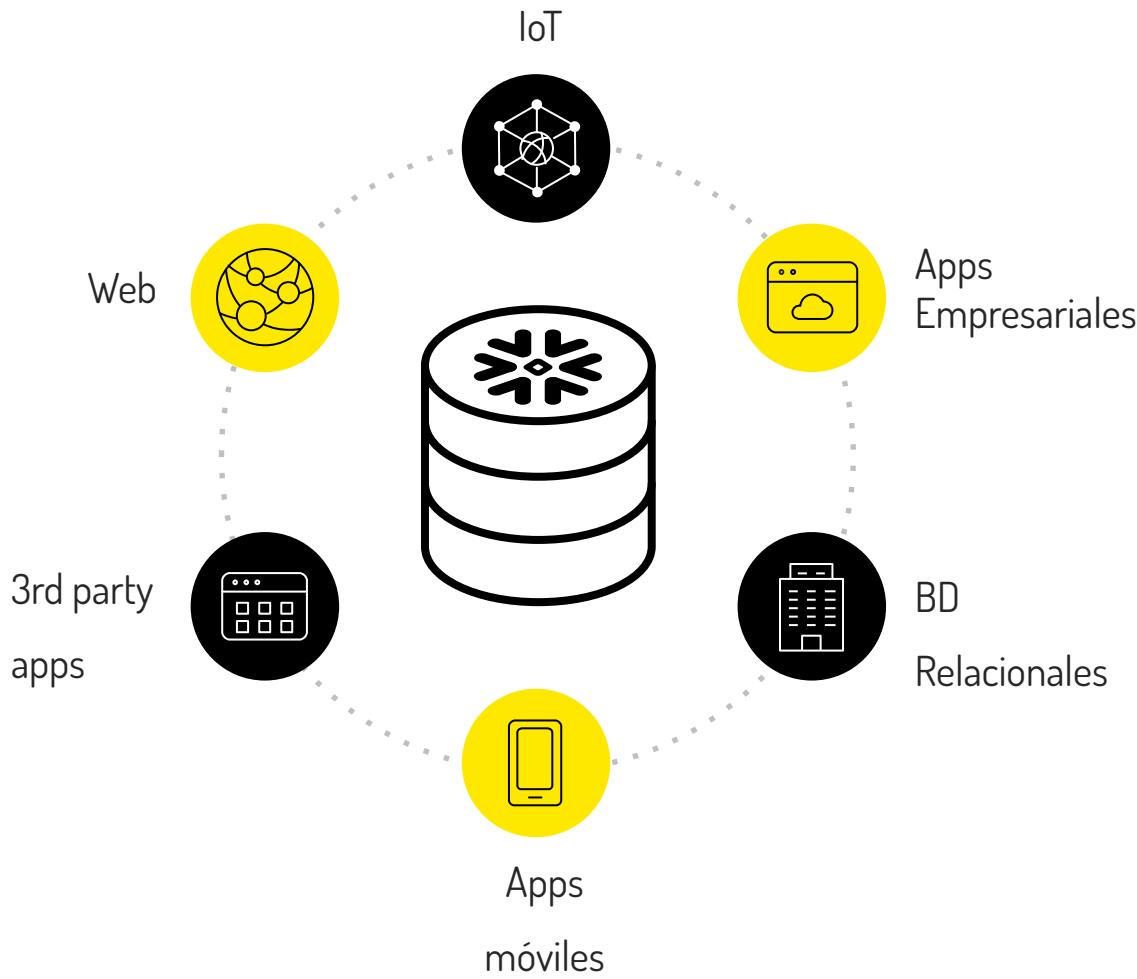
Arquitectura única:
Data multi-cluster compartida en la nube



CERO ADMINISTRACIÓN

- Carga datos y corre queries,
Snowflake se encarga del resto.
- Cero administración de costos o infraestructura.
- Seguro y altamente disponible
- Completamente administrado
- No:
 - Indices
 - Particionamiento
 - Vacuuming

TODA TU INFORMACIÓN



- Archivos estructurados y semiestructurados (JSON, XML, Avro)
- Almacenamiento centralizado, accesible por cualquier usuario y aplicación
- Escalamiento Multi-petabyte

TODOS TUS USUARIOS



Aplicaciones para usos múltiples

- Hasta 200x más rápido que soluciones no diseñadas para la nube.
- Mantenimiento consistente de SLA – Recursos crecen y decrecen automáticamente
- Carga de datos no impacta desempeño de consultas
- Acceso simultáneo de diversos grupos de usuarios no implica degradación en desempeño.
- Soporte ilimitado de usuarios simultáneos.

CACHING ADAPTATIVO

Metadatos

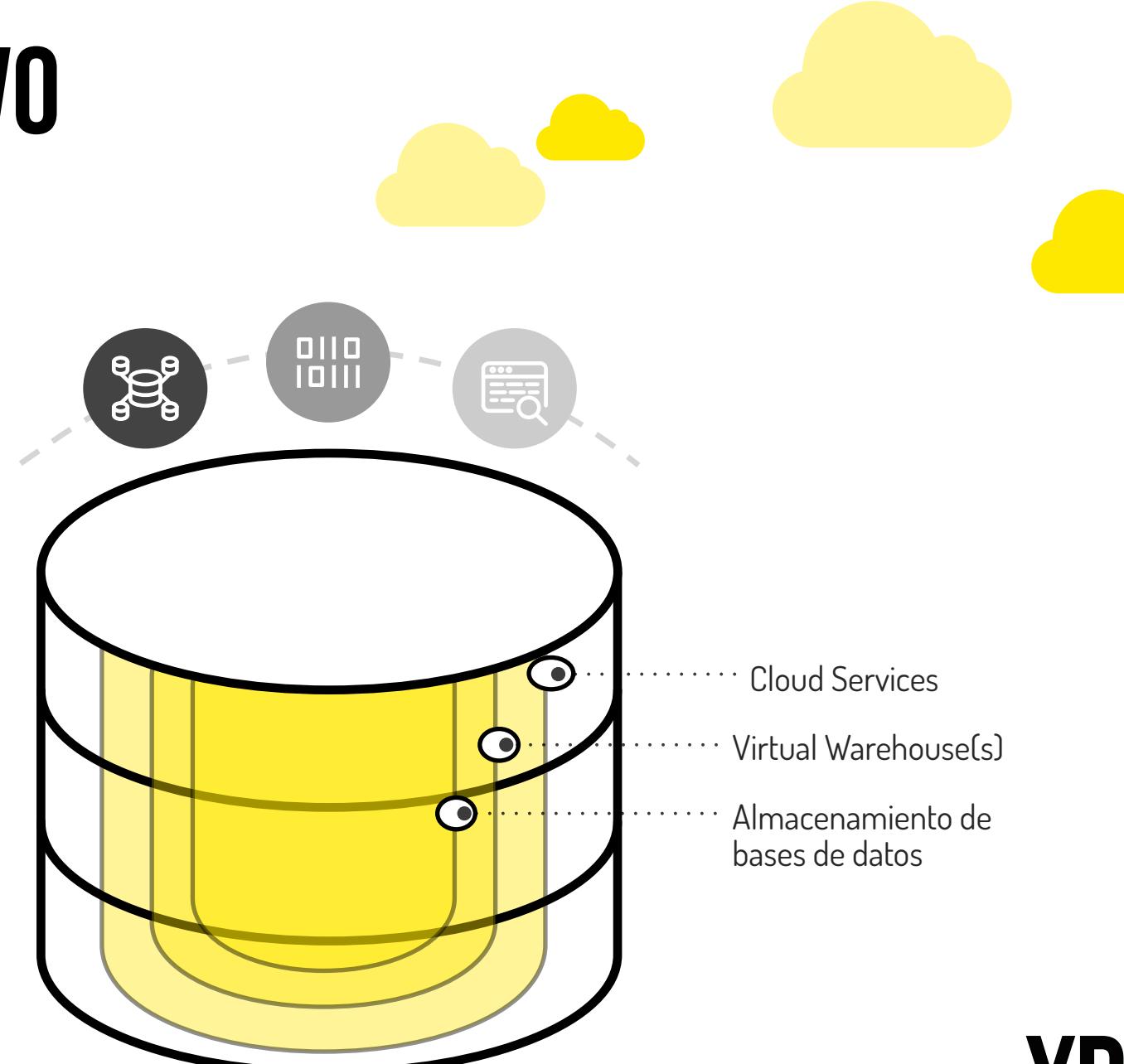
Metadatos en caché para rápido acceso durante planeación y optimización de consultas.

Datos

Set de datos activo, transparentemente almacenados en caché en Warehouses virtuales SSD.

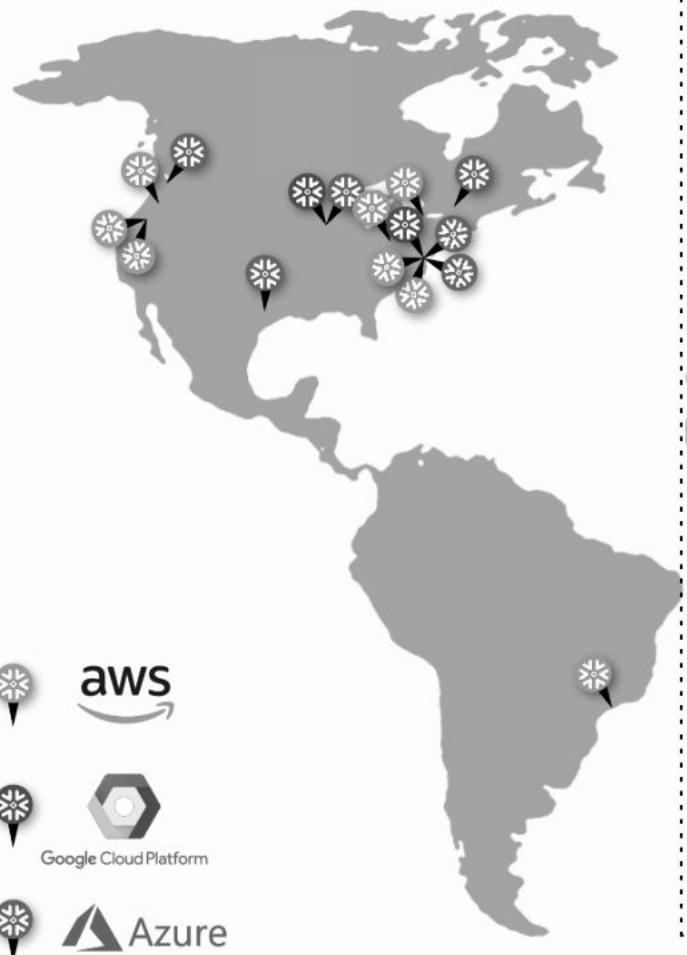
Resultados de Consultas

Resultados guardados en caché para reuso sin utilizar capacidad de reprocesamiento.
(e.g., consultas de dashboards estáticos)



REGIONES DE SNOWFLAKE

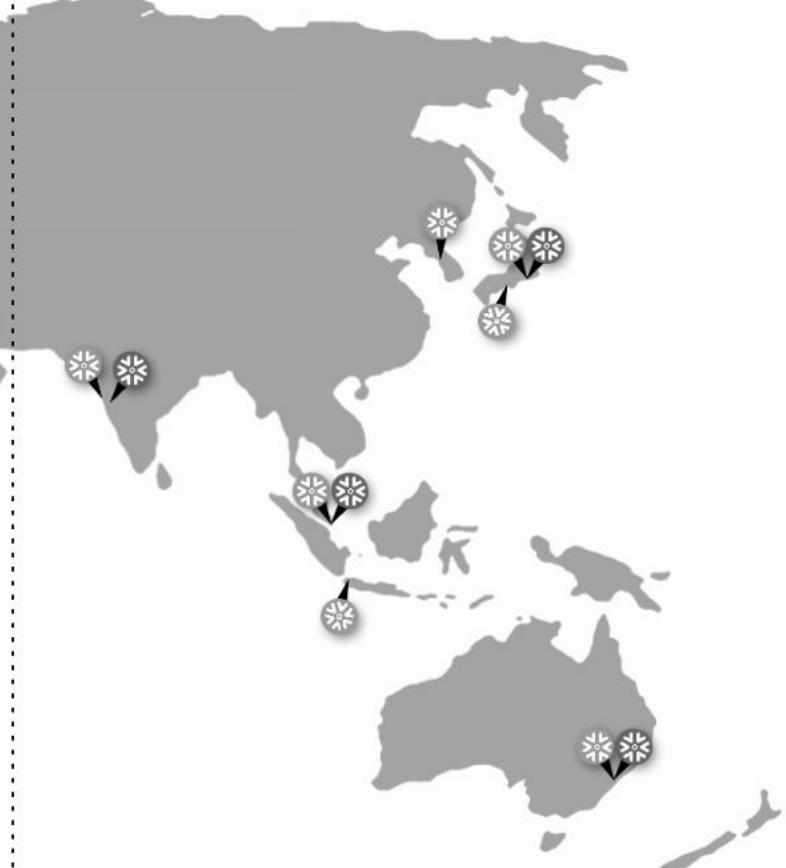
North & South America



Europe & Middle East



Asia Pacific





PLAY STUDIOS



coinbase



T.RowePrice®



GARANTIZADO POR MÁS DE 1000 COMPAÑÍAS TRANSNACIONALES



MADISON REED®



acoordantmedia

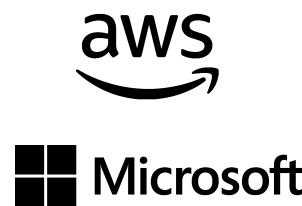


LEANPLUM



XD

Platform



BI/Analytics



ETL



ECOSISTEMA DE DATOS E INTEGRACIONES

Data Science

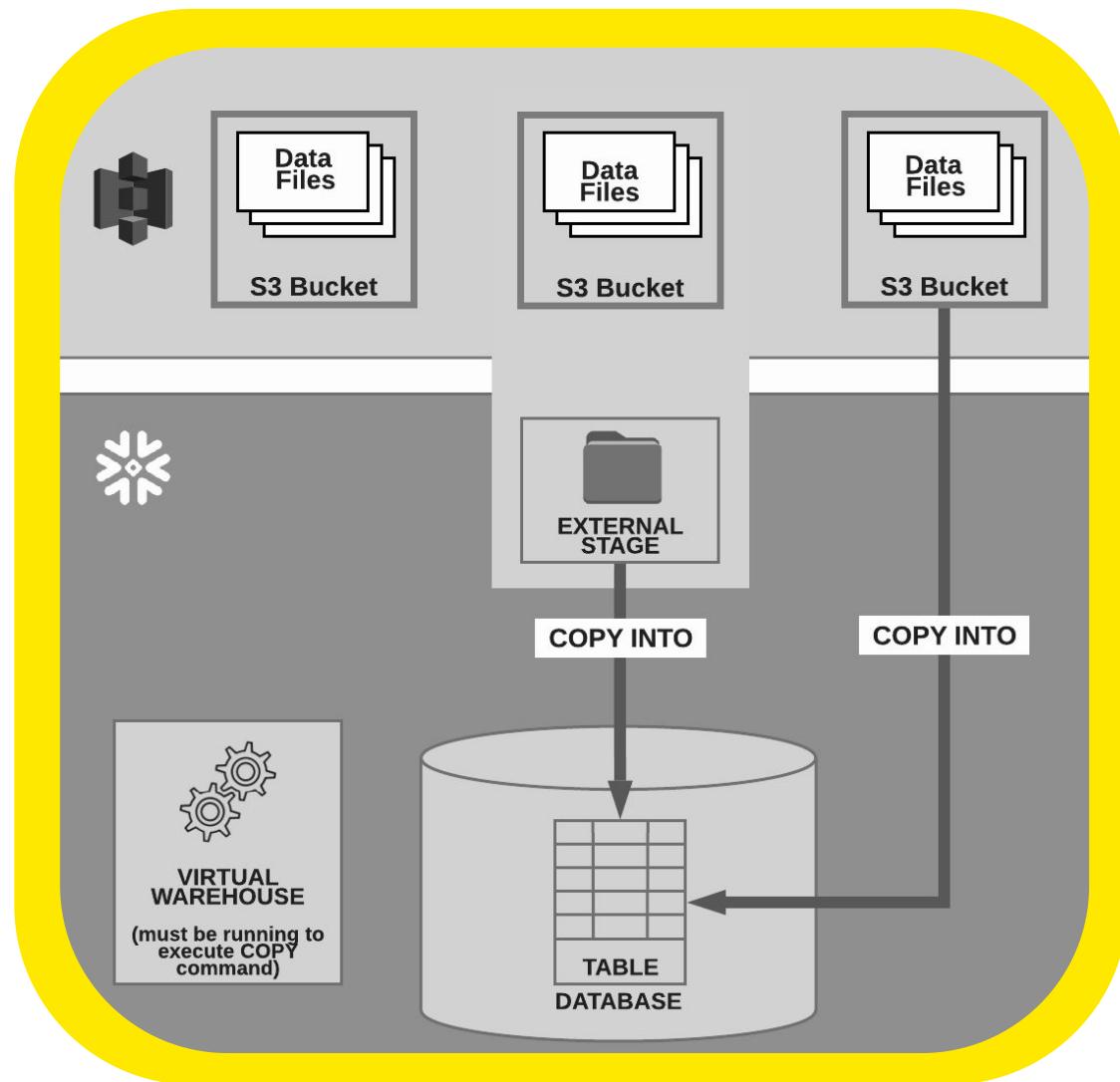


Services

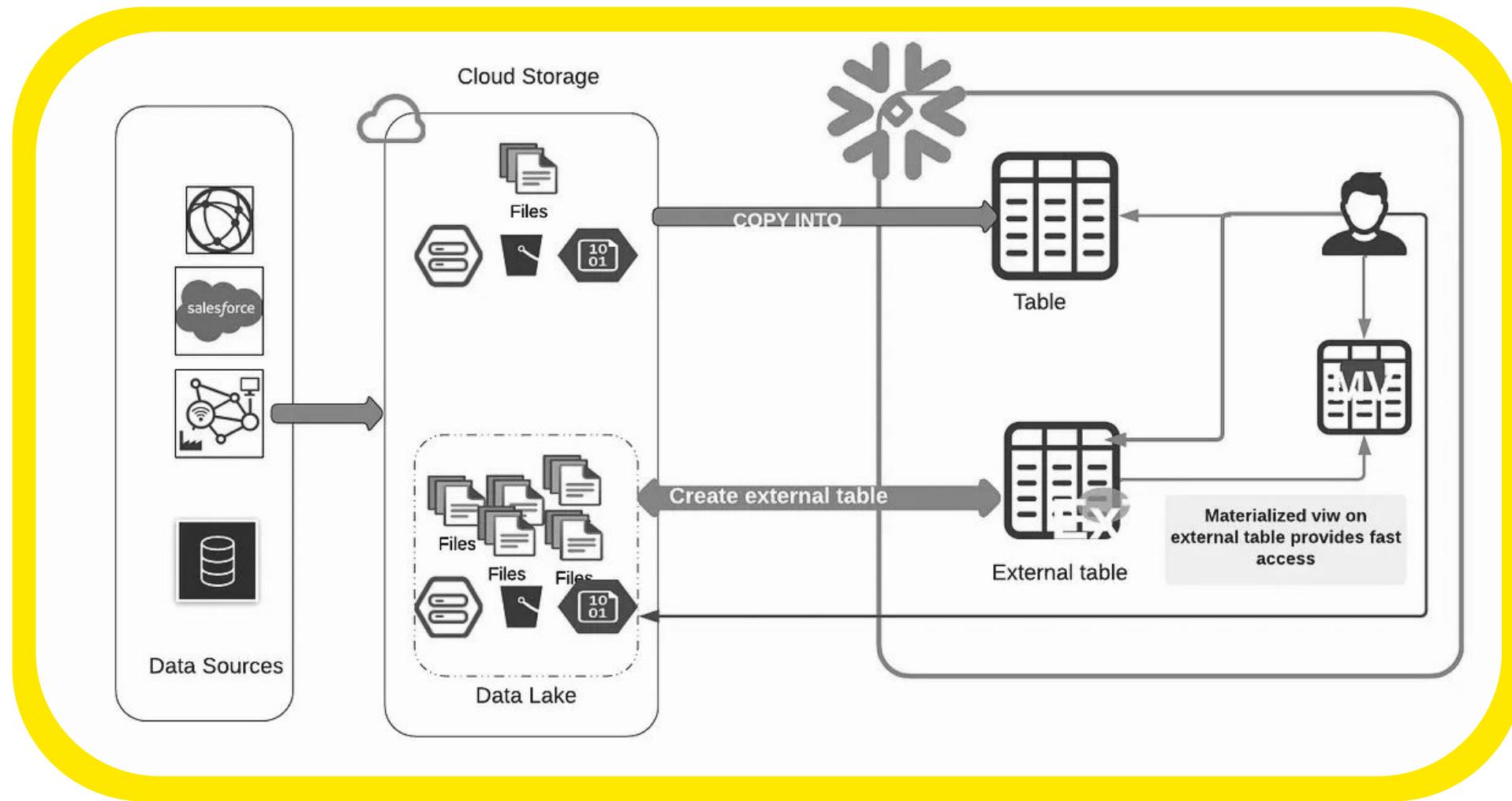


MÉTODOS DE INGESTA

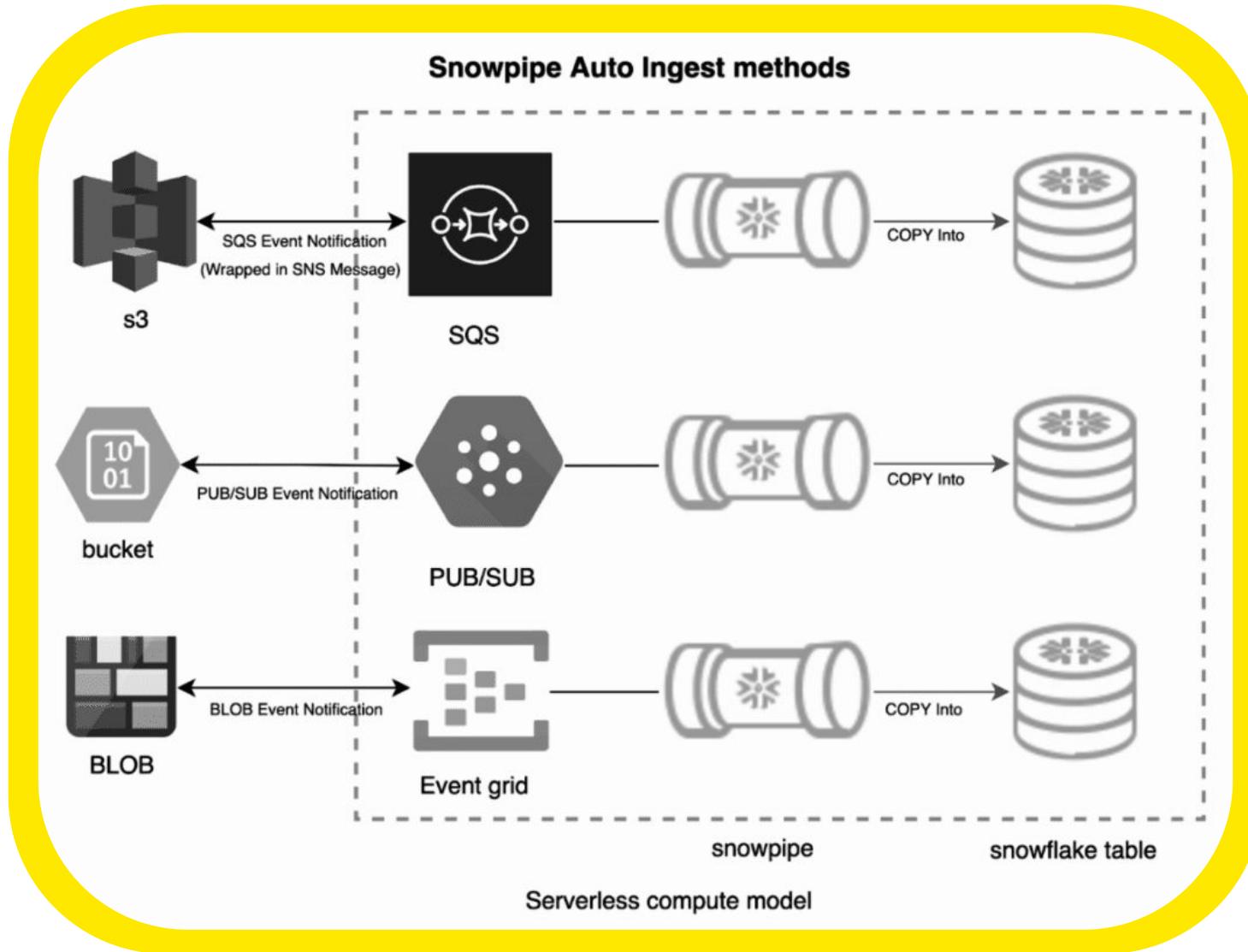
COPIADO MANUAL (COPY INTO)

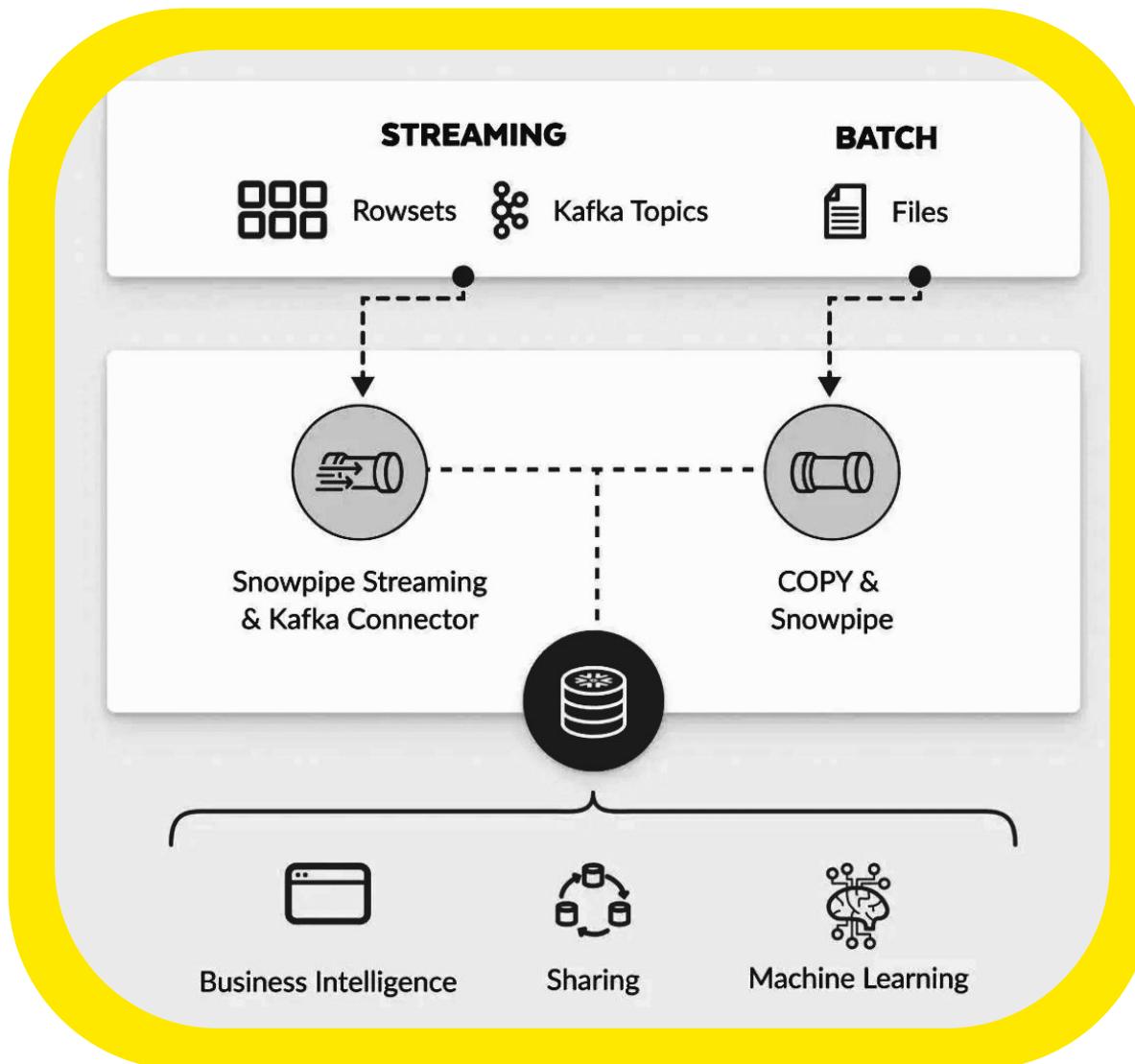


TABLAS EXTERNAS

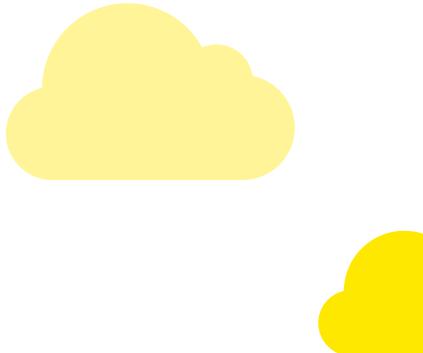


SNOWPIPE





BATCH VS STREAMING



CASOS DE USO

SNOWPARK

SNOWPARK API

Transformaciones en procesamiento distribuido a través de Dataframes.

User-Defined Functions (UDFs)

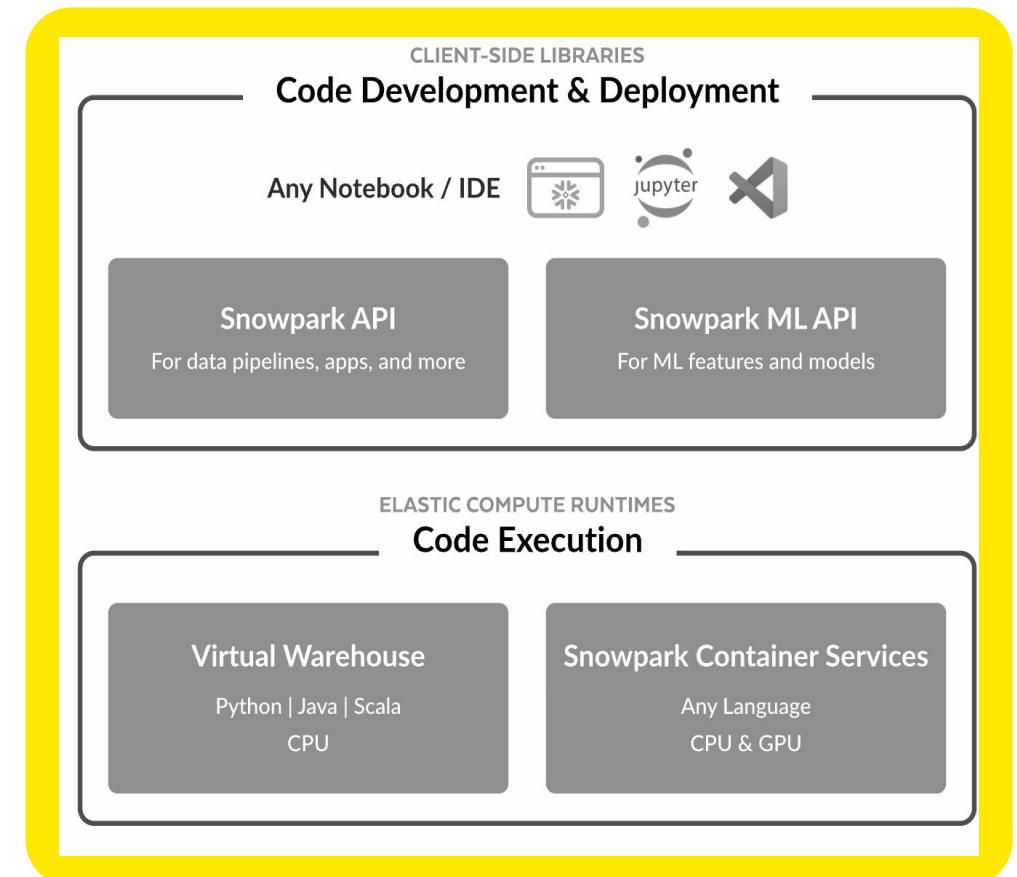
Funciones definidas por usuario a través de lenguajes de programación como Python, Scala y Java

Integraciones con Terceros

Interacción con IDEs como VS Code, Jupyter Notebooks, Contenedores de Docker, o directamente desde la consola de Snowflake. Todo completamente administrado.

Rendimiento y Costos Optimizados

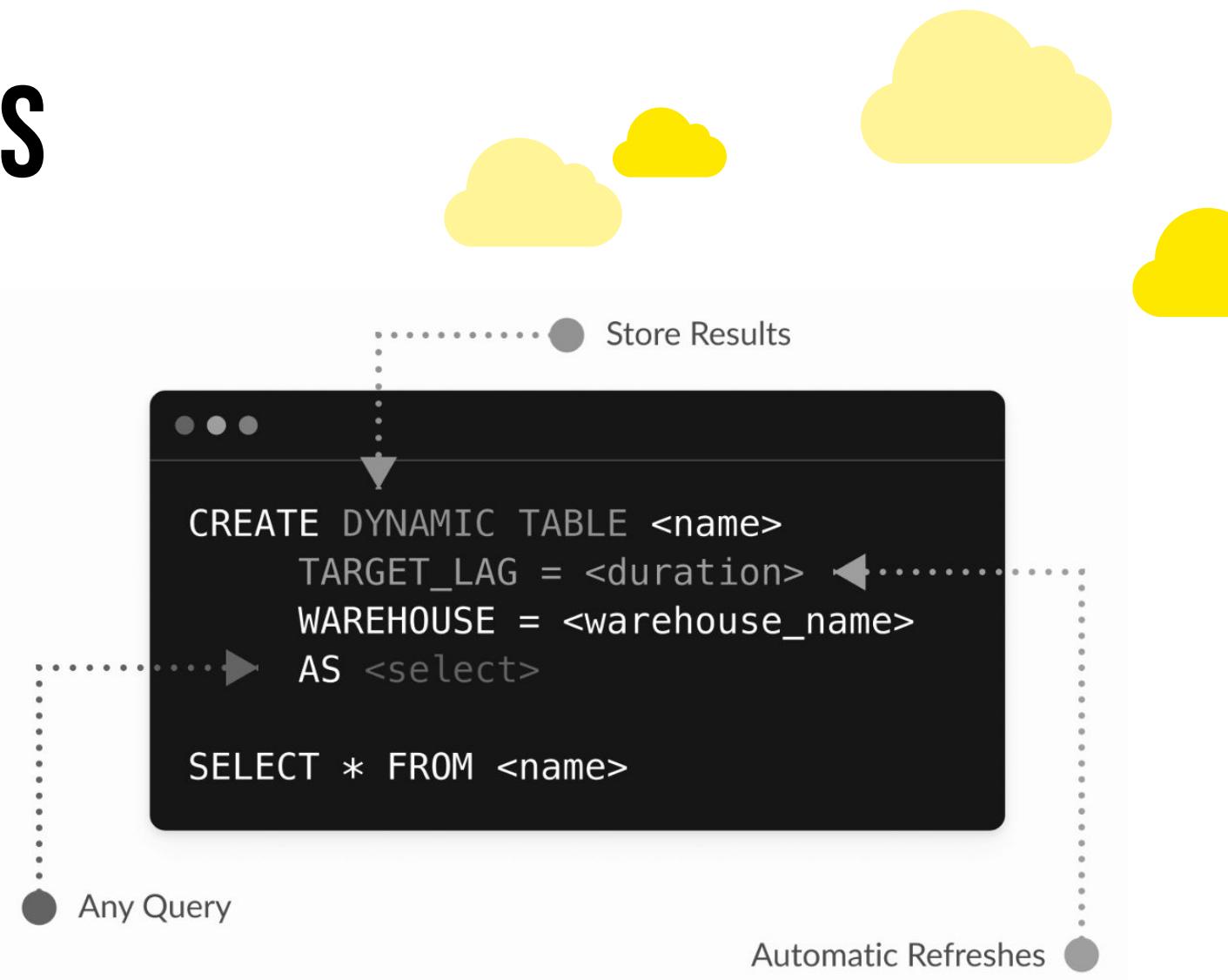
Rendimiento 3.5x más rápido y costos 34% inferiores en comparación con las soluciones gestionadas de Spark [1].



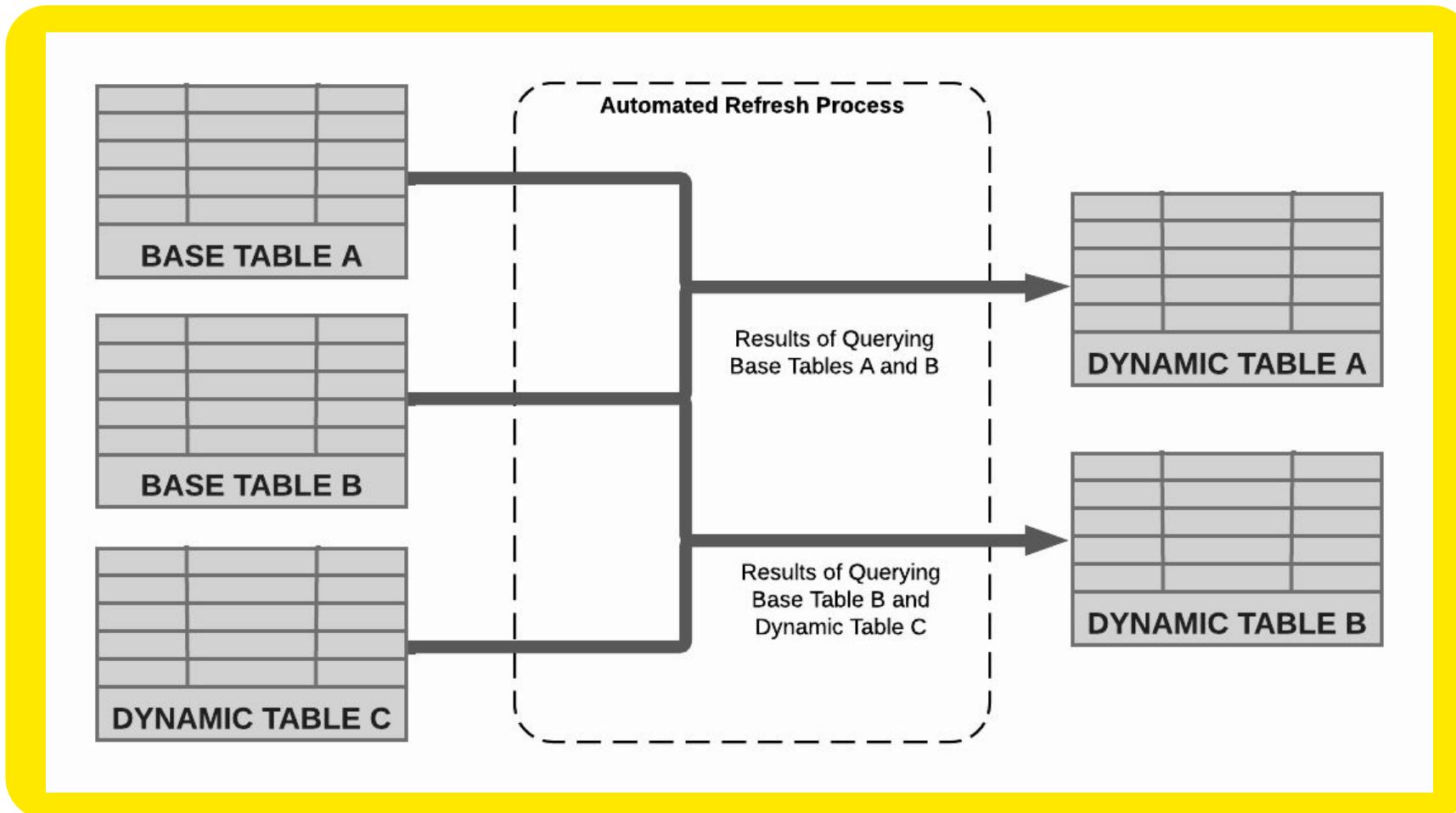
DYNAMIC TABLES

AJUSTA LA LATENCIA CAMBIANDO UN SOLO PARÁMETRO

- Utilizar SQL o Python para definir transformaciones de datos de forma declarativa.
- Snowflake gestionará las dependencias y materializa automáticamente los resultados en función de tus objetivos de actualización.
- Solo usa los datos que han cambiado desde la última actualización para hacer que los grandes volúmenes de datos y los flujos complejos sean más sencillos y rentables.

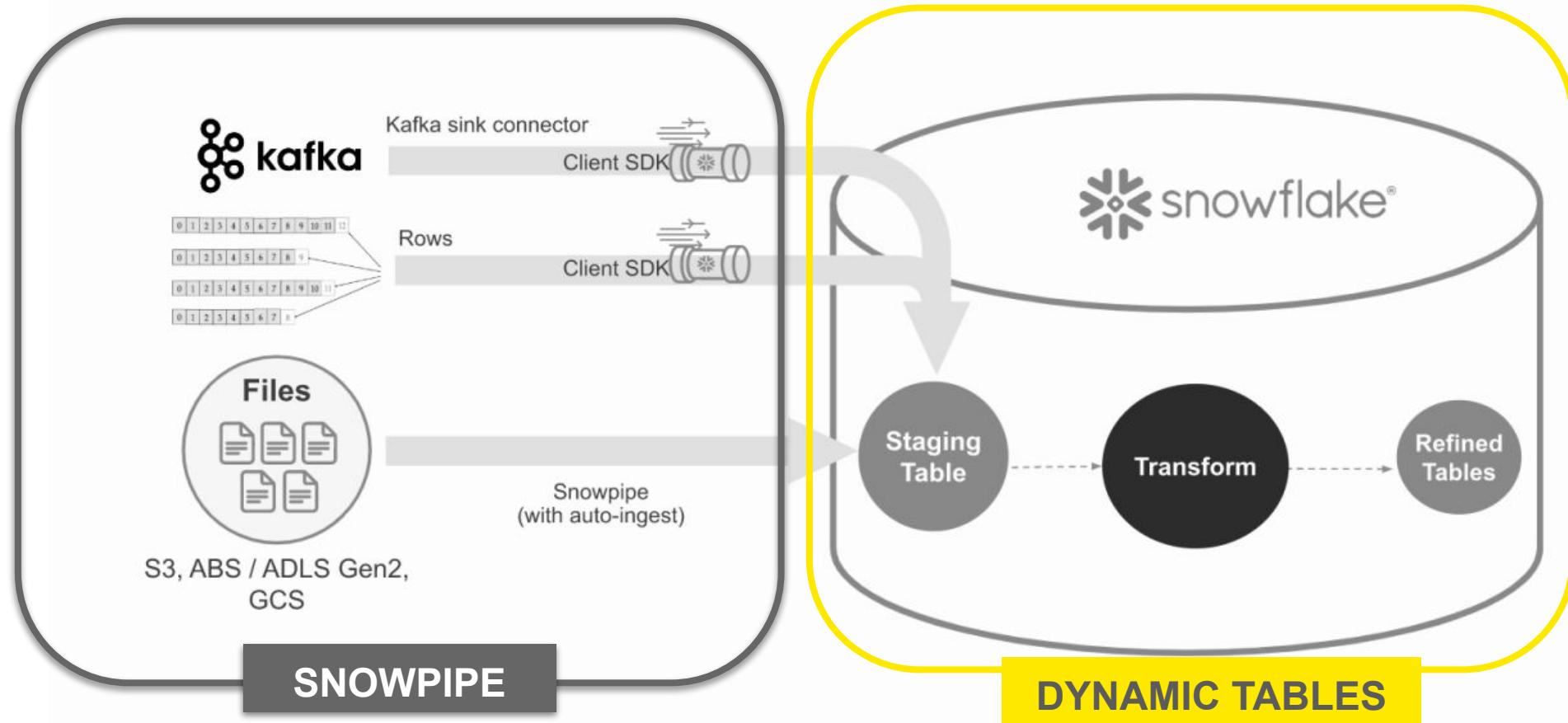


DYNAMIC TABLES - EJEMPLO

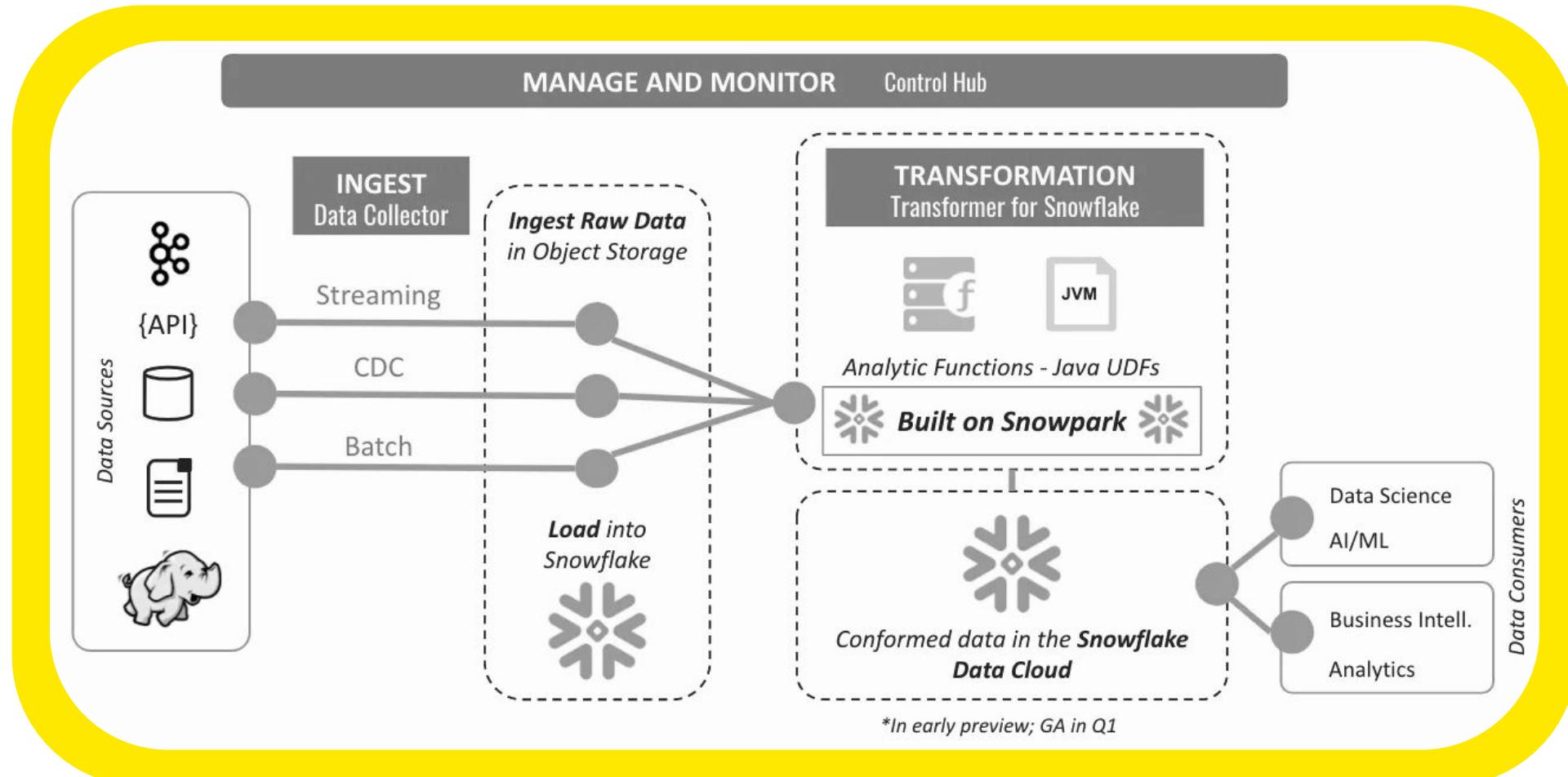


SNOWPIPE + DYNAMIC TABLES

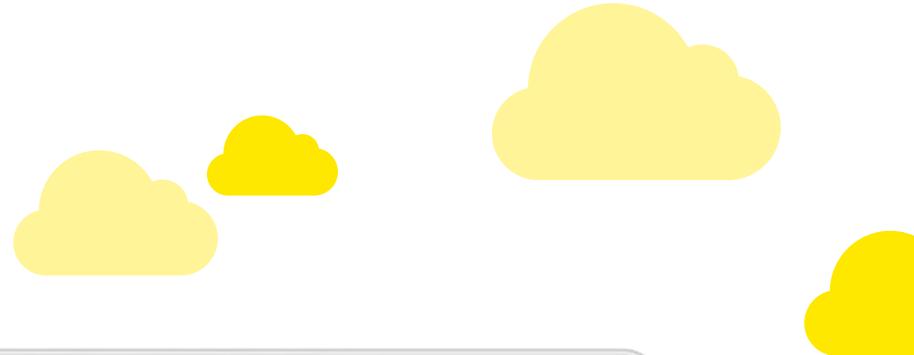
Arquitectura en tiempo real



DATA PIPELINES CON SNOWFLAKE

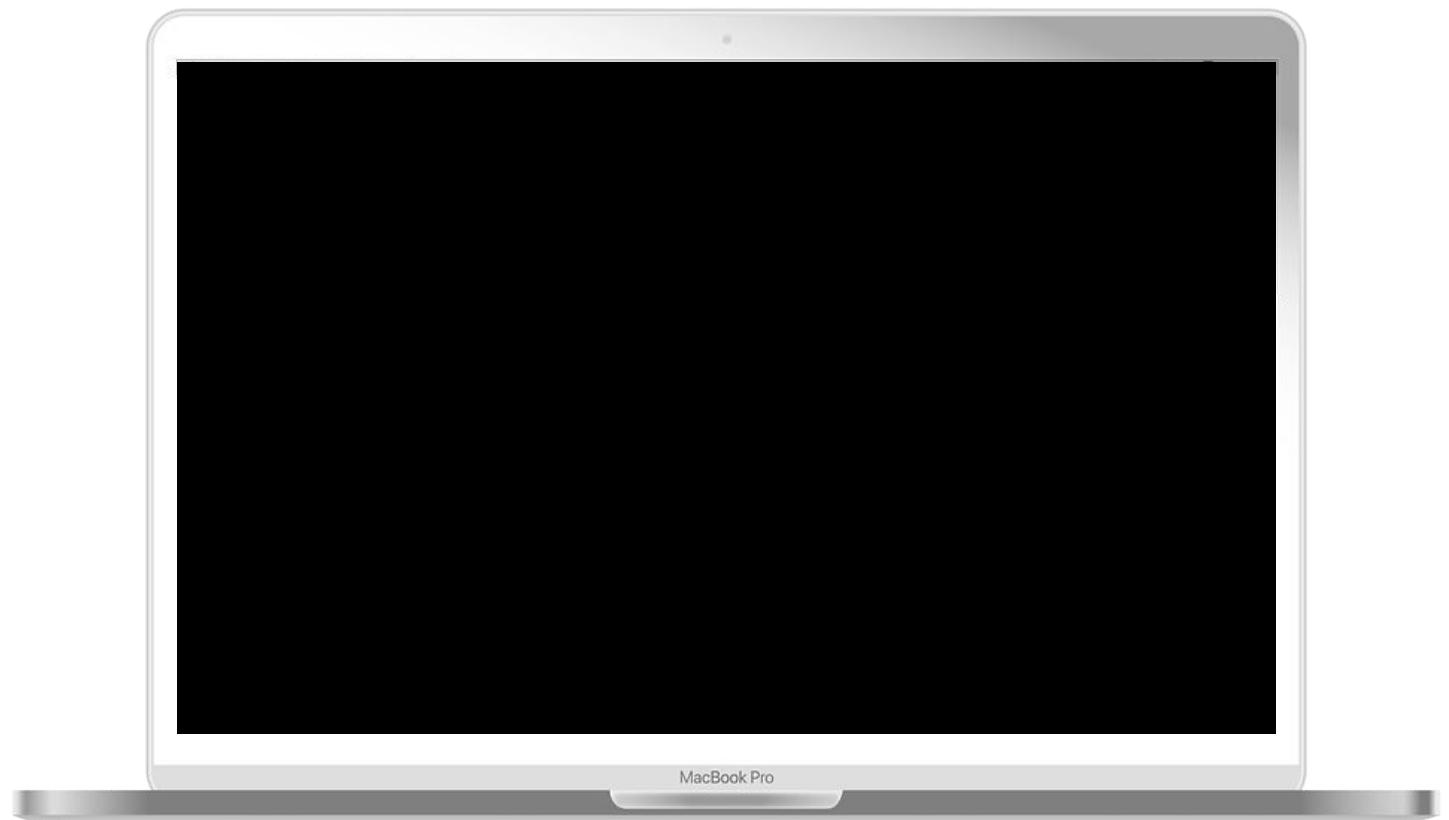


STREAMLIT

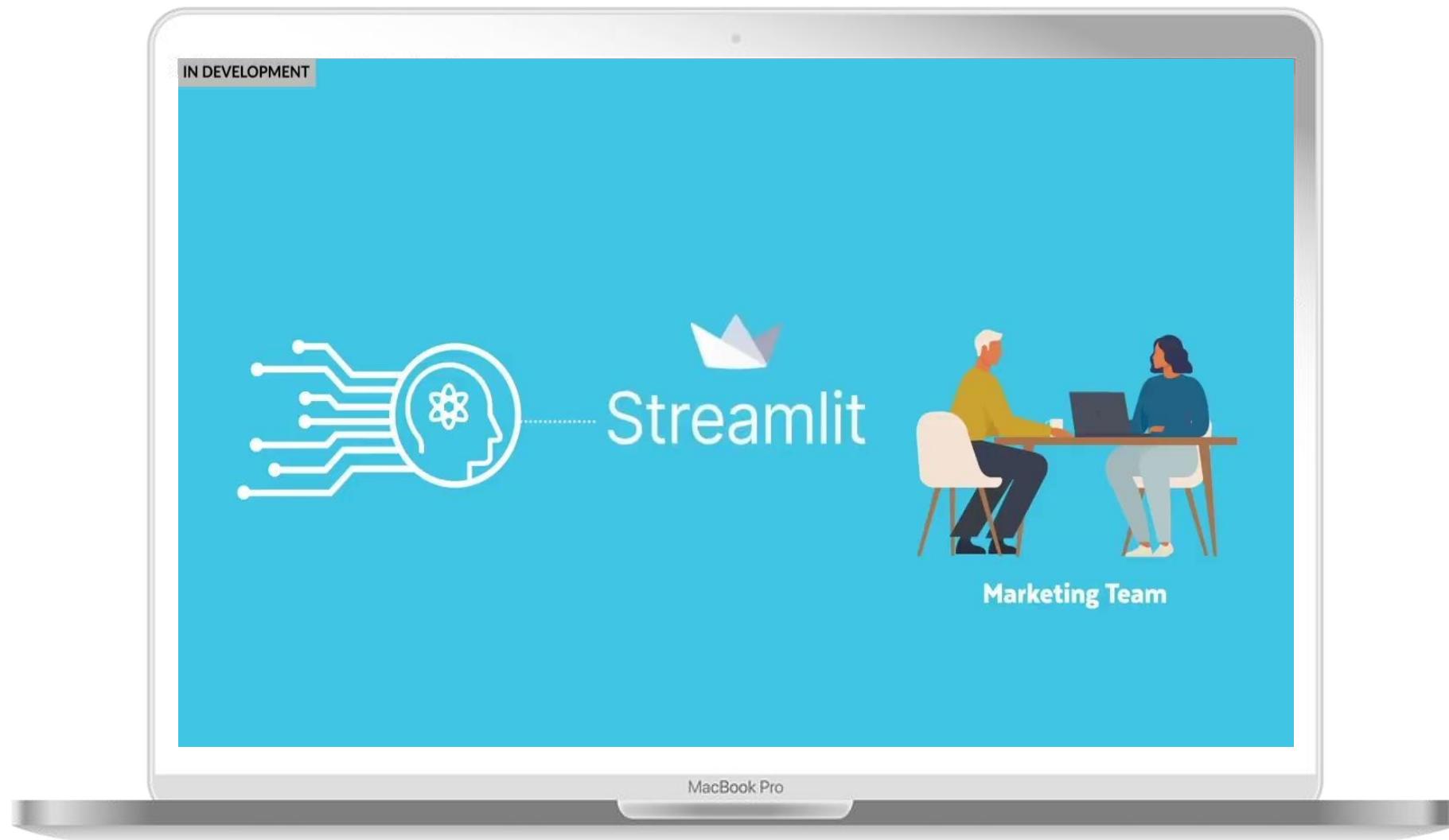


¿Qué es Streamlit?

- Framework web de código bajo en crecimiento rápido para Python, te permite construir aplicaciones web interactivas y basadas en datos con un código mínimo.
- Científicos de datos y desarrolladores Python ahora pueden combinar librerías Python código abierto, componentes de Streamlit, y escalabilidad, rendimiento y seguridad de la plataforma Snowflake.



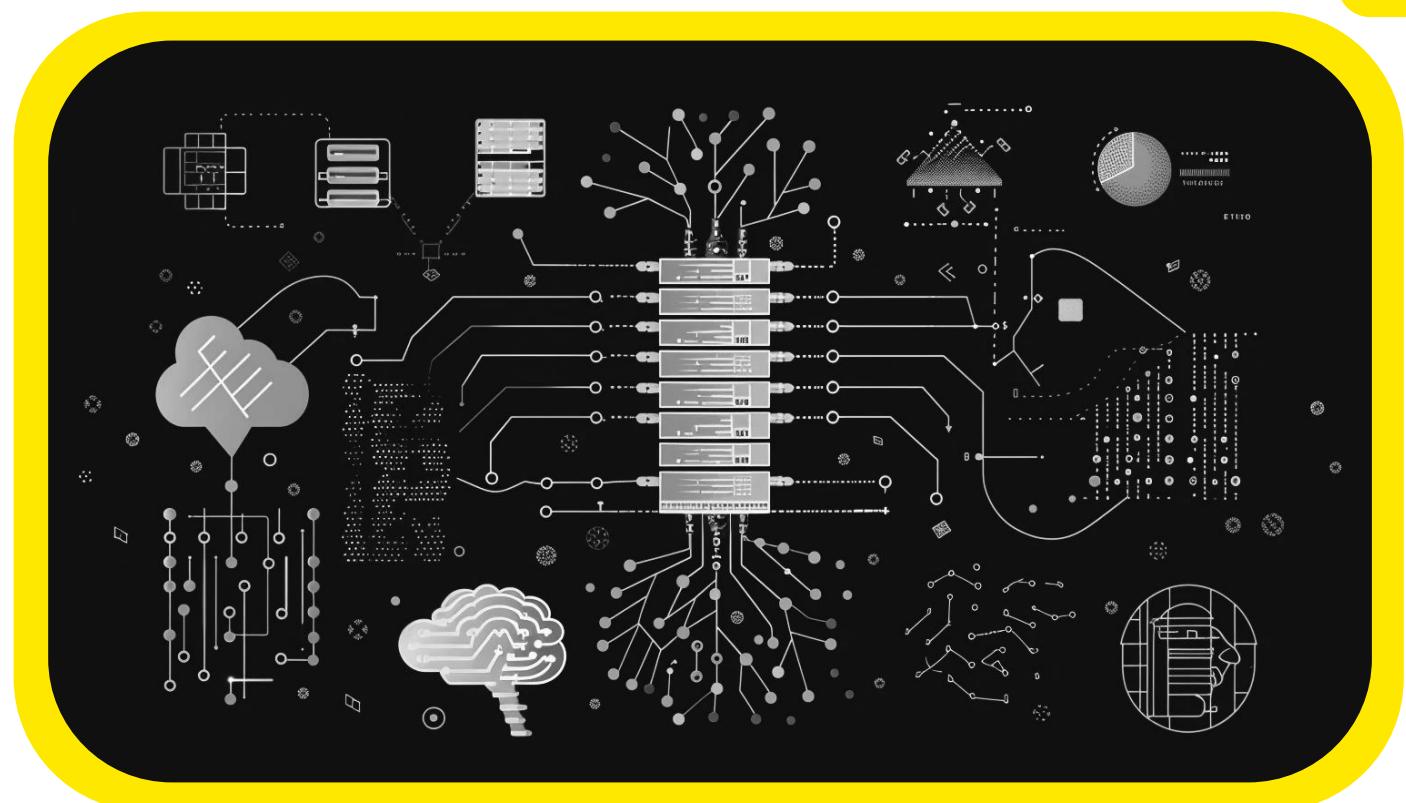
STREAMLIT DEMO



LARGE LANGUAGE MODELS (LLMS)

Modelos de lenguaje de gran tamaño

- Modelos de aprendizaje profundo muy grandes que se pre entrena n con grandes cantidades de datos - Internet público
- Deep Learning - Redes Neuronales
- Los transformadores LLM son capaces de entrenarse sin supervisión, aunque una explicación más precisa es que los transformadores llevan a cabo un autoaprendizaje.
- Transformadores aprenden a entender la gramática, los idiomas y los conocimientos básicos.
- GPT 3.5, GPT 4.0, Llama 2, Google BERT



SNOWFLAKE CORTEX



XALDIGITAL

Empowering Data