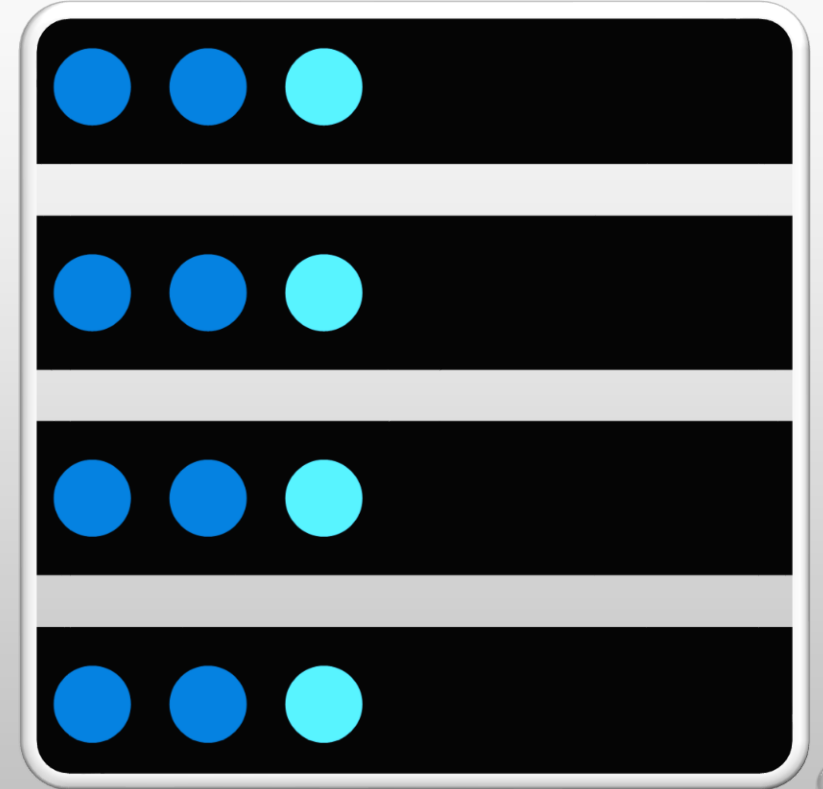


ALMACENAMIENTO Y PROCESAMIENTO DE DATOS EN AZURE



OBJETIVOS



- Describir las opciones de procesamiento de datos para realizar análisis en Azure.

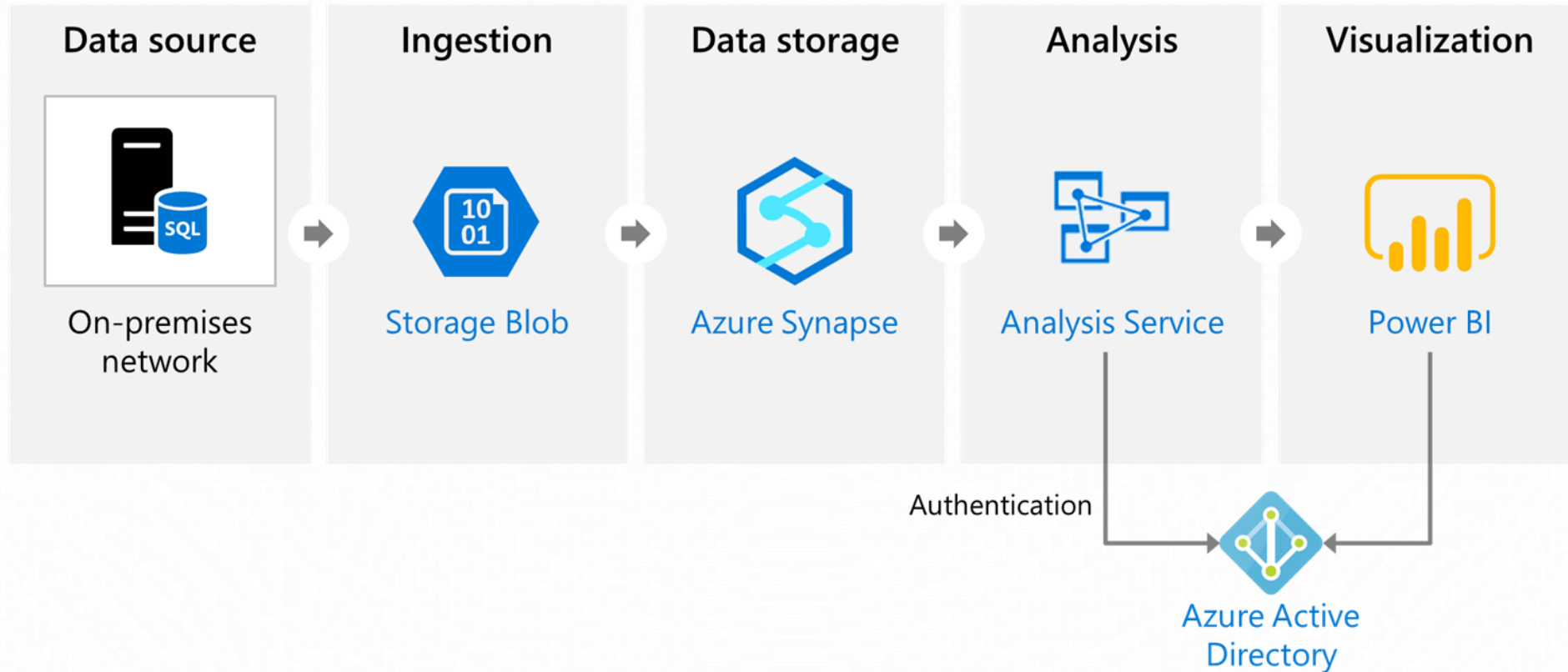


- Explore Azure Synapse Analytics.

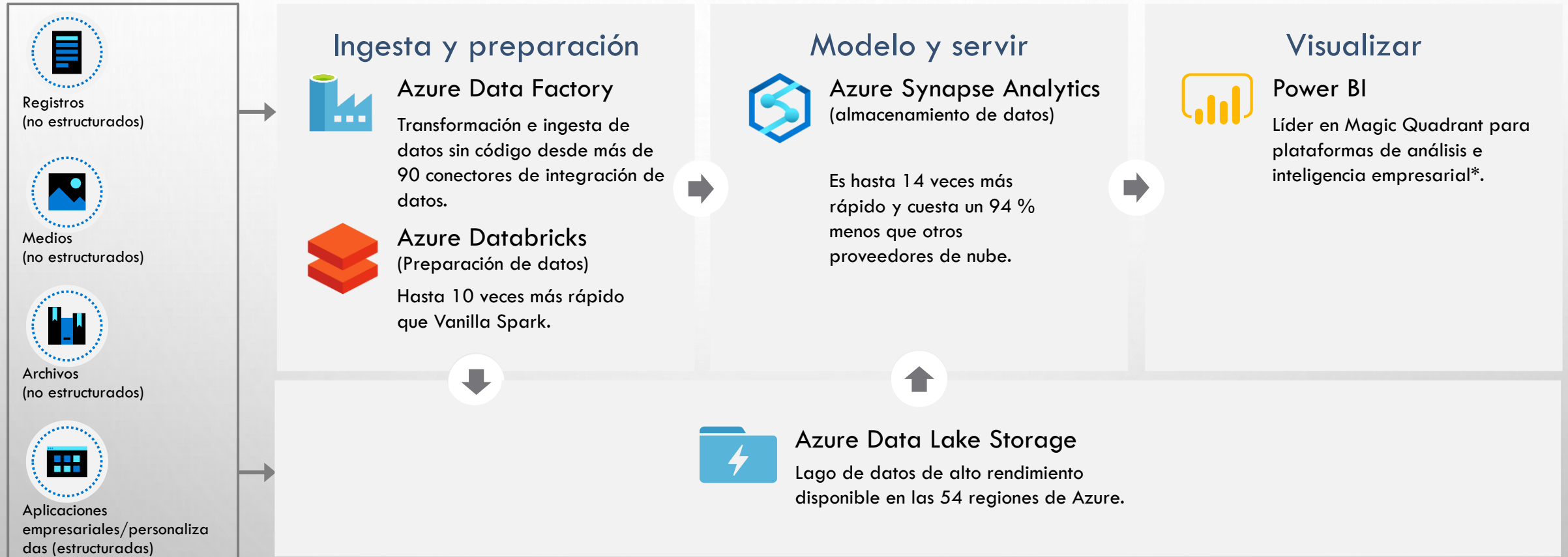
¿QUÉ ES EL ALMACENAMIENTO DE DATOS MODERNO?

- Un almacén de datos moderno puede contener una combinación de datos relacionales y no relacionales, incluidos archivos, transmisiones de redes sociales y datos de sensores de internet de las cosas (IOT).

PROCESO DE UN ALMACENAMIENTO MODERNO



COMBINAR EL PROCESAMIENTO POR LOTES Y EN FLUJO



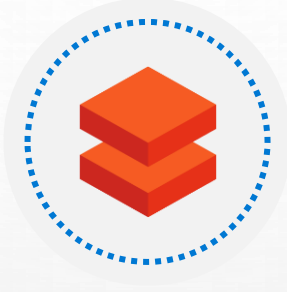
AZURE PROPORCIONA ALGUNAS SOLUCIONES DE ALMACENAMIENTO DE DATOS, COMO :



Azure Data
Factory



Data
Lake Store



Azure Databricks



Azure Synapse
Análisis



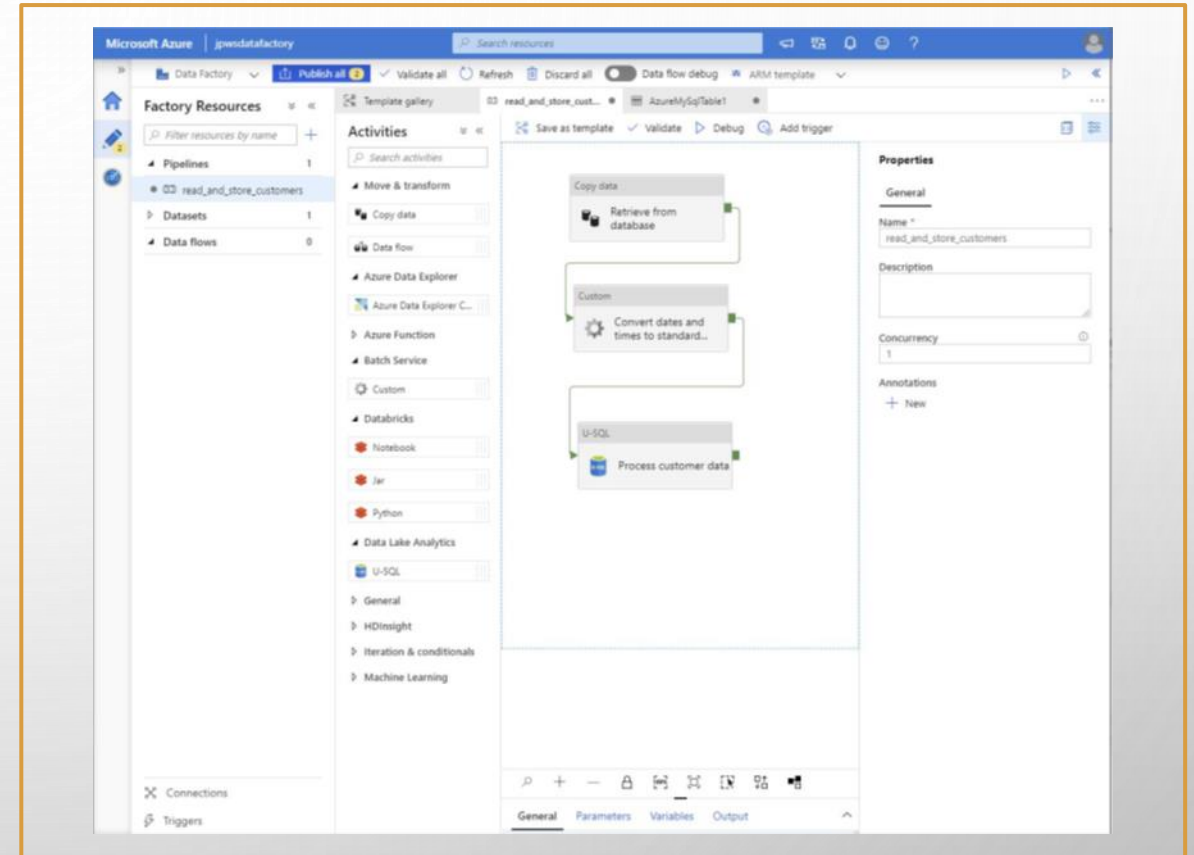
Azure HDInsight

El siguiente vídeo explica la interacción entre cada una de estas soluciones de Azure:

microsoft.com/videooplayer/embed/RE4A3RR

¿QUÉ ES AZURE DATA LAKE STORAGE?

- ✓ Organiza los datos en directorios para mejorar el acceso a los archivos.
- ✓ Es compatible con el sistema de archivos distribuido HADOOP.
- ✓ Un repositorio de datos para su almacenamiento de datos moderno.
- ✓ Admite permisos POSIX y RBAC.



¿QUÉ ES AZURE DATA LAKE STORAGE ?

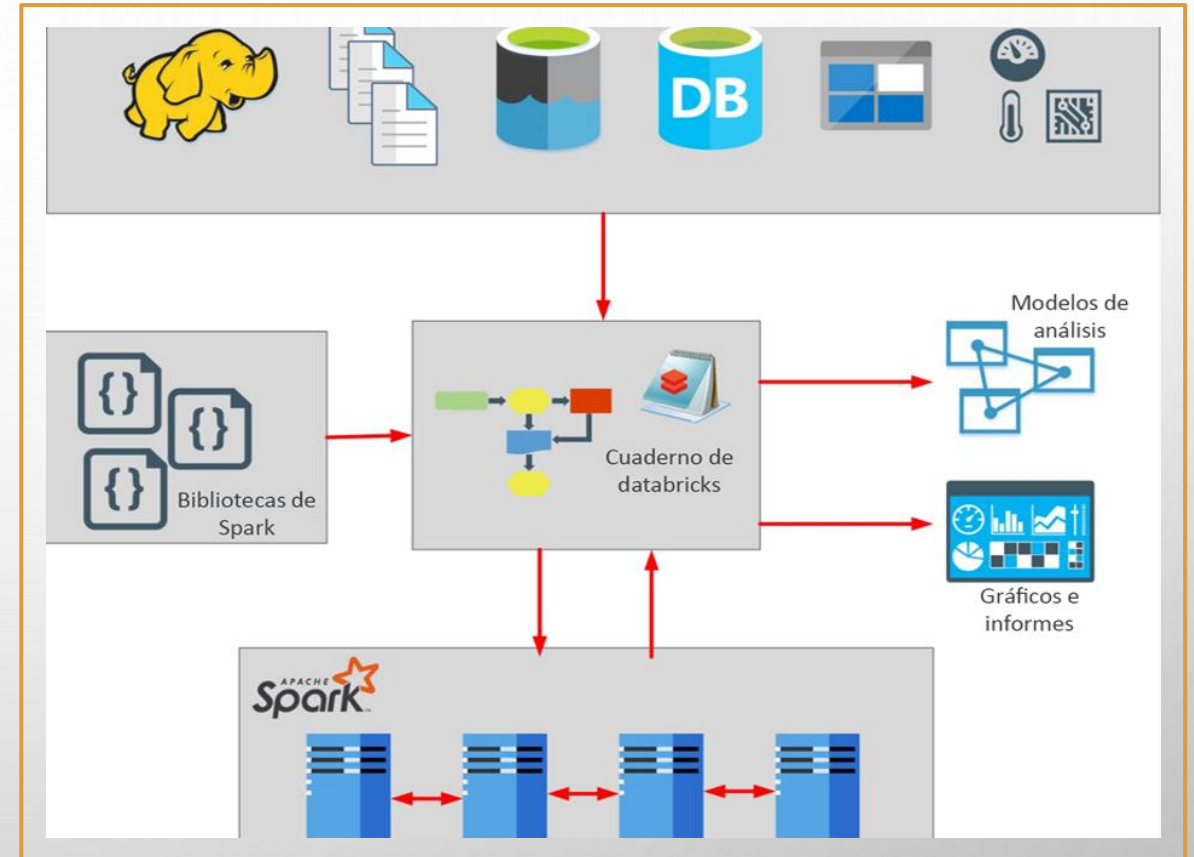
Combina la estructura de directorios jerárquica y la semántica del sistema de archivos, de un sistema de archivos tradicional con la seguridad y la escalabilidad que proporciona azure.

Tiene las siguientes características:

Organiza sus archivos en directorios y subdirectorios para mejorar la organización de los archivos.

Admite los permisos de directorio y de archivo de portable Operating system interface (POSIX) para habilitar el control de acceso, basado en rol (RBAC) granular en sus datos.

Es compatible con el sistema de archivos distribuido de HADOOP (hdfs).



¿Qué es Azure Databricks?



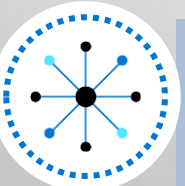
Integración con los servicios de Azure:

Puede integrarse con una variedad de servicios de plataforma de datos de Azure y Power BI.



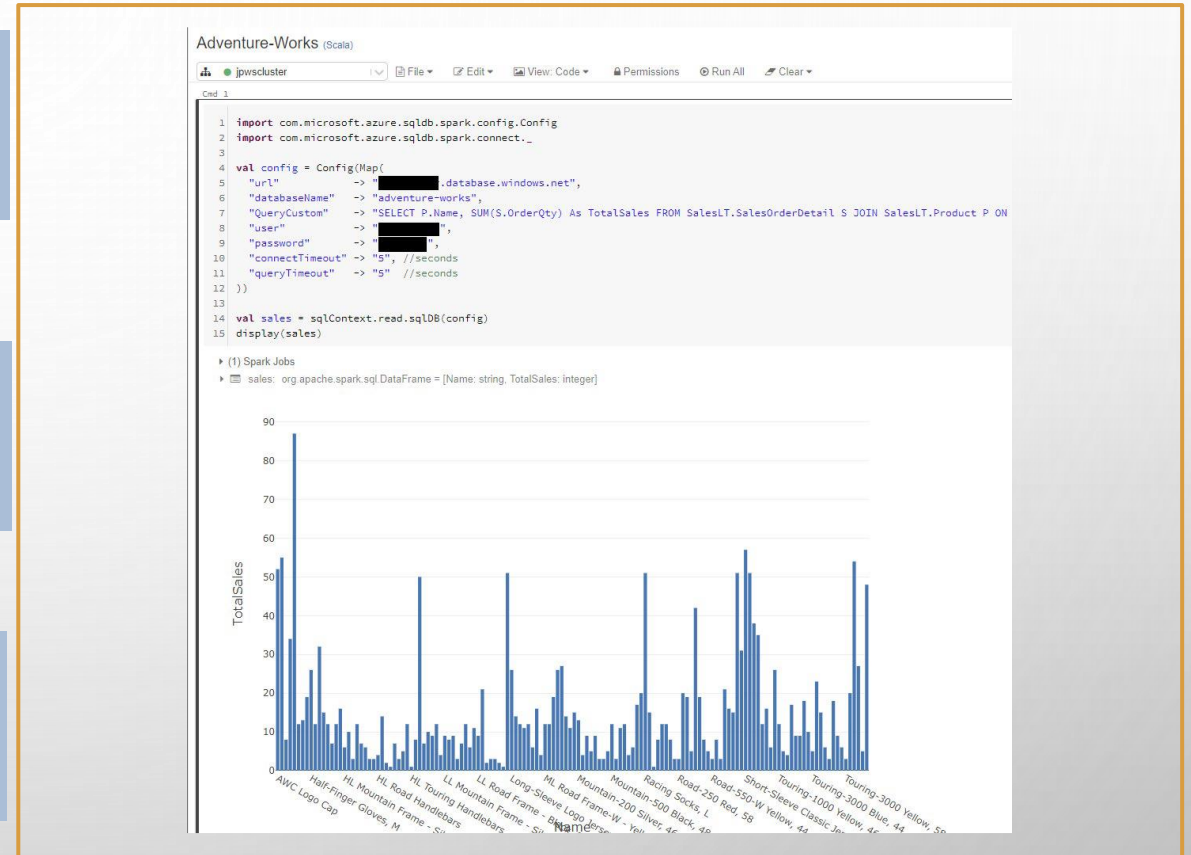
Seguridad empresarial:

Utiliza las capacidades de seguridad de Azure.



Plataforma basada en Apache Spark:

Simplifica el aprovisionamiento y la colaboración de las soluciones analíticas basadas en Apache Spark.

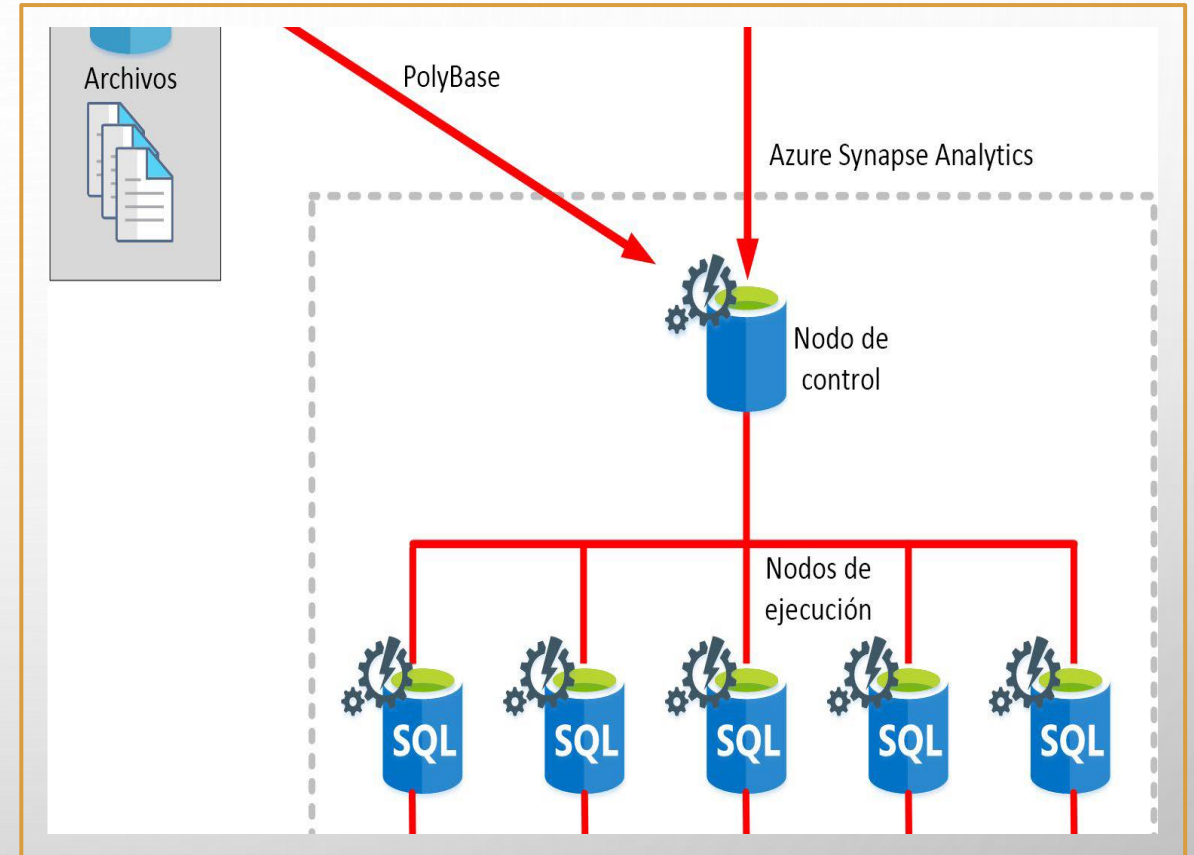


¿QUÉ ES AZURE SYNAPSE ANALYTICS?

Es un motor de análisis. Está diseñado para procesar grandes cantidades de datos muy rápidamente.

Con Synapse Analytics, Puede ingerir datos de orígenes externos, como archivos planos, Azure Data Lake u otros sistemas de administración de bases de datos, y luego transformar y agregar estos datos en un formato adecuado para el procesamiento de análisis. Puede realizar consultas complejas sobre estos datos y generar informes, grafos y tablas.

Aprovecha una arquitectura de procesamiento masivo paralelo (MPP). Esta arquitectura incluye un nodo de control y un grupo de nodos de ejecución.



¿QUÉ ES AZURE ANALYSIS SERVICES?

- Permite crear modelos tabulares para admitir consultas de procesamiento analítico en línea (OLAP). Puede combinar datos de varios orígenes, incluidos Azure SQL Database, Azure Synapse Analytics, Azure Data Lake Store, Azure Cosmos DB Y muchos otros.
- Un modelo es esencialmente un conjunto de consultas y expresiones que recuperan datos de varios orígenes de datos y generan resultados. Los resultados pueden almacenarse en memoria caché para su uso posterior, o pueden calcularse de forma dinámica directamente desde los orígenes de datos subyacentes.
- Analysis Services incluye un diseñador gráfico para ayudarlo a conectar orígenes de datos y definir consultas que combinen, filtren y agreguen datos. Puede explorar estos datos desde Analysis Services o puede usar una herramienta como Microsoft POWER BI para visualizar los datos presentados por estos modelos.



Synapse Pipelines



Synapse Link



Synapse Studio



Synapse SQL



Azure Synapse
Analytics



Synapse Spark

¿QUÉ ES AZURE HDINSIGHT?

Es un servicio de procesamiento de macrodatos que proporciona una plataforma para tecnologías como SPARK en un entorno de Azure. Hdinsight implementa un modelo agrupado que distribuye el procesamiento entre un conjunto de equipos. Este modelo es similar al que usa Synapse Analytics, excepto que los nodos ejecutan el motor de procesamiento de Spark en lugar de Azure SQL Database.

Puede usar Azure Hdinsight junto con Azure Synapse Analytics o en su lugar. Además, de Spark, Hdinsight admite tecnologías de streaming como Apache Kafka y el modelo de procesamiento Apache Hadoop. La siguiente imagen muestra dónde puede usar los componentes de hdinsight en una solución de almacenamiento de datos.

