

4: Exploración de los fundamentos de análisis de datos



Agenda



- Análisis de datos a gran escala
- Streaming y análisis en tiempo real
- Visualización de datos

Objetivos de aprendizaje

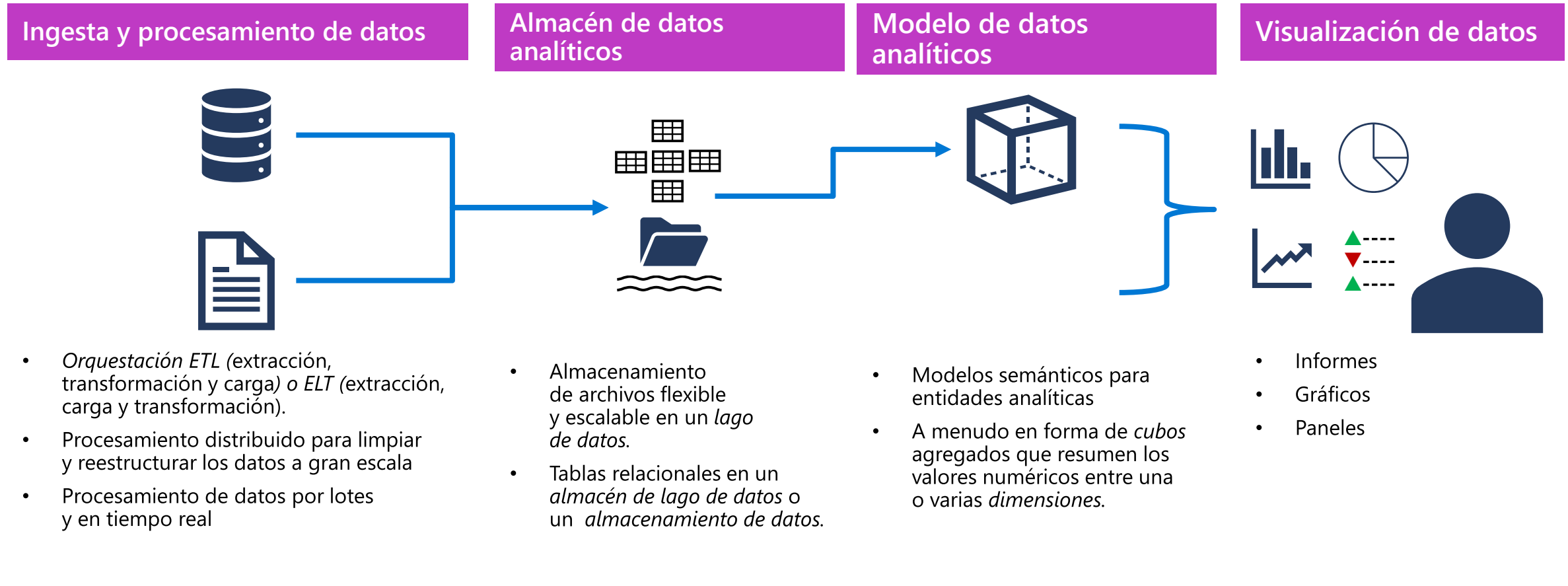
Tras finalizar este módulo, podrá:

- 1 Describir las características comunes de las soluciones analíticas a gran escala.
- 2 Describir las características comunes de las soluciones analíticas en tiempo real.
- 3 Descripción de consideraciones para la visualización de datos

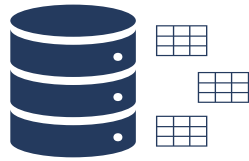
1: Análisis de datos a gran escala



Elementos de una solución de análisis de datos a gran escala



Procesamiento de datos en análisis a gran escala



Base de datos relacional

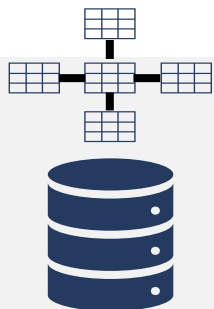
- Modelo bien establecido para el almacenamiento y el procesamiento de datos relacionales.
- Compatibilidad total con el lenguaje SQL para consultas y la manipulación de datos.



Apache Spark

- Plataforma de código abierto para el procesamiento de datos escalable y distribuido.
- Código de procesamiento de datos multilenguaje (Python, Scala, Java, SQL...).

Arquitecturas de almacén de datos analíticos



Data Warehouse

- Los datos se almacenan en una base de datos relacional y se consultan mediante un motor de consultas SQL.
- Los datos se *desnormalizan* para optimizar las consultas
 - Normalmente como un esquema de estrella o copo de nieve de *hechos* numéricos que se pueden agregar mediante *dimensiones*.



Almacén de lago de datos

- Los archivos de datos se almacenan en un sistema de archivos distribuido (un *lago de datos*) y se suelen procesar con Apache Spark.
- Los metadatos se usan para definir tablas que proporcionan una interfaz SQL relacional a los datos del archivo.
 - Normalmente se usa un formato *delta lake* para proporcionar funcionalidad de base de datos transaccional.

Servicios de análisis de datos PaaS



Azure Synapse Analytics

- Solución unificada para el almacenamiento de datos relacionales y el análisis del lago de datos
- Procesamiento escalable y consulta mediante varios entornos de ejecución de análisis
 - SQL de Synapse
 - Apache Spark
 - Synapse Data Explorer
- Experiencia interactiva en Azure Synapse Studio
- Integración de canalización integrada para la ingesta y el procesamiento de datos

Úselo para una única solución analítica unificada a gran escala en Azure



Azure Databricks

- Implementación basada en Azure de la plataforma de análisis en la nube Databricks
- Spark y consulta de SQL escalables para el análisis del lago de datos
- Experiencia interactiva en el área de trabajo de Azure Databricks
- Use Azure Data Factory para implementar canalizaciones de ingesta y procesamiento de datos

Úselo para aprovechar las aptitudes de Databricks y para la portabilidad en la nube



Azure HDInsight

- Implementación basada en Azure de marcos de "macrodatos" comunes de Apache basados en un lago de datos
 - Hadoop: consulta de archivos de lago de datos mediante tablas de Hive
 - Spark: use las API de Spark para consultar datos y abstraer el almacenamiento de archivos subyacente como tablas
 - Kafka: procesamiento de eventos en tiempo real
 - Storm: procesamiento de flujos
 - HBase: almacén de datos NoSQL

Úselo cuando necesite admitir varias plataformas de código abierto

Análisis de datos SaaS con Microsoft Fabric



Laboratorio: Exploración del análisis de datos en Microsoft Fabric



En este ejercicio, explorará la ingesta y el análisis de datos en un almacén de lago de Microsoft Fabric.

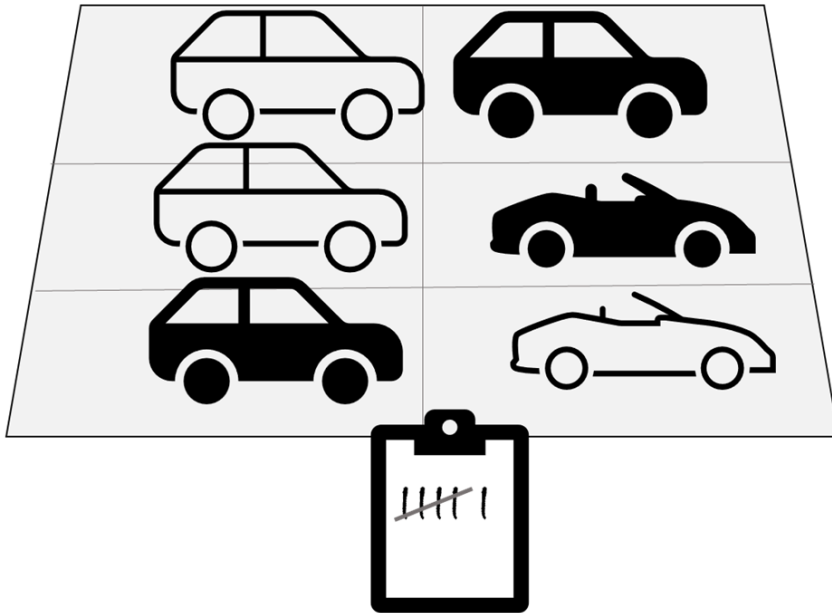
1. **Inicie la máquina virtual para este laboratorio**
o vaya a la página del ejercicio en <https://aka.ms/dp900-fabric-lab>.
2. **Siga las instrucciones para completar el ejercicio en Microsoft Learn**
Use la suscripción de Azure proporcionada para este laboratorio.

2: Streaming y análisis en tiempo real



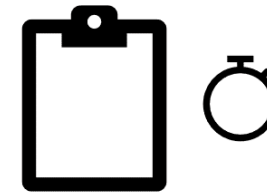
Procesamiento por lotes frente a procesamiento de flujos

Procesamiento por lotes



Los datos se recopilan y procesan a intervalos regulares

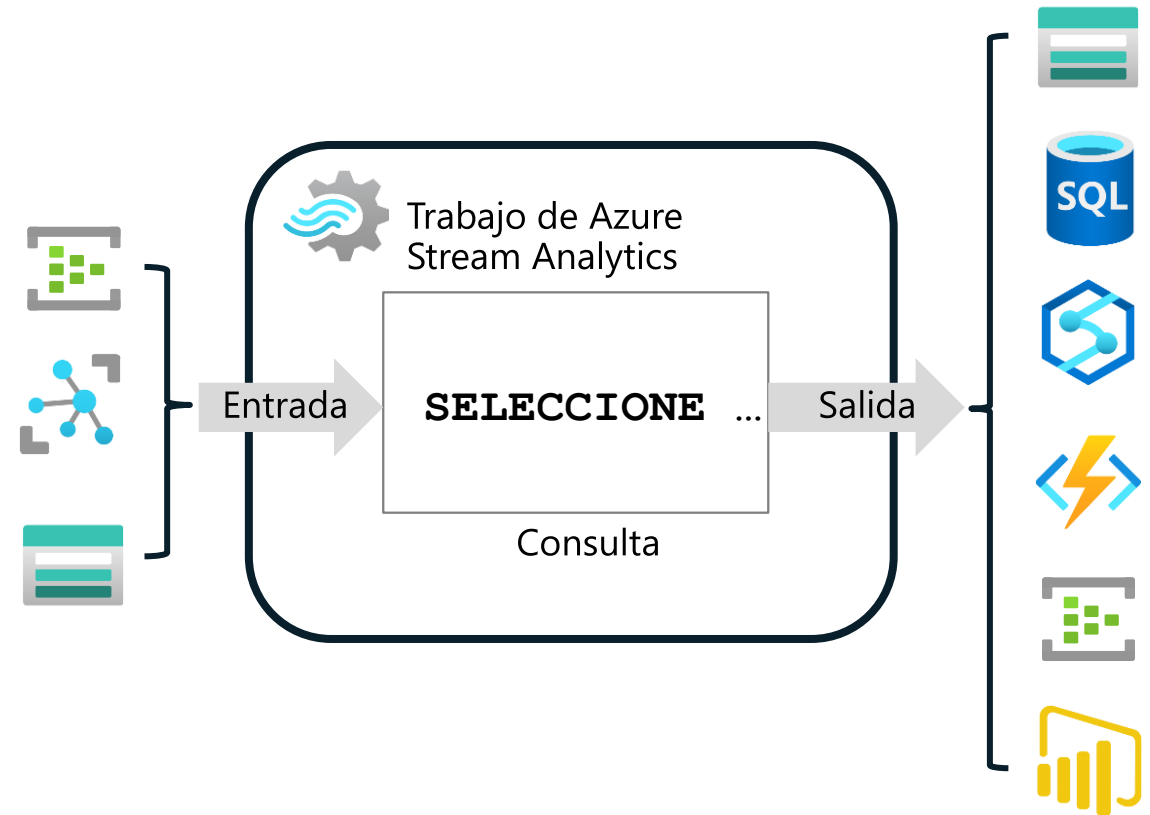
Procesamiento de flujos



Los datos se procesan en tiempo casi real conforme llegan

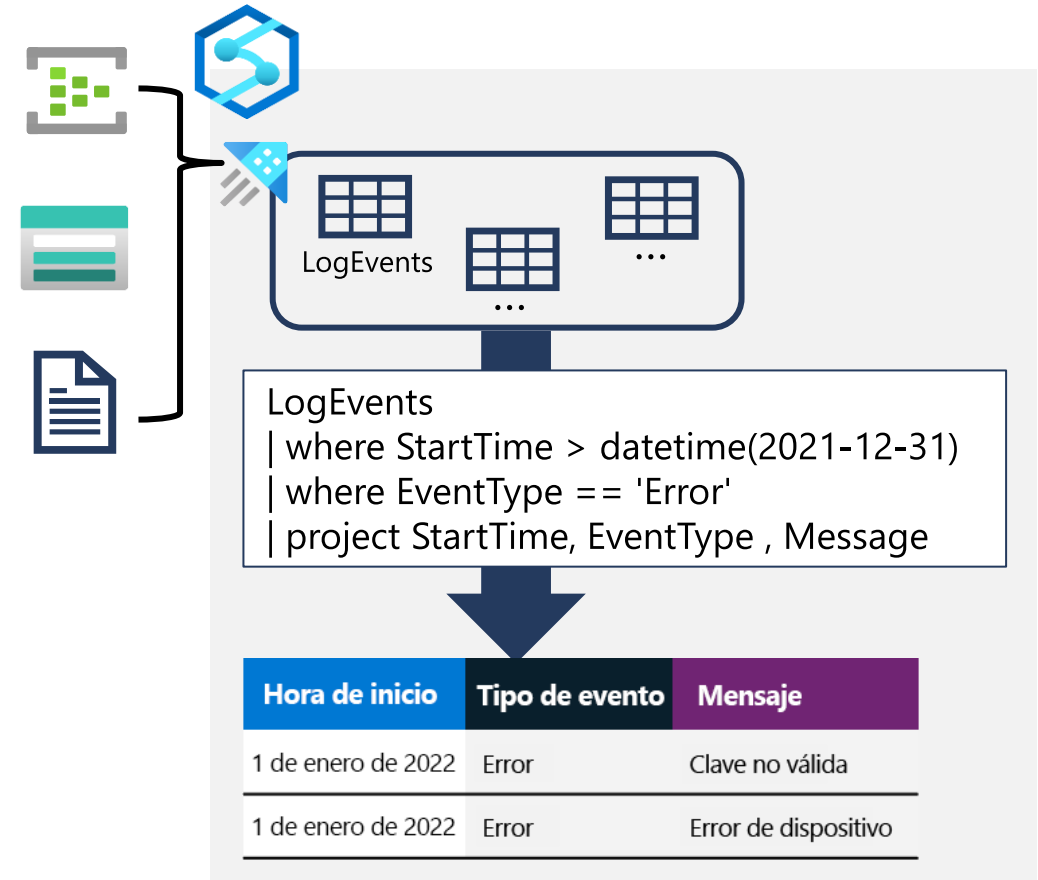
Procesamiento de datos en tiempo real con Azure Stream Analytics

- Cree un *trabajo* o un *clúster* de Azure Stream Analytics.
- Ingera datos de una *entrada*, por ejemplo:
 - Azure Event Hubs
 - Azure IoT Hub
 - Azure Blob Storage
 - ...
- Procese datos con una *consulta perpetua*.
- Envíe los resultados a una *salida*, por ejemplo:
 - Azure Blob Storage
 - Azure SQL Database
 - Azure Synapse Analytics
 - Función de Azure
 - Azure Event Hubs
 - Power BI
 - ...



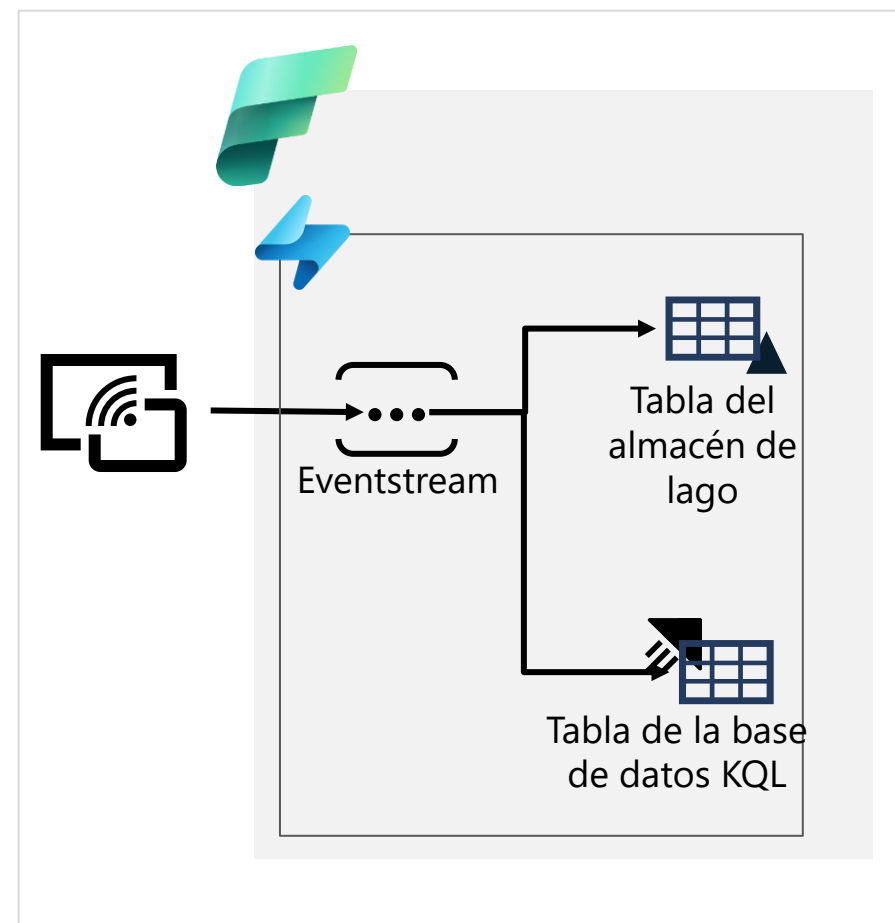
Análisis de telemetría y registro en tiempo real con Azure Data Explorer

- Alto rendimiento, servicio escalable para datos por lotes y de streaming
 - **Servicio** Azure Data Explorer dedicado
 - **Entorno de ejecución de** Azure Synapse Data Explorer en Azure Synapse Analytics
- Los datos se ingieren desde el streaming y los orígenes por lotes en tablas de una base de datos
- Las tablas se pueden consultar con el *Lenguaje de consulta Kusto* (KQL):
 - Sintaxis intuitiva para consultas de solo lectura
 - Optimizado para datos de serie temporal y telemetría sin procesar



Análisis en tiempo real en Microsoft Fabric

- Compatibilidad con la ingesta continua de datos de varios orígenes.
- Captura de datos de streaming en un **Eventstream**.
- Escritura de datos en tiempo real en una tabla en un almacén de lago o una base de datos KQL.
- Consulta de datos en tiempo real con SQL o KQL.
- Creación de visualizaciones en tiempo real.



Laboratorio: Exploración del análisis en tiempo real en Microsoft Fabric



En este laboratorio, explorará el análisis en tiempo real en Microsoft Fabric.

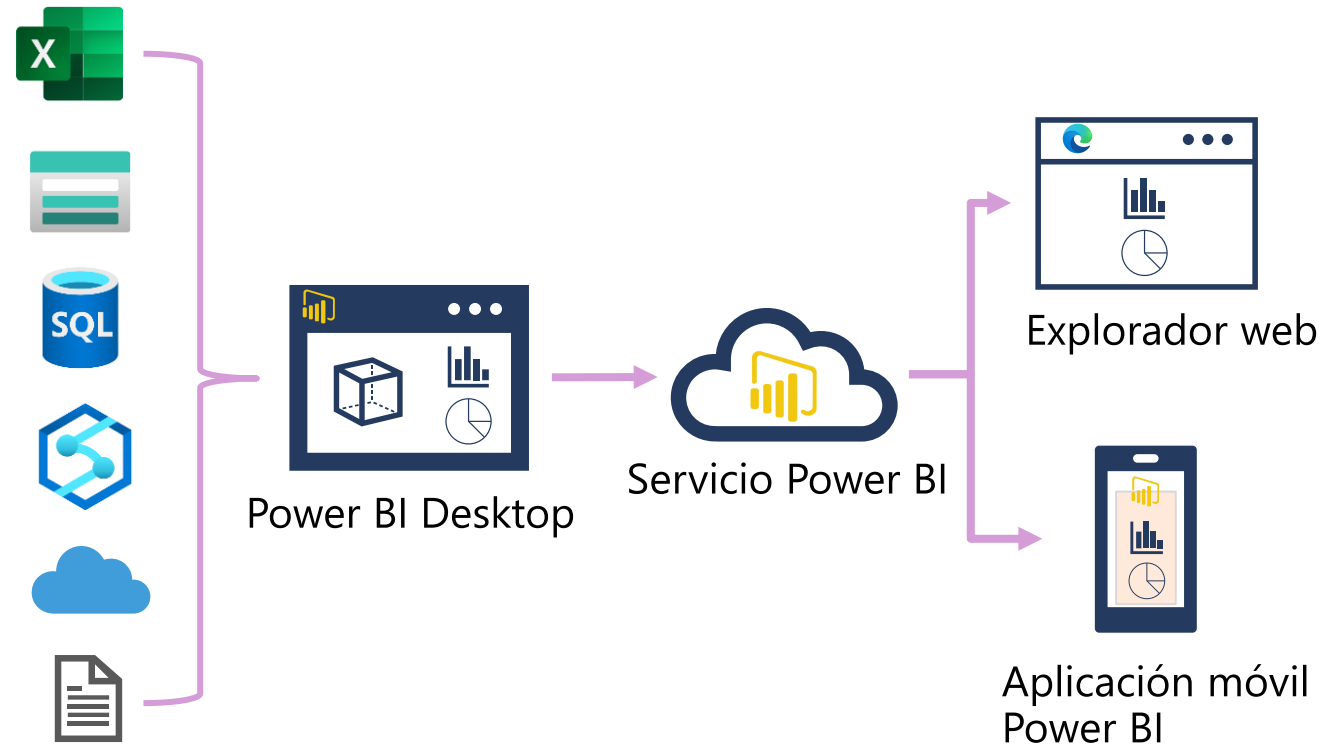
- 1. Inicie la máquina virtual para este laboratorio**
o vaya a la página del ejercicio en <https://aka.ms/dp900-fabric-rta-lab>.
- 2. Siga las instrucciones para completar el ejercicio en Microsoft Learn**
Use la suscripción de Azure proporcionada para este laboratorio y una instancia de Cloud Shell en Azure Portal.

3: Visualización de datos

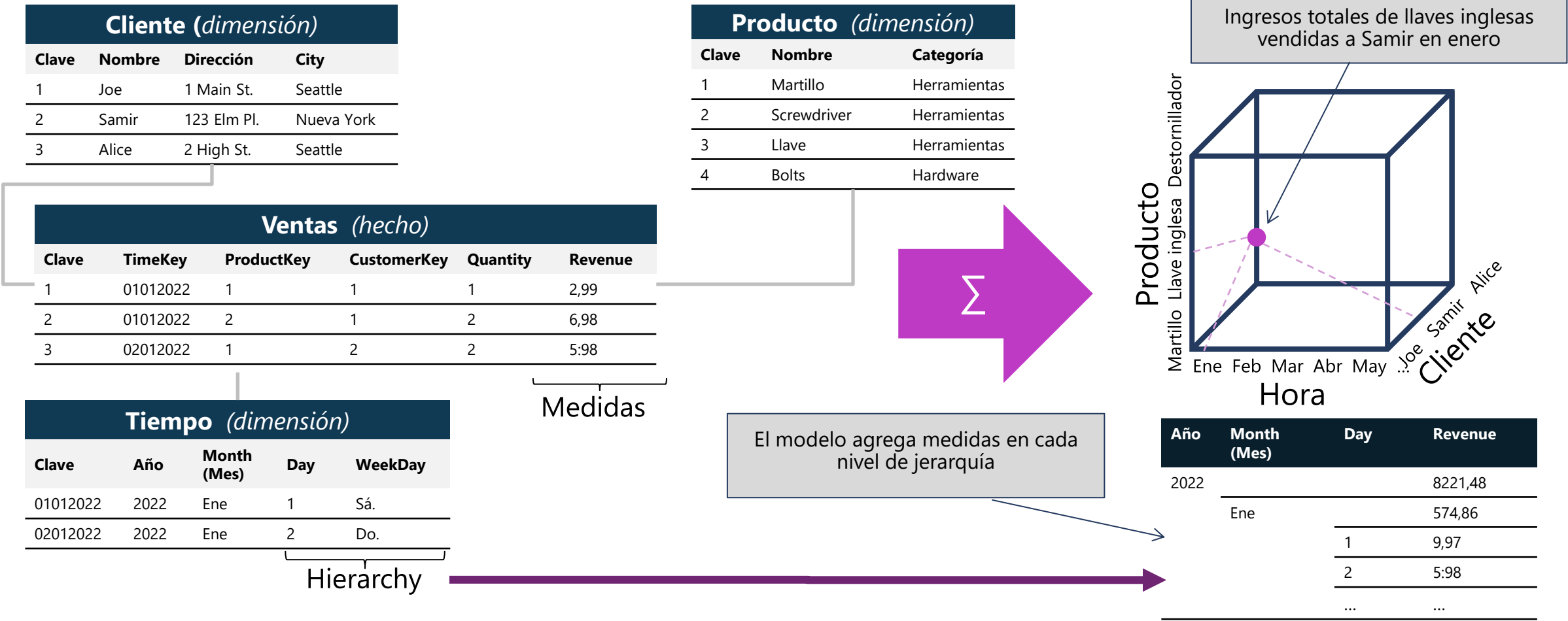


Introducción a la visualización de datos con Power BI

- Empiece con Power BI Desktop
 - Importación de datos desde uno o varios orígenes
 - Definición de un modelo de datos
 - Creación de visualizaciones en un informe
- Publicación en el servicio Power BI
 - Programación de actualizaciones de datos
 - Creación de paneles y aplicaciones
 - Uso compartido con otros usuarios
- Interacción con informes publicados
 - Explorador web
 - Aplicación móvil de Power BI



Modelado de datos analíticos



Visualizaciones de datos comunes en informes

Tablas y texto

Product Sales		
Name	Quantity	
Bolts	2	\$302.91 Revenue
Hammer	4	
Nails	1	
Screwdriver	2	
Screws	2	
Wrench	4	
Total	15	

Gráfico de barras o columnas

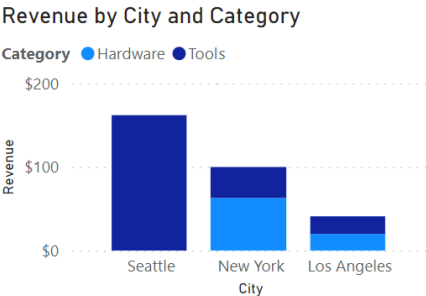


Gráfico de líneas

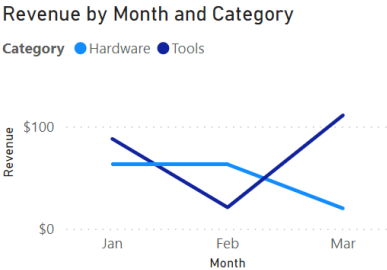


Gráfico circular

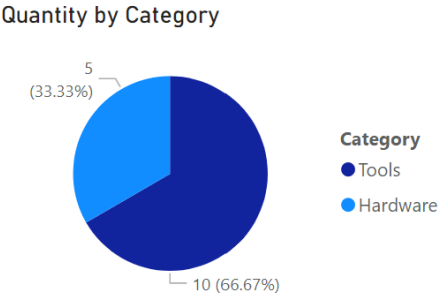
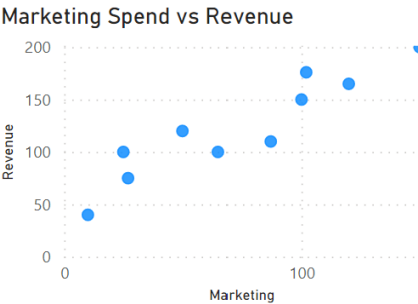
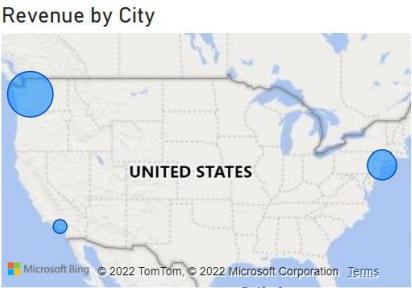


gráfico de dispersión



Asignar



Laboratorio: Visualización de datos con Power BI



En este laboratorio, usará Power BI Desktop para crear un modelo de datos y un informe

- 1. Inicie la máquina virtual para este laboratorio**
o vaya a la página del ejercicio en <https://aka.ms/dp900-pbi-lab>.
- 2. Siga las instrucciones para completar el ejercicio en Microsoft Learn**
Use la suscripción de Azure proporcionada para este laboratorio.

