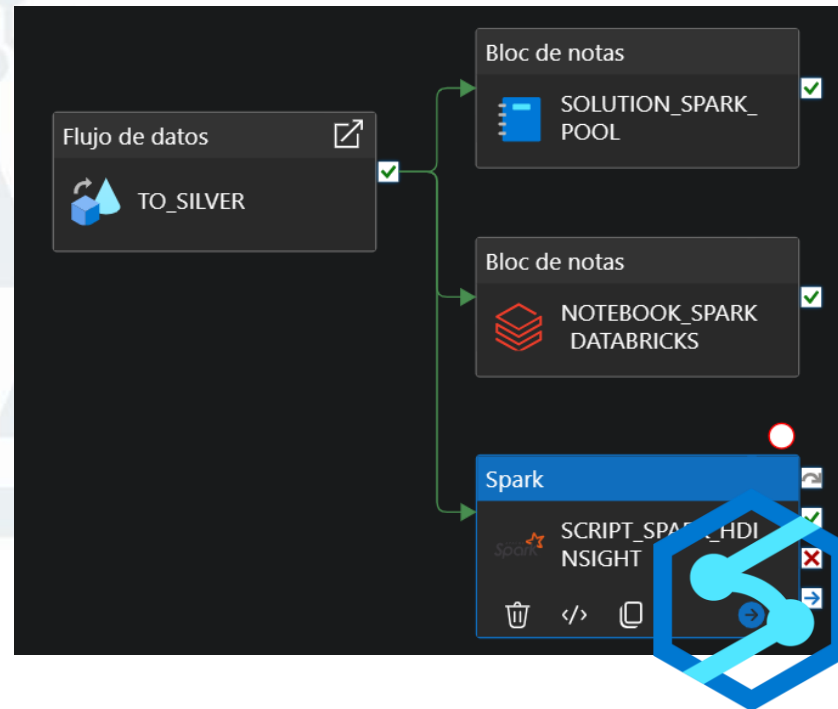

A large, light blue watermark of the Big Data Academy logo and name is centered in the background of the slide.

Synapse Analytics

Pipelines de Procesamiento para Big Data

Synapse Analytics es una **herramienta de procesamiento para entornos de Big Data** que permite dibujar flujos de procesamiento (Data Factory y Dataflow), escribir código (Spark) y explotar los datos (SQL)



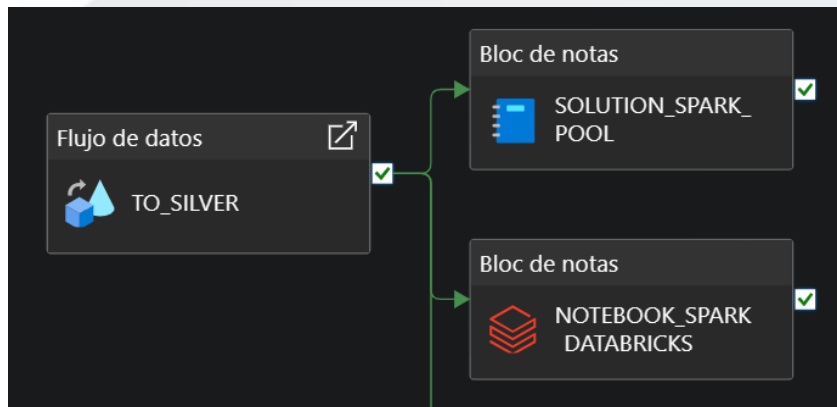
Azure
Synapse Analytics

Se usa generalmente para la **construcción de DATA LAKES y DELTA LAKES**

Herramientas visuales para Big Data

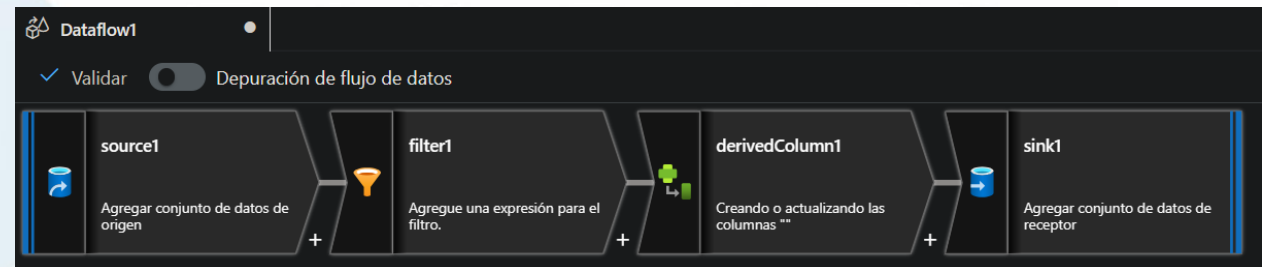
Data Factory

Permite crear la malla de procesamiento



Dataflow

Permite dibujar procesos simples de limpieza de datos



IMPORTANTE: si los procesos son más complejos, por ejemplo, modelamiento semi-estructurado, la limpieza de datos se hace con una herramienta de código (Spark)

Herramientas de código para Big Data

En Synapse existen tres motores de procesamiento de datos basados en SPARK

SPARK POOL

Crea y ejecuta códigos de Spark desde un Notebook, se encuentra incluido dentro de Synapse y es la opción recomendada por Azure



databricks

Crea y ejecuta códigos de Spark desde un Notebook, no se encuentra incluido dentro de Synapse, debe configurarse para ser incluido



Azure HDInsight

Crea y ejecuta códigos de Spark, pero no desde un Notebook, sino desde un Script

Herramientas de explotación de datos

Synapse SQL nos permite ver a los archivos de datos en los BLOB STORAGES como si fueran tablas de bases de datos clásicas y consultarlas con código SQL

```
-- Creamos la BASE DE DATOS del delta lake
-- Tendrá soporte para caracteres especiales (Latin1_General_100_BIN2_UTF8)
CREATE DATABASE DELTALAKE COLLATE Latin1_General_100_BIN2_UTF8;

-- Cambiamos a esta base de datos para trabajar sobre ella
USE DELTALAKE;

-- Creamos la tabla de texto plano "PERSONA" en la capa "BRONZE"
CREATE EXTERNAL TABLE BRONZE.PERSONA(
  ID VARCHAR(8),
  NOMBRE VARCHAR(128),
  TELEFONO VARCHAR(128),
  CORREO VARCHAR(128),
  FECHA_INGRESO VARCHAR(128),
  EDAD INT
) WITH (
  LOCATION = 'bronze/persona/*.data',
  DATA_SOURCE = DELTALAKE_STORAGE,
  FILE_FORMAT = DELTALAKE_FORMAT_TXT
);
```