

Definiendo el Source Source settings

Opciones de esquema

Dado que un "inline dataset" se define dentro del data flow, no hay un schema definido asociado al inline dataset. En la pestaña Projection (Proyección), puede importar el schema de datos de origen y almacenar ese schema como proyección de origen. En esta pestaña, encontrará un botón "Schema options" que le permite definir el comportamiento del servicio de descubrimiento de esquemas de ADF.

Utilizar schema proyectado: Esta opción es útil cuando tiene un gran número de archivos fuente que ADF escaneará como su fuente. El comportamiento por defecto de ADF es descubrir el schema de cada archivo fuente. Pero si tiene una proyección predefinida ya almacenada en su transformación de fuentes, puede activar esta opción y ADF omitirá el autodescubrimiento de cada schema. Con esta opción activada, la transformación de origen puede leer todos los archivos de una forma mucho más rápida, aplicando el esquema predefinido a cada archivo.

Allow schema drift (Permitir el cambio de esquema): Active esta opción para que el data flow permita nuevas columnas que no estén ya definidas en el schema de origen. Esto es útil si el schema cambia a menudo.

Validate schema: Activar esta opción hará que el 'data flow' falle si alguna columna y tipo definidos en la proyección no coinciden con el schema descubierto de los datos de origen.

Infer drifted column types: Cuando ADF identifique nuevas columnas derivadas, esas nuevas columnas se convertirán al tipo de datos apropiado utilizando la inferencia automática de tipos de ADF.

Source settings

Después de añadir una source, configúrela a través de la pestaña "Source settings". Aquí puedes elegir o crear el dataset al que apunta tu source. También puedes seleccionar el schema y las opciones de sampling para tus datos.

Los valores de desarrollo para los parámetros del conjunto de datos pueden configurarse en los ajustes de depuración. (El "Debug mode" debe estar activado).

Output stream name: El nombre de la transformación del source.

Source type (Tipo de source): Elija si desea utilizar un inline dataset o un objeto dataset existente.

Test connection: Pruebe si el servicio Spark del data flow puede conectarse correctamente al linked service utilizado en su source dataset. El Debug mode debe estar activado para que esta función esté habilitada.

Schema drift: Schema drift es la capacidad del servicio para manejar de forma nativa esquemas flexibles en sus data flows sin necesidad de definir explícitamente los cambios de columna.

- Seleccione la casilla **Allow schema drift** si las columnas source van a cambiar a menudo. Esta configuración permite que todos los campos source entrantes fluyan a través de las transformaciones hacia el sink. Esto es útil si el schema cambia a menudo.
- La selección de **Infer drifted column types** indica al servicio que detecte y defina tipos de datos para cada nueva columna descubierta. Con esta opción **desactivada**, todas las columnas derivadas serán de tipo **"string"**.

Validate schema: Si se selecciona Validate schema, el data flow no se ejecutará si los datos source entrantes no coinciden con el schema definido del dataset.

Skip line count: El campo Skip line count especifica cuántas líneas se deben ignorar al principio del dataset.

Sampling (Muestreo): Active Sampling para limitar el número de filas de su source. Utilice esta opción cuando pruebe o muestree datos de su source con fines de debugging. Esto es muy útil cuando se ejecutan data flows en debug mode desde un pipeline.

Para validar que su source está configurado correctamente, active el "Debug Mode" y obtenga una "Data Preview".