

Definiendo el Source Dataset vs Inline

Vamos a ver cómo podemos crear "Data Sources" utilizando "Mapping Data Flow". Así que vamos a empezar con "Source settings" y la primera y más importante decisión que usted tiene que hacer es que "Data Source type" tenemos que utilizar. Tenemos "Dataset" y también "Inline". "Dataset" es sólo una entidad que se puede utilizar a través de los dataflows y también pipelines. Si recuerdas todos esos datasets, los creamos antes para nuestras demostraciones. Así que eso son sólo entidades. Aquí, "Inline" es algo más dedicado para cada lógica de dataflow o workflow. Hay beneficios en el uso de ambos, pero hay casos en que ambos no serán compatibles. Por ejemplo, en el caso de los "Inline datasets", tenemos algunos "out of the box datasets" entre los que podemos elegir, y esos "datasets" funcionarían con algunos conectores. No con todos. Por ejemplo, esos conjuntos de datos Inline no podrían utilizarse con el conector Azure Cosmos DB. Para tener una lista más definida de cuándo utilizar el Inline Source type o el Dataset Source type, recomendaría echar un vistazo a la documentación.

<https://learn.microsoft.com/en-us/azure/data-factory/data-flow-source>

Un elemento importante de utilizar Dataset o Inline es el hecho de que Inline es un tipo de fuente nativo para Spark, mientras que Dataset no lo es.

En resumen, vamos a tener que ir probando cual nos proporciona mejor rendimiento.