

***SCD TYPE 2 IMPLEMENTATION IN AZURE SYNAPSE ANALYTICS***

* Create a table in SQL Database

CREATE TABLE CUST\_SCDTYPE2

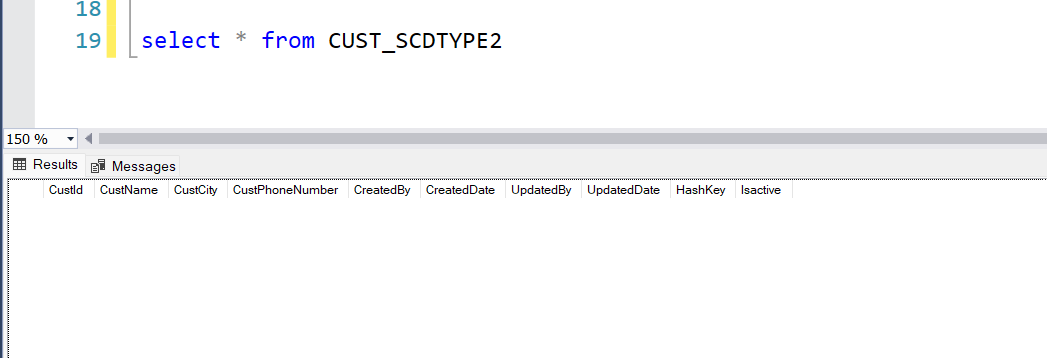
(

CustId Int, CustName Varchar(100),CustCity Varchar(100),CustPhoneNumber Bigint,

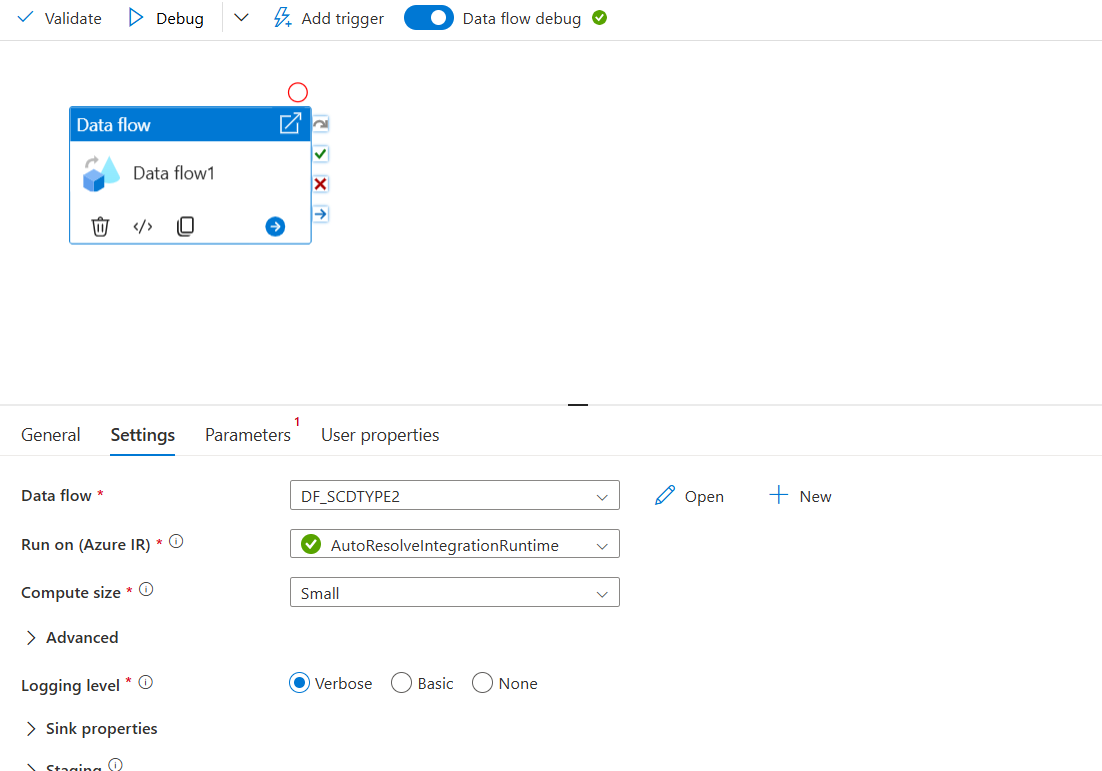
CreatedBy varchar(100),CreatedDate datetime,UpdatedBy varchar(100),

UpdatedDate datetime,HashKey Bigint, Isactive INT

)

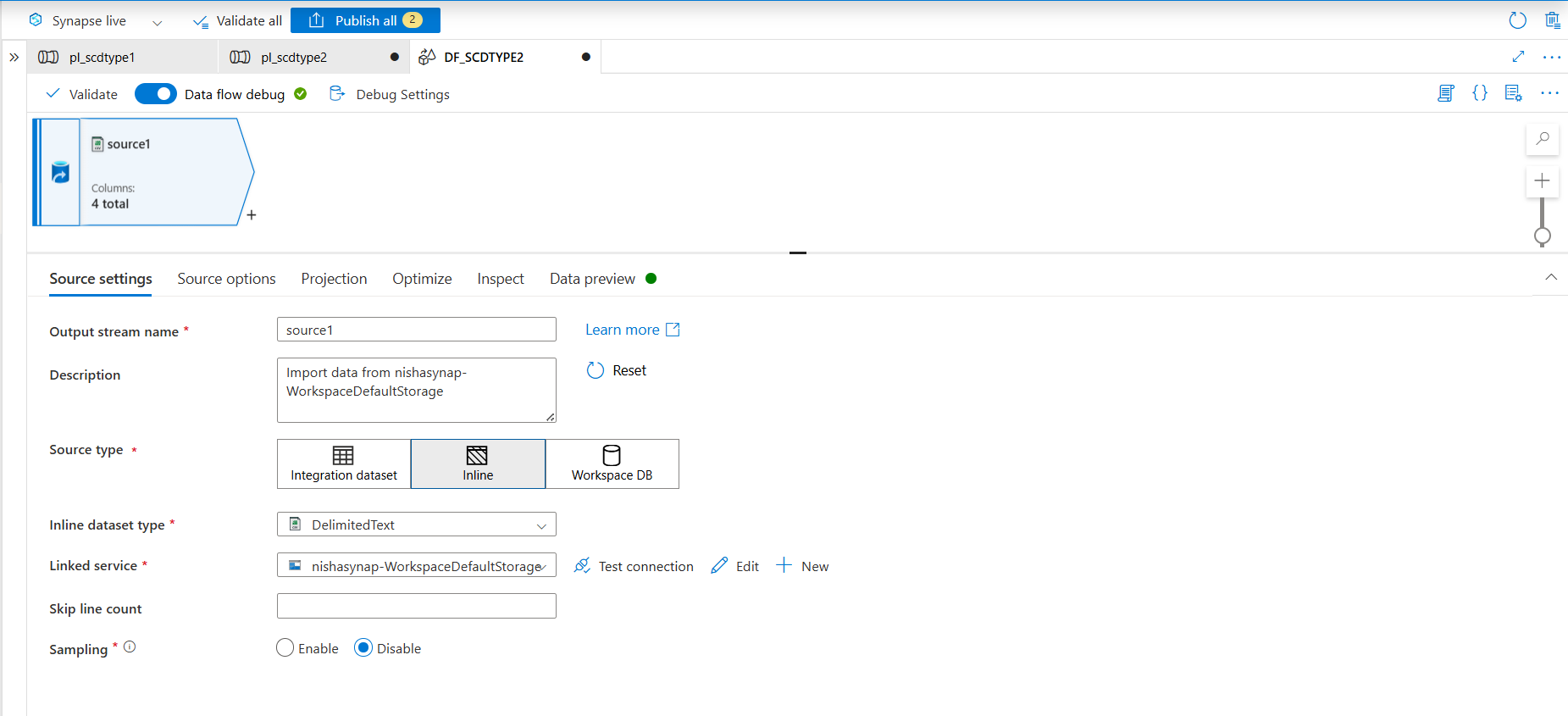


* Create a new pipeline for SCD Type 2 activity in azure synapse analytics, open a dataflow

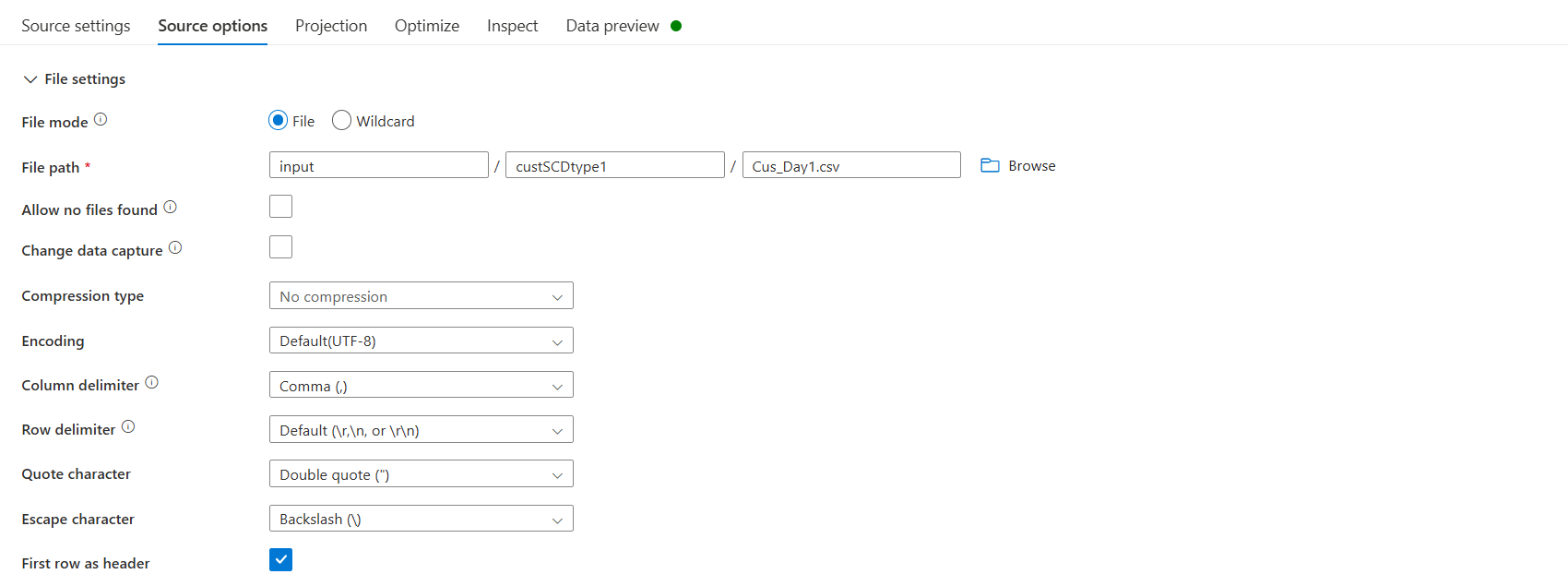


*Open the data flow and start creating the activity*

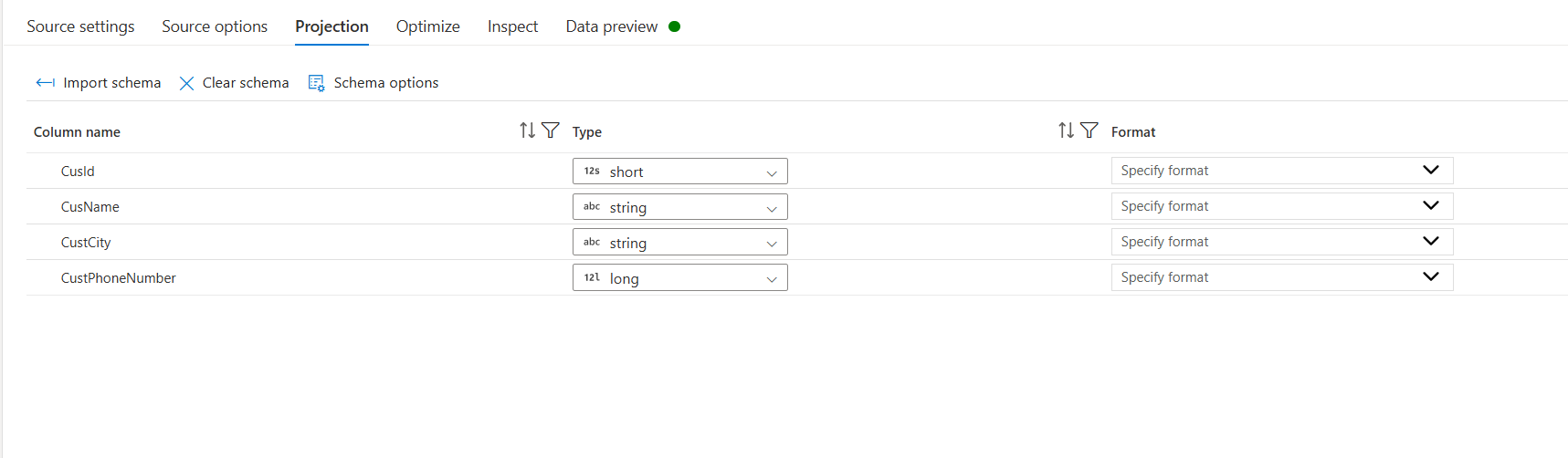
* Create a Source inside the dataflow activity, **the source type** should be **Inline** and **Inline data type** should be **Delimited Text**

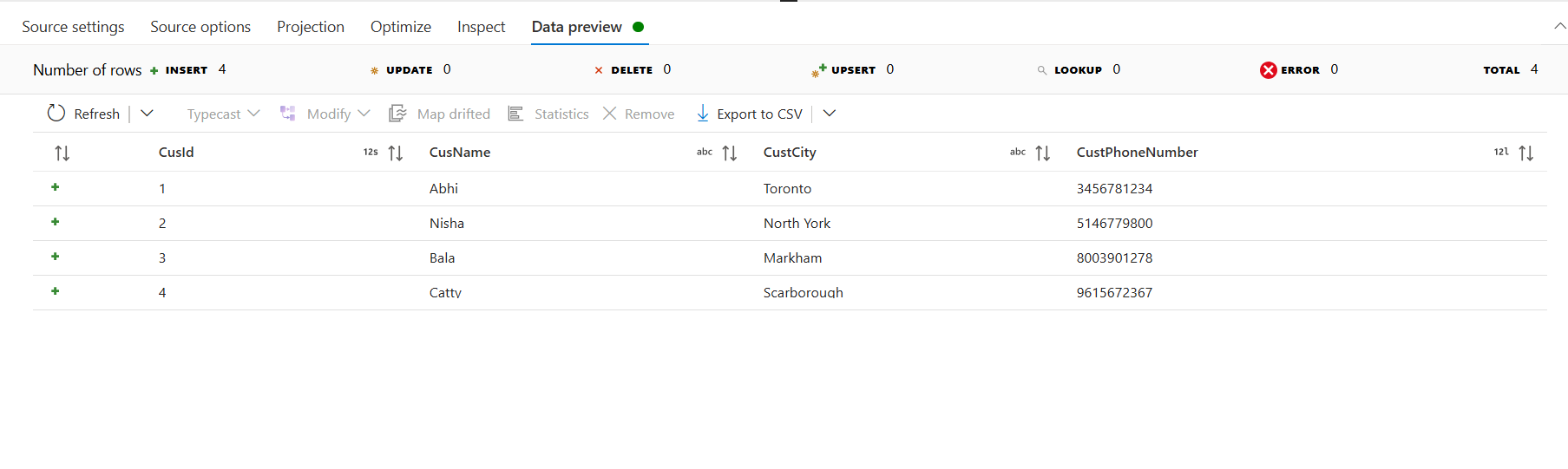


* In the source options choose the Day 1 file and **select** the **first row as header**

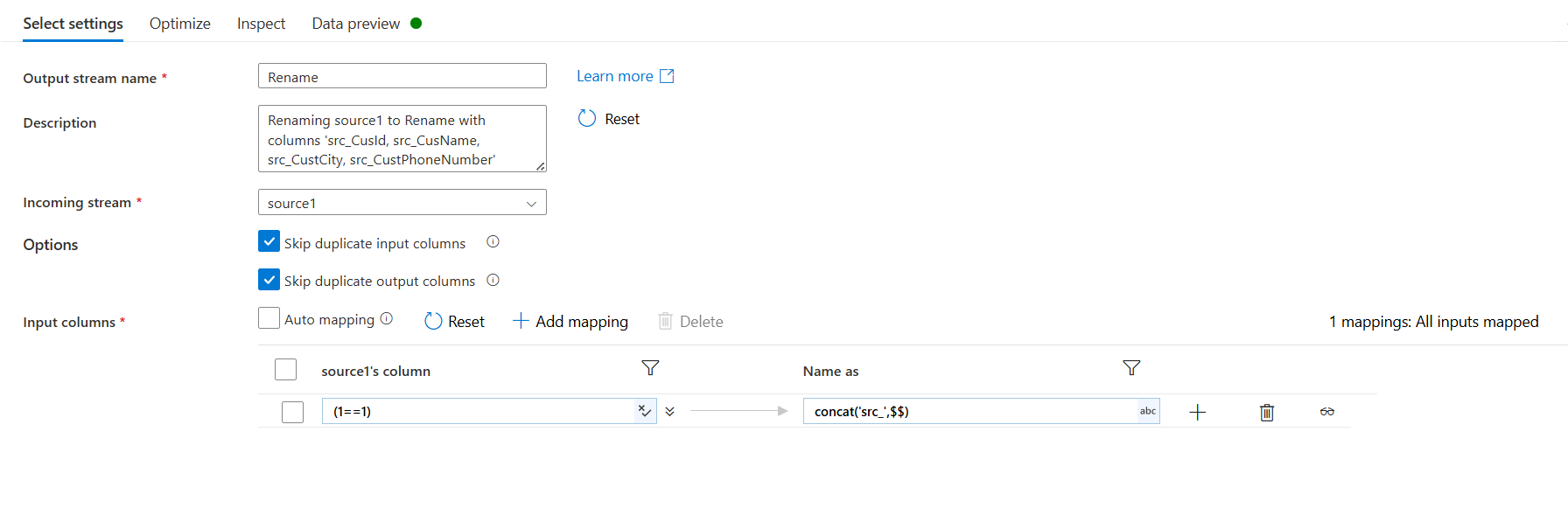


* Import the schema and preview the data

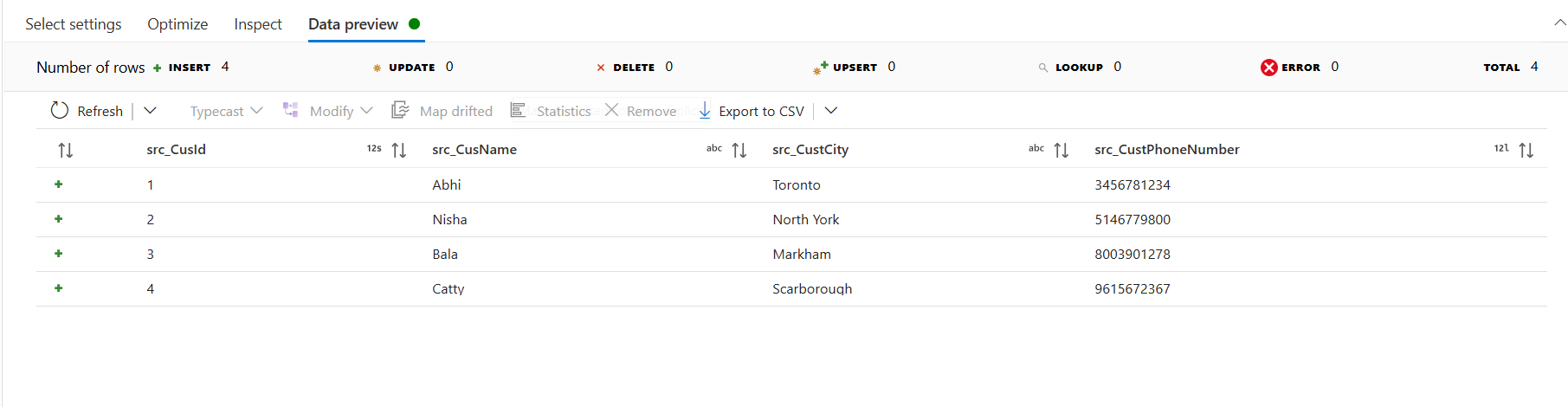




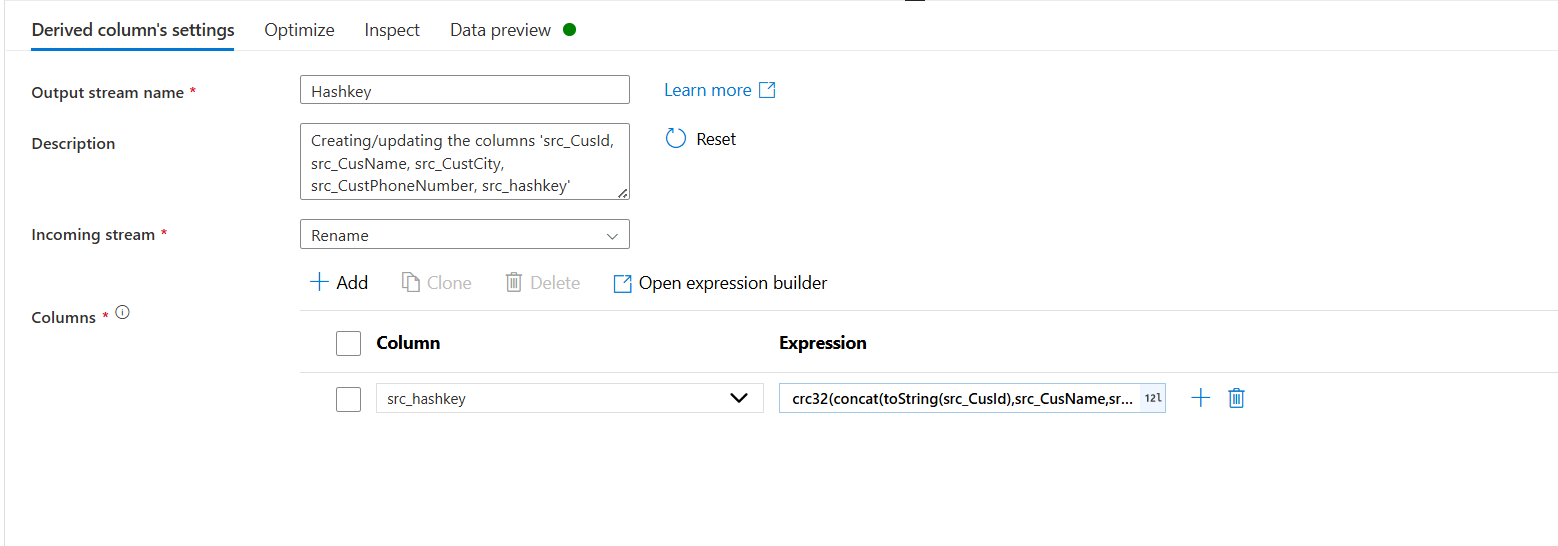
* Add select settings and rename it “Rename”, to rename the columns first delete all the columns and open **Rule based mapping** in **source** field give **(1==1)** and in **expression** field **concat(‘src\_’,$$)**

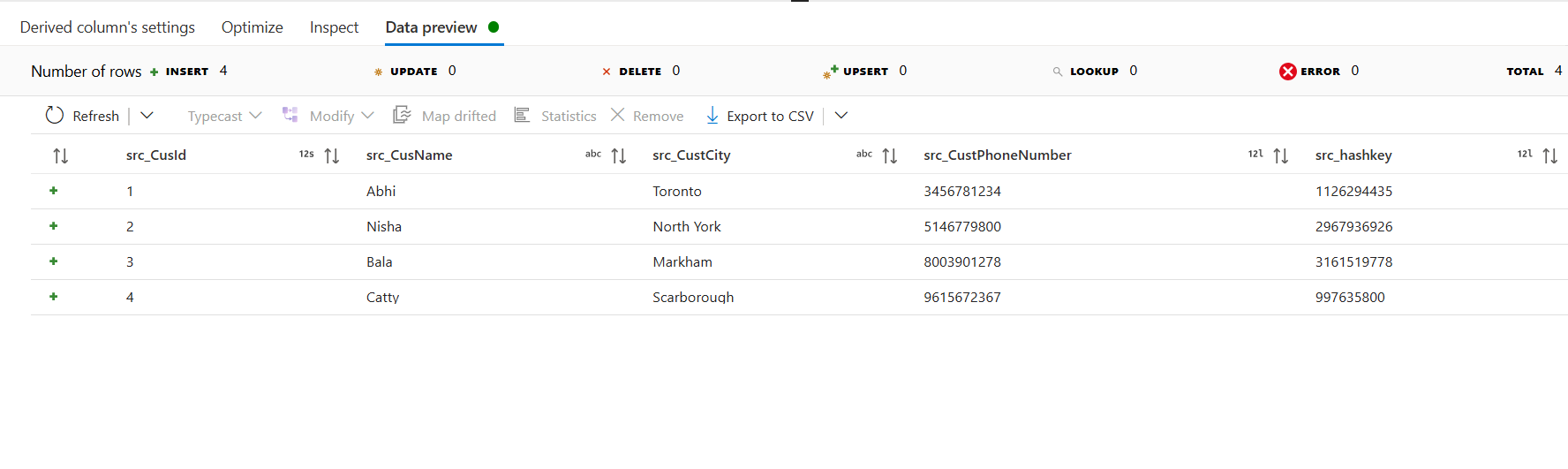


* And preview the data to check whether the column name has been changed or not

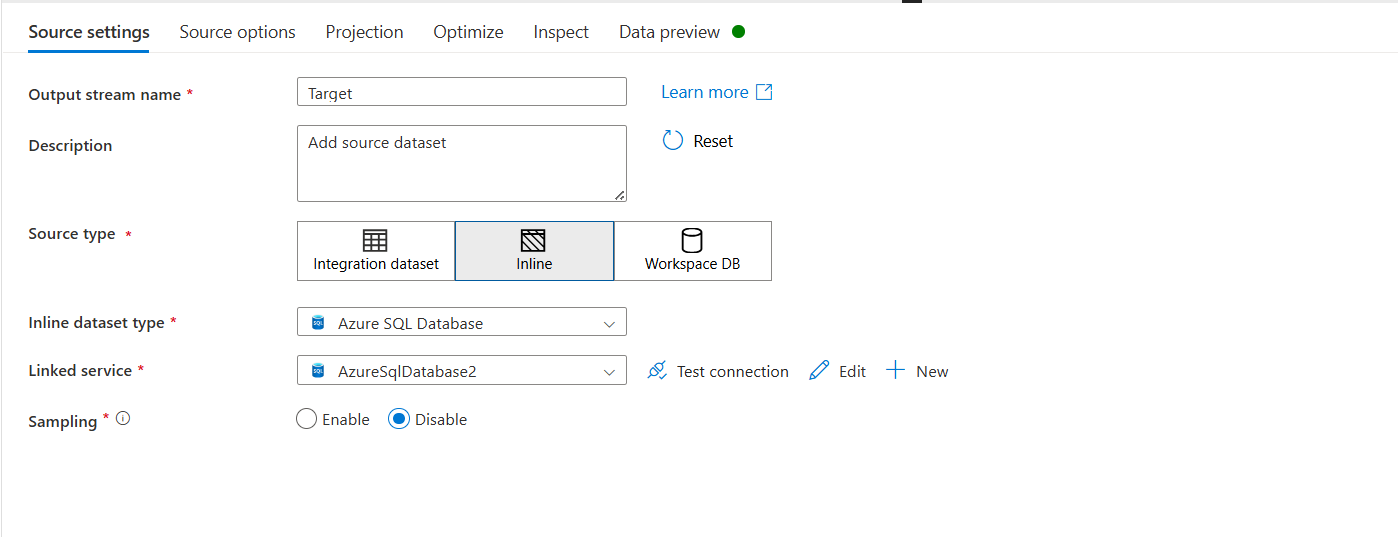


* Add derived column settings and rename it to “Hashkey” in columns field give src\_hashkey and in expression field give crc32(concat(toString(src\_CusId),src\_CusName,src\_CustCity,toString(src\_CustPhoneNumber))) and preview the data



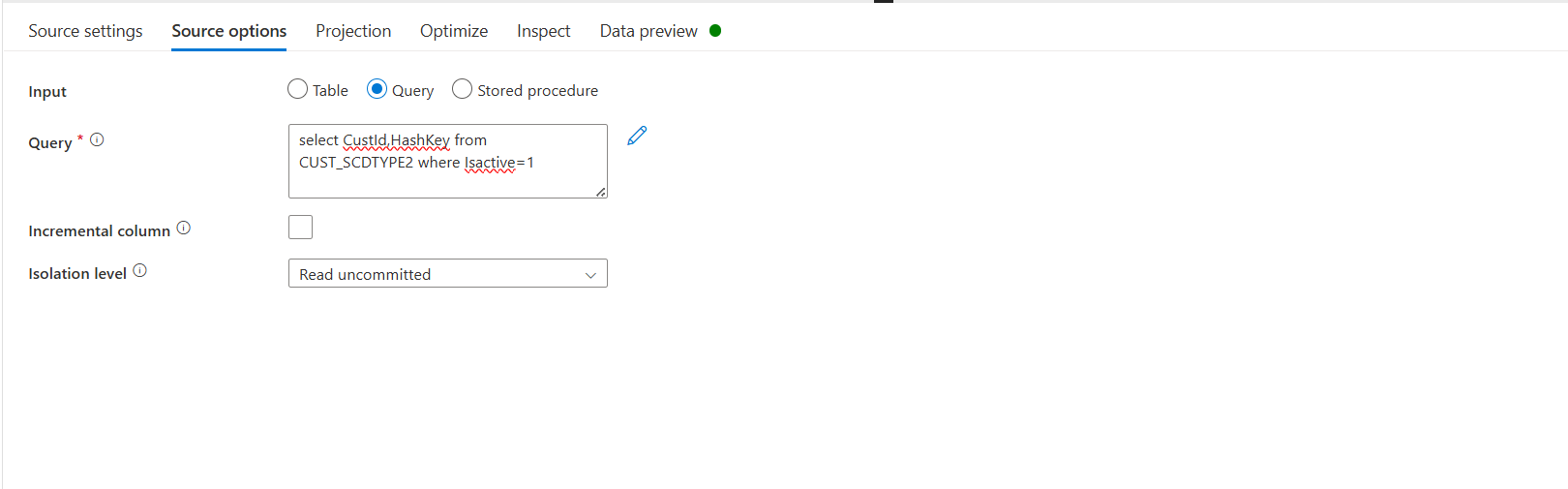


* Add a source and rename it to target and choose the **datatype** as **inline** and the **inline dataset type** as **Azure SQL Database**

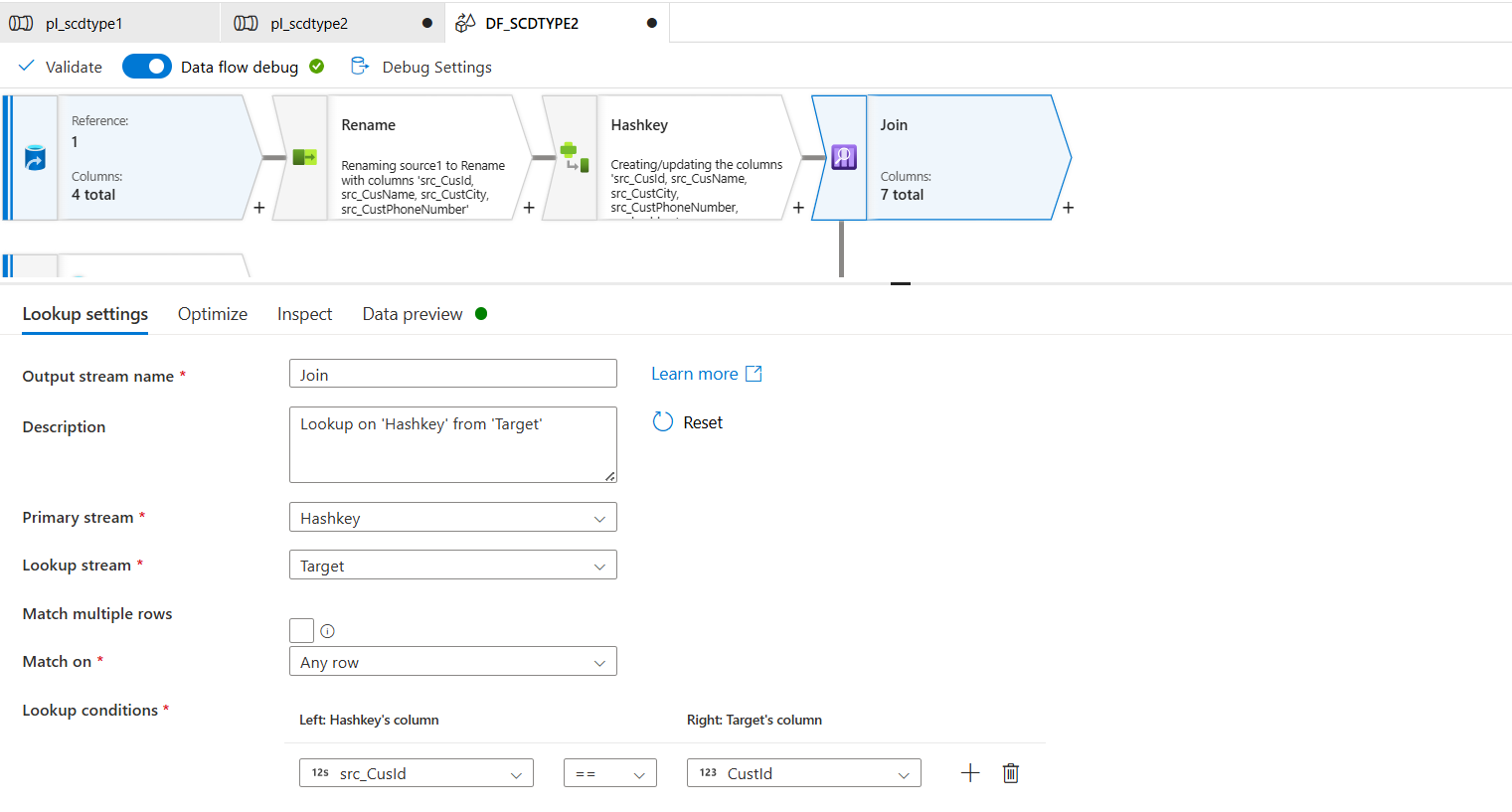


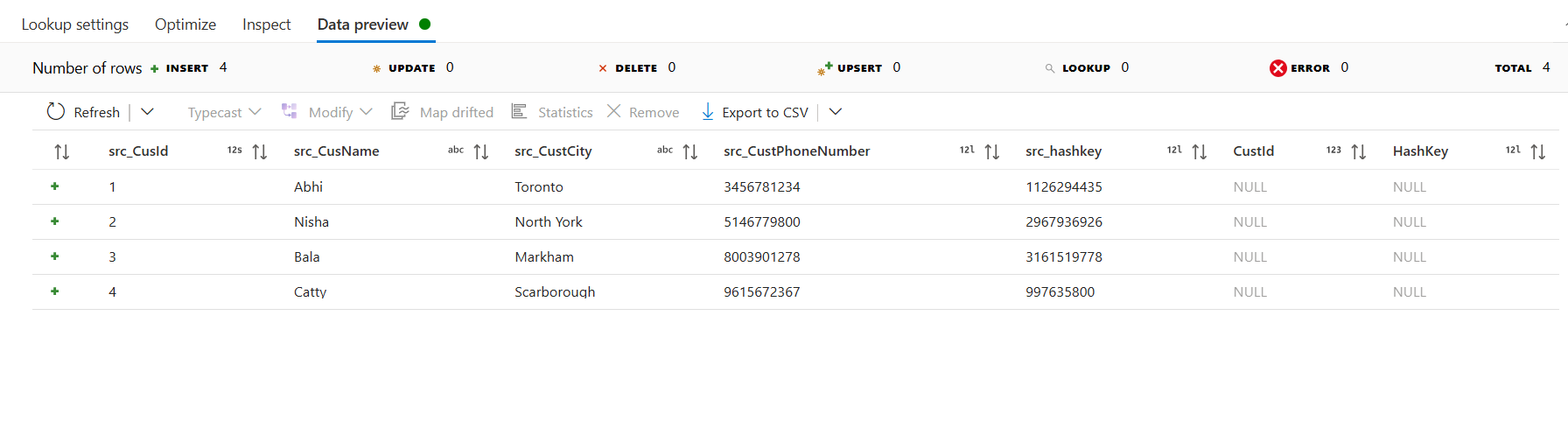
* In **source options** choose **query** as input and write the below query in the query field

select CustId,HashKey from CUST\_SCDTYPE2 where Isactive=1



* Import the schema and preview the data
* Next add Lookup and rename it to “Join”, in **Primary** stream choose **Hashkey** and in **Lookup** Stream choose **target**
* In hashkey column choose src\_custid and in target column choose custId and preview the data it should have Null values in CustId and Hashkey

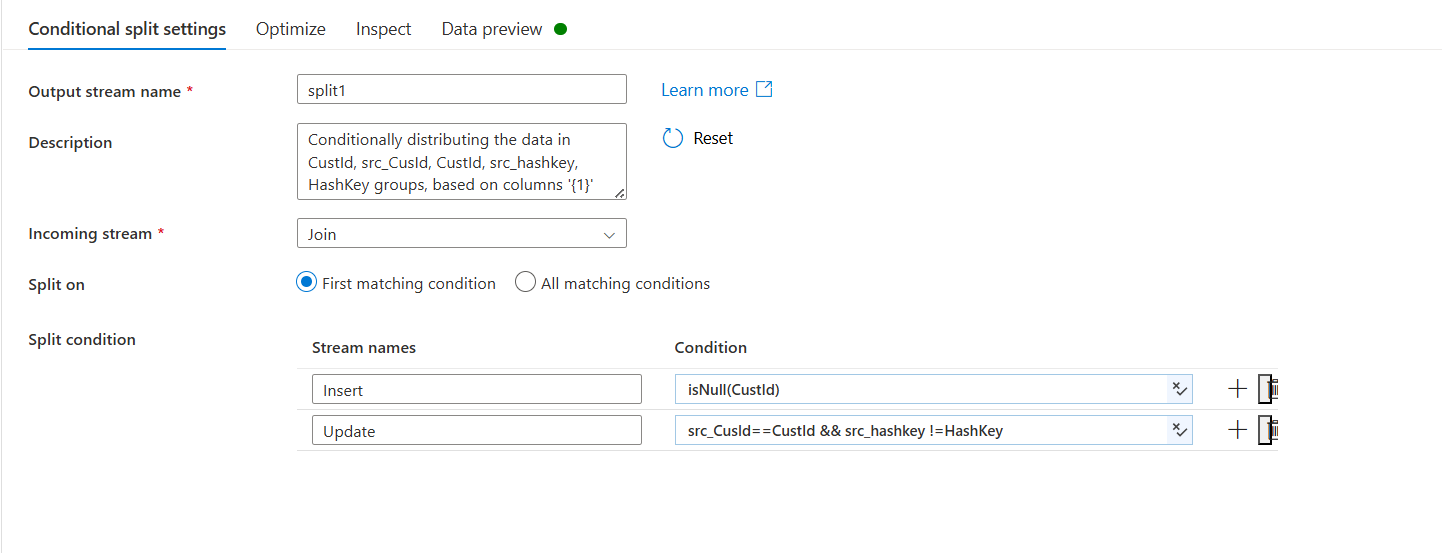




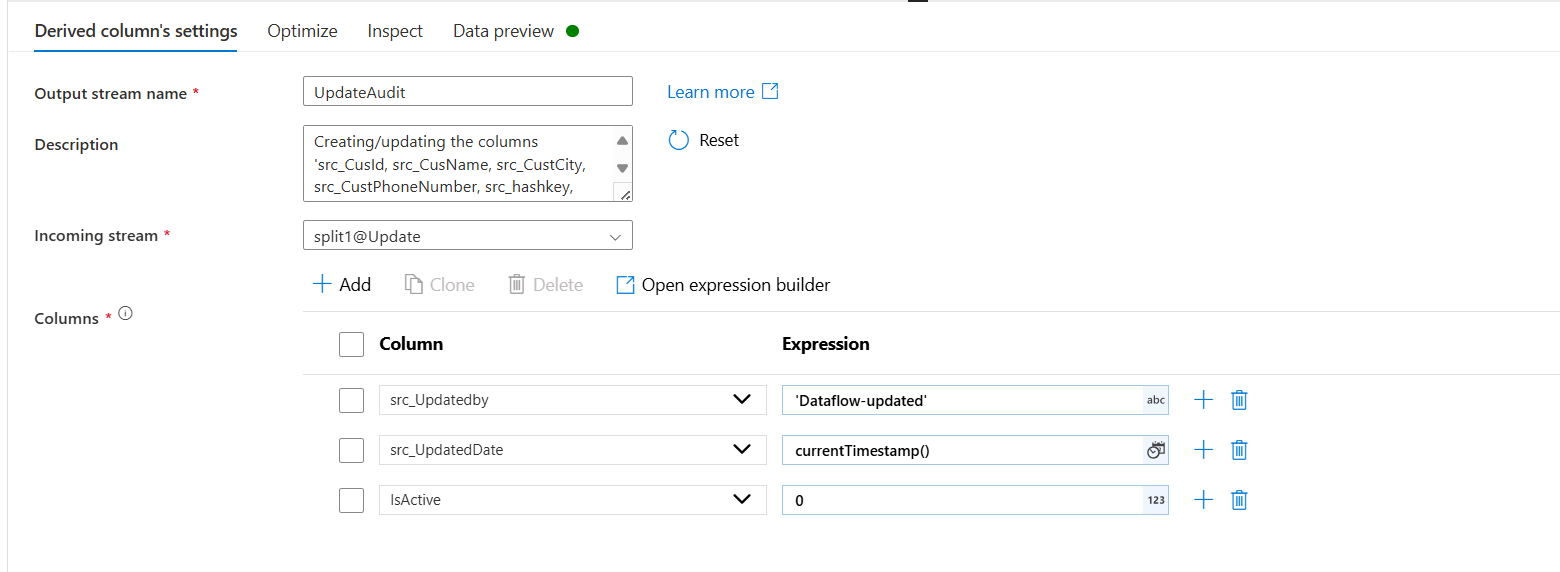
* Add conditional split and keep the stream names as Insert and Update

For Insert the Condition is isNull(CustId)

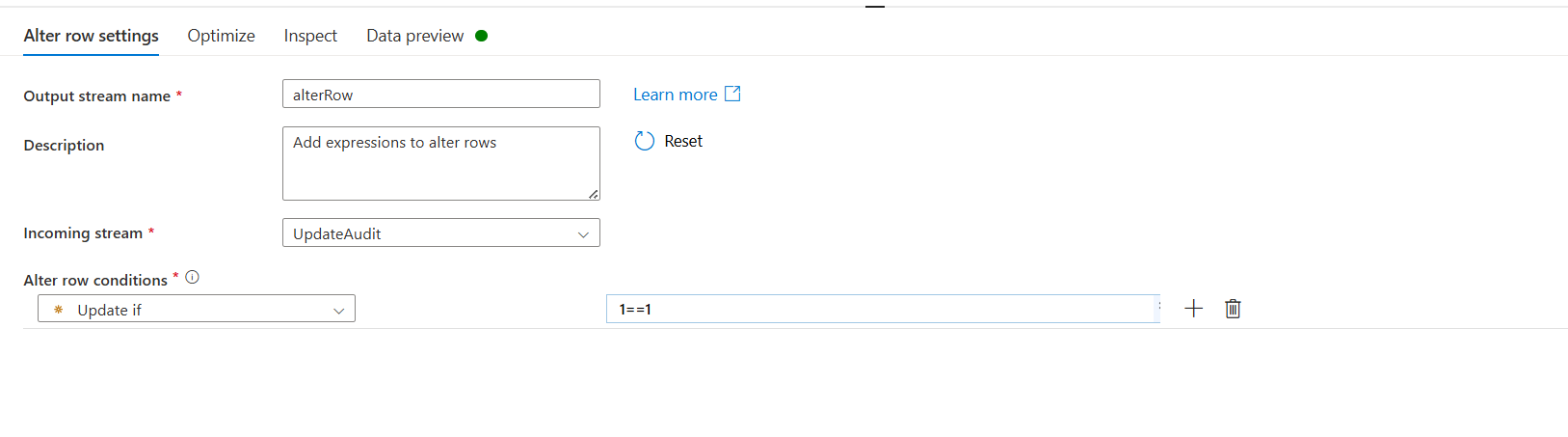
For Update the Condition is src\_CusId==CustId && src\_hashkey !=HashKey



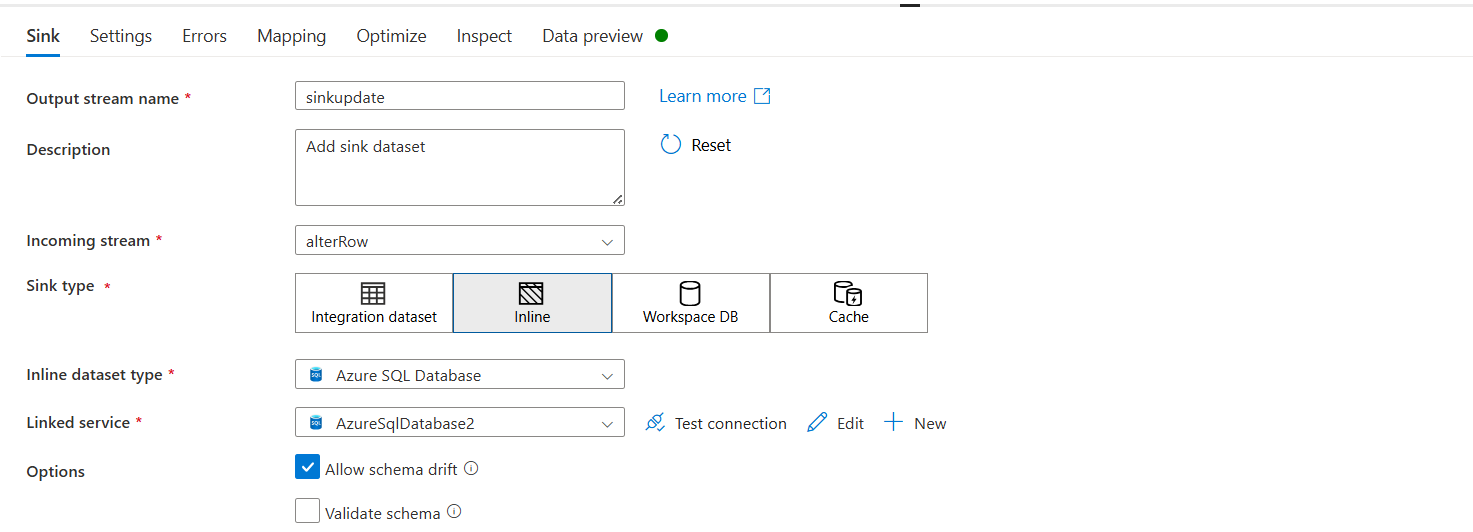
* In Update side add a derived column and rename it to update Audit and give the Column name and expression as below



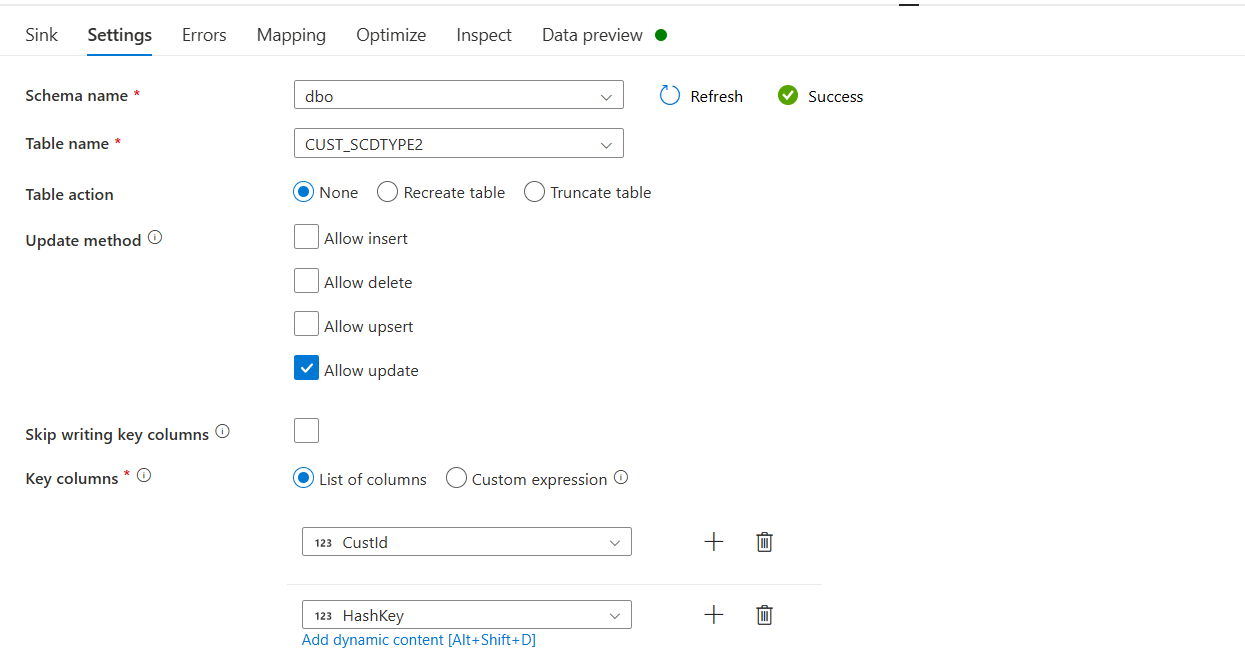
* Add alter row to the Derived column and set the conditions as below



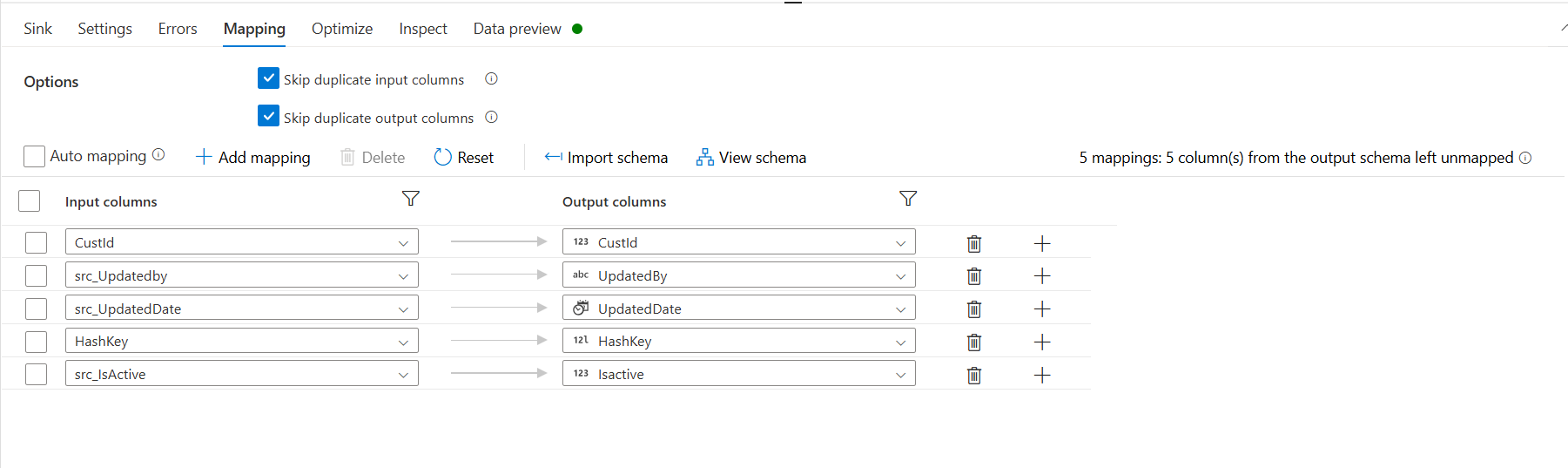
* Add a sink to the above one and rename it to “sinkupdate” and choose the sink type as Inline and Inline dataset type as Azure SQL Database in Sink Tab



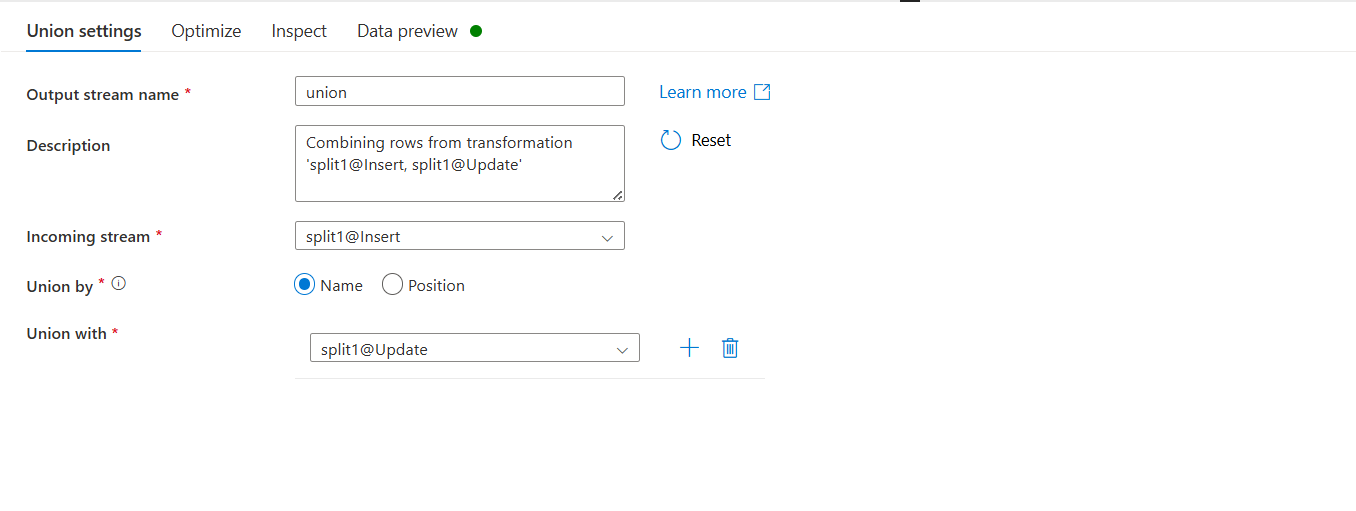
* In settings choose the correct table and keep the update method as follows and in key columns give custid and hashkey



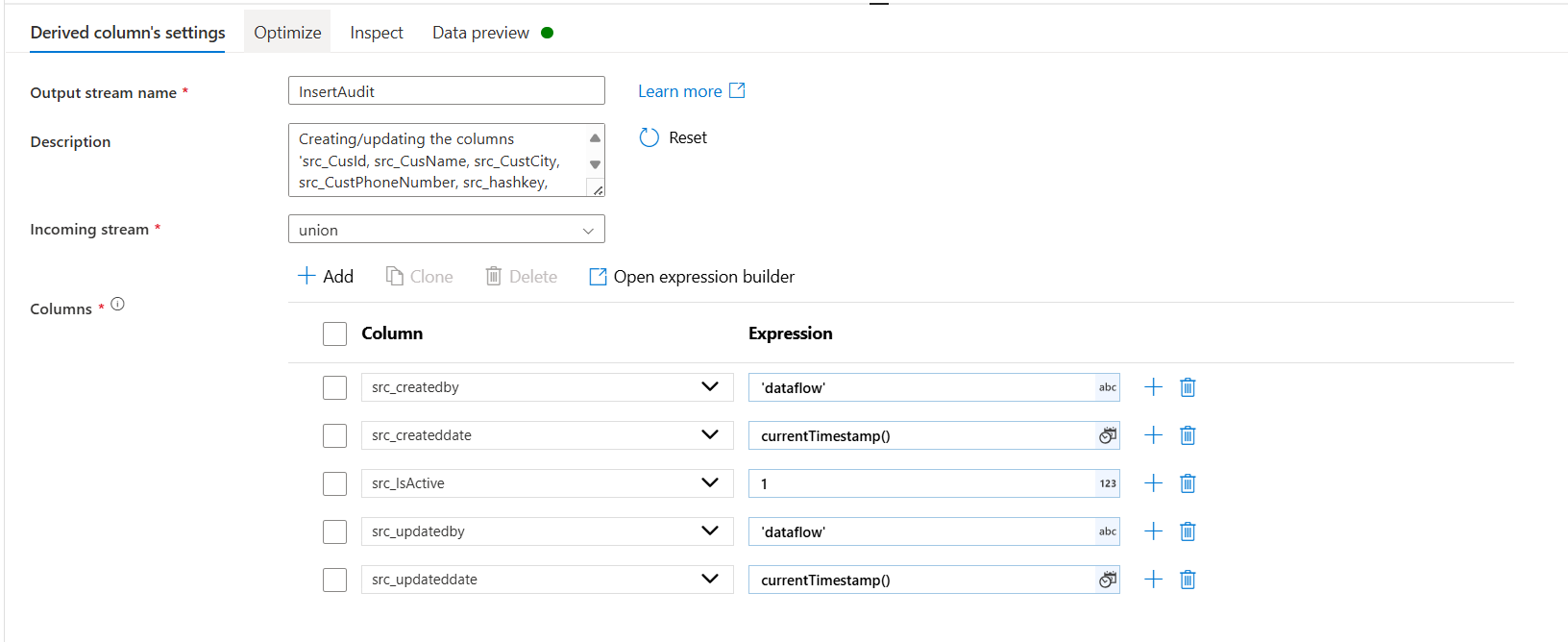
* In mapping just choose and map the columns which will be updated



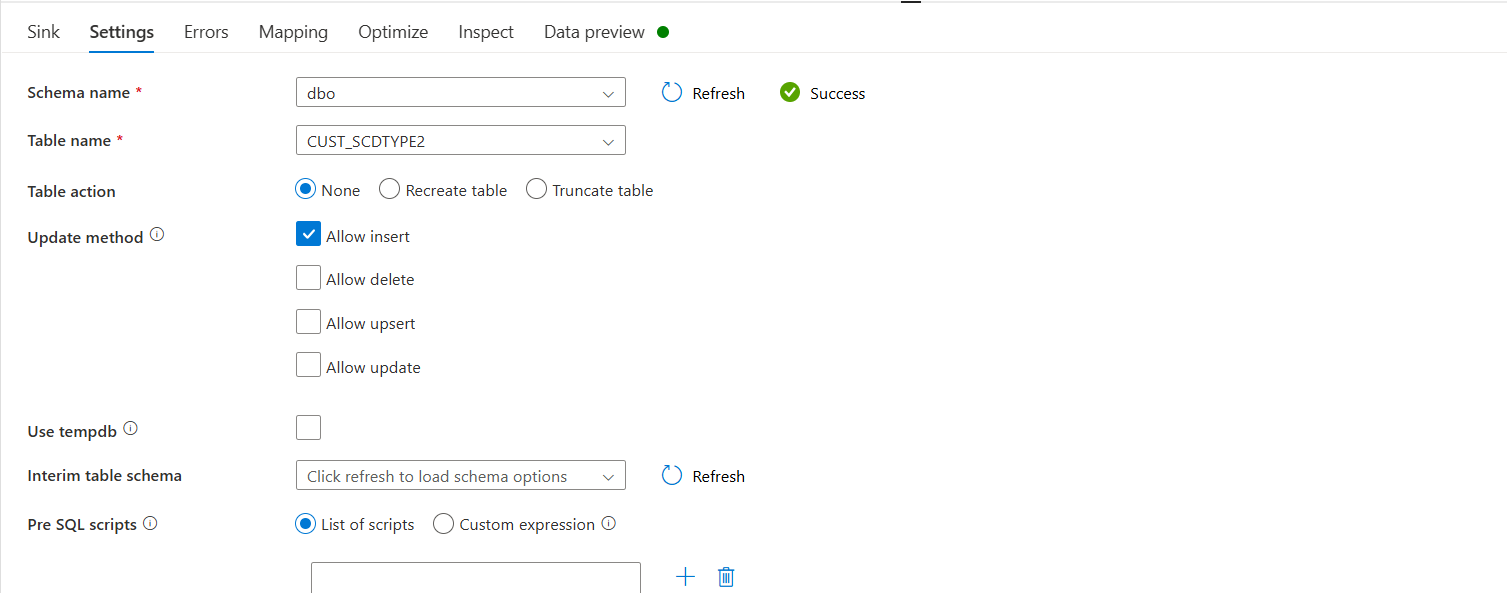
* A union transformation has been added to the insert to merge into a single flow



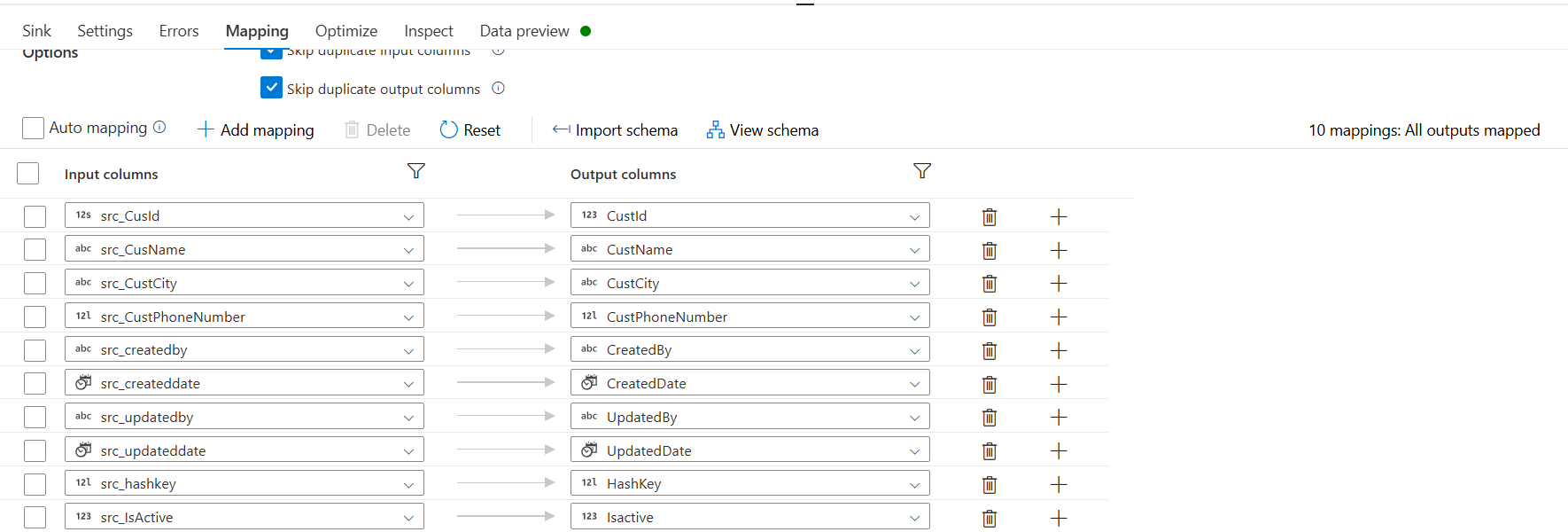
* Add a derived column and rename it to Insert Audit and give the Column name and expression as below

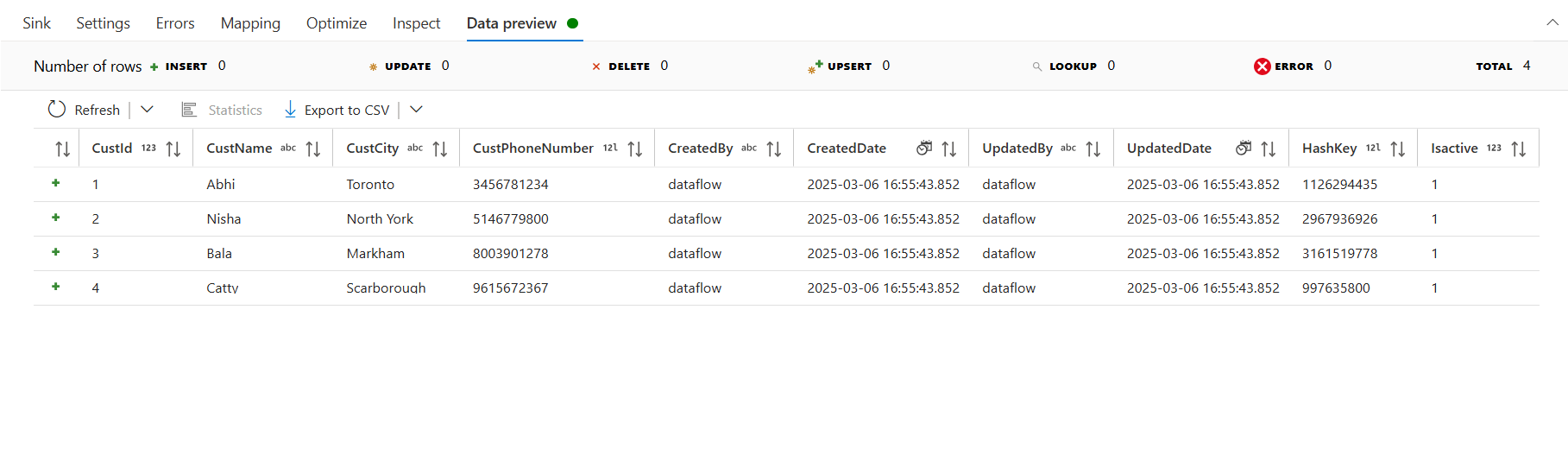


* In settings choose the correct table and keep the update method as follows

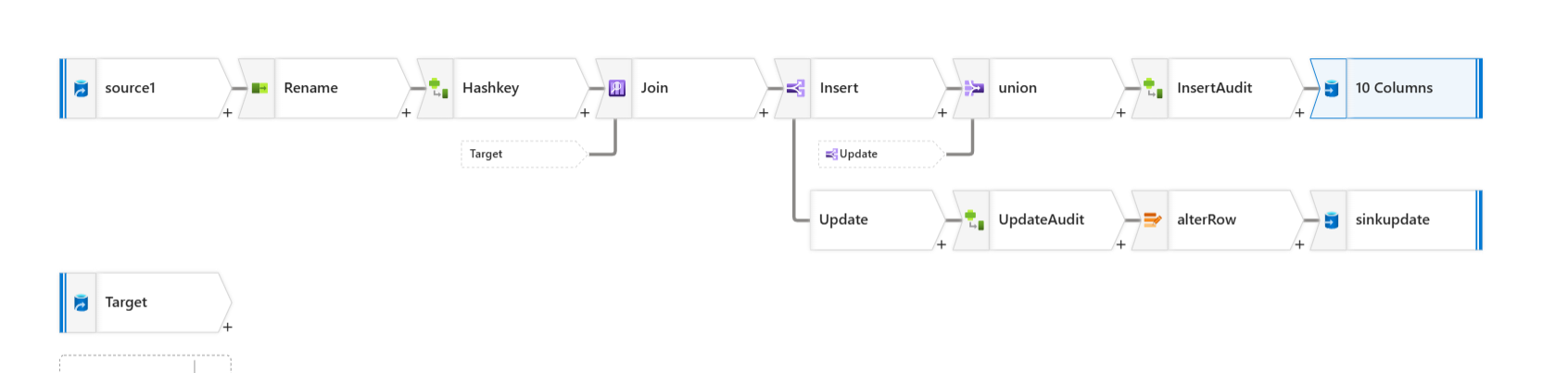


* In mapping just choose and map the columns which will be updated

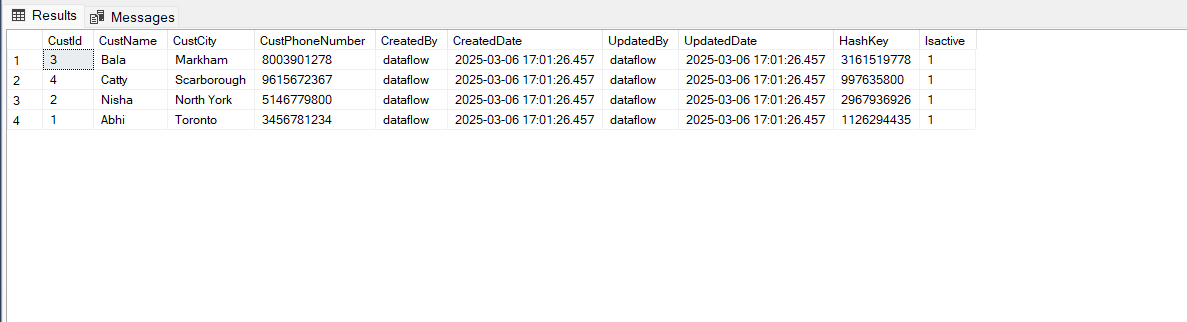




* The overall pipeline looks like below



* Below is the output if we are running the day 1 file



* Below is the output if we are running the day 2 file with new values and some changes in the existing data

