DATA FLOW CONCEPT (USING SCD TYPE 2)

SCD Type 2 (Slowly Changing Dimension Type 2) is a technique used in data warehousing to manage historical changes in dimension tables. It ensures that historical data is preserved when changes occur in attributes.

THE TRANSFORMATION PROPOSE

SELECT 🡪 IT WILL UPDATE THE COLUMN NAME, RENAME THE COLUMN, REMOVE UNNECESSARY COLUMNS

DERIVED COLUMN 🡪 IT WILL HELP TO GENERATE THE HASHKEY

LOOKUP 🡪IT WILL HELP TO MERGE DATA LIKE LEFT JOIN

CONDITIONAL SPLIT 🡪 USED TO ROUTE DATA INTO DIFFERENT STREAMS BASED ON CONDITIONS

ALTER ROW 🡪 USED TO DYNAMICALLY CONTROL ROW LEVEL ACTIONS (INSERT, UPDATE, DELETE, UPSERT(UPDATE+INSERT)

UNION 🡪 USED TO MERGE MULTIPLE COLLECTIONS INTO SINGLE COLLECTION. COMBINE DATA FROM DIFFERENT SOURCES OR DIFFERENT PROCESSING BRANCGES IN YOUR PIPELINE.

Create customer table

create table cust\_scdtype2

(custid int, custname varchar(100), custcity varchar(100), custphonenumber bigint, createdby varchar(100), createddate datetime, updateby varchar(100), updatedate datetime, hashkey bigint, isactive int)

A screenshot of a computer

AI-generated content may be incorrect.

Create a csv file in notepad

My day 1 file is

A screenshot of a computer

AI-generated content may be incorrect.

My day 2 file is

A screenshot of a computer

AI-generated content may be incorrect.

My day 3 file is

A screenshot of a phone number

AI-generated content may be incorrect.

Im gonna fetch the data from the csv to sqldb for that im gonna upload the file in my synapse workspace

A screenshot of a computer

AI-generated content may be incorrect.

STEPS TO CREATE DATA FLOW IN SYNAPSE

STEP 1 : To create dataflow before starting the process click on data flow debug

A screenshot of a computer

AI-generated content may be incorrect.

STEP 2 : drag and drop the data flow and in setting 🡪dataflow 🡪click for new 🡪give name for the data flow(sourcegen)

A screenshot of a computer

AI-generated content may be incorrect.

STEP 3 : My source is gen2 storage account which I have upload the 3 csv files

In sources setting 🡪source type click for inline 🡪inline data type change into delimited text 🡪 and in linked services give synapse storage account where you upload 3 csv files.

A screenshot of a computer

AI-generated content may be incorrect.

STEP 4 : in source options 🡪I gave my file path where my files are locatedA screenshot of a computer

AI-generated content may be incorrect.

STEP 5: In source options check first row as header

A black text on a white background

AI-generated content may be incorrect.

STEP 6 : In projection 🡪 click on import the schema

A screenshot of a computer

AI-generated content may be incorrect.

STEP 7 : Click on data preview (I have imported day1 csv file)

A screenshot of a computer

AI-generated content may be incorrect.

STEP 8 : By clicking + in the sourcesgen Im gonna start new transformation called select

In select transformation I wanna rename the column name by doing this step

Step1 : delete all the column in here

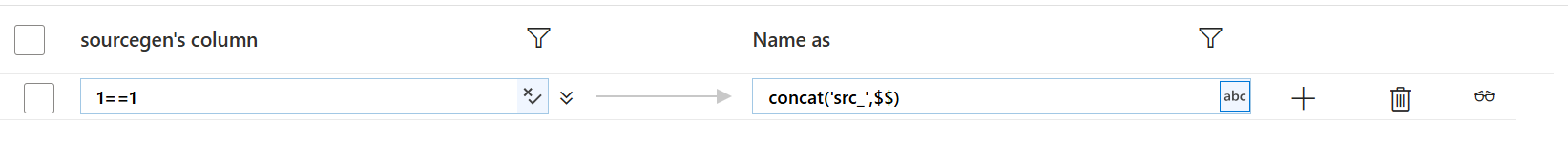
A screenshot of a computer

AI-generated content may be incorrect.

Step 2 : to do this delete all the column and click on + symbol click on rule-bases mapping.

Concat(‘src\_’,$$)

By doing this it will change all the column name into src\_id,src\_name,src\_city,src\_phonenumber



STEP 9 : click on preview data and refresh (column name changed)

A screenshot of a computer

AI-generated content may be incorrect.

STEP 10 : By clicking + in the **select**(rename) Im gonna start new transformation called drivedcolumn

A screenshot of a computer

AI-generated content may be incorrect.

STEP 11 : in derived column , have to create a new column (hashkey) and use concat expressions

crc32(concat(toString(src\_id),src\_name,src\_city,toString(src\_phonenumber)))

A screenshot of a computer

AI-generated content may be incorrect.

STEP 12 : I wanna add my target (sink)

In source setting 🡪source type 🡪inline 🡪inline dataset type 🡪sql db 🡪linked services 🡪servername (ssms)

A screenshot of a computer

AI-generated content may be incorrect.

STEP 13 : I wanna compare this 2 column whether this is existing or not

select custid,hashkey from cust\_scdtype2 where isactive=1

A screenshot of a computer

AI-generated content may be incorrect.

STEP 14 : click on projection 🡪click on import schema

A screenshot of a computer

AI-generated content may be incorrect. STEP 15: By clicking + in the derived(hashkey)🡪 start new transformation called lookup

A screenshot of a computer

AI-generated content may be incorrect.

STEP 16 : in lookup setting 🡪lookup stream (select target (sql db) and in lookup conditions 🡪 select src\_id for left and custid for right

A screenshot of a computer

AI-generated content may be incorrect.

STEP 17 : in data preview🡪refresh

A screenshot of a computer

AI-generated content may be incorrect.

In the above picture I can only see the src\_id,src\_name,src\_city,src\_phonenumber this is my csv file and in the custid,hashkey is null(I don’t inserted data in the sqldb), to bring data use conditional split

STEP 18 : By clicking + in the lookup🡪 start new transformation called conditional split

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

STEP 19 : if the custid is null consider this as a new record

isNull(custid)

If the custid is not null consider this as a existing record and update this

Hashkey is already generated and it will generate another hashkey use the below code code

src\_id==custid && src\_hashkey!=hashkey

A screenshot of a computer

AI-generated content may be incorrect.

STEP 20 : By clicking + in the update Im gonna start new transformation called derived column (to update audit column)

A screenshot of a computer

AI-generated content may be incorrect.

STEP 21 : I wanna update audit column (src\_updateby, src\_updatedate)

Add 🡪new columns

A screenshot of a computer

AI-generated content may be incorrect.

STEP 22 : By clicking + in the derivedcolumn Im gonna start new transformation called alter row

A computer screen shot of a diagram

AI-generated content may be incorrect.

STEP 23 : Permission to modify the data

A screenshot of a computer

AI-generated content may be incorrect.

Updateif 1==1

STEP 24 : By clicking + in the alterrow Im gonna start new transformation called sink

A screenshot of a computer

AI-generated content may be incorrect.

STEP 25 : give sql db and server name (ssms)

A screenshot of a computer

AI-generated content may be incorrect.

STEP 26 : in setting 🡪give schema and table name

A screenshot of a computer

AI-generated content may be incorrect.

STEP 27 : in mapping 🡪uncheck auto mapping 🡪click on importschema and click reset

A screenshot of a computer

AI-generated content may be incorrect.

STEP 28 : delete custname, custcity, custphonenumber, src\_createdby, src\_createddate I only need updateby, updatedate, hashkey, src\_active

A screenshot of a computer

AI-generated content may be incorrect.

STEP 29 : in setting give key column as custid (is my keycolumn based on custid I will update data)

A screenshot of a computer

AI-generated content may be incorrect.

STEP 30 : By clicking + in the insert Im gonna start new transformation called union

A screenshot of a computer

AI-generated content may be incorrect.

STEP 31 : In the union settings 🡪 union with 🡪 give split@update

A screenshot of a computer

AI-generated content may be incorrect.

STEP 32 : By clicking + in the union Im gonna start new transformation called derviedcolumn (insert audit column)

A diagram of a computer

AI-generated content may be incorrect.

STEP 33 : I wanna insert audit column (createdby, createddate, updateby, updatedate)

Add 🡪new columns

A screenshot of a computer

AI-generated content may be incorrect.

STEP 34 : By clicking + in the derived column (insert audit column) add new transformation called sink

A diagram of a software company

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

STEP 35 : In sink 🡪sink type click inline 🡪 inline dataset type click sqldb 🡪in linked services click your ssms(server)

A screenshot of a computer

AI-generated content may be incorrect.

STEP 36 : publish

A screenshot of a computer

AI-generated content may be incorrect.

STEP 37 : output of ssms

A screenshot of a computer

AI-generated content may be incorrect.