testcase12_parquet_parquet_mismatch_manual schema comparison

1. Run Summary

Application Name	:etl_pipeline_goldlayer
Protocol Name	:DATFDemo
Protocol File Path	:/Workspace/Shared/QE_ATF_Enhanced_03/datf_core/test/testprotocol/traversedtestprotocol_schema_loman.xlsx
Testcase Name	:testcase12_parquet_parquet_mismatch_manual
Testcase Type	:schema
Test Environment	:SIT
Start Time	:10-Apr-2025 07:53:26 UTC
End Time	:10-Apr-2025 07:56:49 UTC
Run Time	:0:03:23
Test Result	:Failed
Reason	:Content mismatched

2. Configuration Details

Compare Type	:likeobjectcompare
testquerygenerationmode	:Auto
Testcase Type	:schema
Source Connection Name	:RawS3Bucket
Source Connection Type	:aws-s3
Source Connection Value	:RawS3Bucket
Source Format	:parquet
Source Name	:patients_source_parquet
Source Path	:test/data/source
Target Connection Name	:CuratedS3Bucket
Target Connection Type	:aws-s3
Target Connection Value	:CuratedS3Bucket
Target Format	:parquet
Target Name	:patients_source_parquet
Target Path	:test/data/source
S2T Path	:test/s2t/s2t_1_parquet_parquet_mismatch.xlsx
Primary Keys	:id

3. Schema Summary

Test Result	:Failed
No. of matched columns	:12
No. of columns compared	:10
No. of cols in Source but not in Target	:0
No. of cols in Target but not in Source	:0
No. of rows in Source	:50
No. of distinct rows in Source	:50
No. of duplicate rows in Source	:0
No. of rows in Target	:50
No. of distinct rows in Target	:50
No. of duplicate rows in Target	:0
No. of matched rows	:0
No. of mismatched rows	:50
No. of rows in Source but not in Target	:0
No. of rows in Target but not in Source	:0

4. SQL Queries

4.1 Source Query

select * from s2tmappingsheet.schema where tabletype in ('source','target')

4.2 Target Query

5. Sample Mismatches 8 rows

5.1 Keys in source but not in target

None

5.2 Keys in target but not in source

None

5.3 Keys having one or more unequal column values

S.No	Key Columns
1	tabletype=source, columnname= ADDRESS
2	tabletype=source, columnname= BIRTHDATE
3	tabletype=source, columnname= BIRTHPLACE
4	tabletype=source, columnname= CITY
5	tabletype=source, columnname= COUNTY
6	tabletype=source, columnname= DEATHDATE
7	tabletype=source, columnname= DRIVERS
8	tabletype=source, columnname= ETHNICITY

6. Columnwise Mismatch Summary

No. of rows in Source   :50
No. of rows in Target   :50

S.No	Source Column Name	Target Column Name	Rows in Source	Rows in Target	Rows with Common Keys	Rows Matched	Rows Mismatch
1	tablename	tablename	50	50	0	0	50
2	columnindex	columnindex	50	50	0	0	50
3	datatype	datatype	50	50	0	46	4
4	primarykey	primarykey	50	50	0	48	2

7. Columnwise Mismatch Details

7.1 tablename

S.No	Key Columns	Source value	Target value
1	tabletype=source, columnname= id	sourcedata	rawpatients
2	tabletype=target, columnname= COUNTY	targetdata	curatedpatients
3	tabletype=source, columnname= LAST	sourcedata	rawpatients
4	tabletype=source, columnname= COUNTY	sourcedata	rawpatients
5	tabletype=source, columnname= RACE	sourcedata	rawpatients
6	tabletype=source, columnname= ETHNICITY	sourcedata	rawpatients
7	tabletype=target, columnname= MAIDEN	targetdata	curatedpatients
8	tabletype=target, columnname= CITY	targetdata	curatedpatients

7.2 columnindex

S.No	Key Columns	Source value	Target value
1	tabletype=source, columnname= id	1	
2	tabletype=source, columnname= LAST	9	
3	tabletype=target, columnname= COUNTY	20	
4	tabletype=source, columnname= COUNTY	20	
5	tabletype=source, columnname= RACE	13	

6	tabletype=source, columnname= ETHNICITY	14	
7	tabletype=target, columnname= CITY	18	
8	tabletype=target, columnname= MAIDEN	11	

7.3 datatype

S.No	Key Columns	Source value	Target value
1	tabletype=source, columnname= DEATHDATE	datetime	date
2	tabletype=target, columnname= DEATHDATE	datetime	date
3	tabletype=source, columnname= BIRTHDATE	datetime	date
4	tabletype=target, columnname= BIRTHDATE	datetime	date

7.4 primarykey

S.No	Key Columns	Source value	Target value
1	tabletype=source, columnname= id	Y	
2	tabletype=target, columnname= id	Y	