testcase13_csv_parquet_match_manuel

content 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	:testcase13_csv_parquet_match_manuel
Testcase Type	:content
Test Environment	:SIT
Start Time	:10-Apr-2025 08:41:36 UTC
End Time	:10-Apr-2025 08:45:38 UTC
Run Time	:0:04:01
Test Result	:Failed
Reason	:Content mismatched

2. Configuration Details

Compare Type	:s2tcompare
testquerygenerationmode	:Manual
Testcase Type	:content
Source Connection Name	:RawS3Bucket
Source Connection Type	:aws-s3
Source Connection Value	:RawS3Bucket
Source Format	:delimited
Source Name	:patients_source.csv
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/source_csv_target_parquet_s2t_5.xlsx
Primary Keys	:id

3. Content Summary

Test Result	:Failed
No. of matched columns	:25
No. of columns compared	:24
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	:1,171
No. of distinct rows in Source	:1,171
No. of duplicate rows in Source	:0
No. of rows in Target	:1,171
No. of distinct rows in Target	:1,171
No. of duplicate rows in Target	:0
No. of matched rows	:1,170
No. of mismatched rows	:0
No. of rows in Source but not in Target	:1
No. of rows in Target but not in Source	:1

4. SQL Queries

4.1 Source Query

SELECT src.DEATHDATE as DEATHDATE, src.HEALTHCARE_EXPENSES as HEALTHCARE_EXPENSES, src.SUFFIX as SUFFIX, src.PASSPORT as PASSPORT, src.CITY as CITY, src.BIRTHPLACE as BIRTHPLACE, src.MAIDEN as MAIDEN, src.MARITAL as MARITAL, src.DRIVERS as DRIVERS, src.STATE as STATE, src.COUNTY as COUNTY, src.ADDRESS as ADDRESS, src.PREFIX as PREFIX, src.ZIP as ZIP, src.LON as LON, src.SSN as SSN, src.LAST as LAST, src.BIRTHDATE as BIRTHDATE, src.RACE as RACE, src.ETHNICITY as ETHNICITY, src.GENDER as GENDER, src.id as

id, src.HEALTHCARE_COVERAGE as HEALTHCARE_COVERAGE, src.LAT as LAT, src.FIRST as FIRST FROM
rawpatients src

4.2 Target Query

SELECT tgt.DEATHDATE as DEATHDATE, tgt.HEALTHCARE_EXPENSES as HEALTHCARE_EXPENSES, tgt.SUFFIX
as SUFFIX, tgt.PASSPORT as PASSPORT, tgt.CITY as CITY, tgt.BIRTHPLACE as BIRTHPLACE, tgt.MAIDEN as
MAIDEN, tgt.MARITAL as MARITAL, tgt.DRIVERS as DRIVERS, tgt.STATE as STATE, tgt.COUNTY as COUNTY,
tgt.ADDRESS as ADDRESS, tgt.PREFIX as PREFIX, tgt.ZIP as ZIP, tgt.LON as LON, tgt.SSN as SSN, tgt.LAST as LAST,
tgt.BIRTHDATE as BIRTHDATE, tgt.RACE as RACE, tgt.ETHNICITY as ETHNICITY, tgt.GENDER as GENDER, tgt.id as
id, tgt.HEALTHCARE_COVERAGE as HEALTHCARE_COVERAGE, tgt.LAT as LAT, tgt.FIRST as FIRST FROM
curatedpatients tgt

5. Sample Mismatches 5 rows

5.1 Keys in source but not in target

S.No	Key Columns
1	id=1d604da9-9a81-4ba9-80c2-de3375d59b40

5.2 Keys in target but not in source

S.No	Key Columns
1	id=1d604da9-9a81-4ba9-80c2-de3375d59b40

5.3 Keys having one or more unequal column values

None

6. Columnwise Mismatch Summary

None

7. Columnwise Mismatch Details

None