

testcase12_parquet_parquet_mismatch_manual
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.xlsx
Testcase Name	:testcase12_parquet_parquet_mismatch_manual
Testcase Type	:content
Test Environment	:SIT
Start Time	:08-Apr-2025 06:19:41 UTC
End Time	:08-Apr-2025 06:22:58 UTC
Run Time	:0:03:16
Test Result	:Failed
Reason	:Content mismatched

2. Configuration Details

Compare Type	:likeobjectcompare
testquerygenerationmode	:Manual
Testcase Type	:content
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
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	:0
No. of mismatched rows	:1,171
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 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, upper(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 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	id=00185faa-2760-4218-9bf5-db301acf8274
2	id=0042862c-9889-4a2e-b782-fac1e540ecb4
3	id=0047123f-12e7-486c-82df-53b3a450e365
4	id=010d4a3a-2316-45ed-ae15-16f01c611674
5	id=01207ecd-9dff-4754-8887-4652eda231e2
6	id=0149d553-f571-4e99-867e-fcb9625d07c2
7	id=01e1f394-7219-4189-bceb-3cbd90cff90b
8	id=023a7d29-32b3-4db5-89c8-b88bd7582ec0

6. Columnwise Mismatch Summary

No. of rows in Source :1171
No. of rows in Target :1171

S.No	Source Column Name	Target Column Name	Rows in Source	Rows in Target	Rows with Common Keys	Rows Matched	Rows Mismatch
1	CITY	CITY	1171	1171	0	0	1171

7. Columnwise Mismatch Details

7.1 CITY

S.No	Key Columns	Source value	Target value
1	id=58100357-85b6-4334-bc2a-8998c24a4812	Lunenburg	LUNENBURG
2	id=7b51cca8-284b-42a5-8117-cc21f48454f5	Cheshire	CHESHIRE
3	id=cf12424d-63da-4fbb-b93f-22308793cafa	Needham	NEEDHAM
4	id=d95287b4-213f-4b3b-b68e-b23b9f3bac90	Fairhaven	FAIRHAVEN
5	id=eedb8836-d94c-4a70-8491-bea52efc4ce8	Millis	MILLIS
6	id=9e5a9e59-cfe3-42de-8c2e-76ef7e2efa5f	West Springfield	WEST SPRINGFIELD
7	id=a60c2937-a6a7-4974-8c9b-f34370c29b7e	Tewksbury	TEWKSBURY
8	id=431971f5-a39f-44f8-9ef9-10502db973c9	Lynn	LYNN