

testcase12_parquet_parquet_mismatch_manual null comparison

1. Run Summary

Application Name :etl_pipeline_goldlayer
Protocol Name :DATFDemo
Protocol File Path :datf_core//test/testprotocol/testselection_template.xlsx
Testcase Name :testcase12_parquet_parquet_mismatch_manual
Testcase Type :null
Test Environment :SIT
Start Time :11-Mar-2025 14:37:59 UTC
End Time :11-Mar-2025 14:39:37 UTC
Run Time :0:01:37
Test Result :Failed
Reason :Nulls are not matching for each column

2. Configuration Details

Compare Type :likeobjectcompare
testquerygenerationmode :Manual
Testcase Type :null
Source Connection Name :RawS3Bucket
Source Connection Type :aws-s3
Source Connection Value :RawS3Bucket
Source Format :parquet
Source Name :patient_data_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 :patient_data_target_parquet
Target Path :test/data/target
Primary Keys :id

3. Null Summary

Test Result :Failed
No of column has null in source :5
No of column has null in target :5
No of columns has null count match :0
No of columns has null count mismatch :5

4. SQL Queries

4.1 Source Query

```
SELECT  
ADDRESS,BIRTHDATE,BIRTHPLACE,CITY,COUNTY,DEATHDATE,DRIVERS,ETHNICITY,FIRST,GENDER,HEALTH  
CARE_COVERAGE,HEALTHCARE_EXPENSES,LAST,LAT,LON,MAIDEN,MARITAL,PASSPORT,PREFIX,RACE,SSN,  
STATE,SUFFIX,ZIP,id FROM rawpatients
```

4.2 Target Query

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

5. Columns having nulls at source and target

5.1 Columns having nulls in source

S.No	Column	Count
------	--------	-------

1	DEATHDATE	900877
2	DRIVERS	200539
3	MAIDEN	800319
4	PASSPORT	300807
5	SUFFIX	800558

5.2 Columns having nulls in target

S.No	Column	Count
1	DEATHDATE	901273
2	DRIVERS	200008
3	MAIDEN	800625
4	PASSPORT	299989
5	SUFFIX	800405

5.3 Columns having null count mismatch between source and target

S.No	Source Column Name	Src Null Count	Target Column Name	Tgt Null Count
1	DEATHDATE	900877	DEATHDATE	901273
2	DRIVERS	200539	DRIVERS	200008
3	MAIDEN	800319	MAIDEN	800625
4	PASSPORT	300807	PASSPORT	299989
5	SUFFIX	800558	SUFFIX	800405