

testcase2_parquet_parquet_match
duplicate comparison

1. Run Summary

Application Name	:sample
Protocol Name	:DATFDemo
Protocol File Path	:/Workspace/Repos/sahil.gupta@tigeranalytics.com/QE_ATF/datf_core/test/testprotocol/testprotocol.xlsx
Testcase Name	:testcase2_parquet_parquet_match
Testcase Type	:duplicate
Test Environment	:dev
Start Time	:30-Aug-2024 12:00:25 UTC
End Time	:30-Aug-2024 12:00:35 UTC
Run Time	:0:00:10
Test Result	:Failed
Reason	:Duplicates

2. Configuration Details

Compare Type	:s2tcompare
testquerygenerationmode	:Auto
Testcase Type	:duplicate
Source Connection Type	:aws-s3
Source Format	:parquet
Source Path	:/Workspace/Repos/sahil.gupta@tigeranalytics.com/QE_ATF/datf_core/test/data/source/patients_source_parquet
Target Connection Type	:aws-s3
Target Format	:delta
Target Path	:hive_metastore.default.patients_target_modified_new
S2T Path	:test/s2t/source_parquet_target_delta_auto.xlsx
Primary Keys	:id_new

3. Duplicate Summary

Test Result	:Failed
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	:1,169
No. of distinct rows in Target	:1,169
No. of duplicate rows in Target	:0

4. SQL Queries

4.1 Source Query

SELECT Count(DISTINCT id_new) FROM
parquet.`/Workspace/Repos/sahil.gupta@tigeranalytics.com/QE_ATF/datf_core/test/data/source/patients_source_parquet`

4.2 Target Query

SELECT Count(DISTINCT id_new) FROM delta.`hive_metastore.default.patients_target_modified_new`