

testcase12_parquet_parquet_mismatch_manual

duplicate 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	:duplicate
Test Environment	:SIT
Start Time	:18-Mar-2025 04:44:41 UTC
End Time	:18-Mar-2025 04:45:08 UTC
Run Time	:0:00:27
Test Result	:Failed
Reason	:Duplicates

2. Configuration Details

Compare Type	:likeobjectcompare
testquerygenerationmode	:Manual
Testcase Type	:duplicate
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. Duplicate Summary

Test Result	:Failed
No. of rows in Source	:1,001,000
No. of distinct rows in Source	:1,000,000
No. of duplicate rows in Source	:1,000
No. of rows in Target	:1,001,000
No. of distinct rows in Target	:1,000,000
No. of duplicate rows in Target	:1,000

4. SQL Queries

4.1 Source Query

SELECT Count(DISTINCT id) FROM rawpatients

4.2 Target Query

SELECT Count(DISTINCT id) FROM curatedpatients

testcase23_parquet_snowflake_mismatch

duplicate comparison

1. Run Summary

Application Name	:etl_pipeline_goldlayer
Protocol Name	:DATFDemo
Protocol File Path	:datf_core//test/testprotocol/testselection_template.xlsx
Testcase Name	:testcase23_parquet_snowflake_mismatch
Testcase Type	:duplicate
Test Environment	:SIT
Start Time	:18-Mar-2025 04:45:08 UTC
End Time	:18-Mar-2025 04:45:18 UTC
Run Time	:0:00:09
Test Result	:Passed
Reason	:No Duplicates

2. Configuration Details

Compare Type	:likeobjectcompare
testquerygenerationmode	:Auto
Testcase Type	:duplicate
Source Connection Name	:RawS3Parquet
Source Connection Type	:aws-s3
Source Connection Value	:RawS3Parquet
Source Format	:parquet
Source Name	:patients_source_parquet
Source Path	:test/data/source
Source Exclude Columns	:BIRTHDATE,DEATHDATE
Target Connection Name	:raw_snowflake_sql_connection
Target Connection Type	:snowflake
Target Connection Value	:raw_snowflake_sql_connection
Target Format	:table
Target Name	:SAMPLE_SRC_PATIENTS
Target Exclude Columns	:BIRTHDATE,DEATHDATE
Primary Keys	:id

3. Duplicate Summary

Test Result	:Passed
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

4. SQL Queries

4.1 Source Query

SELECT Count(DISTINCT id) FROM actualpatients

4.2 Target Query

SELECT Count(DISTINCT id) FROM SAMPLE_SRC_PATIENTS

testcase24_snowflake_snowflake_etljob

duplicate comparison

1. Run Summary

Application Name	:etl_pipeline_goldlayer
Protocol Name	:DATFDemo
Protocol File Path	:datf_core//test/testprotocol/testselection_template.xlsx
Testcase Name	:testcase24_snowflake_snowflake_etljob
Testcase Type	:duplicate
Test Environment	:SIT
Start Time	:18-Mar-2025 04:45:18 UTC
End Time	:18-Mar-2025 04:45:21 UTC
Run Time	:0:00:03
Test Result	:Passed
Reason	:No Duplicates

2. Configuration Details

Compare Type	:likeobjectcompare
testquerygenerationmode	:Auto
Testcase Type	:duplicate
Source Connection Name	:raw_snowflake_sql_connection
Source Connection Type	:snowflake
Source Connection Value	:raw_snowflake_sql_connection
Source Format	:table
Source Name	:SAMPLE_SRC_PATIENTS
Source Exclude Columns	:BIRTHDATE,DEATHDATE
Target Connection Name	:raw_snowflake_sql_connection
Target Connection Type	:snowflake
Target Connection Value	:raw_snowflake_sql_connection
Target Format	:table
Target Name	:SAMPLE_TGT_PATIENTS
Target Exclude Columns	:BIRTHDATE,DEATHDATE
Primary Keys	:ID

3. Duplicate Summary

Test Result	:Passed
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) FROM SAMPLE_SRC_PATIENTS

4.2 Target Query

SELECT Count(DISTINCT ID) FROM SAMPLE_TGT_PATIENTS