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	:18-Mar-2025 04:46:22 UTC
End Time	:18-Mar-2025 04:48:18 UTC
Run Time	:0:01:56
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

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	:null
Test Environment	:SIT
Start Time	:18-Mar-2025 04:48:52 UTC
End Time	:18-Mar-2025 04:49:42 UTC
Run Time	:0:00:49
Test Result	:Failed
Reason	:Nulls are not matching for each column

2. Configuration Details

Compare Type	:likeobjectcompare
testquerygenerationmode	:Auto
Testcase Type	:null
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. Null Summary

Test Result	:Failed
No of column has null in source	:7
No of column has null in target	:7
No of columns has null count match	:1
No of columns has null count mismatch	:8

4. SQL Queries
4.1 Source Query

SELECT
ADDRESS,BIRTHPLACE,CITY,COUNTY,DRIVERS,ETHNICITY,FIRST,GENDER,HEALTHCARE_COVERAGE,HEALT
HCARE_EXPENSES,LAST,LAT,LON,MAIDEN,MARITAL,PASSPORT,PREFIX,RACE,SSN,STATE,SUFFIX,ZIP,id
FROM actualpatients

4.2 Target Query

SELECT
ADDRESS,BIRTHPLACE,CITY,COUNTY,DRIVERS,ETHNICITY,FIRST,GENDER,HEALTHCARE_COVERAGE,HEALT
HCARE_EXPENSES,ID,LAST,LAT,LON,MAIDEN,MARITAL,PASSPORT,PREFIX,RACE,SSN,STATE,SUFFIX,ZIP
FROM SAMPLE_SRC_PATIENTS

5. Columns having nulls at source and target
5.1 Columns having nulls in source

S.No	Column	Count

1	DRIVERS	213
2	MAIDEN	840
3	MARITAL	380
4	PASSPORT	273
5	PREFIX	244
6	SUFFIX	1159
7	ZIP	543

5.2 Columns having nulls in target

S.No	Column	Count
1	DRIVERS	213
2	MAIDEN	840
3	MARITAL	380
4	PASSPORT	273
5	PREFIX	244
6	SUFFIX	1159
7	ZIP	543

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	MAIDEN	840	LON	0
2	MARITAL	380	MAIDEN	840
3	PASSPORT	273	MARITAL	380
4	PREFIX	244	PASSPORT	273
5	RACE	0	PREFIX	244
6	SUFFIX	1159	STATE	0
7	ZIP	543	SUFFIX	1159
8	id	0	ZIP	543

testcase24_snowflake_snowflake_etljob

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	:testcase24_snowflake_snowflake_etljob
Testcase Type	:null
Test Environment	:SIT
Start Time	:18-Mar-2025 04:50:13 UTC
End Time	:18-Mar-2025 04:51:04 UTC
Run Time	:0:00:50
Test Result	:Failed
Reason	:Nulls are not matching for each column

2. Configuration Details

Compare Type	:likeobjectcompare
testquerygenerationmode	:Auto
Testcase Type	:null
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. Null Summary

Test Result	:Failed
No of column has null in source	:7
No of column has null in target	:9
No of columns has null count match	:3
No of columns has null count mismatch	:6

4. SQL Queries

4.1 Source Query

SELECT
ADDRESS,BIRTHPLACE,CITY,COUNTY,DRIVERS,ETHNICITY,FIRST,GENDER,HEALTHCARE_COVERAGE,HEALT
HCARE_EXPENSES,ID,LAST,LAT,LON,MAIDEN,MARITAL,PASSPORT,PREFIX,RACE,SSN,STATE,SUFFIX,ZIP
FROM SAMPLE_SRC_PATIENTS

4.2 Target Query

SELECT
ADDRESS,BIRTHPLACE,CITY,COUNTY,DRIVERS,ETHNICITY,FIRST,GENDER,HEALTHCARE_COVERAGE,HEALT
HCARE_EXPENSES,ID,LAST,LAT,LON,MAIDEN,MARITAL,PASSPORT,PREFIX,RACE,SSN,STATE,SUFFIX,ZIP
FROM SAMPLE_TGT_PATIENTS

5. Columns having nulls at source and target

5.1 Columns having nulls in source

S.No	Column	Count
1	DRIVERS	213

2	MAIDEN	840
3	MARITAL	380
4	PASSPORT	273
5	PREFIX	244
6	SUFFIX	1159
7	ZIP	543

5.2 Columns having nulls in target

S.No	Column	Count
1	DRIVERS	213
2	ETHNICITY	6
3	MAIDEN	839
4	MARITAL	381
5	PASSPORT	273
6	PREFIX	244
7	RACE	5
8	SUFFIX	1157
9	ZIP	541

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	ETHNICITY	0	ETHNICITY	6
2	MAIDEN	840	MAIDEN	839
3	MARITAL	380	MARITAL	381
4	RACE	0	RACE	5
5	SUFFIX	1159	SUFFIX	1157
6	ZIP	543	ZIP	541