

# testcase1\_parquet\_parquet\_mismatch

## 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	:testcase1_parquet_parquet_mismatch
Testcase Type	:null
Test Environment	:SIT
Start Time	:11-Mar-2025 14:36:25 UTC
End Time	:11-Mar-2025 14:37:27 UTC
Run Time	:0:01:01
Test Result	:Failed
Reason	:Nulls are not matching for each column

### 2. Configuration Details

Compare Type	:s2tcompare
testquerygenerationmode	:Auto
Testcase Type	:null
Source Connection Type	:aws-s3
Source Format	:parquet
Source Path	:datf_core/test/data/source/patients_source_parquet
Target Connection Type	:aws-s3
Target Format	:parquet
Target Path	:datf_core/test/data/target/patients_target_parquet_mismatch
S2T Path	:test/s2t/s2t_1_parquet_parquet_mismatch.xlsx
Primary Keys	:id

### 3. Null Summary

Test Result	:Failed
No of column has null in source	:8
No of column has null in target	:8
No of columns has null count match	:5
No of columns has null count mismatch	:3

### 4. SQL Queries

#### 4.1 Source Query

```
["readdatadf=spark.read.format('parquet').load('datf_core/test/data/source/patients_source_parquet')",  
"readdatadf.createOrReplaceTempView('dataview')", 'spark.sql("SELECT src.ADDRESS as ADDRESS, src.BIRTHDATE as  
BIRTHDATE, src.BIRTHPLACE as BIRTHPLACE, src.CITY as CITY, src.COUNTY as COUNTY, src.DEATHDATE as  
DEATHDATE, src.DRIVERS as DRIVERS, src.ETHNICITY as ETHNICITY, src.FIRST as FIRST, src.GENDER as  
GENDER, src.HEALTHCARE_COVERAGE as HEALTHCARE_COVERAGE, src.HEALTHCARE_EXPENSES as  
HEALTHCARE_EXPENSES, src.LAST as LAST, src.LAT as LAT, src.LON as LON, src.MAIDEN as MAIDEN,  
src.MARITAL as MARITAL, src.PASSPORT as PASSPORT, src.PREFIX as PREFIX, src.RACE as RACE, src.SSN as SSN,  
src.STATE as STATE, src.SUFFIX as SUFFIX, src.ZIP as ZIP, src.id as id FROM dataview src ")', "]
```

#### 4.2 Target Query

```
["readdatadf=spark.read.format('parquet').load('datf_core/test/data/target/patients_target_parquet_mismatch')",  
"readdatadf.createOrReplaceTempView('dataview')", 'spark.sql("SELECT ADDRESS, BIRTHDATE, BIRTHPLACE, CITY,  
COUNTY, DEATHDATE, DRIVERS, ETHNICITY, FIRST, GENDER, HEALTHCARE_COVERAGE,  
HEALTHCARE_EXPENSES, LAST, LAT, LON, MAIDEN, MARITAL, PASSPORT, PREFIX, RACE, SSN, STATE,  
SUFFIX, ZIP, id FROM dataview tgt ")', "]
```

### 5. Columns having nulls at source and target

#### 5.1 Columns having nulls in source

S.No	Column	Count
1	DEATHDATE	1000

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

### 5.2 Columns having nulls in target

S.No	Column	Count
1	DEATHDATE	998
2	DRIVERS	213
3	MAIDEN	838
4	MARITAL	380
5	PASSPORT	273
6	PREFIX	244
7	SUFFIX	1157
8	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	DEATHDATE	1000	DEATHDATE	998
2	MAIDEN	840	MAIDEN	838
3	SUFFIX	1159	SUFFIX	1157

# 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

# testcase22\_bigquery\_bigquery\_match

## 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	:testcase22_bigquery_bigquery_match
Testcase Type	:null
Test Environment	:SIT
Start Time	:11-Mar-2025 14:40:06 UTC
End Time	:11-Mar-2025 14:41:14 UTC
Run Time	:0:01:08
Test Result	:Passed
Reason	:Nulls are matching for each column

### 2. Configuration Details

Compare Type	:likeobjectcompare
testquerygenerationmode	:Auto
Testcase Type	:null
Source Connection Name	:raw_bigquery_credentials
Source Connection Type	:bigquery
Source Connection Value	:raw_bigquery_credentials
Source Format	:table
Source Name	:driven-slice-423611-i9.gcp_schema.gcp_patients_src
Source Path	:driven-slice-423611-i9
Source Exclude Columns	:BIRTHDATE,DEATHDATE
Target Connection Name	:raw_bigquery_credentials
Target Connection Type	:bigquery
Target Connection Value	:raw_bigquery_credentials
Target Format	:table
Target Name	:driven-slice-423611-i9.gcp_schema.gcp_patients_tgt
Target Path	:driven-slice-423611-i9
Target Exclude Columns	:BIRTHDATE,DEATHDATE
Primary Keys	:id

### 3. Null Summary

Test Result	:Passed
No of column has null in source	:7
No of column has null in target	:7
No of columns has null count match	:7
No of columns has null count mismatch	:0

### 4. SQL Queries

#### 4.1 Source Query

```
SELECT
ADDRESS,BIRTHPLACE,CITY,COUNTY,DRIVERS,ETHNICITY,FIRST,GENDER,HEALTHCARE_COVERAGE,HEALTHCARE_EXPENSES,LAST,LAT,LON,MAIDEN,MARITAL,PASSPORT,PREFIX,RACE,SSN,STATE,SUFFIX,ZIP,id
FROM driven-slice-423611-i9.gcp_schema.gcp_patients_src
```

#### 4.2 Target Query

```
SELECT
ADDRESS,BIRTHPLACE,CITY,COUNTY,DRIVERS,ETHNICITY,FIRST,GENDER,HEALTHCARE_COVERAGE,HEALTHCARE_EXPENSES,LAST,LAT,LON,MAIDEN,MARITAL,PASSPORT,PREFIX,RACE,SSN,STATE,SUFFIX,ZIP,id
FROM driven-slice-423611-i9.gcp_schema.gcp_patients_tgt
```

### 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

None