

testcase34_delta_parquet_auto content 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/fe wtestprotocol.xlsx |
| Testcase Name | :testcase34_delta_parquet_auto |
| Testcase Type | :content |
| Test Environment | :dev |
| Start Time | :30-Aug-2024 12:36:57 UTC |
| End Time | :30-Aug-2024 12:37:37 UTC |
| Run Time | :0:00:39 |
| Test Result | :Failed |
| Reason | :Content mismatched |

2. Configuration Details

| | |
|-------------------------|--|
| Compare Type | :s2tcompare |
| testquerygenerationmode | :Auto |
| Testcase Type | :content |
| Source Connection Name | :LandingLocation |
| Source Connection Type | :aws-s3 |
| Source Format | :parquet |
| Source Path | :/Workspace/Repos/sahil.gupta@tigeranalytics.com/QE_ATF/datf_core/test/data/source/pat ients_source_parquet |
| Target Connection Name | :LandingLocation |
| Target Connection Type | :aws-s3 |
| Target Format | :delta |
| Target Path | :/Workspace/Repos/sahil.gupta@tigeranalytics.com/QE_ATF/datf_core/dbfs:/user/hive/war ehouse/healthcare.db/delta_target |
| S2T Path | :test/s2t/s2t_34_delta_parquet_auto.xlsx |
| Primary Keys | :Id_New |

3. Content Summary

| | |
|---|---------|
| Test Result | :Failed |
| No. of matched columns | :25 |
| No. of columns compared | :24 |
| No. of cols in Source but not in Target | :0 |
| No. of cols in Target but not in Source | :0 |
| 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 |
| No. of matched rows | :0 |
| No. of mismatched rows | :1,171 |
| No. of rows in Source but not in Target | :0 |
| No. of rows in Target but not in Source | :0 |

4. SQL Queries

4.1 Source Query

```
readdatadf=spark.read.format('parquet').load('file:/Workspace/Repos/sahil.gupta@tigeranalytics.com/QE_ATF/datf_core/test/dat  
a/source/patients_source_parquet')
```

```
readdatadf.createOrReplaceTempView('dataview')
```

```
spark.sql("SELECT src.Id as Id_New, src.BIRTHDATE as BIRTHDATE, src.DEATHDATE as DEATHDATE, src.SSN as
```

SSN, src.DRIVERS as DRIVERS, src.PASSPORT as PASSPORT, src.PREFIX as PREFIX, src.FIRST as FIRST, src.LAST as LAST, src.SUFFIX as SUFFIX, src.MAIDEN as MAIDEN, src.MARITAL as MARITAL, src.RACE as RACE, src.ETHNICITY as ETHNICITY, src.GENDER as GENDER, src.BIRTHPLACE as BIRTHPLACE, src.ADDRESS as ADDRESS, src.CITY as CITY, src.STATE as STATE, src.COUNTY as COUNTY, src.ZIP as ZIP, src.LAT as LAT, src.LON as LON, src.HEALTHCARE_EXPENSES as HEALTHCARE_EXPENSES, src.HEALTHCARE_COVERAGE as HEALTHCARE_COVERAGE FROM dataview src ")

4.2 Target Query

readdatadf=spark.read.format('delta').load('dbfs:/user/hive/warehouse/healthcare.db/delta_target')

readdatadf.createOrReplaceTempView('dataview')

spark.sql("SELECT Id_New, BIRTHDATE, DEATHDATE, SSN, DRIVERS, PASSPORT, PREFIX, FIRST, LAST, SUFFIX, MAIDEN, MARITAL, RACE, ETHNICITY, GENDER, BIRTHPLACE, ADDRESS, CITY, STATE, COUNTY, ZIP, LAT, LON, HEALTHCARE_EXPENSES, HEALTHCARE_COVERAGE FROM dataview tgt ")

5. Sample Mismatches 8 rows

5.1 Keys in source but not in target

None

5.2 Keys in target but not in source

None

5.3 Keys having one or more unequal column values

| S.No | Key Columns |
|------|---|
| 1 | Id_New=00185faa-2760-4218-9bf5-db301acf8274 |
| 2 | Id_New=0042862c-9889-4a2e-b782-fac1e540ecb4 |
| 3 | Id_New=0047123f-12e7-486c-82df-53b3a450e365 |
| 4 | Id_New=010d4a3a-2316-45ed-ae15-16f01c611674 |
| 5 | Id_New=01207ecd-9dff-4754-8887-4652eda231e2 |
| 6 | Id_New=0149d553-f571-4e99-867e-fcb9625d07c2 |
| 7 | Id_New=01e1f394-7219-4189-bceb-3cbd90cff90b |
| 8 | Id_New=023a7d29-32b3-4db5-89c8-b88bd7582ec0 |

6. Columnwise Mismatch Summary

None

7. Columnwise Mismatch Details

None