

testcase12_parquet_parquet_mismatch_manual

duplicate comparison

1. Run Summary

| | |
|--------------------|---|
| Application Name | :etl_pipeline_goldlayer |
| Protocol Name | :DATFDemo |
| Protocol File Path | :/Workspace/Shared/QE_ATF_Latest/datf_core/test/testprotocol/traversedtestprotocol.xlsx |
| Testcase Name | :testcase12_parquet_parquet_mismatch_manual |
| Testcase Type | :duplicate |
| Test Environment | :SIT |
| Start Time | :16-Jan-2025 13:23:44 UTC |
| End Time | :16-Jan-2025 13:23:48 UTC |
| Run Time | :0:00:03 |
| Test Result | :Passed |
| Reason | :No 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 | :patients_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 | :patients_source_parquet |
| Target Path | :/test/data/source |
| 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 rawpatients

4.2 Target Query

SELECT Count(DISTINCT id) FROM curatedpatients

testcase15_parquet_parquet_match_likeobject

duplicate comparison

1. Run Summary

| | |
|--------------------|---|
| Application Name | :etl_pipeline_goldlayer |
| Protocol Name | :DATFDemo |
| Protocol File Path | :/Workspace/Shared/QE_ATF_Latest/datf_core/test/testprotocol/traversedtestprotocol.xlsx |
| Testcase Name | :testcase15_parquet_parquet_match_likeobject |
| Testcase Type | :duplicate |
| Test Environment | :SIT |
| Start Time | :16-Jan-2025 13:23:48 UTC |
| End Time | :16-Jan-2025 13:23:50 UTC |
| Run Time | :0:00:02 |
| Test Result | :Passed |
| Reason | :No Duplicates |

2. Configuration Details

| | |
|-------------------------|--------------------------|
| Compare Type | :likeobjectcompare |
| testquerygenerationmode | :Auto |
| Testcase Type | :duplicate |
| Source Connection Name | :RawS3Bucket |
| Source Connection Type | :aws-s3 |
| Source Connection Value | :RawS3Bucket |
| Source Format | :parquet |
| Source Name | :patients_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 | :patients_source_parquet |
| Target Path | :/test/data/source |
| 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 expectedpatients