testcase21_csv_bigquery_match

content comparison

1. Run Summary

Application Name	:sample
Protocol Name	:DATFDemo
Protocol File Path	:/app/test/testprotocol/testprotocol.xlsx
Testcase Name	:testcase21_csv_bigquery_match
Testcase Type	:content
Test Environment	:dev
Start Time	:27-May-2024 07:30:40 UTC
End Time	:27-May-2024 07:35:01 UTC
Run Time	:0:04:20
Test Result	:Failed
Reason	:Content mismatched

2. Configuration Details

Compare Type	:s2tcompare
testquerygenerationmode	:Auto
Testcase Type	:content
Source Connection Type	:aws-s3
Source Format	:delimitedfile
Source Name	:patients_target.csv
Source Path	:/app/test/data/target/patients_target.csv
Source Exclude Columns	:BIRTHDATE,DEATHDATE
Target Connection Type	:bigquery
Target Connection Value	:raw_bigquery_credentials
Target Format	:table
Target Name	:driven-slice-423611-i9.gcp_schema.gcp_patients_src
Target Path	:driven-slice-423611-i9.gcp_schema.gcp_patients_src/driven-slice-423611-i9
Target Exclude Columns	:BIRTHDATE,DEATHDATE
S2T Path	:/app/test/s2t/s2t_21_csv_bigquery_match.xlsx
Primary Keys	:id

3. Content Summary

Test Result	:Failed
No. of matched columns	:23
No. of columns compared	:22
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	:543
No. of mismatched rows	:628
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('delimitedfile').option('delimiter',,).option('header','true').load('/app/test/data/target/patients_target.csv')

readdatadf.createOrReplaceTempView('dataview')

spark.sql("SELECT src.ADDRESS as ADDRESS, src.BIRTHPLACE as BIRTHPLACE, src.CITY as CITY, src.COUNTY as COUNTY, 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

spark.sql("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.driven-slice-423611-i9.gcp_schema.gcp_patients_src 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=00185faa-2760-4218-9bf5-db301acf8274
2	id=0042862c-9889-4a2e-b782-fac1e540ecb4
3	id=010d4a3a-2316-45ed-ae15-16f01c611674
4	id=0149d553-f571-4e99-867e-fcb9625d07c2
5	id=023a7d29-32b3-4db5-89c8-b88bd7582ec0
6	id=02ea2f1a-ddcf-4809-8279-dde7a62e0318
7	id=02f9aadd-72de-4b20-b381-f4c3b1cf7aa3
8	id=0325261f-61eb-46f8-acc6-89d15053fecd

6. Columnwise Mismatch Summary

None

7. Columnwise Mismatch Details

None

testcase22_bigquery_bigquery_match

content comparison

1. Run Summary

Application Name	:sample
Protocol Name	:DATFDemo
Protocol File Path	:/app/test/testprotocol/testprotocol.xlsx
Testcase Name	:testcase22_bigquery_bigquery_match
Testcase Type	:content
Test Environment	:dev
Start Time	:27-May-2024 07:35:08 UTC
End Time	:27-May-2024 07:35:43 UTC
Run Time	:0:00:35
Test Result	:Failed
Reason	:Content mismatched

2. Configuration Details

Compare Type	:likeobjectcompare
testquerygenerationmode	:Auto
Testcase Type	:content
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. Content Summary

Test Result	:Failed
No. of matched columns	:23
No. of columns compared	:22
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	:543
No. of mismatched rows	:628
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

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 driven-slice-423611-i9.gcp_schema.gcp_patients_src

4.2 Target 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 driven-slice-423611-i9.gcp_schema.gcp_patients_tgt
```

5. Sample Mismatches 5 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=00185faa-2760-4218-9bf5-db301acf8274
2	id=0042862c-9889-4a2e-b782-fac1e540ecb4
3	id=010d4a3a-2316-45ed-ae15-16f01c611674
4	id=0149d553-f571-4e99-867e-fcb9625d07c2
5	id=023a7d29-32b3-4db5-89c8-b88bd7582ec0

6. Columnwise Mismatch Summary

None

7. Columnwise Mismatch Details

None