

testcase24_snowflake_snowflake_etljob

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/testprotocol.xlsx
Testcase Name	:testcase24_snowflake_snowflake_etljob
Testcase Type	:content
Test Environment	:dev
Start Time	:22-Aug-2024 16:15:09 UTC
End Time	:22-Aug-2024 16:21:37 UTC
Run Time	:0:06:28
Test Result	:Failed
Reason	:Content mismatched

2. Configuration Details

Compare Type	:likeobjectcompare
testquerygenerationmode	:Auto
Testcase Type	:content
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. 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	:1,169
No. of distinct rows in Target	:1,169
No. of duplicate rows in Target	:0
No. of matched rows	:537
No. of mismatched rows	:632
No. of rows in Source but not in Target	:2
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,HEALTHCARE_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. Sample Mismatches 8 rows

5.1 Keys in source but not in target

S.No	Key Columns
1	ID=88ea8573-863c-47e3-b144-b810c63156a0
2	ID=0982ef39-7ff9-4c24-8239-e9fc0667e8ca

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-acc6-89d15053fecb

6. Columnwise Mismatch Summary

No. of rows in Source :1171
No. of rows in Target :1169

S.No	Source Column Name	Target Column Name	Rows in Source	Rows in Target	Rows with Common Keys	Rows Matched	Rows Mismatch
1	ETHNICITY	ETHNICITY	1171	1169	537	1163	6
2	MARITAL	MARITAL	1171	1169	537	1167	2
3	RACE	RACE	1171	1169	537	1164	5
4	ZIP	ZIP	1171	1169	537	541	628

7. Columnwise Mismatch Details

7.1 ETHNICITY

S.No	Key Columns	Source value	Target value
1	ID=73fc00f8-a341-4f62-be27-07c35c3460bc	nonhispanic	None
2	ID=8a9a9b41-2e52-4805-92ed-3be1c2972c41	nonhispanic	None
3	ID=1bacaab1-708b-42fe-8f12-fbe97a725440	nonhispanic	None
4	ID=35e1afdd-e938-441d-b54e-db15db0bbcbf	nonhispanic	None
5	ID=c1a6f986-01c4-4581-ba74-bf22d28f8ace	nonhispanic	None
6	ID=d9988dce-ec05-4138-8f04-b0c77473a260	nonhispanic	None

7.2 MARITAL

S.No	Key Columns	Source value	Target value
1	ID=034e9e3b-2def-4559-bb2a-7850888ae060	M	None

2	ID=1d604da9-9a81-4ba9-80c2-de3375d59b40	M	None
---	---	---	------

7.3 RACE

S.No	Key Columns	Source value	Target value
1	ID=1fc8dafd-90e5-44d6-9179-81156c0ea083	white	None
2	ID=fc817953-cc8b-45db-9c85-7c0ced8fa90d	white	None
3	ID=6c2b1cfd-887a-43f8-bee0-000ce2fc6ef0	white	None
4	ID=6d048a56-edb8-4f29-891d-7a84d75a8e78	white	None
5	ID=fca3178e-fb68-41c3-8598-702d3ca68b96	white	None

7.4 ZIP

S.No	Key Columns	Source value	Target value
1	ID=cf12424d-63da-4fbb-b93f-22308793cafa	02492	2492
2	ID=73fc00f8-a341-4f62-be27-07c35c3460bc	01027	1027
3	ID=4f9dcdaa-fb47-412b-8288-03f8cbbc321e	01462	1462
4	ID=431971f5-a39f-44f8-9ef9-10502db973c9	01905	1905
5	ID=5f08cc74-e542-4b93-a565-bf802a03d529	01510	1510
6	ID=ca3330c5-bbbc-47e7-addb-302f2e069986	01105	1105
7	ID=304e3c61-cd44-4251-8186-0537150ff62c	02740	2740
8	ID=92709d5b-63d2-4e47-b857-dccb09724ea3	02145	2145

testcase31_delta_parquet_manual

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/testprotocol.xlsx
Testcase Name	:testcase31_delta_parquet_manual
Testcase Type	:content
Test Environment	:dev
Start Time	:22-Aug-2024 16:22:26 UTC
End Time	:22-Aug-2024 16:23:34 UTC
Run Time	:0:01:08
Test Result	:Failed
Reason	:Content mismatched

2. Configuration Details

Compare Type	:s2tcompare
testquerygenerationmode	:Manual
Testcase Type	:content
Source Connection Name	:RawS3Bucket
Source Connection Type	:aws-s3
Source Connection Value	:RawS3Bucket
Source Format	:delta
Source Path	:dbfs:/user/hive/warehouse/healthcare.db/delta_patients
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
S2T Path	:test/s2t/s2t_31_delta_parquet_manual.xlsx
Primary Keys	:id

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

SELECT src.DEATHDATE as DEATHDATE, src.HEALTHCARE_EXPENSES as HEALTHCARE_EXPENSES, src.SUFFIX as SUFFIX, src.PASSPORT as PASSPORT, src.CITY as CITY, src.BIRTHPLACE as BIRTHPLACE, src.MAIDEN as MAIDEN, src.MARITAL as MARITAL, src.DRIVERS as DRIVERS, src.STATE as STATE, src.COUNTY as COUNTY, src.ADDRESS as ADDRESS, src.PREFIX as PREFIX, src.ZIP as ZIP, src.LON as LON, src.SSN as SSN, src.LAST as LAST, src.BIRTHDATE as BIRTHDATE, src.RACE as RACE, src.ETHNICITY as ETHNICITY, src.GENDER as GENDER, src.id as id, src.HEALTHCARE_COVERAGE as HEALTHCARE_COVERAGE, src.LAT as LAT, src.FIRST as FIRST FROM

```
`delta`.`dbfs:/user/hive/warehouse/healthcare.db/delta_patients` src;
```

4.2 Target Query

SELECT tgt.DEATHDATE as DEATHDATE, tgt.HEALTHCARE_EXPENSES as HEALTHCARE_EXPENSES, tgt.SUFFIX as SUFFIX, tgt.PASSPORT as PASSPORT, upper(tgt.CITY) as CITY, tgt.BIRTHPLACE as BIRTHPLACE, tgt.MAIDEN as MAIDEN, tgt.MARITAL as MARITAL, tgt.DRIVERS as DRIVERS, tgt.STATE as STATE, tgt.COUNTY as COUNTY, tgt.ADDRESS as ADDRESS, tgt.PREFIX as PREFIX, tgt.ZIP as ZIP, tgt.LON as LON, tgt.SSN as SSN, tgt.LAST as LAST, tgt.BIRTHDATE as BIRTHDATE, tgt.RACE as RACE, tgt.ETHNICITY as ETHNICITY, tgt.GENDER as GENDER, tgt.id as id, tgt.HEALTHCARE_COVERAGE as HEALTHCARE_COVERAGE, tgt.LAT as LAT, tgt.FIRST as FIRST FROM curatedpatients 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=0047123f-12e7-486c-82df-53b3a450e365
4	id=010d4a3a-2316-45ed-ae15-16f01c611674
5	id=01207ecd-9dff-4754-8887-4652eda231e2
6	id=0149d553-f571-4e99-867e-fcb9625d07c2
7	id=01e1f394-7219-4189-bceb-3cbd90cff90b
8	id=023a7d29-32b3-4db5-89c8-b88bd7582ec0

6. Columnwise Mismatch Summary

None

7. Columnwise Mismatch Details

None