testcase12_parquet_parquet_mismatch_manual
content 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	:content
Test Environment	:SIT
Start Time	:10-Mar-2025 09:23:50 UTC
End Time	:10-Mar-2025 09:28:37 UTC
Run Time	:0:04:47
Test Result	:Failed
Reason	:Content mismatched

2. Configuration Details

Compare Type	:likeobjectcompare
testquerygenerationmode	:Manual
Testcase Type	:content
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. 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	:1,001,000
No. of distinct rows in Source	:1,000,000
No. of duplicate rows in Source	:1,000
No. of rows in Target	:1,001,000
No. of distinct rows in Target	:1,000,000
No. of duplicate rows in Target	:1,000
No. of matched rows	:0
No. of mismatched rows	:1,000,000
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,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. 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=0
2	id=1
3	id=2
4	id=3
5	id=4
6	id=5
7	id=6
8	id=7

6. Columnwise Mismatch Summary

No. of rows in Source   :1001000
No. of rows in Target   :1001000

S.No	Source Column Name	Target Column Name	Rows in Source	Rows in Target	Rows with Common Keys	Rows Matched	Rows Mismatch
1	ADDRESS	ADDRESS	1001000	1001000	0	0	1000000
2	BIRTHDATE	BIRTHDATE	1001000	1001000	0	16	999984
3	BIRTHPLACE	BIRTHPLACE	1001000	1001000	0	0	1000000
4	CITY	CITY	1001000	1001000	0	0	1000000
5	COUNTY	COUNTY	1001000	1001000	0	0	1000000
6	DEATHDATE	DEATHDATE	1001000	1001000	0	810273	189727
7	DRIVERS	DRIVERS	1001000	1001000	0	40178	959822
8	FIRST	FIRST	1001000	1001000	0	0	1000000
9	HEALTHCARE_COVERAGE	HEALTHCARE_COVERAGE	1001000	1001000	0	0	1000000
10	HEALTHCARE_EXPENSES	HEALTHCARE_EXPENSES	1001000	1001000	0	0	1000000
11	LAST	LAST	1001000	1001000	0	0	1000000
12	LAT	LAT	1001000	1001000	0	0	1000000
13	LON	LON	1001000	1001000	0	0	1000000
14	MAIDEN	MAIDEN	1001000	1001000	0	639470	360530
15	MARITAL	MARITAL	1001000	1001000	0	0	1000000
16	PASSPORT	PASSPORT	1001000	1001000	0	89850	910150
17	PREFIX	PREFIX	1001000	1001000	0	0	1000000
18	RACE	RACE	1001000	1001000	0	0	1000000
19	SSN	SSN	1001000	1001000	0	0	1000000
20	STATE	STATE	1001000	1001000	0	0	1000000
21	SUFFIX	SUFFIX	1001000	1001000	0	639383	360617
22	ZIP	ZIP	1001000	1001000	0	0	1000000

7. Columnwise Mismatch Details

7.1 ADDRESS

S.No	Key Columns	Source value	Target value
1	id=46546	ADXfCnxCEXocEiN	BMdCtYUwnZPBSLO
2	id=61332	ADXfCnxCEXocEiN	BMdCtYUwnZPBSLO
3	id=320974	ADXfCnxCEXocEiN	BMdCtYUwnZPBSLO
4	id=435108	ADXfCnxCEXocEiN	BMdCtYUwnZPBSLO
5	id=494091	ADXfCnxCEXocEiN	BMdCtYUwnZPBSLO
6	id=497769	ADXfCnxCEXocEiN	BMdCtYUwnZPBSLO
7	id=562226	ADXfCnxCEXocEiN	BMdCtYUwnZPBSLO
8	id=603026	ADXfCnxCEXocEiN	BMdCtYUwnZPBSLO

7.2 BIRTHDATE

S.No	Key Columns	Source value	Target value
1	id=17615	2017-04-16	1997-05-17
2	id=122353	1986-05-14	1959-08-12
3	id=180848	1968-01-11	2021-11-08
4	id=185666	1954-01-20	2021-07-19
5	id=187674	1938-09-19	1979-10-10
6	id=215556	1996-08-27	1934-11-25
7	id=218332	2009-08-21	1986-02-27
8	id=266703	1953-12-27	1989-11-14

7.3 BIRTHPLACE

S.No	Key Columns	Source value	Target value
1	id=21947	dlniiUYFPs	iMfoNgqGOg
2	id=33997	dlniiUYFPs	iMfoNgqGOg
3	id=34016	dlniiUYFPs	iMfoNgqGOg
4	id=63136	dlniiUYFPs	iMfoNgqGOg
5	id=75935	dlniiUYFPs	iMfoNgqGOg
6	id=114210	dlniiUYFPs	iMfoNgqGOg
7	id=317582	dlniiUYFPs	iMfoNgqGOg
8	id=365500	dlniiUYFPs	iMfoNgqGOg

7.4 CITY

S.No	Key Columns	Source value	Target value
1	id=69093	LSWJdAweZl	vyZdOaPDSC
2	id=165631	LSWJdAweZl	vyZdOaPDSC
3	id=267032	LSWJdAweZl	vyZdOaPDSC
4	id=274049	LSWJdAweZl	vyZdOaPDSC
5	id=478405	LSWJdAweZl	vyZdOaPDSC
6	id=491425	LSWJdAweZl	vyZdOaPDSC
7	id=573412	LSWJdAweZl	vyZdOaPDSC
8	id=625930	LSWJdAweZl	vyZdOaPDSC

7.5 COUNTY

S.No	Key Columns	Source value	Target value
1	id=64688	oOOquzvTyq	VKAAvLZtXp
2	id=172215	oOOquzvTyq	VKAAvLZtXp

3	id=277726	oOOquzvTyq	VKAAvLZtXp
4	id=377814	oOOquzvTyq	VKAAvLZtXp
5	id=477641	oOOquzvTyq	VKAAvLZtXp
6	id=523001	oOOquzvTyq	VKAAvLZtXp
7	id=582307	oOOquzvTyq	VKAAvLZtXp
8	id=663317	oOOquzvTyq	VKAAvLZtXp

7.6 DEATHDATE

S.No	Key Columns	Source value	Target value
1	id=440577	None	2028-04-18
2	id=624341	None	2100-03-22
3	id=932675	2018-07-04	None
4	id=221445	None	2074-07-18
5	id=570462	2009-07-22	None
6	id=677670	2018-01-19	None
7	id=205669	None	2014-02-17
8	id=546944	None	2052-09-21

7.7 DRIVERS

S.No	Key Columns	Source value	Target value
1	id=42049	None	eraEkPDvck
2	id=149228	KRKBvUYNsj	eraEkPDvck
3	id=208852	KRKBvUYNsj	eraEkPDvck
4	id=230448	KRKBvUYNsj	None
5	id=233988	KRKBvUYNsj	eraEkPDvck
6	id=268087	KRKBvUYNsj	eraEkPDvck
7	id=374651	KRKBvUYNsj	eraEkPDvck
8	id=416916	None	eraEkPDvck

7.8 FIRST

S.No	Key Columns	Source value	Target value
1	id=114938	qwkUmh	ndHrDn
2	id=171877	qwkUmh	ndHrDn
3	id=215644	qwkUmh	ndHrDn
4	id=314809	qwkUmh	ndHrDn
5	id=315357	qwkUmh	ndHrDn
6	id=398282	qwkUmh	ndHrDn
7	id=443778	qwkUmh	ndHrDn
8	id=445008	qwkUmh	ndHrDn

7.9 HEALTHCARE_COVERAGE

S.No	Key Columns	Source value	Target value
1	id=150992	4100.30029296875	7101.109375
2	id=213759	73.8502197265625	9384.5458984375
3	id=322426	6645.10693359375	4066.121337890625
4	id=408507	4553.41552734375	3408.560302734375
5	id=415035	9367.6669921875	1368.4429931640625
6	id=437093	5617.67822265625	8774.6669921875
7	id=484215	1395.2606201171875	9147.6787109375

8	id=484493	9041.43359375	4974.8486328125

7.10 HEALTHCARE_EXPENSES

S.No	Key Columns	Source value	Target value
1	id=365405	11130.833984375	85889.6796875
2	id=376085	21235.736328125	26268.728515625
3	id=434579	52990.37890625	67087.328125
4	id=469565	53279.04296875	53571.1953125
5	id=684015	32409.7265625	72221.875
6	id=684748	50905.81640625	78977.5546875
7	id=33047	81924.671875	6306.19091796875
8	id=101957	90876.4765625	30188.859375

7.11 LAST

S.No	Key Columns	Source value	Target value
1	id=10993	URMXGQla	mhsJsvKd
2	id=20709	URMXGQla	mhsJsvKd
3	id=119379	URMXGQla	mhsJsvKd
4	id=134501	URMXGQla	mhsJsvKd
5	id=171642	URMXGQla	mhsJsvKd
6	id=171877	URMXGQla	mhsJsvKd
7	id=174026	URMXGQla	mhsJsvKd
8	id=176126	URMXGQla	mhsJsvKd

7.12 LAT

S.No	Key Columns	Source value	Target value
1	id=10283	-55.76679229736328	5.004130840301514
2	id=48019	80.94187927246094	85.09967041015625
3	id=314051	-61.48698806762695	-78.38497161865234
4	id=357877	74.77754211425781	-58.97312927246094
5	id=374248	68.99937774658203	-75.47573852539062
6	id=382500	67.81210327148438	52.548465728759766
7	id=450304	30.263118743896484	21.774572372436523
8	id=566689	20.632762908935547	-51.11421585083008

7.13 LON

S.No	Key Columns	Source value	Target value
1	id=175161	76.24896240234375	-40.097869873046875
2	id=193562	-133.38967895507812	-120.52035522460938
3	id=201919	105.75178527832031	69.48912811279297
4	id=213525	-101.05460357666016	4.300279140472412
5	id=289405	-117.97879028320312	-85.13064575195312
6	id=466681	-93.8414535522461	-1.2613072395324707
7	id=506266	148.07571411132812	84.01607513427734
8	id=590964	50.7124137878418	-26.970840454101562

7.14 MAIDEN

S.No	Key Columns	Source value	Target value

1	id=133058	HOslOtnn	None
2	id=269113	None	dhuqqoeC
3	id=345184	HOslOtnn	dhuqqoeC
4	id=734484	HOslOtnn	None
5	id=75422	None	dhuqqoeC
6	id=721700	HOslOtnn	None
7	id=758031	None	dhuqqoeC
8	id=876604	HOslOtnn	None

7.15 MARITAL

S.No	Key Columns	Source value	Target value
1	id=273859	Widowed	Divorced
2	id=300275	Widowed	Divorced
3	id=304598	Widowed	Divorced
4	id=364726	Widowed	Divorced
5	id=464275	Widowed	Divorced
6	id=590492	Widowed	Divorced
7	id=620543	Widowed	Divorced
8	id=650850	Widowed	Divorced

7.16 PASSPORT

S.No	Key Columns	Source value	Target value
1	id=70479	uSKsADpZo	None
2	id=116482	uSKsADpZo	None
3	id=138240	uSKsADpZo	uFHdivQyy
4	id=143936	uSKsADpZo	uFHdivQyy
5	id=163559	None	uFHdivQyy
6	id=180507	uSKsADpZo	None
7	id=215644	uSKsADpZo	uFHdivQyy
8	id=300360	uSKsADpZo	None

7.17 PREFIX

S.No	Key Columns	Source value	Target value
1	id=63960	Mr.	Ms.
2	id=165729	Mr.	Ms.
3	id=187674	Mr.	Ms.
4	id=198747	Mr.	Ms.
5	id=233911	Mr.	Ms.
6	id=428291	Mr.	Ms.
7	id=499370	Mr.	Ms.
8	id=783239	Mr.	Ms.

7.18 RACE

S.No	Key Columns	Source value	Target value
1	id=67612	Other	Asian
2	id=114727	Other	Asian
3	id=156519	Other	Asian
4	id=185444	Other	Asian
5	id=260920	Other	Asian

6	id=276665	Other	Asian
7	id=279235	Other	Asian
8	id=295075	Other	Asian

7.19 SSN

S.No	Key Columns	Source value	Target value
1	id=13257	167-89-2702	266-70-6856
2	id=57200	167-89-2702	266-70-6856
3	id=81995	167-89-2702	266-70-6856
4	id=146482	167-89-2702	266-70-6856
5	id=181728	167-89-2702	266-70-6856
6	id=232338	167-89-2702	266-70-6856
7	id=256022	167-89-2702	266-70-6856
8	id=510771	167-89-2702	266-70-6856

7.20 STATE

S.No	Key Columns	Source value	Target value
1	id=3771	FL	NY
2	id=16745	FL	NY
3	id=127456	FL	NY
4	id=197534	FL	NY
5	id=198059	FL	NY
6	id=247713	FL	NY
7	id=365679	FL	NY
8	id=397380	FL	NY

7.21 SUFFIX

S.No	Key Columns	Source value	Target value
1	id=146931	Sr.	III
2	id=824384	Sr.	None
3	id=909811	None	III
4	id=994992	Sr.	III
5	id=99302	Sr.	None
6	id=172513	Sr.	None
7	id=333482	Sr.	III
8	id=373737	None	III

7.22 ZIP

S.No	Key Columns	Source value	Target value
1	id=22273	17395	88546
2	id=135580	17395	88546
3	id=158691	17395	88546
4	id=195559	17395	88546
5	id=200619	17395	88546
6	id=349071	17395	88546
7	id=357799	17395	88546
8	id=375082	17395	88546