

[Delta] Sink does not support Schema drift with upsert or update

Last updated by | Jackie Huang | Jan 4, 2022 at 12:24 AM PST

Contents

- [Issue](#)
- [Symptom](#)
- [Error:](#)
- [Recommendation:](#)

Issue

The customer mentioned that Delta Sink in Mapping Data Flows does not support schema drift with upsert/update. The problem is that schema drift does not work when Delta is the target in a mapping data flow and user-configured an update/upsert.

Symptom

If a column is added to the source after an “initial” load to Delta, the subsequent jobs just fail with an error that it cannot find the new column. This happens when the customer does an upsert/update with the Alter Row. It seems to work for inserts only.

Error:

DF-SYS-01 at Sink 'SnkDeltaLake': org.apache.spark.sql.AnalysisException: cannot resolve target.BICC_RV in UPDATE clause given columns target.

Recommendation:

This is a known issue for delta format due to the limitation of io delta library we used in data flow runtime. PG has a backlog item to support the merge schema option for delete/update/upsert as well.

For now, the workaround will be to update the schema first then write the data, basically

#1 Create one data flow that includes an insert-only delta sink with the merge schema option to update the schema.

#2 . After **#1** , use delete/upsert/update to modify the target sink as you want without schema changing.

Reference: [120120726004532](#) 

How good have you found this content?

