Write Functions Reference

These are the write functions you can use in the json settings files. This is a reference document and does not show how to use these functions as well as one of the procedures or guides.

Functions

|  |  |
| --- | --- |
| **Function Name** | **Description** |
| overwrite\_table | overwrites a Delta table using df.write.mode("overwrite") |
| upsert\_table | merges a DataFrame into the destination table |
| batch\_upsert\_scd2 | thin wrapper around upsert\_table for SCD2 batch upserts |
| write\_upsert\_snapshot | merges a single-row-per-key snapshot into the destination table |
| stream\_write\_table | write a streaming DataFrame directly to a Delta table |
| stream\_upsert\_table | apply an upsert function to each streaming micro-batch |
| microbatch\_upsert\_fn | wrapper used in foreachBatch for standard upserts |
| microbatch\_upsert\_scd2\_fn | wrapper used in foreachBatch for SCD2 style upserts |

Settings

Functions that end in upsert\_fn return a function for use inside the foreachbatch() function. They are not true write functions but if simple settings is set to false it will be necessary to put them as the upsert\_function.

|  |  |
| --- | --- |
| **Function Name** | **Key(s) Required** |
| overwrite\_table |  |
| upsert\_table | business\_key, surrogate\_key |
| batch\_upsert\_scd2 | business\_key, surrogate\_key |
| write\_upsert\_snapshot | business\_key |
| stream\_write\_table |  |
| stream\_upsert\_table | business\_key, surrogate\_key, upsert\_function |
| microbatch\_upsert\_fn | business\_key, surrogate\_key |
| microbatch\_upsert\_scd2\_fn | business\_key, surrogate\_key |

Example

"write\_function": "functions.write.write\_upsert\_snapshot"

"business\_key": [

"id"

],

"surrogate\_key": []

SCD2 Example

This shows that the record identified by LFT 1072 - Edmund Mahon has “changed” three times since 6/25. Each time it changes the SCD2 table adds another record so that history is tracked. SCD2 tables are commonly used for financial data.

