

Change Data Capture

Learning Objectives

- ▶ Definition of Change data capture
- ▶ Processing Change data capture with DLT

Change Data Capture

- ▶ CDC: Process of identifying changes made to data in the source and delivering those changes to the target



Row-Level changes

- ▶ Inserting new records
- ▶ Updating existing records
- ▶ Deleteing existing records

CDC Feed

► Row data + metadata

Country ID	Country	Vaccination Rate	sequence	operation
FR	France	0,74	2022-11-01 07:00	UPDATE
FR	France	0.75	2022-11-01 08:00	UPDATE
CA			2022-11-01 00:00	DELETE
US	USA	0,5	2022-11-01 00:00	INSERT
IN	India	0.66	2022-11-01 00:00	INSERT

CDC Feed



Country ID	Country	Vaccination Rate
FR	France	0,6
CA	Canada	0,71

Target table

CDC Feed

► Row data + metadata

Country ID	Country	Vaccination Rate	sequence	operation
FR	France	0.74	2022-11-01 07:00	UPDATE
FR	France	0.75	2022-11-01 08:00	UPDATE
CA			2022-11-01 00:00	DELETE
US	USA	0.5	2022-11-01 00:00	INSERT
IN	India	0.66	2022-11-01 00:00	INSERT

CDC Feed



Country ID	Country	Vaccination Rate
FR	France	0.75
US	USA	0.5
IN	India	0.66

Target table

CDC with DLT

► APPLY CHANGES INTO command

```
APPLY CHANGES INTO LIVE.target_table  
FROM STREAM(LIVE.cdc_feed_table)  
KEYS (key_field)  
APPLY AS DELETE WHEN operation_field = "DELETE"  
SEQUENCE BY sequence_field  
COLUMNS *
```

APPLY CHANGES INTO: **Pros**

- ▶ Orders late-arriving records using the sequencing key
- ▶ Default assumption is that rows will contain inserts and updates
- ▶ Can optionally apply deletes (APPLY AS DELETE WHEN condition)
- ▶ Specify one or many fields as the primary key for a table
- ▶ Specify columns to ignore with the EXCEPT keyword
- ▶ Support applying changes as SCD Type 1 (default) or SCD Type 2

APPLY CHANGES INTO: **Cons**

- ▶ Breaks the append-only requirements for streaming table sources
 - ▶ Can Not perform streaming queries against the table