

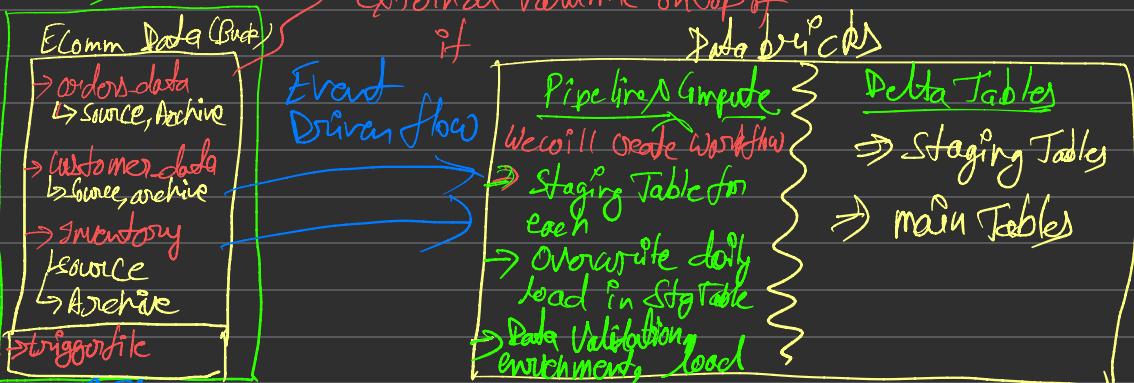
Event Driven Data Ingestion & SCD2 Type maintenance

GCS

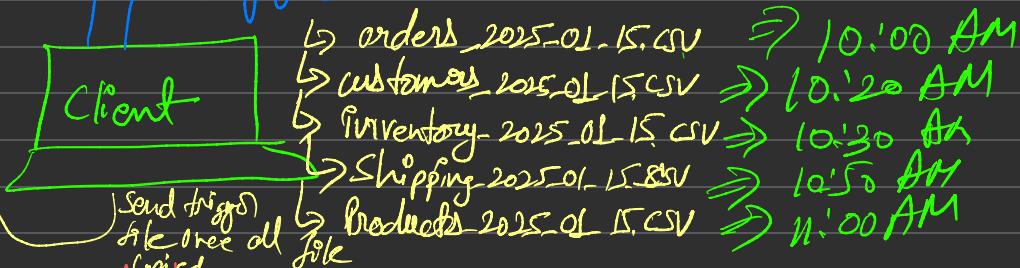
We have created external volume on top of

if

Data bricks

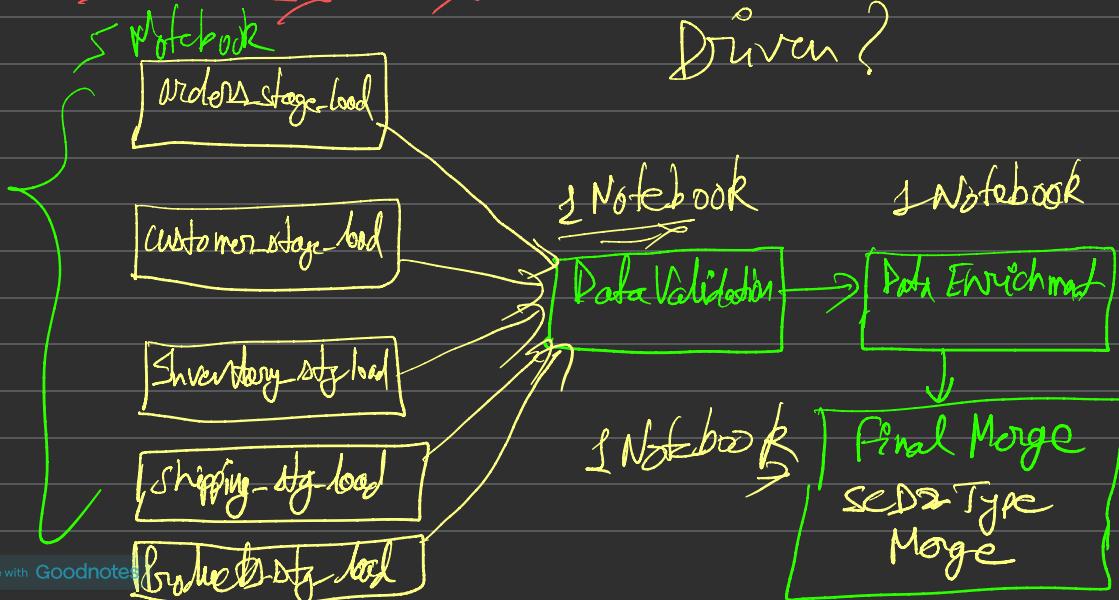


daily files



Data bricks Design flow

How Event
Driven?



Merge Command in Delta Table

↳ Merge command is helpful to perform Conditional Insert/ Update/ Delete into a target table in compare to source table.

Source-tb (L)

| Order-id | address | full_amount |
|----------|---------|-------------|
| 100 | ABC | 1000 |
| 200 | KFG | 3000 |

today's data, delta data etc

target-tb (R)

| Order-id | address | amount |
|----------|---------|--------|
| 100 | ABC | 800 |
| | | |
| | | |
| | | |

Merge Into ~~target delta table more~~

Using ~~Source-table~~ { Sub Query }
ON st.orderid = tt.orderid AND (more exp)

When Matched AND st.full_amount != tt.amount

Then Update Set

tt.address = st.address

tt.amount = st.full_amount

When Not Matched

Then Insert ~~X~~;

SCD 2 Type

↳ Slowly Changing Dimension table

Customer

| cust_id | name | phone | address |
|---------|------|-------|---------|
| 1 | A | 01 | AD1 |
| 2 | B | 02 | AD2 |
| 3 | C | 03 | AD3 |
| 4 | D | 04 | AD4 |
| 5 | E | 05 | AD5 |

cust \Rightarrow 1 changed
phone numb = 01 \Rightarrow 09

Customer (SCD2 Typ)

| id | name | phone | address | is-active | start-dt | end-dt |
|----|------|-------|---------|-----------|----------|--------|
| 1 | A | 01 | AD1 | False | 20-Sep | 24-Sep |
| 1 | A | 09 | AD1 | False | 24-Sep | 27-Sep |
| 1 | A | 09 | AD5 | True | 27-Sep | 999999 |
| | | | | | | |
| | | | | | | |
| | | | | | | |

X

After = today cust 1 changed the phone number
from 01 \Rightarrow 09
 \Rightarrow 3 day address change

Source_tb

| id | name | phone | date |
|----|------|-------|--------|
| 1 | A | 09 | 24-sep |
| 2 | B | 02 | 24-sep |

tgt_tb

| id | name | phone | is active | start date | end date |
|----|------|-------|-----------|------------|----------|
| 1 | A | 01 | True | 20-Sep | 2020 |
| 2 | | | | | |

S C D 2 Positional

Merge into tgt_tb

using src_tb

on tgt_tb.id = src_tb.id and tgt_tb.is active
= True

When Matched AND tgt_tb.phone != src_tb.phone

Then Update SET

tgt.phone = src.phone,
tgt.is active = False,
tgt.end date = src.date

Source_df with Column ("is active", True)

• with Column ("start date?", date)

• with Column ("end date?", "99999")

