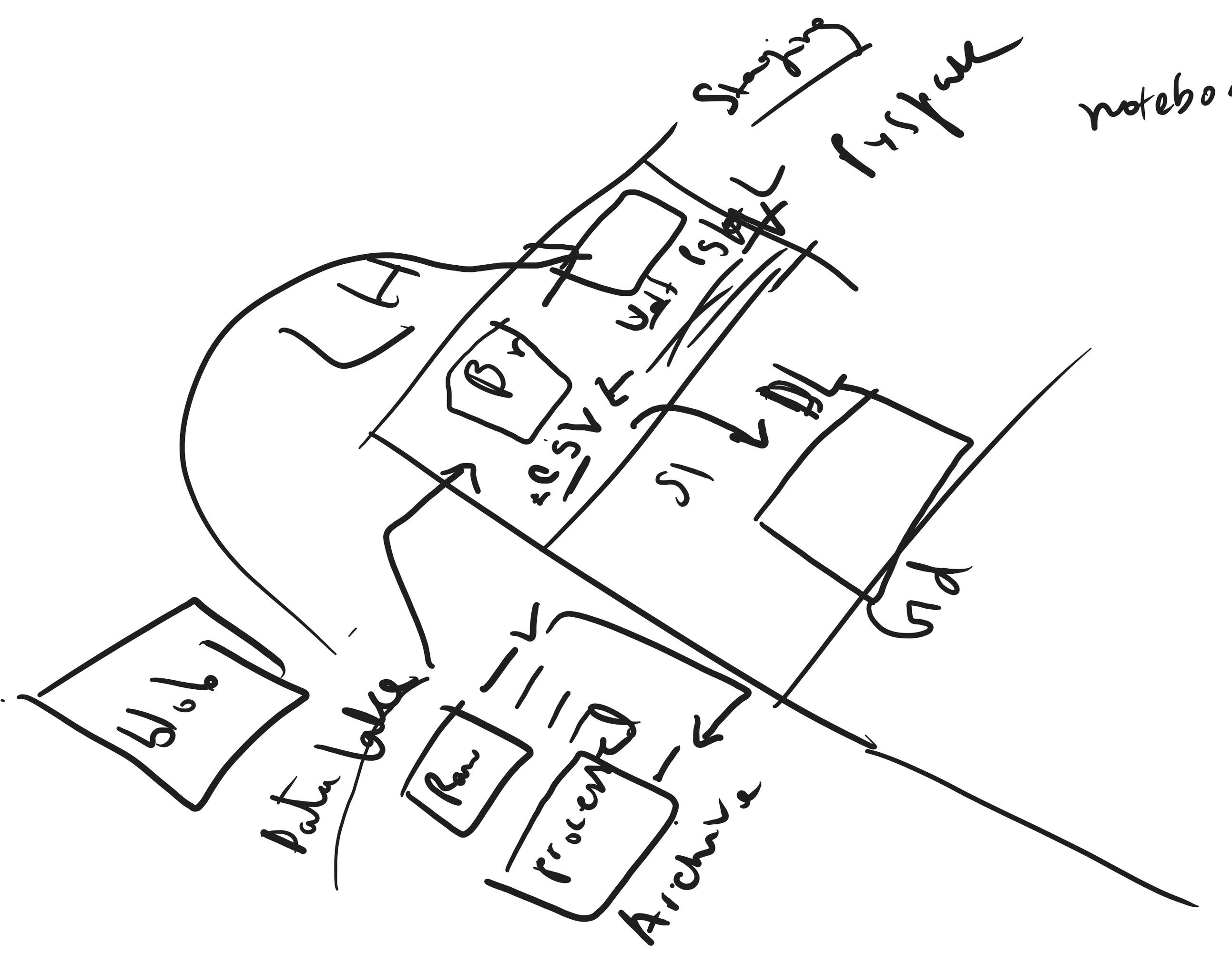

notebookutils

cp
mv 7 datalake

LH <- Notebooks -> SLH
mv
udf
sparksql
pyspark

DeltaTable +Notebooks

Tables | Logic

Adf — Fabric

SCD Type 2

Src ID	Data
1	A
2	B
3	D
4	F
5	F
1	B

] needs to update as already exists
] Exists already
] new values
] changed value

(1) Dedup to get the latest data row with latest changes

ID	Date
2	D
3	D
4	E
1	B

Trg

ID	data	status	Enddate
1	A	1 → 0	
2	B	1	
1	C	0	

(2) update → Enddate = SYSDATETIME()
trg.status = 0

is (src.ID = trg.ID &
trg.status = 1 &
src.data <> trg.data)

if + &

(3) Insert
Leftjoin src.ID = tgt.ID & tgt.status = 1

if tgt.ID IS NULL

(X) wrong

1 → 0
1 → 0
1 → 0

will be inserted 3 times
if ID = ID &
tgt.status = 0

ID = ID & trg.status = 0
or
srcdata <> srcdata
this is wrong
multiple insert for multiple true value

Src ID	tgt ID	tgt Status
2	2	
3	Null	
4	Null	
5	Null	
1	Null	