NOMBA Data Engineer Take-Home

The task is to build a simple data pipeline that moves data from MongoDB and Postgres into a warehouse. The goal is to show how you handle change data capture, build a basic transformation layer, and set up a small CI/CD workflow.

The pipeline should be able to handle new and updated records and run multiple times without creating duplicates. Some of the collections might not have the necessary keys to do CDC, so you'll need to think about how to handle that.

You'll then use dbt to build a few transformations. Create staging models, a couple of dimension tables, and at least one fact table.

This is not meant to be a big production system. A clean, working pipeline with clear logic, good CDC handling, and simple transformations is what matters most.

You can use a local setup for the database or data warehouse, but if you want to go cloud, Mongo Atlas has a free tier cluster, and Supabase/Aiven have free tier Postgres you can use.

Sample Data:

MongoDB

User Collection: Holds basic customer details; it is rumored to be around 20 million records.

{"_id": "6735f0123b1a2c4e876a9bde",

"Uid": "",

"firstName": "",

"lastName": "",

"occupation": "",

"state": ""}

Table: savingsTransaction

Tracks buy and sell activity for savings products.

Column	Type	Description
txn_id	UUID	Unique transaction ID
plan_id	UUID	Foreign key to savings_plan
amount	NUMERIC	Transaction amount
currency	TEXT	Currency of transaction
side	TEXT	buy or sell
rate	NUMERIC	Exchange rate at the time of the transaction
txn_timestamp	TIMESTAMP	Transaction timestamp
updated_at	TIMESTAMP	Last updated time
deleted_at	TIMESTAMP	Soft delete time (nullable)

Table: savings_plan

Represents customer savings plans.

Column	Type	Description
plan_id	UUID	Unique plan ID
product_type	TEXT	Foreign key to savings_product
customer_uid	TEXT	User ID that owns the plan
amount	NUMERIC	Amount saved in the plan
frequency	TEXT	Frequency of savings (daily, weekly, monthly)
start_date	DATE	Date the plan started
end_date	DATE	Date the plan will end
status	TEXT	Current status of the plan (active, completed)
created_at	TIMESTAMP	Record creation time

updated_at	TIMESTAMP	Last updated time
deleted_at	TIMESTAMP	Soft delete time (nullable)

Note: You can add extra tables if you need more tables to build the model