Nubank Analytics Engineer Case

Introduction

You are in the next stage of the Nubank Analytics Engineer hiring process and now it is time to work in a hands-on project case.

The idea here is to show your knowledge based on previous experiences in fields that are relevant to this position applied to a very similar context to what happens at our company.

Data Architecture Overview

Here at Nubank, we have three different environments: Production, Data Warehouse and Reporting (Figure 1).

The Production Environment is where Nubank's services live. In this environment, services are responsible for getting customers' information and storing in their respective databases.

Then, the datasets generated by the services are maintained in the Data Warehouse Environment, where Business Analysts and Analytics Engineers create new tables, improve data models, and work to turn raw data into easily consumable tables for analysis.

Finally, these tables feed into the Reporting Environment, where it is possible to create all sorts of dashboards and visualisations for monitoring and decision-making.
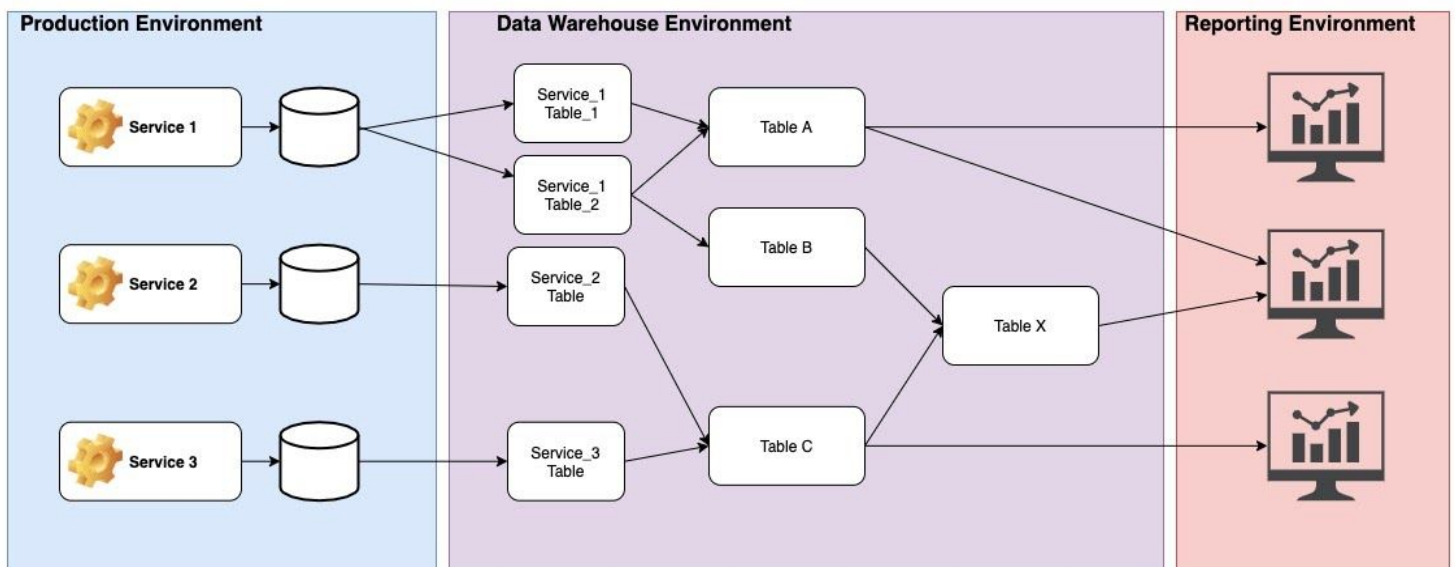

Figure 1. Nubank's environments: Production, Data Warehouse and Reporting.

A slice of the table structure from the Data Warehouse Environment is depicted in Figure 2.

Apart from time (d_time, d_year, d_month, d_week, d_weekday), location (city, state, country), accounts, and customers tables, three tables store the financial movements of the accounts:

- transfer_ins: non PIX transfers made to an account (money arriving)
- transfer_outs: non PIX transfers made from an account (money leaving)
- pix_movements: transfers that are either received by or sent from an account using PIX
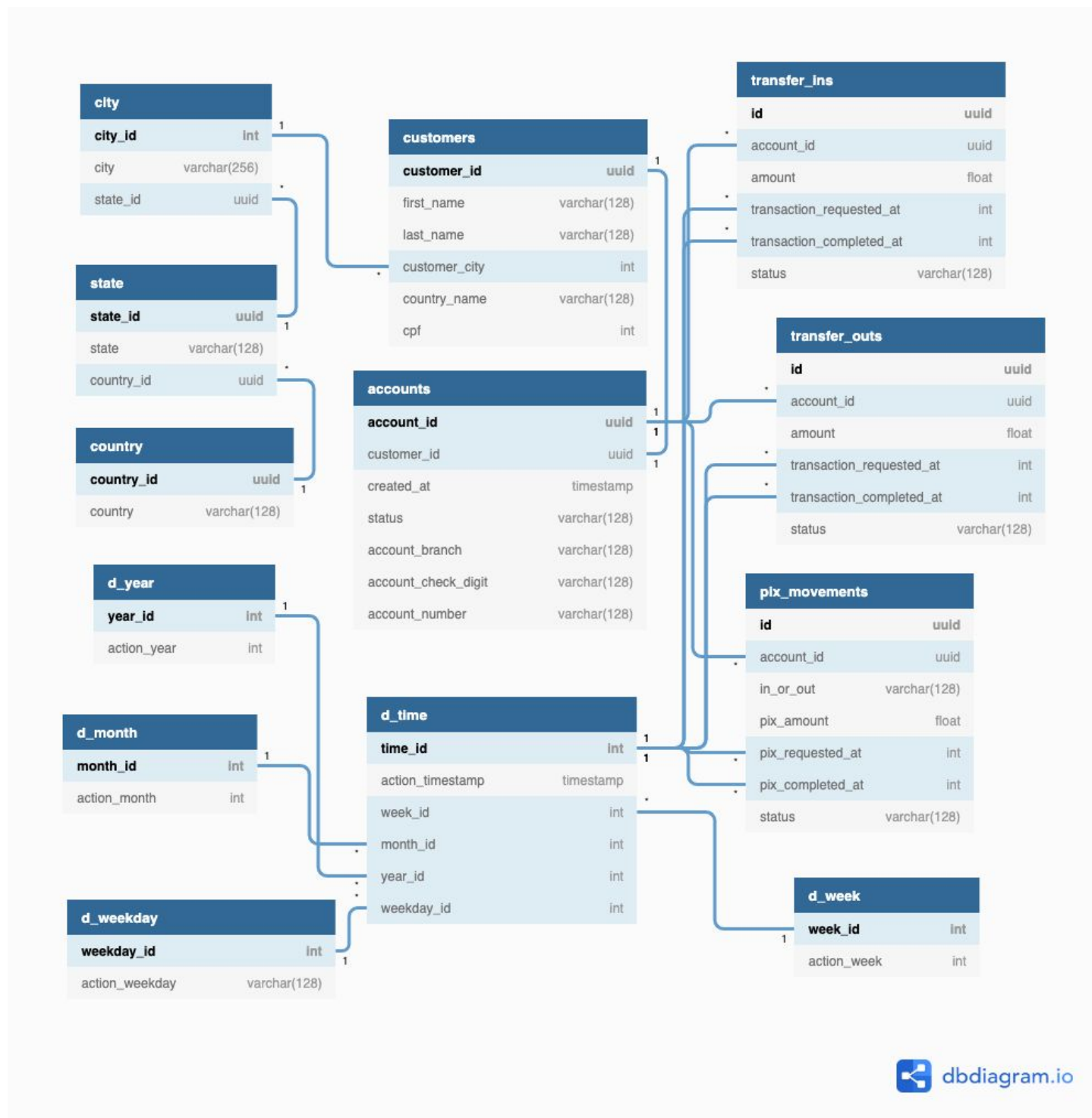

Figure 2. A slice of the table structure from the Data Warehouse Environment (*diagram/ table_diagram.png*). You can check the code used to generate these diagrams using dbdiagram.io on file (*diagram/ table_diagram.txt*).

Business Context

To solve this case you need to be familiar with the concept of "Account Monthly Balance". Account Monthly Balance is the amount of money a customer had in their account at the end of a given month. This information can be calculated by adding all the transfers in and subtracting all the transfers out from the previous account balance. Remember you should also consider transfers made by PIX, the newest Brazilian transfer method. You can see an example below:

Month	Customer	Total Transfer In	Total Transfer Out	Account Monthly Balance
1	A	1000	200	800
1	B	2000	0	2000
2	A	0	200	600
2	B	100	500	1600
2	C	500	100	400

Table 1. An example of account monthly balance data.

Problem Statement

Your colleague Jane Hopper, the analyst in charge of analysing customer behaviour, who directly consumes data from the Data Warehouse Environment, needs to get all the account's monthly balances between Jan/2020 and Dec/2020. She wasn't able to do it alone, and asked for your help. Add to your resolution the SQL query used to retrieve the data needed (the necessary tables in csv format were sent along with this pdf, on folder *tables/*). Feel free to use the dialect of your choice, but please specify the SQL engine.

Imagine now that you could remodel the data warehouse environment freely, keeping in mind that Nubank is always evolving with new products (Whatsapp Payments, PIX, phone recharge, etc) and it is also expanding to new countries, so our data warehouse needs to accommodate all these incoming changes. Which modifications would you propose and why? Remember to consider that other analysts will be using the same structure, so it should be as clear as possible. Feel free to change, remove or add tables/fields to generate a better database design.

In order to change our data warehouse above with your suggestions, we need to come up with a migration plan since many people are consuming data from it. Which strategy would you propose in order to implement those changes?

Jane's friend, Pepino, wants to know how well PIX is doing inside Nubank. For that, he wants your help to come up with indicators that can be used to track the performance of the product. Which metrics would you suggest to track it and why?

Here is a summary of what we expect from you:

1. Create a SQL query to help Jane retrieving the monthly balance of all accounts (this query should be made using the warehouse structure before the changes you propose on 2.)
2. Improve the data warehouse architecture and justify your changes
3. Come up with a strategy to implement the warehouse changes you proposed
4. Propose metrics to track PIX performance and its impact on Nubank. Feel free to come up with any metrics you consider relevant

Here is some tips that might help you create your resolution:

- Keep in mind that if your case gets accepted, you'll need to present it to people
- If you want to send additional files, feel free to do so
- You were presented to the entire data architecture, but for the purposes of this case you can focus specifically on the data warehouse environment