Analytics Case

Introduction

You are in the next stage of the 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

The Architecture for this use case is divided in three different environments: Production, Data Warehouse and Reporting (Figure 1).

The Production Environment is where transactional 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 visualizations for monitoring and decision-making.

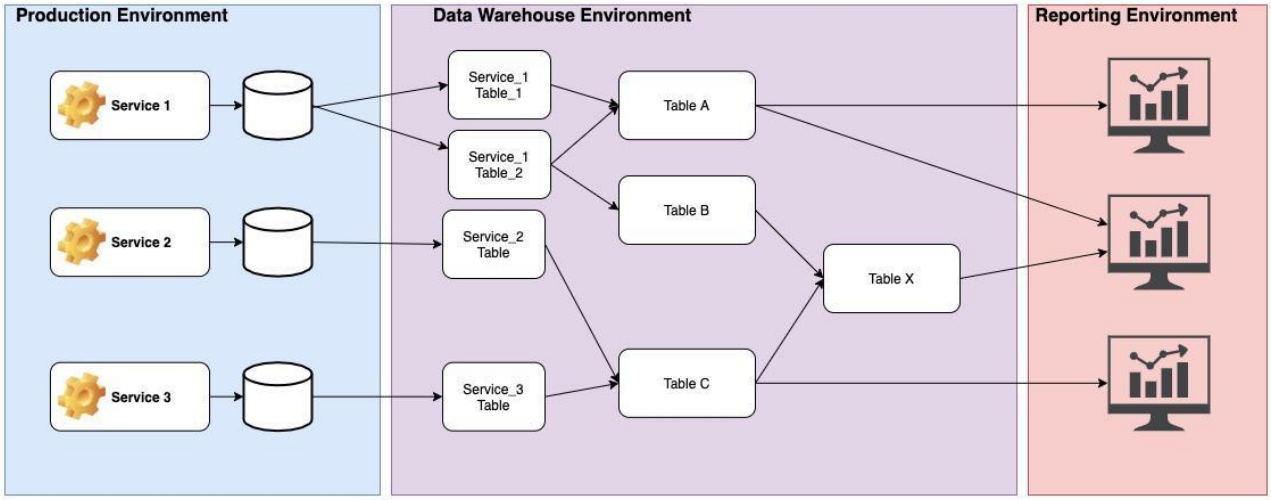

Figure 1. The environments: Production, DataWarehouse and Reporting.A slice of the table structure from the Data Warehouse Environment is depicted in Figure 2.

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

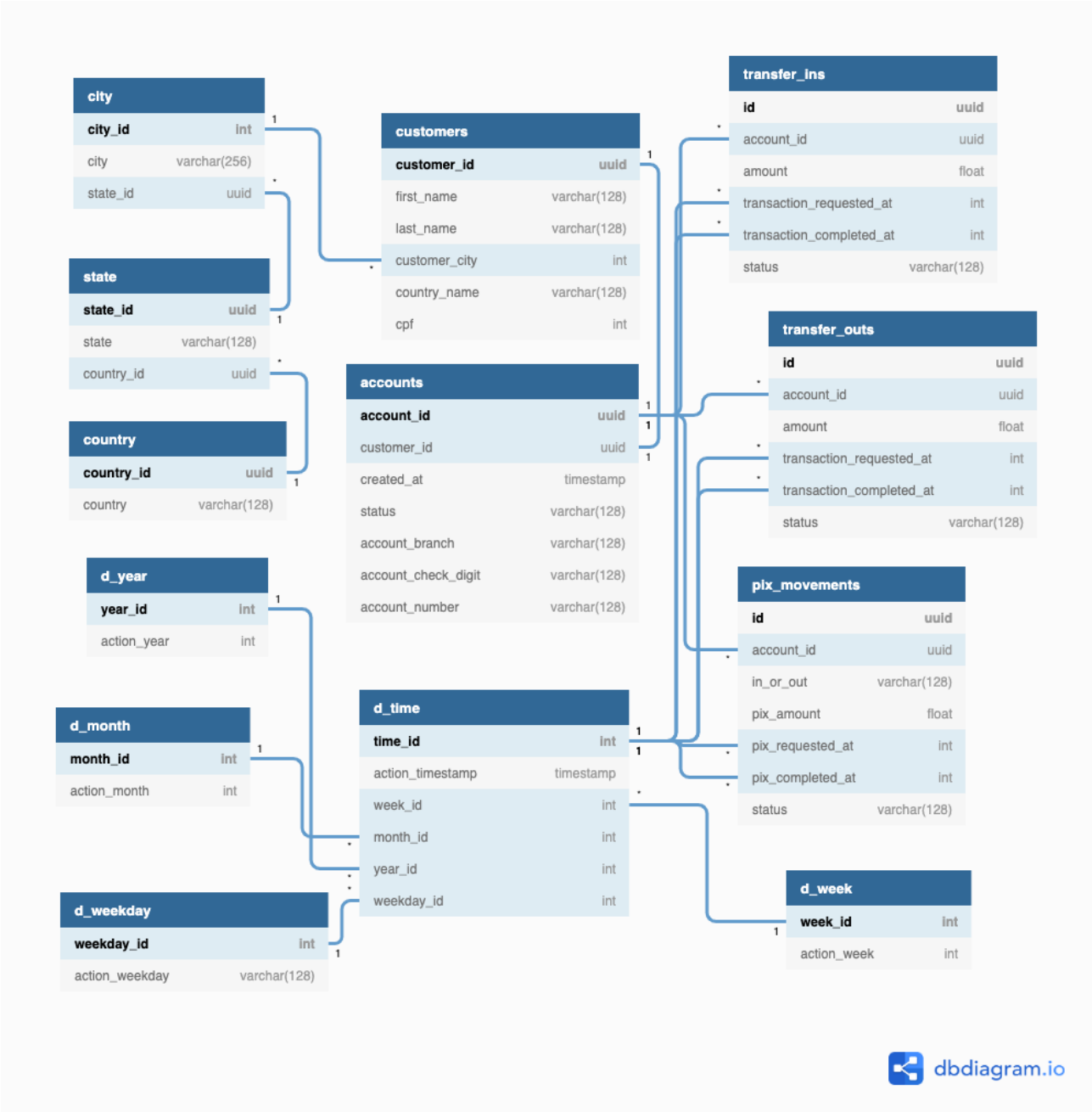

Figure 2. A slice of the table structure from the Data Warehouse Environment (diagram/table_diagram.png). You can check the code used to generate these diagrams on file (diagram/table_diagram.txt).

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 analyzing customer behavior, 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 couldn’t 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 should come with the same resolution but using Python instead (you can still use SQL but make sure it runs inside of your Python notebook).

Here is a summary of what we expect from you:

1. Create a SQL file to help Jane retrieving the monthly balance of all accounts

2. Create a Python notebook that contains the solution for the same monthly balance of all accounts

Here are 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 Analytics team

- If you want to send additional files, feel free to do so.