

8WeekSQLChallenge.com

CASE STUDY #4



DATA BANK

That's money.

DataWithDanny.com

Introduction to the case study Data Bank

There is a new innovation in the financial industry called Neo-Banks: new aged digital only banks without physical branches.

Danny thought that there should be some sort of intersection between these new age banks, cryptocurrency and the data world...so he decides to launch a new initiative - Data Bank!

Data Bank runs just like any other digital bank - but it isn't only for banking activities, they also have the world's most secure distributed data storage platform!

Customers are allocated cloud data storage limits which are directly linked to how much money they have in their accounts. There are a few interesting caveats that go with this business model, and this is where the Data Bank team need help!

The management team at Data Bank want to increase their total customer base - but also need some help tracking just how much data storage their customers will need.

This case study is all about calculating metrics, growth and helping the business analyse their data in a smart way to better forecast and plan for their future developments!

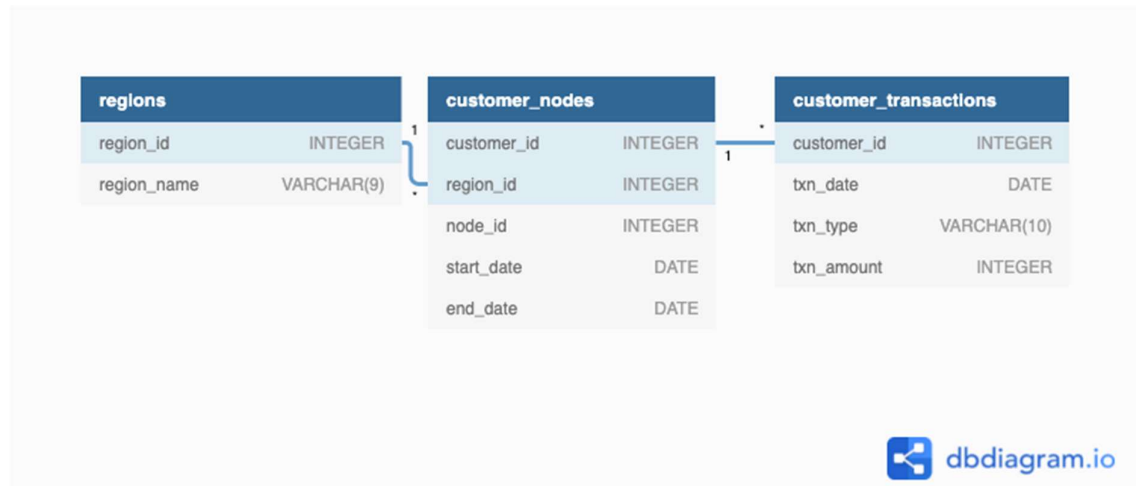
Key Concepts:

- **Neo-Banks:** Digital-only banks that operate without physical branches.
- **Distributed Data Storage Platform:** World-class secure storage system linked to customer balances.
- **Customer Data Allocation:** Customers receive data storage based on the amount in their accounts, blending banking with data management.

AVAILABLE DATA

The Data Bank team have prepared a data model for this case study as well as a few example rows from the complete dataset below to get you familiar with their tables.

Entity Relationship Diagram



Before we dive into the analysis, let's take a moment to understand the different tables that were eventually created in the database.

Table 1: Regions

Just like popular cryptocurrency platforms - Data Bank is also run off a network of nodes where both money and data is stored across the globe. In a traditional banking sense - you can think of these nodes as bank branches or stores that exist around the world.

This regions table contains the region_id and their respective region_name values.

region_id	region_name
1	Africa
2	America
3	Asia
4	Europe
5	Oceania

Table 2: Customer Nodes

Customers are randomly distributed across the nodes according to their region - this also specifies exactly which node contains both their cash and data.

This random distribution changes frequently to reduce the risk of hackers getting into Data Bank's system and stealing customer's money and data!

Below is a sample of the top 10 rows of the `data_bank.customer_nodes`

customer_id	region_id	node_id	start_date	end_date
1	3	4	2020-01-02	2020-01-03
2	3	5	2020-01-03	2020-01-17
3	5	4	2020-01-27	2020-02-18
4	5	4	2020-01-07	2020-01-19
5	3	3	2020-01-15	2020-01-23
6	1	1	2020-01-11	2020-02-06
7	2	5	2020-01-20	2020-02-04
8	1	2	2020-01-15	2020-01-28

customer_id	region_id	node_id	start_date	end_date
9	4	5	2020-01-21	2020-01-25
10	3	4	2020-01-13	2020-01-14

Table 3: Customer Transactions

This table stores all customer deposits, withdrawals and purchases made using their Data Bank debit card.

customer_id	txn_date	txn_type	txn_amount
429	2020-01-21	deposit	82
155	2020-01-10	deposit	712
398	2020-01-01	deposit	196
255	2020-01-14	deposit	563
185	2020-01-29	deposit	626
309	2020-01-13	deposit	995
312	2020-01-20	deposit	485
376	2020-01-03	deposit	706
188	2020-01-13	deposit	601
138	2020-01-11	deposit	520

Case Study Questions

The following case study questions include some general data exploration analysis for the nodes and transactions before diving right into the core business questions and finishes with a challenging final request!

The case study questions are grouped into:

A. Customer Nodes Exploration

B. Customer Transactions

C. Data Allocation Challenge

E. Extra Challenge

F. Extension Request

Goals:

- **Customer Growth:** The Data Bank team is focused on increasing their customer base.
- **Data Storage Management:** The team needs help tracking the **data storage** needed for customers based on their account balances.