

Self-Help Data Warehouse Project

Data Warehouse Exercise:

The exercise is to build a schema for a simple data warehouse that tracks information about bank loans.

First, create the database structure in any popular relational database environment you like (SQL Server Express, PostgreSQL, MySQL, etc).

Then, populate the database with some fake data, and write the reporting queries specified below.

The data:

We have two sets of information about the loans:

(1) Loan origination information

- loan amount
- on what date the loan was made
- on what date the loan matures
- interest rate
- loan product (home mortgage, car loan, etc)
- bank branch (random names / city names are fine)
- name of the loan officer

(2) Loan borrower information

- name
- address
- date of birth
- race
- ethnicity
- gender
- income

Reports:

1. Write a query that shows year-to-date loan production (amount and count) by bank branch and month of close
2. Write a query that shows the top 5 loan products in each state, in terms of loan production, by amount lent. Use the borrower's address to determine the state.
3. Write a query that illustrates whether we are consistently making loans to low-income or minority borrowers across geographies.