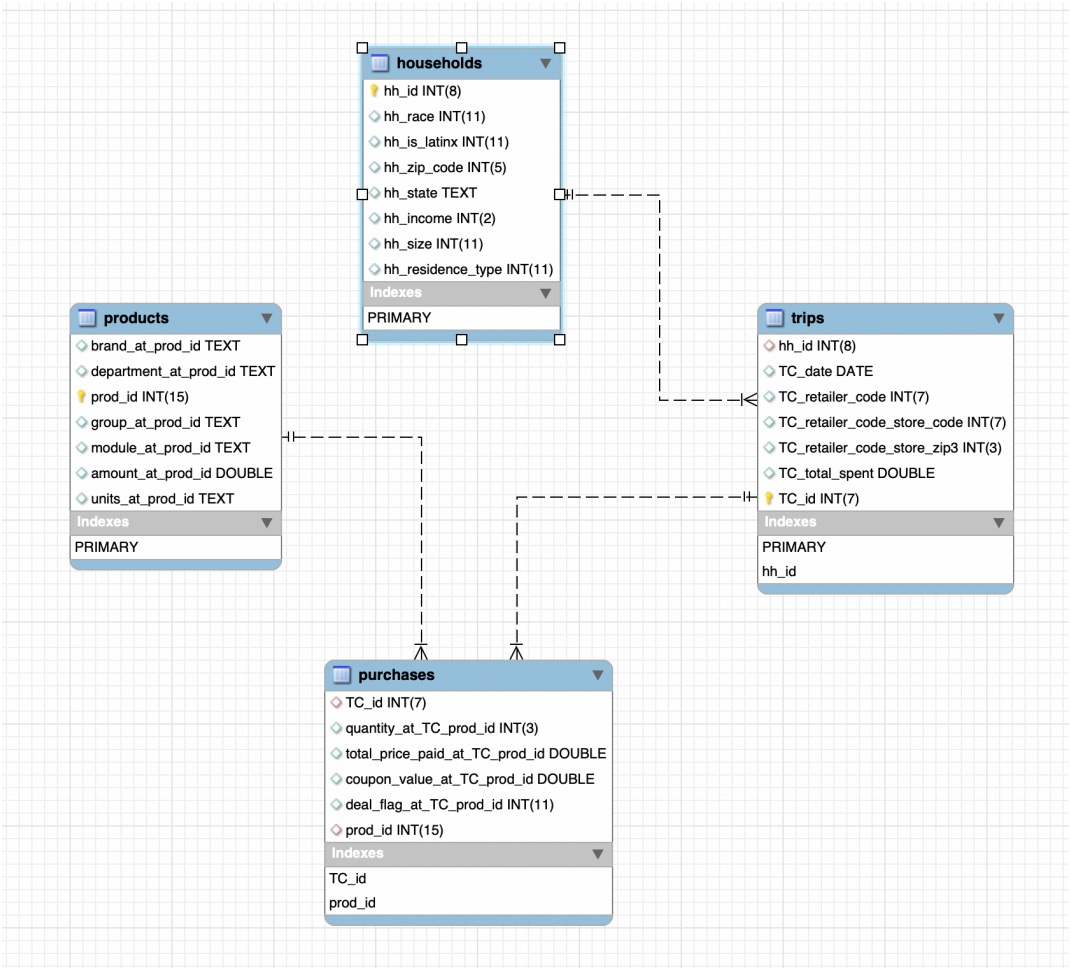
Part 1– Creating a database and Tables



Instructions

We created a schema and tables in workbench first, then imported the raw data into created tables using Python. In this way, we can control the types of every column easily when we are creating tables. If we want to change the type of a column after we imported data, it will be so time-consuming because of the high volumes of data. The code of importing data is attached.

As for the types of columns, we changed the TC_date as DATE format for future operations. As for each column represented by integers, we set the integer length according to the longest value it appears in the data, setting the type of prod_id as INT(15), hh_id as INT(8), hh_zip_code as INT(5), hh_income as INT(2), TC_id as INT(7), TC_retailer_code as INT(7), TC_retailer_code_store_code as INT(7), TC_retailer_code_store_zip3 as INT(3), hh_zip_code as INT(5), hh_income as INT(2), and other integer colums as INT(it is 11 by default).