Our design

We decide to remove country attribute in every table. Because our target is to research about the housing price index in Canada, country attribute will be very redundant.

We also made our data instead of setting date as a timestamp, we set date as string. Because our raw data is raw data, so if we change the date to string, we don’t need to worry that the date data will be the future data. And we focus on monthly or quarterly change of each factor, we will remove day in the date variable. There are two options that we can made for date. One is to get year and month as two attributes, another is to get year-month attribute. To make our schema looks easy to read, we refer primary key is one attribute instead of combination of two attribute. Timestamp variable always need day, so we decide to change the data type to string.

Data clean

We decide to clean the data inside the raw csv data and we decide to delete the unrelated raw, and change the date raw to yyyy-dd manually. Then we input this date into our defined schema.