



CSCI5408

ASSIGNMENT 1 – PROBLEM 3

Opportunities in Halifax -Flexible project with
abstract problem, and dataset)

Dhrumil Amish Shah
B00857606

Problem #3: Opportunities in Halifax (Flexible project with abstract problem, and dataset)

Table 1 consists of a list of entities that are the factors that help real-estate clients to decide the opportunities of having better future scope in Halifax.

Table 1 – Identified entities and their description

No.	Entity Name	Entity Description
1.	schools	The schools entity is responsible for storing all the information related to the schools.
2.	business_agencies	The business_agencies entity is responsible for storing a list of agencies that supports blooming businesses. This entity details information about agencies the real-estate agents to make a wise decision.
3.	trades	The trades entity lists down the producers, brands, and other contact details of local food, craft and artisan products.
4.	cooperatives_information	The cooperatives_information entity helps the real estate agents to analyze various business services available in high demand.
5.	restaurants	The restaurants entity stores information related to food establishments. This is currently one of the dominant industry in the market.

Initial data modelling (Chen model) diagram

Figure 1 is the initial data modelling diagram created on a paper with the use of entities discovered in **Table 1**.

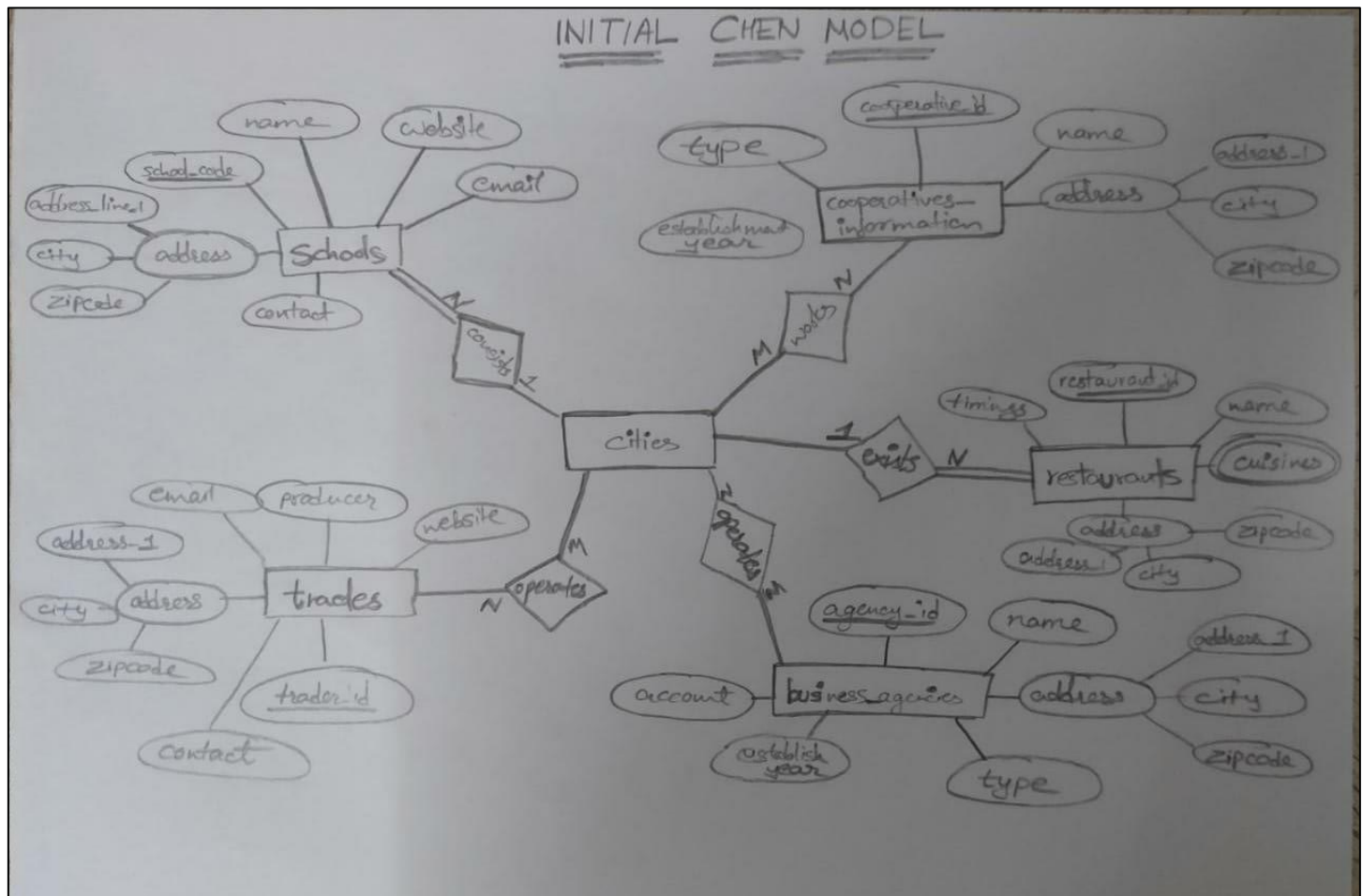


Figure 1 - Initial ERD Diagram(Chen model) designed on a paper

Steps performed for cleaning/transforming data

1. A lot of data was available in all the datasets for different cities in Nova Scotia, but since we were interested in the factors that help the real-estate clients to do business in Halifax and expand their opportunities, data related to Halifax city and other surrounding cities like Bedford, Dartmouth, Wolfville and Truro were considered.
2. Out of the dataset available, we filtered out all the records that had city name not equal to the ones above mentioned.
3. In few datasets, some city names were incorrect. For example, along with Halifax, the dataset also contained Halfax and Haifax. Similarly, Wolfville city was named Wolfvile in few columns. These anomalies were corrected in all the datasets and the cities were named correctly.
4. Few records had empty cells which were replaced with NULL values to avoid data loss.
5. To avoid any data integrity, all the values in the field city were made consistent (i.e., in lowercase). For example, Halifax became halifax.
6. A city dataset was created with unique cities so that it can act as a central entity while developing the database.

⇒ The datasets that were loaded are attached as a folder – **datasets_problem3**

⇒ The dump that was exported – **sqldump_problem3**

Dump files (Folder – sqldump_problem3)

1. csci5408_a1_p3_schools.sql
2. csci5408_a1_p3_business_agencies.sql
3. csci5408_a1_p3_trades.sql
4. csci5408_a1_p3_cooperatives_information.sql
5. csci5408_a1_p3_food_establishments.sql
6. csci5408_a1_p3_cities.sql

Dataset files (Folder – datasets_problem3)

1. schools.csv
2. business_agencies.csv
3. trades.csv
4. cooperatives_information.csv
5. food_establishments.csv
6. cities.csv

Datasets used

1. <https://data.novascotia.ca/Education-Primary-to-Grade-12/Nova-Scotia-Public-School-Enrolment-by-Board-and-S/cjnf-ywvt>
2. <https://data.novascotia.ca/Business-and-Industry/Nova-Scotia-Business-Inc-Export-Development-and-In/Gaac-8xtn>
3. <https://data.novascotia.ca/Agriculture-and-Agri-business/Select-Nova-Scotia/t8ij-8svj>
4. <https://data.novascotia.ca/Business-and-Industry/Nova-Scotia-Co-operatives/k29k-n2db>
5. <https://data.novascotia.ca/Permits-and-Licensing/List-of-Licensed-Food-Establishments/2shh-dv8p>

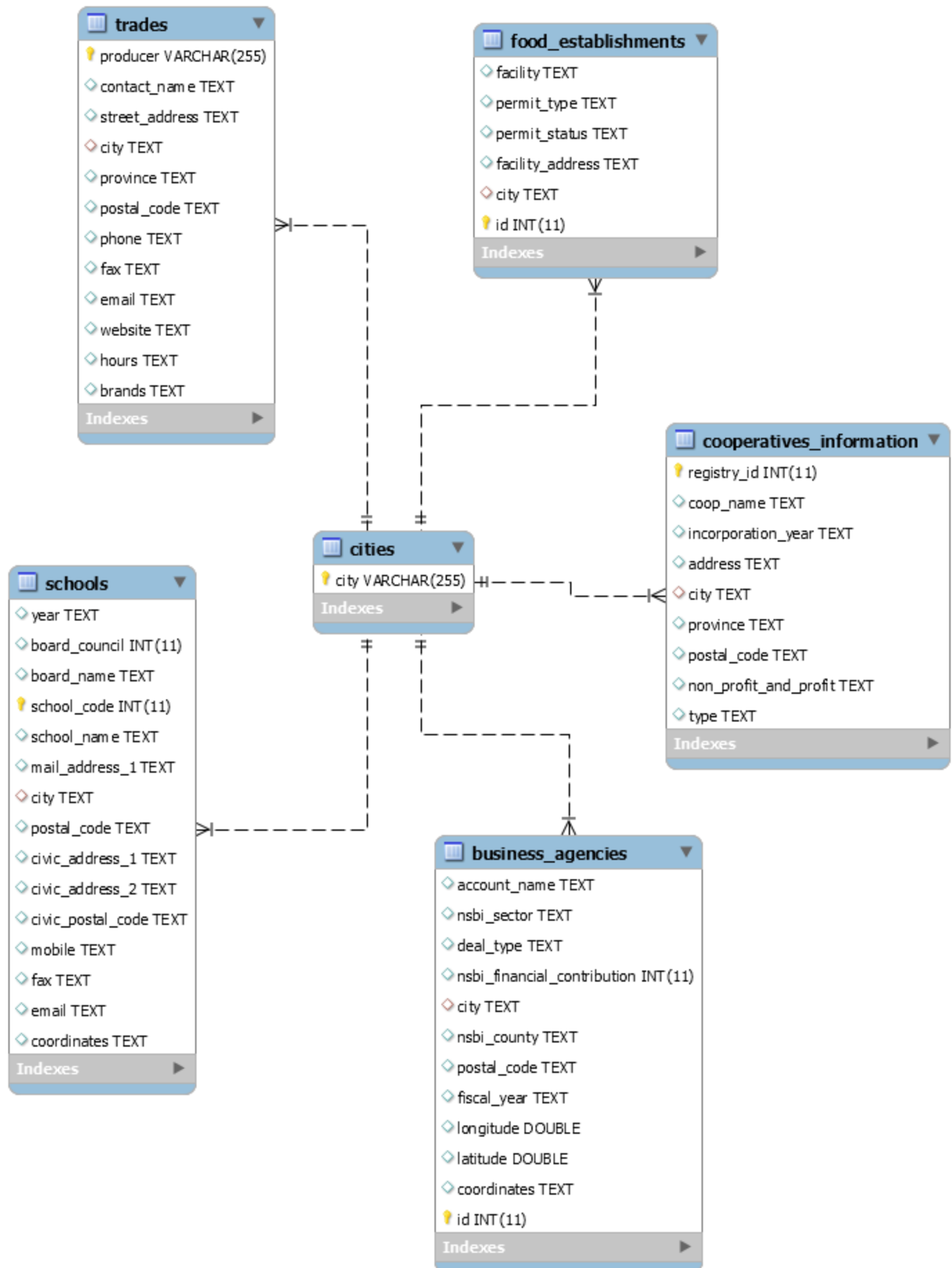


Figure 2 - Final ERD diagram(Crow's foot model) designed in MySQL Workbench