Introduction to Automotive Industry

The automotive industry is one of the largest and most impactful in Canada. It makes up for a significant portion of the economy and therefore would benefit from being streamlined into one database. The automotive industry currently accounts for 12% of Canada’s GDP, this means that it plays a significant role in the development and sustainability of the Canadian economy. On a more global scale, the automotive sector contributes to 25% of Canada’s manufacturing trade generated. This means that of all the manufacturing goods that are exported automobiles are among the most common. Currently, Canada produces about 14.7 million vehicles annually which are sold both locally and abroad. North American automotive production makes up for 23% of global production meaning that it is a critical piece of the Canadian and global economies (Biesebroek). Because of the significant impact on the Canadian economy, we noticed there was a need to create a database that would help users navigate the substantiative Canadian automotive industry. With this information in mind we decided to design a database directory for the Automotive Industry using the various techniques learned in class.

Conceptual Database Design

Firstly, starting off with our main entity type is the dealership entity. We chose to start here for the obvious reason of our directory being an automotive directory for Windsor-Essex County. Upon our starting point of the dealership entity, we began to brainstorm ideas of the main practices of a dealership. We discovered that most dealerships all have a sales, parts and service department. We created an entity for both the parts and service side of the dealership and rather than a sales entity we focused more so on the inventory each dealership had. This information would allow our directory to conveniently navigate the parts available, services dealerships were offering as well as their current vehicle inventory on hand. Our next step was focusing on the staff at the dealerships. This information allows our directory to identify staff members at every dealership quickly and easily. Additionally, with inventory and staff information in our system, we felt current promotions and pricing on vehicles would be key information if this were a real-world directory. For our final entity we decided to approach from the dealership perspective and provide value for them by creating a supplier entity. This information is very important as every dealership’s primary contact is their automotive supplier, not only for new vehicle purchases but also for warranty transactions.

Our data dictionary is also accessible in our “Project Data” spreadsheet for a more clear and visible illustration.

We can begin by first defining what an Entity Relationship Model (ER model) is, an ER model is the visual representation of relationships between different entities. ER diagrams are a simplified view of the database, which provides the basic design of the database while showing the relevant data points that are important and the relationships between these datapoints. Since ER Diagrams are observed through graphical representation, they are easier to understand and thus easier for communication to management regarding the database. The first aspect of the ER diagram is the entity type, the entity type is a business notion that holds meaning to a group of users while an entity is a specific occurrence of an entity type. In our ER model, the entity types can be observed as dealers, suppliers, inventory, promotions, services, parts, and staff. Next, we can define attribute types, an attribute type is a property that falls under an entity type while an attribute is a specific occurrence of a given attribute type. In our ER model we can observe attribute types such as name, address, and email, among others. A key attribute type is a distinct attribute, in our model we can observe key attribute types such as dealer ID, supplier ID, part ID, service ID, employee ID, vehicle ID, part number, and promotion ID. The key attribute is denoted by an underline. We can also observe composite attribute types in our model, a composite attribute type is when an attribute type can be divided into more parts, for example, address can be further expanded into province, city, street name, and unit number. A multi valued attribute type can also be observed within our model. A multivalued attribute type happens when an attribute type may have several values, for example, it is possible for emails to have several values. We denote this by a double circle around the attribute.

Next, we can discuss relationship types within the ER model, a relationship type defines the relationship between two or more entity types. These relationship types can take the form of unary (degree 1), binary (degree 2), and ternary (degree 3). A benefit of employing the relationship type in the ER model is the indication of the direction and significance of the relationship between entity types which in turn make understanding and communication of the model easier. These relationships can be defined by their cardinality, the minimum cardinality can be either 0 or 1 where there is partial participation or total existence dependence respectfully. The maximum cardinality is either 1 or N where they can be defined as 1:1, 1:N, or N:M relationships. It is possible in some cases to have a ternary relationship where three entity types hold a relationship, although it is acceptable to model this by using a binary approach barring any resistance from the model's clarity.

By viewing figure 1 below, we can observe the following entity types: Dealers, Staff, Supplier, Services, Promotions, Parts, and Inventory. These represent the main data points of which we collected data about for our MySQL database. Branching out from each entity type we can observe their corresponding attribute types. Each entity type has its own corresponding key attribute types as well, thus there are no weak entity types found in our model. If we take a closer look at our model, we can see that the entity ‘dealers’ has a key attribute called ‘dealership ID’, furthermore we can see that suppliers  supplier ID, staff  employee ID, inventory  vehicle ID, parts  part ID, promotions  promotion ID, and services  service ID. We also can observe composite attribute types in our model; name can be divided into first name, last name, and address can be divided into province, city, street name, and unit number. Our model also contains multivalued attributes where email can have several potential values.

Figure 1.

Now by taking a closer look at figure 2 below, the entity relationship model, we can better observe the relationships between the entity types by temporarily removing the attributes from the model. Beginning with a ternary relationship between Dealers, Suppliers, and Inventory. In this case there is a 1:N relationship between dealers and suppliers since a dealer will primarily use one supplier for their vehicles while a supplier could potentially supply their product to several dealers, for example, if there are more than one Ford dealership in relative proximity, they will typically use the same supplier for new Ford vehicles so the supplier will supply several dealerships in the area. There is a 1:N relationship between Dealers and Inventory since the dealer will list several vehicles for sale at the same dealership. There is also a 1:N relationship between Supplier and Inventory since the supplier will supply the inventory for the dealerships to sell. Furthermore, we can observe a 1:N relationship between Inventory and Parts, 1:N relationship between Dealers and Staff, 1:N relationship between Dealers and Services, and finally a 1:N relationship between Dealers and Promotions

Figure 2

Limitations:

Although the ER model has many benefits and I would recommend its use in any company, it does have some limitations. For example, an ER model is unable to assign an event to an entity after a given time period has elapsed. Another limitation in question is the ability of ER model to be consistent among multiple relationship types, in regard to our ER model, we cannot specify that certain promotions should be assigned to certain vehicles and how these promotions may differ with a variable cash down option a customer has on any given vehicle. Domains are disclosed from ER models, we were unable to classify an entity or attribute type as > 0 or < 0. Lastly, we were unable to include functions in our model, for example, we were unable to include the average number of inventories sold in the ER model or the number of sales an individual salesperson has had. The ER model can be extremely effective, but it does not provide a level of specificity that might be necessary for a more detailed analysis because it is designed to offer a general but very clear understanding of how the database is structured and how it will operate.

Physical Database Design:

Database Management System:

Next, we began to consider how to bring our conceptual directory database to a database management system. With the options provided of using either Microsoft Access or MySQL we had to make a careful decision to decide which platform to use based on the needs of our target industry.

As discussed earlier, the automotive industry is very vast with multiple relations between dealers, supplier and customers. The industry standard for this scale of operations is making use of enterprise applications which are often developed in other programming languages while having a fast backend database to quickly and efficiently retrieve and update data.

After careful consideration we decided to use MySQL as our database management system for the following reasons:

• MySQL is considered an enterprise DBMS

• MySQL allows connecting to applications in a standardized and secure way

• MySQL is scalable and can be hosted on a separate server while Micrsoft Access is considered a desktop DBMS

• MySQL has multiple extensions which can prove useful in this industry

Once the decision was made, we began to start designing the physical database based on the conceptual entity relation diagram created earlier. All 7 entities were created into specific tables with separate primary keys and attributes.

Attributes such as Phone Numbers were set to be CHAR(10) numbers and would be inputted in (XXXXXXXXXX) format with “-” in between.

Two State Attributes were set to be BOOLEAN/TINYINT(1) data type attributes with 0 indicating their first State and 1 indicating their second State, each of these states are detailed in the data dictionary.

Composite attributes such as address and full name were split into their sub attributes, address was split into Unit Number, Street Name, City, Province while full names were split into First Name and Last Name.

Certain attributes such as Rating, Condition, Odometer were given logical default values to handle scenarios where this information was not provided. For example, Condition and Odometer were defaulted as Null to let dealers know that the vehicle condition and kilometers used is unknown instead of just being set to 0 which would be misleading.

Naming conventions used for attributes are as follows:

• All attribute names are lower case

• Spaces between two words are show by “\_” (ex. “dealer\_id” , “phone\_number” )

• Longer attribute names are given abbreviations (ex. “condition” is “cond”)

• Certain attribute names were adjusted to avoid overlapping with SQL keywords (ex. “condition” to “cond”)

Strings that required large number of characters such as promotional details and part descriptions were given the data type VARCHAR(3000) and VARCHAR(250) to allow storing larger strings.

Lastly for the release year attribute of vehicle's we decided to use the YEAR(4) data type which takes 4-digit year values in the format of (YYYY) as it was more relevant for the automotive industry instead of using a DATETIME data type.

Gathering relevant data and identifying high quality sources:

Our data is gathered from various dealership and supplier websites, nearly every dealership listed was visited on their website to gather data about them while also looking for data such as vehicles they are carrying, promotions they are currently running as well as any parts or services they offer. From their websites we also found suppliers which were added directly to our dataset. Essentially the procedure used for gathering data for our directory directly flows just like our ER diagram where we begin with dealer and the network of relationships continue as we look deeper into the various aspects of each dealership.

In order to ensure we only gathered data from high quality sources, we picked dealerships that are well known in the Windsor region as our team is strongly familiar with the city. In terms of suppliers, all of these are well known companies that manufacture vehicles with a very strong brand image which validates the information on their website is of the highest quality and most recently updated/

Data Preparation, Cleaning and Importing:

Since most of the data was manually gathered, it was already organized into neat lines in an excel spreadsheet. In order to prepare the data for importing, we had to ensure that the columns matched the data type format for each attribute. Money columns had to be set as Numbers in excel and Date columns had to be generalized to 4-digit Year number.

After adjusting formats, each columns had to be verified to be in the exact same order as the attributes in each table to further streamline the importing process.

Lastly any excess columns or rows were deleted to ensure that our next step won't have any extra delimiters.

Next, each worksheet from the excel file was exported as a CSV file with the first row being column headers and the rest holding the data gathered. These CSV files were directly imported in the database through MySQL Workbench’s table data import wizard.

By following these steps and maintaining strong communication between members working on data gathering and database design we were able to simplify the process and ensured we didn’t run into unexpected issues at this step.