Contents

[Abstract: 1](#_Toc529727361)

[Preface: 1](#_Toc529727362)

[Deliverable: Dealer of Pre-Owned Cars ER Diagram 2](#_Toc529727363)

[Dataset considerations 3](#_Toc529727364)

[Deliverable: Car Sales Dealership ER Diagram from Assignment 1 5](#_Toc529727365)

[Deliverable: ER diagrams representing each step of your integration process, with each step accompanied by a description (in narrative prose) of your integration process 6](#_Toc529727366)

[Deliverable: Final ER Diagram 9](#_Toc529727367)

[Difficult Decisions 9](#_Toc529727368)

[Appendix 9](#_Toc529727369)

[3.A Population of Tables for Assignment 3 for the Pre-Owned Car Sales Dealership are found below for reference 9](#_Toc529727370)

[References 12](#_Toc529727371)

## Abstract:

The goal of assignment 3 was to provide ER diagrams that would follow the process of integrating two separate corporations. The datasets both come from car dealerships with the distinct that one sells used cars, and one sells new cars. The data sets are very similar, but we have been asked to use a derivation approach and apply the same schema to both datasets. The first deliverable was to provide an ER diagram of the schema for a Dealership of pre-owned cars, followed by an ER diagram of the database design established for assignment 1 of a car dealership focused on new cars. The assignment requires documenting the process of merging the two schemas. Finally, an ER diagram of the merged schema is provided.

Full page pdfs of the diagrams provided in this document have been made available in a github repository: <https://github.com/megado123/cs-598-Assignment3.git>

## Preface:

As part of the data curation process, it was requested that tables could be identified as entities and relationships. After reading a publication on database design, (Earp, 2003), I have documented entities, weak entities, and relationship tables.

When determining if a table was to be considered an Entity, a Weak Entity or a relationship, the following principals were used. If a table represented a physical entity, or a concept that could exist independently of other tables, it was considered an *entity*. If a table represented a physical entity, or a concept but could not exist independent of additional supporting tables, it was considered a *weak entity*. If a table existed to link information together, it was considered a *relationship*.

## Deliverable: Dealer of Pre-Owned Cars ER Diagram

|  |
| --- |
|  |

Figure 3.1 Create an ER diagram for Pre-owned dealer database, as described in the attached file

A full page pdf of the diagram in Figure 3.1 is available in the following git repository: <https://github.com/megado123/cs-598-Assignment3/blob/master/1%20Deliverable%20Dealer%20of%20Pre-Owned%20Cars%20ER%20Diagram.pdf>

The tables are labels as Entities, Weak Entities and Relationships according to the definitions provided in the preface. Note the relationship tables are colored in blue to provide distinction between them and the entity tables (both Entity and Weak Entity Tables are white).

The customer table is an entity since is is something that physically exists, a customer and has attributes. It has a one to many relationship with the sale table since a customer can participate in many sales. The customerName and phone were selected as a uniqueness constraint on the table.

The associate table is an entity since a sales associate physically exists and contains a name. It could have been considered an attribute of a sale, however this would have resulted in duplicates in the sales table so to avoid duplicate entries into the sales table, it was placed in its own table. One associate may have many sale records.

The vehicle table is an entity since a vehicle physically exists as vehicles can be bought and sold and has an attribute of a sticker price. One vehicle may be involved with many transaction activities, as it can be bought and sold in different transactionActivity records.

The transactionActivity table is considered a weak entity. Conceptually a transaction activity exists and has an attribute of a price associated with it but cannot exist on its own without something to sell or by, so it is a weak entity. One transaction Activity can be involved in one sales to transaction relationship. A sales transaction may have 1 or many transaction activities. This does allow for duplicate entries in the transactionActivity table (for a vehicle and price), but these are in-deed not duplicates (the same car maybe bought or sold at the same price in a separate sale) so the primary key being an identity column will ensure that the transactions are represented in the table. Keeping these duplicates will enable maintaining the lineage of transactions that have occurred for a given vehicle.

The sale table is also considered a weak entity. It represents a group of transactions that contain attributes but cannot exist on its own so it is considered to be a weak entity.

The saleTransactions table is a relationship table since it is tying together the transactions with the sales. One sale may have many transactions. This table could have be merged into a foreign key constraint in the transactionActivity table, but that would not have provided clear delineation between entities and relationships, so it was placed into a separate table.

To establish the model provided in figure 3.1, several data curation methods were required to be applied. The notes column was included twice. This was easily remedied by removing the additional column. In addition, the data required normalization to remove redundancies in the logical data design.

### Dataset considerations

The new datasets introduced during Assignment 3 appeared to have data quality issues as the definitions for the attributes were not well defined. As can be seen from the definitions provided in table 3.1 below, there is no clear definition for when something is not the case, so we may see NULL, no, or even – to represent an attribute value. In addition to unclear attributes, we see in transaction: 10123457 both BUY and TRADE are both “y”, but by the definitions provided below in table 3.1, that should not be an acceptable occurrence. This indicates that the data model currently in use, is not reflecting reality and needs to be adjusted. In addition to the questionable data provided, it is clear that R2-D2 is not a probably name for a sales associate.

In addition to R2-D2, the occurance of the “notes” attribute twice in the sample data made it clear there would be data quality issues that the database design would need to protect against to the quality of data.

|  |  |
| --- | --- |
| Attribute | Definition |
| BUY | If there is a "y", this was a transaction in which the preowned dealership BOUGHT a car, without making a sale |
| TRADE | If there is a "y", this is a transaction in which the preowned dealership both BOUGHT and SOLD a car |
| SALE | If there is a "y", this is a transaction in which the preowned dealership only SOLD a car |

Table 3.1

After analysis of the data, it was clear that capturing the transactions that were occurring at the dealership would accurately capture the information without information loss. If a transaction was a SALE, a TRADE or a BUY could be calculated based on the transactions that occurred. If a negative transaction occurred in the database, that would indicate money leaving the dealership and thus was a BUY provided there was no additional transaction with a positive value. If there were two transactions that occurred, it could be calculated that the sale was a trade. Further a transaction could be determined to be a SALE if there was one transaction that occurred with a positive amount found within the TransactionActivity table.

Sticker was determined to be an attribute of the vehicle itself as it is not required for a transaction, but to give a complete picture for a vehicle if a sticker price applied. (As an example a vehicle sold to the dealership which the dealership had not previously sold would not have a sticker price).

The design of the table provided above accurately enables capturing the data provided within the sample dataset, but does so in a more efficient manner. After handling the data cleansing activities including removing duplicate columns and establishing data normalization into a single schema, the data is better prepared for the integration steps outlined later in this document.

For comprehension analysis of the design, in Appendix A of this assignment, the database design has been implemented with the sample data provided. The a SQL implementation of creating the tables and populating the data can be found within the previously mentioned github repository titled, “[1.1 Insert And Create of Pre-Owned.sql](https://github.com/megado123/cs-598-Assignment3/blob/master/1.1%20Insert%20And%20Create%20of%20Pre-Owned.sql)”

## Deliverable: Car Sales Dealership ER Diagram from Assignment 1

|  |
| --- |
|  |

Figure 3.2 Create a separate ER diagram that reflects the schema you designed for Assignment 1

A full page pdf of the diagram in Figure 3.2 is available in the following git repository: <https://github.com/megado123/cs-598-Assignment3/blob/master/2%20Deliverable%20Car%20Sales%20Dealership%20ER%20Diagram%20from%20Assignment%201.pdf>

The figure above shows an ER diagram of the logical model for the database schema delivered for assignment 1. Assignment 1 provided the justifications for the design as well as sample data applied to the database design.

The tables are labels as Entities, Weak Entities and Relationships according to the definitions provided in the preface. Note the relationship tables are colored in blue to provide distinction between them and the entity tables (both Entity and Weak Entity Tables are white).

The variations on the lines is to make it visualization of the linkages between the tables easier to read.

The a SQL implementation of creating the tables and populating the data can be found within the previously mentioned github repository titled, “[2.1 Table Creation from Assignment 1.sql](https://github.com/megado123/cs-598-Assignment3/blob/master/2.1%20Table%20Creation%20from%20Assignment%201.sql)”

## Deliverable: ER diagrams representing each step of your integration process, with each step accompanied by a description (in narrative prose) of your integration process

The diagram below is the first effort in schema integration between the pre-owned sales dealership and the car dealership from assignment 1.

When attempting to integrate the two schemas, the first component that was considered was the customer table and its integration with the customerPurchasers table from assignment 1. The customer name provided from the pre-owned dealership was considered a composite attribute and was broken down into simple attributes of customerNameFirst, customerNameLast, and a customerNameMI was included. The attribute customerNameMI will be null in the customers from the pre-owned car dealership, but this enables no information loss when combining the two schema. The customerOccupationId was then also added to again ensure no information loss when the two-schemas are merged. The phone from the pre-owned dealship was included into the customerPurchers table which will be null for the car dealership from assignment 1, but it will enable the goal of preventing information loss.

|  |
| --- |
|  |

Figure 3.3 Migration of customer Entity into customerPurchasers Entity

The city and street were removed from the customerPurchasers table. This did impact the table design as shown in the figure 3.3.1 below. There was a one-to-many relationship in assignment 1 between the addressId and the customerAddressId. That was modified into a one to zero-or-many relationship.

|  |  |
| --- | --- |
| Assignment 1 Table Design | Assignment 3 Schema Merge |
|  |  |

Figure 3.3.1 Figure to show-case modification on table design because of merging schema

Finally, the phone number needed to be considered. Also note the primaryPhone was added to the customerPurchasers. It was renamed from phone to primaryPhone in an effort to reduce confusion regarding if it would be a home address, or a cell. This should ideally would be confirmed with the pre-owned car dealership to ensure this is the most appropriate name for the column.

Another modification that is also captured in figure 3.3.1 is that customerOccupationId relationship is also modified from a one-to-many to a one-to-zero-or-many relationship. This enables the common theme of integration without information loss.

|  |
| --- |
|  |

Figure 3.4 Migration of the vehicle and vehicleInventory tables

The figure above shows the process of integrating information pretaining to the vehicle entities. It was noted that vehicle details from assignment 1 could be moved into a details table and as captured below in figure 3.4.1, a one-or-zero to one relationship could link the vehicleInventoryDetails to the vehicle inventory. A concern which should be addressed through communication with the pre-owned sales dealership is that the column “STICKER” was renamed to “MSRP” which is a schema integration synonyms concern. It does need to be confirmed that the sticker price is actually the MSRP, but for the initial implementation, this does seem to be a valid assumption to make during the integration process.

While reviewing the vehicleInventoryDetails table, it can be noted that since the inventoryIngestId is an attribute of a vehicleInventoryDetail, which is a concept it is considered to be a weak entity rather than a relationship since it does have it’s own attribute but requests ids from other tables for its existence.

|  |
| --- |
|  |

Figure 3.4.1

|  |
| --- |
|  |

Figure 3.5 Migration of sale, saleTransaction and TransactionActivty with saleTransaction entity

As Figure 3.5 captures, the final effort required was to merge the sale information between the two schemas. This was a challenge due to the representational heterogeneity. As captured in the diagram above, different modeling choices were made between the two schemas.

## Deliverable: Final ER Diagram

|  |
| --- |
|  |

## Difficult Decisions

Based on the information provided by the pre-owned dealership, it was difficult to justify adjusting the table design to accommodate an environment that did not appear place a high value on data quality. Revising the logical model to accommodate the additional dataset would at minimum require the views that were created to be re-created to allow for existing reports for the car dealership introduced during assignment 1.

## Appendix

### 3.A Population of Tables for Assignment 3 for the Pre-Owned Car Sales Dealership are found below for reference

|  |  |  |
| --- | --- | --- |
| vehicleId | VIN | Sticker |
| 1 | 1BJ38LO45129JUT4I | 9000 |
| 2 | 25D9MEI2NMDLPDK85 | NULL |
| 3 | 6S58W2S3F6G8G4D1D | 7000 |
| 4 | 74EHF4F8YT56SMZA9 | NULL |
| 5 | 1E02D58GMZ5CP9D87 | 11000 |
| 6 | 81S2Q4JFMEWL54218 | NULL |
| 7 | 526DOEM78D9E124DL | 8500 |
| 8 | 5UD5LODK8W62DLKIEM | 9700 |
| 9 | 256DKEM74DOLF8521 | 12500 |
| 10 | 71DE6E55R2F3Q4A1Z | 11000 |

Figure 3.A.1 vehicle table

|  |  |
| --- | --- |
| associateId | associateName |
| 1 | Kylo Ren |
| 2 | Padme Amidala |
| 3 | Leia Organa |
| 4 | Anakin Skywalker |
| 5 | R2-D2 |

Figure 3.A.2 associate table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| customerId | customerName | phone | street | city |
| 1 | Baggins, Frodo | 202-555-0109 | 7405 Oak Meadow Road | Elk Grove Village |
| 2 | Gamgee, Samwise | 701-555-0109 | 9372 Stillwater Ave. | Champaign |
| 3 | Took, Peregrin | 202-555-0182 | 24 West Beechwood Drive | Urbana |
| 4 | Brandybuck, Meriadoc | 202-555-0147 | 8 Hall Lane | Savoy |
| 5 | Wormtongue, Grima | 701-555-0136 | 628 Center Rd. | Zionsville |
| 6 | Bolger, Fredegar | 202-555-0179 | 9827 Morris Ave. | Bloomington |
| 7 | Goatleaf, Harry | 701-555-0137 | 6 Blue Spring Court | Des Plaines |
| 8 | Willow, Old Man | 701-555-0192 | 7186 Wintergreen St. | Champaign |
| 9 | Angmar, Witch-King of | 701-555-0190 | 12 Rockaway Street | Urbana |
| 10 | Gandalf | 701-555-0172 | 7390 E. Glenridge Rd. | Rantoul |

Figure 3.A.3 customer table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| saleId | salesLog | saleDate | associateId | customerId | note |
| 1 | 10123456 | 2/4/2016 | 1 | 1 | NULL |
| 2 | 10123461 | 9/6/2016 | 1 | 6 | Discount applied: Autumn sales event |
| 3 | 10123464 | 9/27/2017 | 2 | 9 | Financing given |
| 4 | 10123457 | 8/1/2016 | 3 | 2 | NULL |
| 5 | 10123462 | 5/2/2017 | 4 | 7 | Discount applied: senior citizen |
| 6 | 10123463 | 6/3/2017 | 4 | 8 | NULL |
| 7 | 10123458 | 7/6/2017 | 4 | 3 | NULL |
| 8 | 10123459 | 6/5/2017 | 5 | 4 | Financing given |
| 9 | 10123460 | 1/5/2016 | 2 | 5 | NULL |
| 10 | 10123465 | 1/1/2016 | 3 | 10 | Discount applied: repeat customer |

Figure 3.A.4 sale table

|  |  |  |
| --- | --- | --- |
| transactionActivityId | vehicleId | price |
| 1 | 1 | -6200.00 |
| 2 | 5 | -1450.00 |
| 3 | 6 | -3500.00 |
| 4 | 8 | 9500.00 |
| 5 | 5 | 9995.00 |
| 6 | 9 | 11999.00 |
| 7 | 2 | -1205.00 |
| 8 | 3 | 6800.00 |
| 9 | 3 | -4200.00 |
| 10 | 1 | 8600.00 |
| 11 | 4 | -1025.00 |
| 12 | 7 | 8000.00 |
| 13 | 7 | -5500.00 |
| 14 | 10 | 10100.00 |

Figure 3.A.5 TransactionActivity table

|  |  |  |
| --- | --- | --- |
| salesTransactionsId | saleId | transactionActivityId |
| 1 | 1 | 1 |
| 2 | 2 | 2 |
| 3 | 3 | 3 |
| 4 | 4 | 4 |
| 5 | 5 | 5 |
| 6 | 6 | 6 |
| 7 | 7 | 7 |
| 8 | 7 | 8 |
| 9 | 8 | 9 |
| 10 | 8 | 10 |
| 11 | 9 | 11 |
| 12 | 9 | 12 |
| 13 | 10 | 13 |
| 14 | 10 | 14 |

Figure 3.A.6 saleTransactions table

# References

Earp, S. B. (2003). *Database Design Using Entity-Relationship Diagrams.* Auerbach Publications .