Instructions

1. [10 points] Create an ER diagram for Pre-owned dealer database, as described in the attached file.
2. [10 points] Create a separate ER diagram that reflects the schema you designed for Assignment 1. You may update the schema based on feedback you received from instructors after submitting Assignment 1.
3. Create 1 to 3 intermediate ER diagrams that showcase your integration process [15 points]. These diagrams should be accompanied by narrative prose (either in a separate document or as annotations directly to the diagram) that describe each of the integration steps taken on the diagram [20 points]. See the integration process described in the data integration slides for examples of what this might look like, and follow the example shown in the "Schema integration: an example" lecture. There is no one right way to do this, but your decisions should be justifiable, and should minimize the potential for information loss. Be sure to justify your design decisions in your narrative prose! Discussion of both various curatorial objectives and the pros and cons of various integration steps is necessary.
4. Finally, create one integrated ER diagram represented the merged schemas of the two dealerships (i.e., the final product of the integration process) [10 points]. Be sure to describe any final integration steps taken at this point (as described in step 3 above) [10 points]. Be sure to justify your design decisions in your narrative prose! Discussion of both various curatorial objectives and the pros and cons of the final design is necessary. Consider how you could have done things differently and in which areas the design can still be improved.
5. Submit your documents to Assignment 3 Peer Review for peer grading. Each student will be required to grade the submissions of 5 of their peers. Submissions will be graded based on the following criteria. Write a constructive and professional review and post to the course forum replying to the individual's submission.

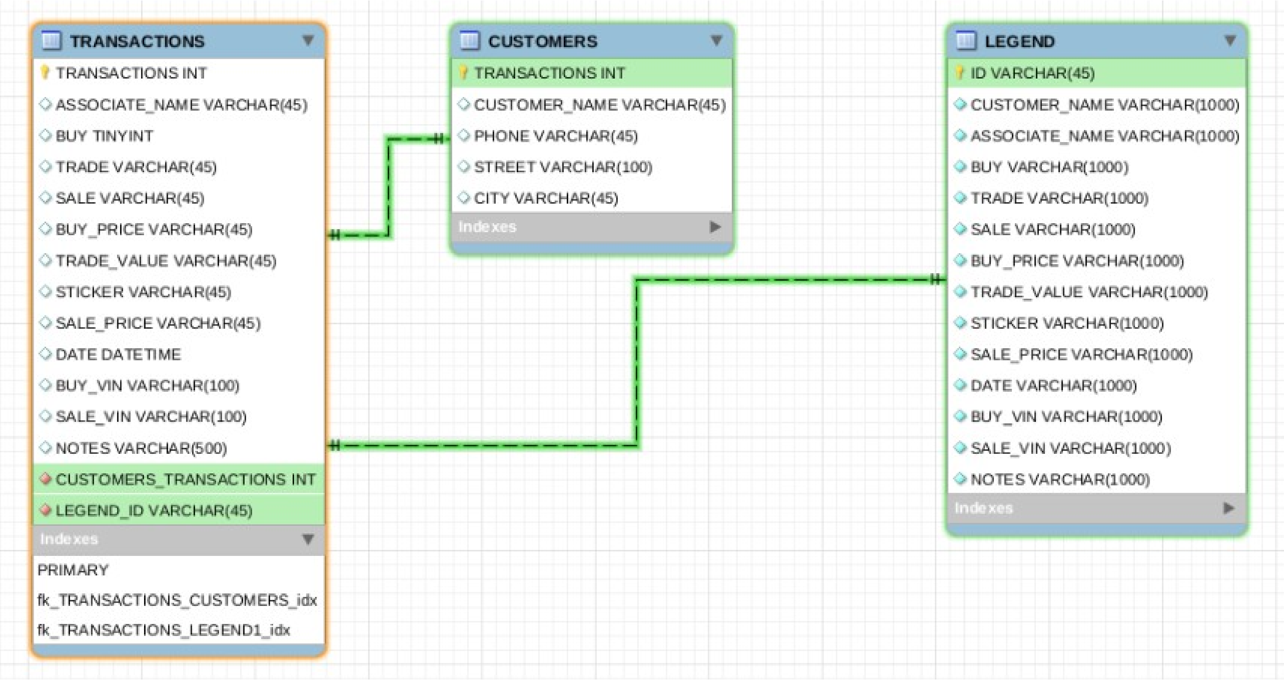
- Is everything represented?
- Is it clearly written? Are the ER diagrams and integration process presented?
- What are the pros and cons of the representation in the integrated ER diagram?
- What are the pros and cons of this integration process?
- How could it be done differently? How could it be improved?

6. Using your peer reviews, revise and submit all documents to Assignment 3 Submission for instructor grading.

If you have questions about the assignment, use the course Forum. This is a great place to ask questions and also help your fellow classmates.

Assignment 3: Ontologies / ER Diagram Design Exercise

1. **Pre-owned dealer database (MDS_Exercise3_FileA.xlsx) contains 3 tables. TRANSACTIONS, CUSTOMERS, LEGEND tables. Below is the ER diagram and narrative on the database entries.**



TRANSACTIONS
TRANSACTIONS INT
ASSOCIATE_NAME VARCHAR(45)
BUY TINYINT
TRADE VARCHAR(45)
SALE VARCHAR(45)
BUY_PRICE VARCHAR(45)
TRADE_VALUE VARCHAR(45)
STICKER VARCHAR(45)
SALE_PRICE VARCHAR(45)
DATE DATETIME
BUY_VIN VARCHAR(100)
SALE_VIN VARCHAR(100)
NOTES VARCHAR(500)
CUSTOMERS_TRANSACTIONS INT
LEGEND_ID VARCHAR(45)
Indexes
PRIMARY
fk_TRANSACTIONS_CUSTOMERS_idx
fk_TRANSACTIONS_LEGEND1_idx

CUSTOMERS
TRANSACTIONS INT
CUSTOMER_NAME VARCHAR(45)
PHONE VARCHAR(45)
STREET VARCHAR(100)
CITY VARCHAR(45)
Indexes

LEGEND
ID VARCHAR(45)
CUSTOMER_NAME VARCHAR(1000)
ASSOCIATE_NAME VARCHAR(1000)
BUY VARCHAR(1000)
TRADE VARCHAR(1000)
SALE VARCHAR(1000)
BUY_PRICE VARCHAR(1000)
TRADE_VALUE VARCHAR(1000)
STICKER VARCHAR(1000)
SALE_PRICE VARCHAR(1000)
DATE VARCHAR(1000)
BUY_VIN VARCHAR(1000)
SALE_VIN VARCHAR(1000)
NOTES VARCHAR(1000)
Indexes

Write a Narrative on MDS_Ex3_FileA contents and tables
MDS_Ex3_FileA contains three tables. Below is the explanation of three table. For this assignment we used SQL Workbench as a software tool for generating ER diagrams.

Transactions

- This table seems to have any transaction information of a customer be it be BUY, SELL or a TRADE.
- Entries in this table seems to be uniquely identified with a TRANSACTION Number. We can safely assume that this could be used as a Primary Key in our SQL Schema generation process.
- There are two NOTES files in this table, we can remove the duplicate column entry and use one NOTES entry with the information from both the columns.
- Another important column to notice is TRADE (BUY AND SALE) entry. If this value is set to TRUE, we can safely assume that this transaction involves both BUY and SELL of a customer.
- We seem to have VIN values for both BUY and SELL. We can safely assume that BUY_VIN and SELL_VIN entries are valid if it is a TRADE. Otherwise we can use either use BUY_VIN or SELL_VIN depending on the transaction type i.e BUY or SELL.
- Informate related to “repeat cusotmer” from NOTES column needs to pulled in to Customers Tables from Assignment1.

CUSTOMERS

- This table seems to contain information on customers who has some Transactions with the dealer.

- This table contains address, phone number information.
- With respect to primary key, looks like we cannot use any existing column as primary key for our SQL Schema generation process.
- It is safe to introduce a new column entry such as CUSTOMER_ID to uniquely identify a customer.
- It is very much possible that a single customer could have done multiple transactions. Hence it is a good idea to maintain CUSTOMER_ID from above as primary key for the table.

LEGEND

- This table contains information describing each column in the table.
- I think it is a good idea to have a reference of this table in all the SQL SCHEMA tables to gain more information on what kind of fields we are using across all the tables.
- Hence we intend to create a table with all the columns description and pass it as a foreign key (reference) to all the tables during the SQL Schema generation process.

For reference below is a snapshot of MDS_Exercise3_FileA.xls file.

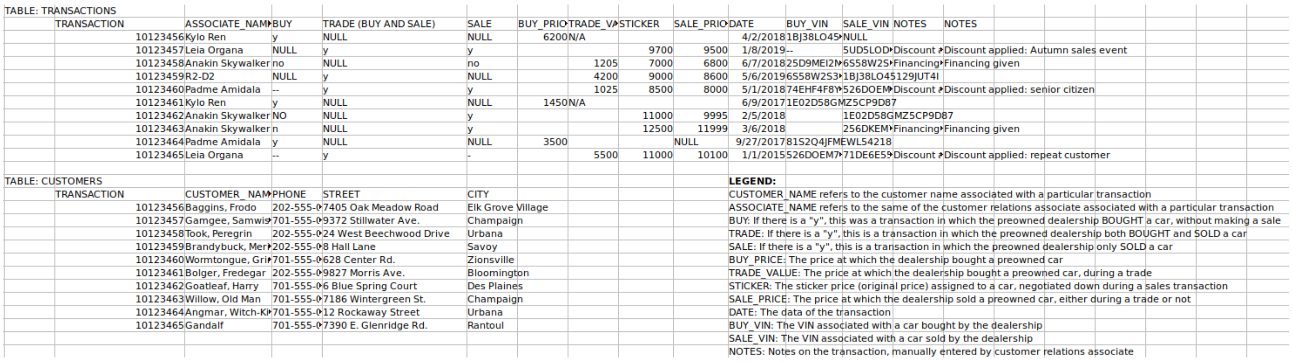


TABLE: TRANSACTIONS

TRANSACTION	ASSOCIATE_NAM	BUY	TRADE (BUY AND SALE)	SALE	BUY_PRIO	TRADE_VA	STICKER	SALE_PRIO	DATE	BUY_VIN	SALE_VIN	NOTES	NOTES
10123456	Kylo Ren	y	NULL	NULL	6200	N/A			4/2/2018	1BJ38LO45	NULL		
10123457	Leia Organa	NULL	y	y			9700	9500	1/8/2019	--	5UD5LOD	Discount a	Discount applied: Autumn sales event
10123458	Anakin Skywalker	no	NULL	no		1205	7000	6800	6/7/2018	25D9MEI2N	6S58W2S	Financing	Financing given
10123459	R2-D2	NULL	y	NULL		4200	9000	8600	5/6/2019	6S58W2S3	1BJ38LO45129JUT4I		
10123460	Padme Amidala	--	y	y		1025	8500	8000	5/1/2018	74EHF4F8Y	526DOEM	Discount a	Discount applied: senior citizen
10123461	Kylo Ren	y	NULL	NULL	1450	N/A			6/9/2017	1E02D58GMZ5CP9D87			
10123462	Anakin Skywalker	NO	NULL	y			11000	9995	2/5/2018		1E02D58GMZ5CP9D87		
10123463	Anakin Skywalker	n	NULL	y			12500	11999	3/6/2018		256DKEM	Financing	Financing given
10123464	Padme Amidala	y	NULL	NULL	3500			NULL	9/27/2017	81S2Q4JFMEWL54218			
10123465	Leia Organa	--	y	-		5500	11000	10100	1/1/2015	526DOEM7	71DE6E5	Discount a	Discount applied: repeat customer

TABLE: CUSTOMERS

TRANSACTION	CUSTOMER_ NAM	PHONE	STREET	CITY
10123456	Baggins, Frodo	202-555-0	7405 Oak Meadow Road	Elk Grove Village
10123457	Gamgee, Samwis	701-555-0	9372 Stillwater Ave.	Champaign
10123458	Took, Peregrin	202-555-0	24 West Beechwood Drive	Urbana
10123459	Brandybuck, Mer	202-555-0	8 Hall Lane	Savoy
10123460	Wormtongue, Gri	701-555-0	628 Center Rd.	Zionsville
10123461	Bolger, Fredegar	202-555-0	9827 Morris Ave.	Bloomington
10123462	Goatleaf, Harry	701-555-0	6 Blue Spring Court	Des Plaines
10123463	Willow, Old Man	701-555-0	7186 Wintergreen St.	Champaign
10123464	Angmar, Witch-Ki	701-555-0	12 Rockaway Street	Urbana
10123465	Gandalf	701-555-0	7390 E. Glenridge Rd.	Rantoul

LEGEND:

CUSTOMER_NAME refers to the customer name associated with a particular transaction

ASSOCIATE_NAME refers to the same of the customer relations associate associated with a particular transaction

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

BUY_PRICE: The price at which the dealership bought a preowned car

TRADE_VALUE: The price at which the dealership bought a preowned car, during a trade

STICKER: The sticker price (original price) assigned to a car, negotiated down during a sales transaction

SALE_PRICE: The price at which the dealership sold a preowned car, either during a trade or not

DATE: The data of the transaction

BUY_VIN: The VIN associated with a car bought by the dealership

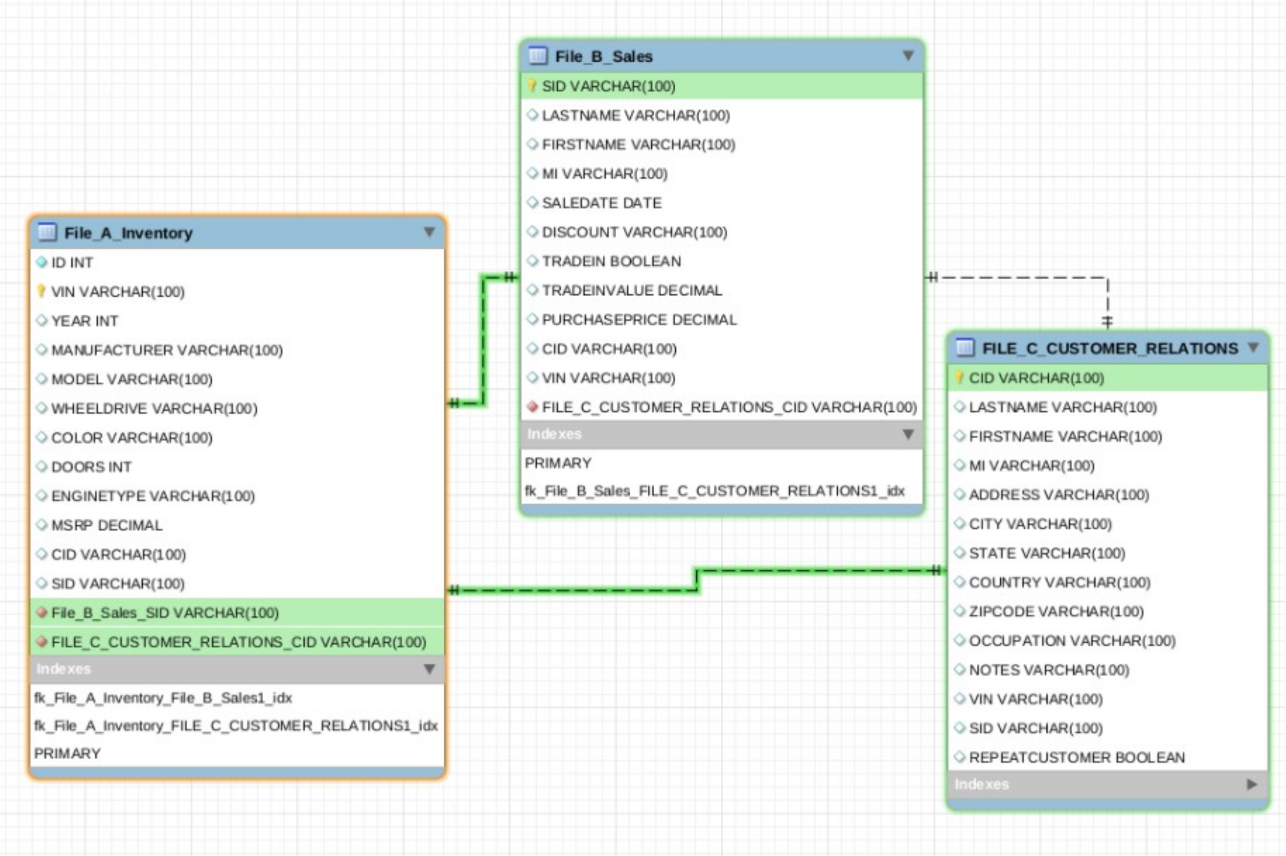
SALE_VIN: The VIN associated with a car sold by the dealership

NOTES: Notes on the transaction, manually entered by customer relations associate

In the following sections we explore on the integration process between different tables for this assignment purpose.

2. ER diagram for Assignment1 (FileA, FileB, FileC)

In Assignment1, we created there tables namely Inventory, Sales and Customer_Relations. Below is the ER diagrams explaining the relationship between the tables and a narrative of each table.



File_A_Inventory
ID INT
VIN VARCHAR(100)
YEAR INT
MANUFACTURER VARCHAR(100)
MODEL VARCHAR(100)
WHEELDRIVE VARCHAR(100)
COLOR VARCHAR(100)
DOORS INT
ENGINETYPE VARCHAR(100)
MSRP DECIMAL
CID VARCHAR(100)
SID VARCHAR(100)
File_B_Sales_SID VARCHAR(100)
FILE_C_CUSTOMER_RELATIONS_CID VARCHAR(100)
Indexes
fk_File_A_Inventory_File_B_Sales1_idx
fk_File_A_Inventory_FILE_C_CUSTOMER_RELATIONS1_idx
PRIMARY

File_B_Sales
SID VARCHAR(100)
LASTNAME VARCHAR(100)
FIRSTNAME VARCHAR(100)
MI VARCHAR(100)
SALEDATE DATE
DISCOUNT VARCHAR(100)
TRADEIN BOOLEAN
TRADEINVALUE DECIMAL
PURCHASEPRICE DECIMAL
CID VARCHAR(100)
VIN VARCHAR(100)
FILE_C_CUSTOMER_RELATIONS_CID VARCHAR(100)
Indexes
PRIMARY
fk_File_B_Sales_FILE_C_CUSTOMER_RELATIONS1_idx

FILE_C_CUSTOMER_RELATIONS
CID VARCHAR(100)
LASTNAME VARCHAR(100)
FIRSTNAME VARCHAR(100)
MI VARCHAR(100)
ADDRESS VARCHAR(100)
CITY VARCHAR(100)
STATE VARCHAR(100)
COUNTRY VARCHAR(100)
ZIPCODE VARCHAR(100)
OCCUPATION VARCHAR(100)
NOTES VARCHAR(100)
VIN VARCHAR(100)
SID VARCHAR(100)
REPEATCUSTOMER BOOLEAN
Indexes

Write a Narrative on File A, File B and File C.

Inventory

- This table has information about the purchased car.
- VIN is used as Unique Key/ID to identify a record.
- Year, Manufacturer, Model, WheelDrive, Color, Doors, EngineType, MSRP are identified as the attributes related to a car. Hence they are group in this table.
- We added foreign keys CID(Customer ID), SID(Sales_ID) to refer the customer and Sales information.

Sales

- This table indentifies a person who needs to buy/bought a car.
- The inventory details need to brought from File A if required, based on the created VIN number.
- We have some person details related to the person and some sales related information on the purchased item.
- VIN Is identified as foreign key that can be used to retrieve car information from File A table.
- We haveSID(Sales_ID) as a Unique identifier for a entry in the table. This is used as Primary Key for the table. Same key can be used as a Foreign key to File C and File A to retrieve customer related information.

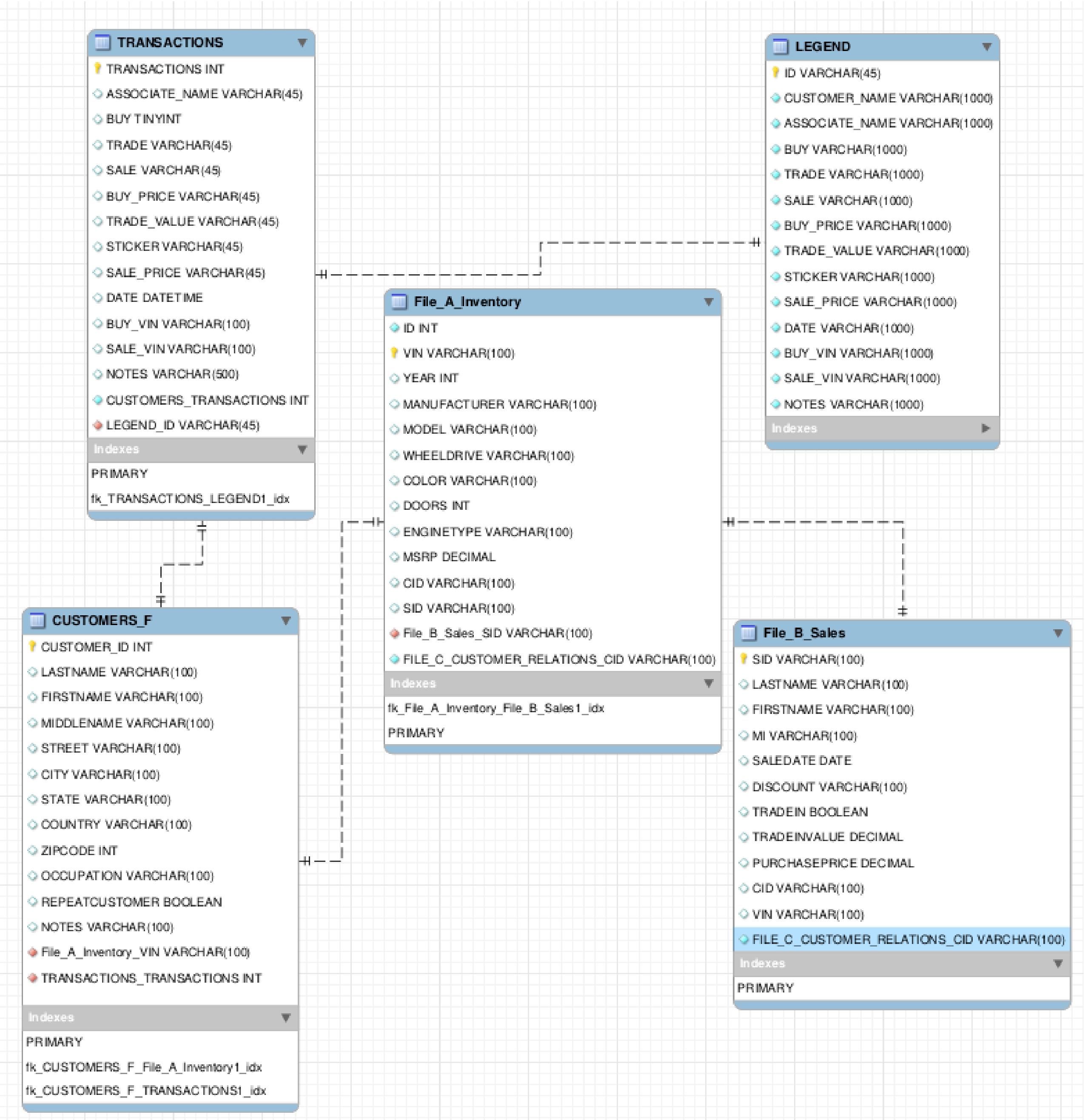
Customer Relations

- This table has information related to the person, and the kind of customer he is to the company.

In the following steps will detail on the merging/integration of these two databases into one database as per the assignment tasks.

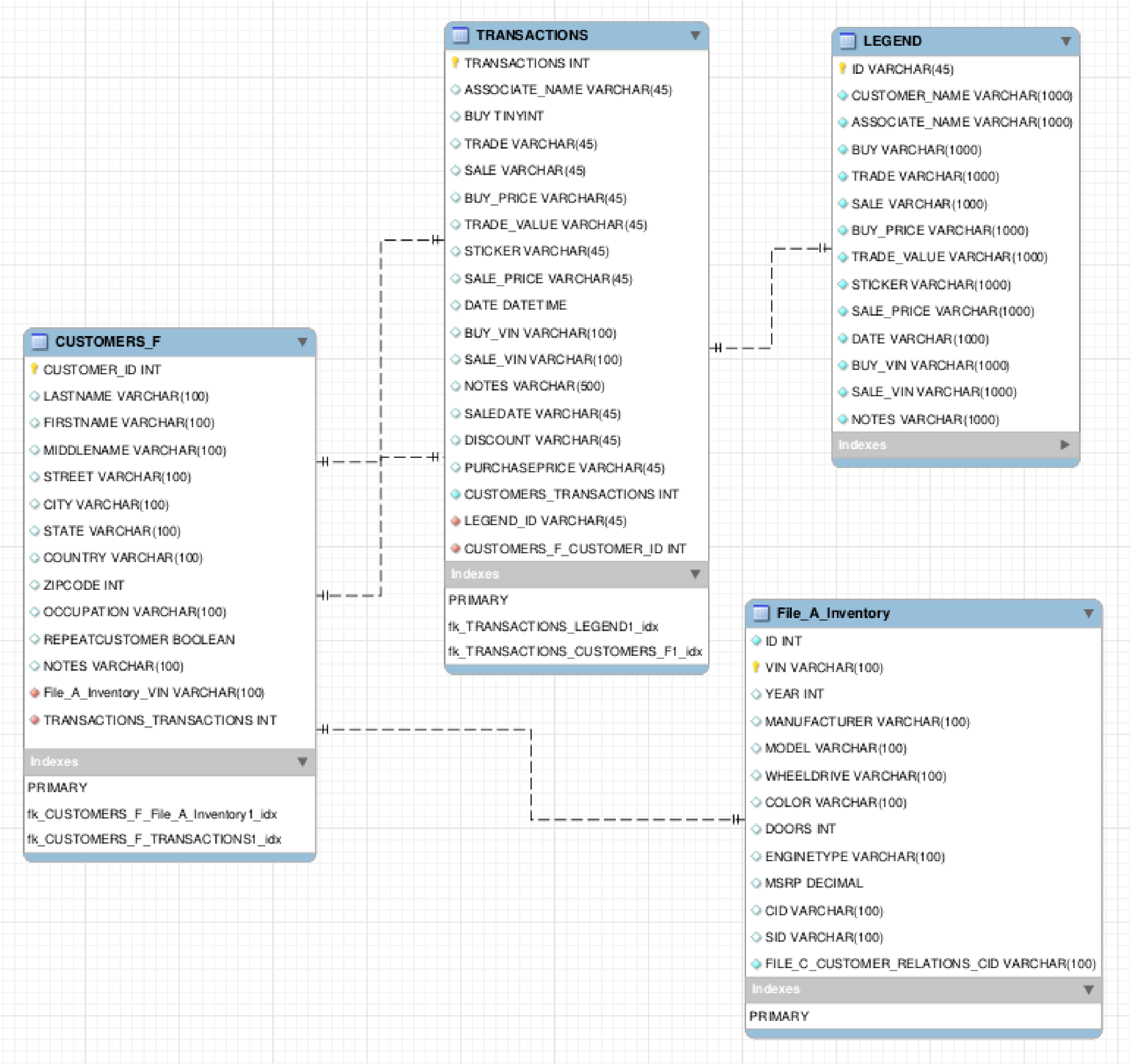
3. Merging ER1 and ER2 is performed in 3 steps. In the first step we create a integrated CUSTOMERS table. In Step two we integrate Sales and Transactions table. In the final step we integrate Inventory, Legend table with all the other table and generate final integrated ER diagram.

I. **Integrate CUSTOMER_RELATIONS table from Assignment1 and CUSTOMERS table from MDS_Exercise3_FileA.**



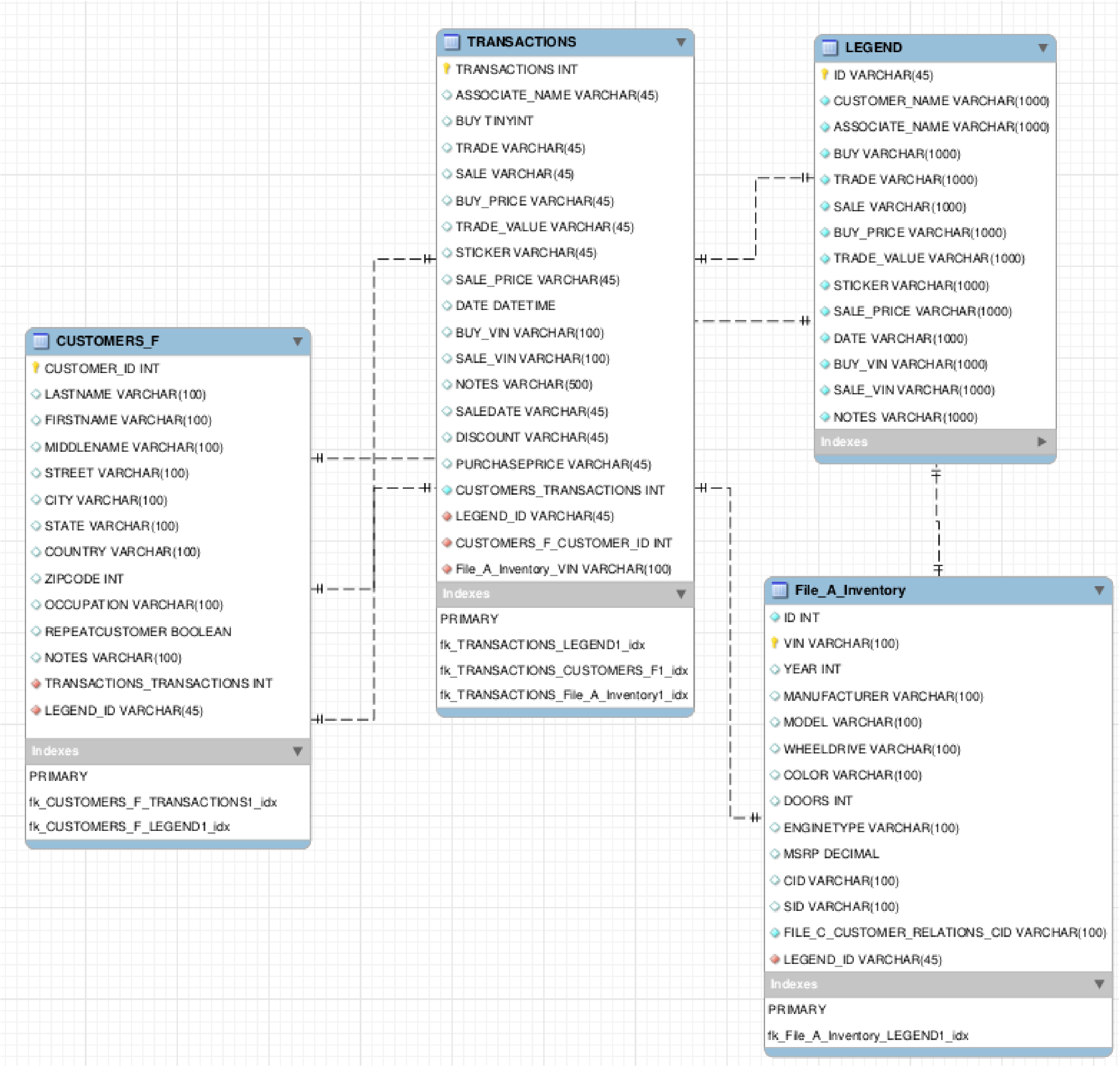
- In this step, we need to consider creating a single table CUSTOMERS_F that contains all the information in regard with the transactions between the customer and the dealers. We need to create a CUSTOMER_ID to uniquely identify the customer from the table. Hence CUSTOMER_ID is used a the primary key for this table.
- We listed LASTNAME, FIRSTNAME, MIDDLENAME, STREET, CITY, STATE, COUNTRY, ZIPCODE, OCCUPATION, REPEATED CUSOTMER and NOTES as the attributes for this table.
- We believe that this information can give us all the information we need from the customers.
- In order to access other tables from CUSTOMER_F table we need to create enough foreign keys to access other tables and vice versa.

II. Integrate SALES table from Assignment with TRANSACTIONS table from MDS_Exercise3_FileA. Main considerations for these tables integration is the relationship establishment between TRANSACTIONS table and CUSTOMER_F table.



- In this step, we need to integrate SALES table from Assignment 1 to TRANSACTIONS table from Assignment 3.
- Major points to consider during the integration process is to have unique key to identify a entry in this table. Hence we choose that to be TRANSACTION_ID. At the same time this table should be accessed by CUSTOMER_F table hence CUSTOMER_F table uses TRANSACTIONS_ID as foreign key.
- In-case we need to go back-wards, It is a good idea to used CUSOTMER_ID as foreign key in this table so that we can back-trace the customer-info of the transaction.
- It is worth to note that we still maintain the THREE attributes BUY, SELL and TRADE as different entries and have corresponding BUY_VIN and SELL_VIN as reference to Inventory table.

III. Integrate Inventory and Legend table with every other table, in other words Final Integrated ER diagram.



In this step, we focus more on the relationship establish between information flow from the customer to various other tables.

- We have taken extra care not to duplicate data, minimize data read/write transaction in multiple tables, safely and uniquely identify primary keys in various tables and step toward the final ER diagram.
- We intend to add reference of LEGEND table to all the other tables so that we get a general idea on what even table items means and at any every point of time. **I am not sure if this is a good idea or not.**
- We created Inventory_ID as a primary key to uniquely access attributes data from the table. Naturally this ID is used a foreign key in other tables that needs to access this table.
- Note that there is not direct access from CUSTOMER_F Table to this table. This is to achieve data abstraction.

4. **Narrative, Steps and Decisions taken during the Integration of ER Diagrams and cleanup activities and Final ER diagram**

Below are the steps taken to generate the final ER diagram.

Step 1:

- Create CUSTOMER_ID to uniquely identify attributes from the CUSTOMERS_F table. This is going to be sole information that is required by a customer to get his information related to any transactions he made with the dealer.
- We group all the attributes related to address of customer in to one generic group of attributes such as STREET, CITY, COUNTRY, ZIPCODE and phone number.
- REPEATED_CUSTOMER information is added in this table as a quality of service required can be determined as early as possible during the data retrieval process.
- NOTES attribute is added in this table for any extra information in regard with Customer or a transaction the customer made.

Step 2:

- TRADE_VALUE in Transactions table from MDS_Exercise3_FileA and TRADE_IN_VALUE from Sales table from Assignment1 are duplicated. We can discard one.
- TRADE_VALUE in Transactions table from MDS_Exercise3_FileA and PURCHASE_PRICE from Sales table from Assignment1 are duplicated. We can discard one.
- LASTNAME, FIRSTNAME,MI from Sales table from Assignment1 can be replaced with CUSTOMER_ID from customers table.
- TRADE in Transactions table from MDS_Excerise3_FileA and TRADE_IN from Sales table from Assignment1 are duplicated. We can discard one.

Step 3:

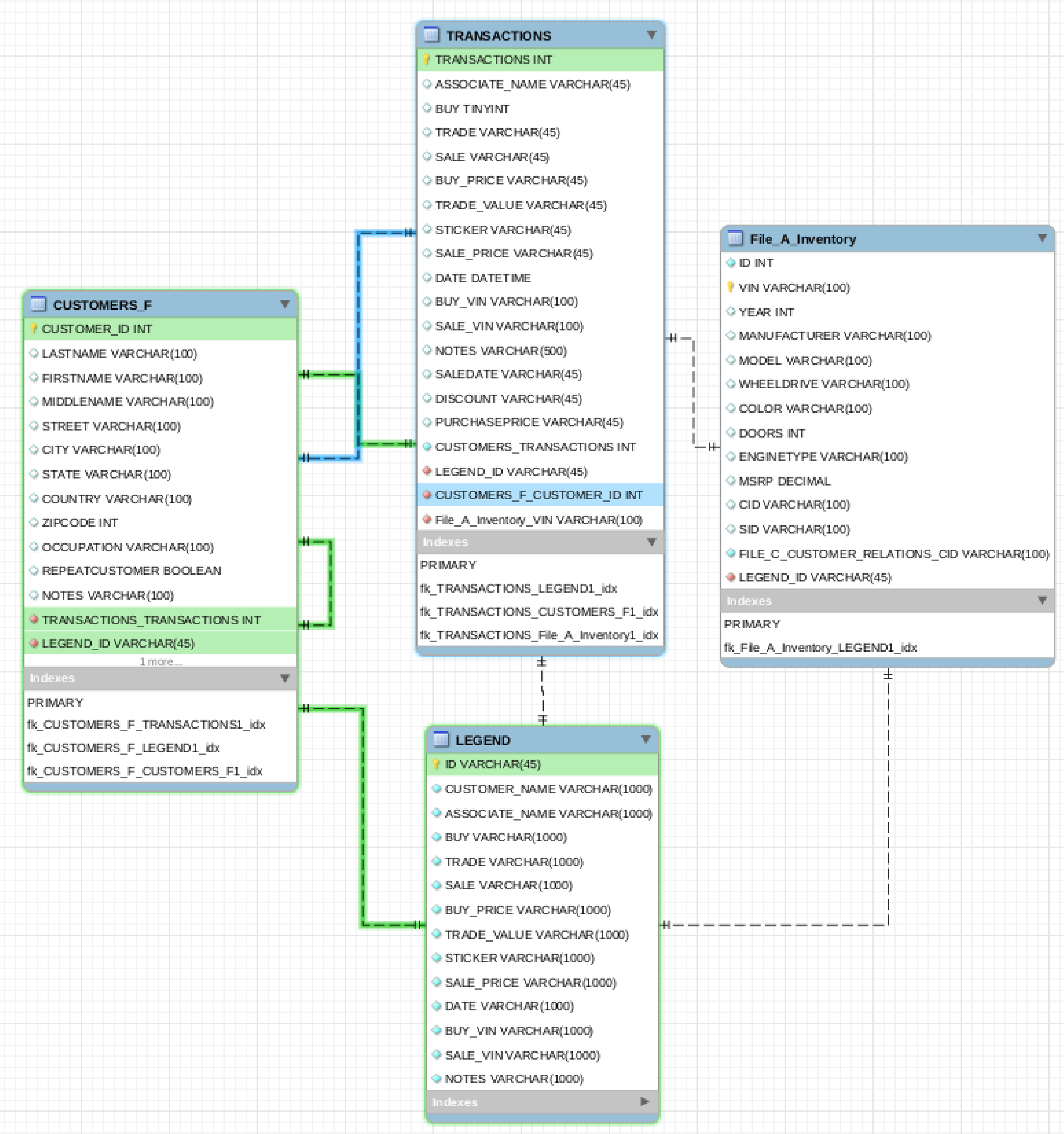
- Inventory table should be accessible only from Transaction table based on the SALE_VIN or BUY_VIN attribute.
- Disconnect connection between CUSTOMERS_F table and Inventory table.
- Inventory Table doesn’t need Customer ID entry. Transaction ID is a better way to arrive at Inventory Information.
-

All together:

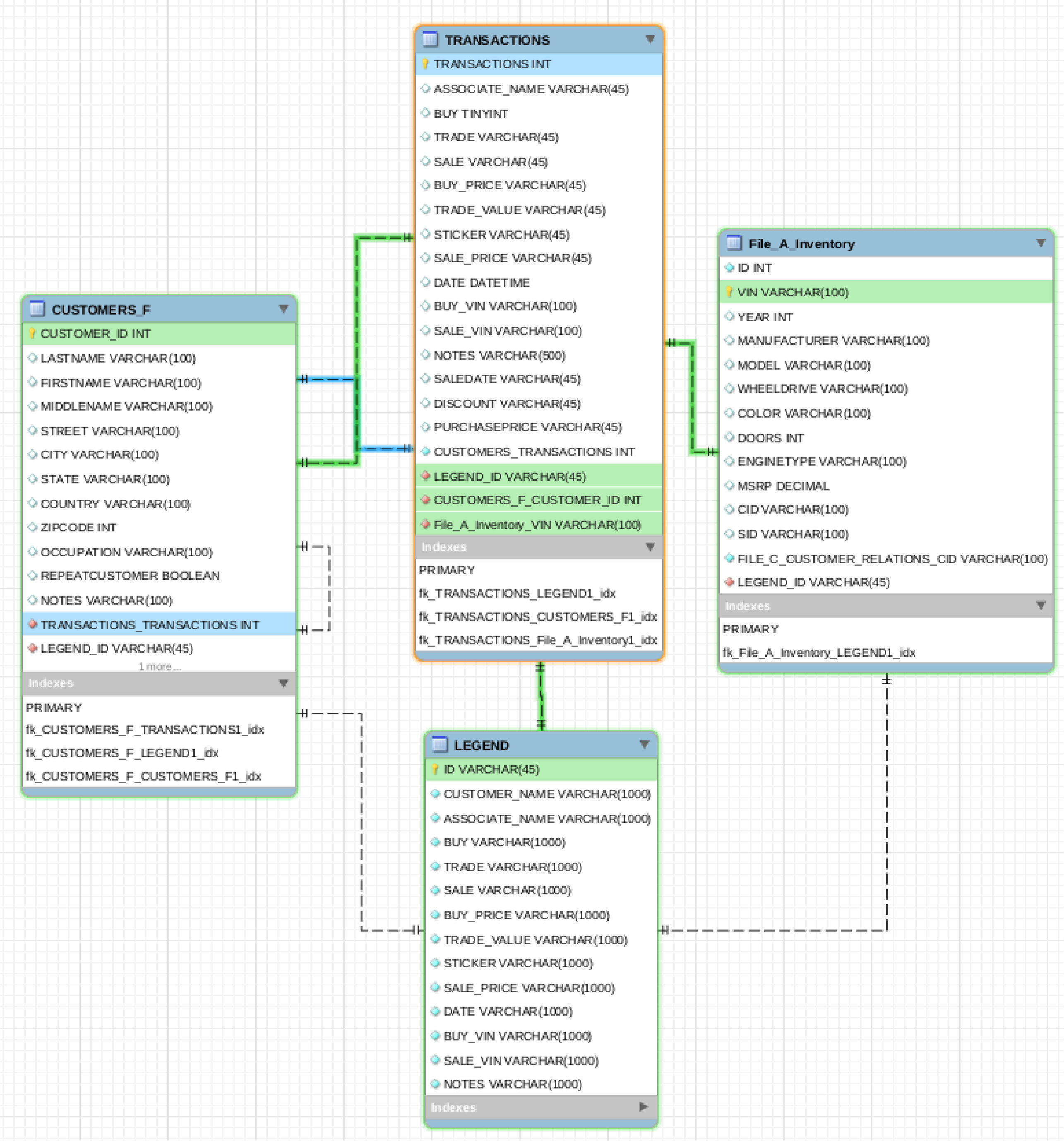
- Customer ID from Customers Table is the primary entry point from Customer, Sales representative point of view.
- Customer Tables is linked with Transactions table via TRANSACTION_ID and to Legend Table via LEGEND_ID.
- Transactions table will have information on the kind of sale i.e BUY, SALE or TRADE based on the respective attributes from Transactions table. We have a corresponding SALE_VIN and BUY_VIN unique entries for BUY and SALE transactions.
- SALE_VIN and BUY_VIN are the entry points to the INVENTORY Table.
- LEGEND_ID is linked to INVENTORY Table via LEGEND_ID for information reference.
- We removed redundant NOTES from Transactions table.
- During the Clean up process, we have taken extra to maintain no-duplicity and data abstraction.

After performing above steps we arrive at the final ER diagram as shown below three snapshots clearly explain on how the data retrieval is performed and how each tables is inter-connected. We have three snapshots explaining three possible paths of data retrieval per table with different colors.

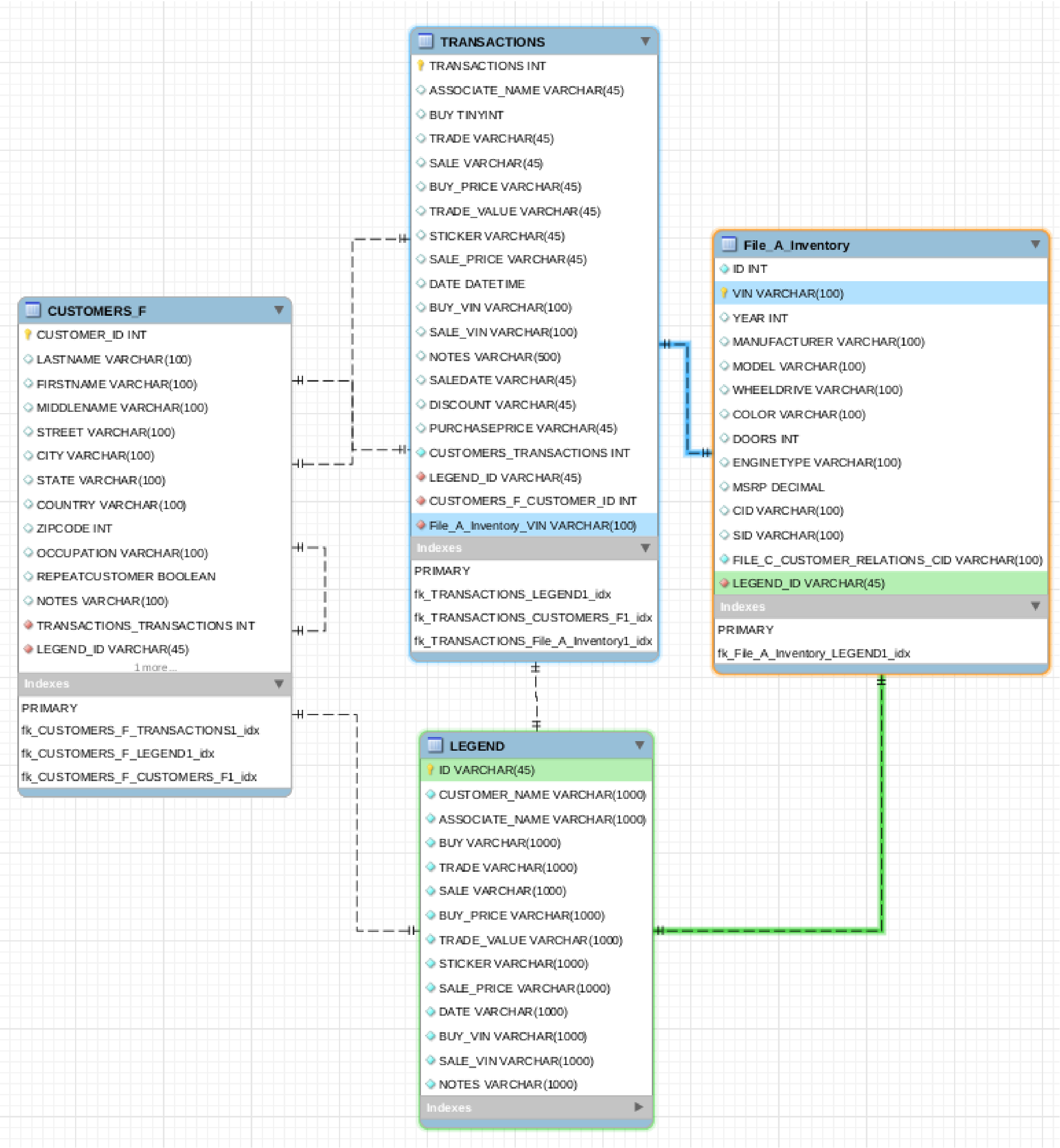
Case 1: CUSTOMER_F table interface with other tables.



Case 2: TRANSACTIONS Table interface with other tables



Case 3: INVENTORY Table interface with other tables



5. Justify design decisions, various curatorial objectives and the pros and cons for final design
Most of the decision making reasons and curatorial objectives have been in the integration process. Below is a one-liner or summary of various decision taken in the process.

- **Design and Integration process decision**

  I. No multiple paths to retrieve a particular attribute
  We have taken extra so that no multiple paths exists to retrieve a particular information from different tables. This helps in maintaining data consistency and helps establish a clean way/process of extracting information from the data sets.

  II. No data redundancy
  We made sure that no data is duplicated or replicated in multiples tables. In doing so we reduced the changes of data redundancy and data consistency in the process of retrieving data from the data set.

  III. Data abstraction
  Attributes in the entities and relationships are integrated in way to maintain data abstraction. For instance a customer without a TRANSACTION_ID cannot get information from the INVENTORY table. Although he might get information from LEGEND with I deem to be harmless.

  IV. Clearly defined attributes
  All the attributes and entities are clearing defined and describe. For this purpose we are using LEGEND table. **I am not sure if this is a good idea or not.**

  V. DataTypes and Size of VARCHAR
  DataTypes of variable attributes in the entities are arguable and purely implementation or requirements depending. As a general rule we used VARCHAR(45) or VARCHAR(100) or VARCHAR(1000) for different attributes depending on the usage. This can be adjusted as needed.

- **Pros, Cons and Improvement areas**

  I. ***Pros***
    - SQL Schema looks simple and easy to understand. Please find SQL schema definitions in the SQL Script folder in submissions.
    - Information retrieval is easy and consistent. We have taken extra care to remove redundancy.
    - Access a specific information leads to minimal queries because of the structure of the tables and SQL Schema design.

  II. ***Cons***
    - Some might say, this database design is too simple and doesn’t cater to wide range of requirement. And I agree with that.
    - We could split up BUY, SALE, TRADE fields from TRANSACTIONs table and create another table with this information. This helps in creating flexibility in handling more information related to the type of transactions.
    - We could a new table to add NOTES or extra information for transaction tables. This could help us in adding more information to the transaction details and helps in understanding more specifics of a transaction at a later stage (in future).

  III. ***Improvement Areas***
    We could have done a better job in almost every aspect of the design process to cater to wide range of requirement. I believe that this database isn’t commercial ready.

- ***Data Curation Objectives and Curatorial activities***

I think we addressed Data curation objectives efficiently and effectively so far in this assignment. Below are the areas of the curatorial activities that we have addressed so far.
Activities like Collection (support and acquisition of data), Organization (employed appropriate data model and use appropriate standards), Storage(support reliable and effective storage in the form of tabular-hierarchical structure), Preservation(data is understandable and reusable), Discoverability (ability to search for and located relevant data), Sharing(SQL schema and data can be shared and easily understood), Modification (support management of corrections and updates) have been addressed.
Activities like Access(ability to retireve and distribute data), Workflow(ability to systematize work with data), , Identification (ability to identify, authenticate and validate data), , Integration(integration of data from different sources using different models), , Reformatting (reformatting for use by different tools or to match new format standards), Reproducible (reproduce results ensuring scientific validity and reliability), Security, Compliance, Provenance (support identifying what inputs, calculations and actions are responsible for data values), Communication (support representation, publishing and visualization that could provide insight) have not been addressed in this assignment.

Note:

- **Please find SQL script for the tables in SQL folder.**
- **Please find ER diagrams benchmark diagrams in ER folder.**