Database Integration & ER Diagram Design

SHRASHTI SINGHAL
Data Curation- IS 531
Department of Information Science
The University of Illinois at Urbana Champaign

Abstract

The setting is an auto dealer. In this company, there are three departments, including Inventory, Sales and Customer relations. We have a database schema in place for this auto dealer, for these three departments in a shared database. These three databases are easily maintained and searchable by foreign and primary keys. We have an ER Diagram/Conceptual Schema for it as well.

The task is to integrate this database with a new database of a dealer of pre-owned cars. We have to design a merged database, as well as a complete ER Diagram for the integrated database.

The solution is to design and populate an integrated database such that that is effective and efficient for all departments of both the dealers, and that will also support combining data from different departments.

I. Auto Dealer Database

The below are the ER Diagram & Database Schema for the Auto Dealer Database, which is already in place.

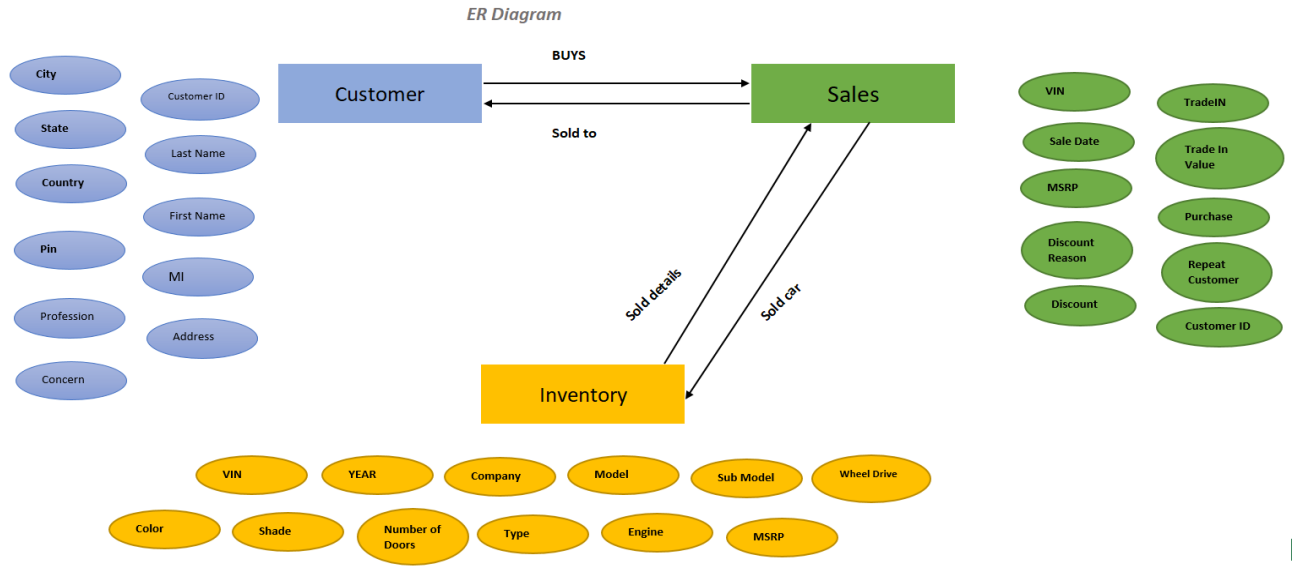
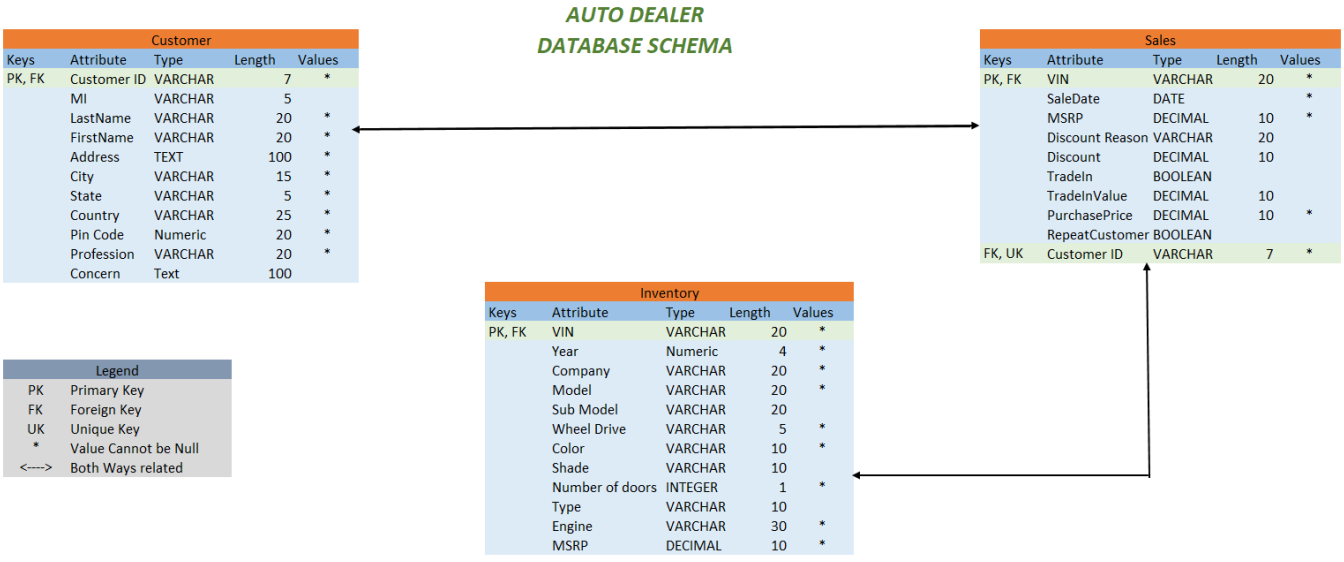


Image 1: ER Diagram for AutoDealer



Customer				
Keys	Attribute	Type	Length	Values
PK, FK	Customer ID	VARCHAR	7	*
	MI	VARCHAR	5	
	LastName	VARCHAR	20	*
	FirstName	VARCHAR	20	*
	Address	TEXT	100	*
	City	VARCHAR	15	*
	State	VARCHAR	5	*
	Country	VARCHAR	25	*
	Pin Code	Numeric	20	*
	Profession	VARCHAR	20	*
	Concern	Text	100	

Sales				
Keys	Attribute	Type	Length	Values
PK, FK	VIN	VARCHAR	20	*
	SaleDate	DATE		*
	MSRP	DECIMAL	10	*
	Discount Reason	VARCHAR	20	
	Discount	DECIMAL	10	
	TradeIn	BOOLEAN		
	TradeInValue	DECIMAL	10	
	PurchasePrice	DECIMAL	10	*
	RepeatCustomer	BOOLEAN		
FK, UK	Customer ID	VARCHAR	7	*

Inventory				
Keys	Attribute	Type	Length	Values
PK, FK	VIN	VARCHAR	20	*
	Year	Numeric	4	*
	Company	VARCHAR	20	*
	Model	VARCHAR	20	*
	Sub Model	VARCHAR	20	
	Wheel Drive	VARCHAR	5	*
	Color	VARCHAR	10	*
	Shade	VARCHAR	10	
	Number of doors	INTEGER	1	*
	Type	VARCHAR	10	
	Engine	VARCHAR	30	*
	MSRP	DECIMAL	10	*

Legend	
PK	Primary Key
FK	Foreign Key
UK	Unique Key
*	Value Cannot be Null
<---->	Both Ways related

Image 2: Tabular Relational Schema

II. Pre-Owned Car Dealer Database

The pre-owned Dealer Database is in form of an excel sheet. This excel sheet contains 3 tables. The one table is Transactions, Customers and the third table is a legend.

Below is the database of Pre-owned cars as shown:

TABLE: TRANSACTIONS

TRANSACTION	ASSOCIATE_NAME	BUY	TRADE (BUY AND SALE)	SALE	BUY_PRICE	TRADE_VALUE	STICKER	SALE_PRICE	DATE	BUY_VIN	SALE_VIN	NOTES	NOTES
10123456	Kylo Ren	y	NULL	NULL	6200	N/A			04-02-2016	1BJ38LO45129JUT4I	NULL		
10123457	Leia Organa	NULL	y	y			9700	9500	01-08-2016	--	5UD5LODK8W62DLKIEM	Discount applied: Autumn sales event	Discount applied: Autumn sales even
10123458	Anakin Skywalker	no	NULL	no		1205	7000	6800	06-07-2017	25D9MEI2NMDLPDK85	6S58W2S3F6G8G4D1D	Financing given	Financing given
10123459	R2-D2	NULL	y	NULL		4200	9000	8600	05-06-2017	6S58W2S3F6G8G4D1D	1BJ38LO45129JUT4I		
10123460	Padme Amidala	--	y	y		1025	8500	8000	05-01-2016	74EHF4F8YT56SMZA9	526DOEM78D9E124DL	Discount applied: senior citizen	Discount applied: senior citizen
10123461	Kylo Ren	y	NULL	NULL	1450	N/A			06-09-2016	1E02D58GMZ5CP9D87			
10123462	Anakin Skywalker	NO	NULL	y			11000	9995	02-05-2017		1E02D58GMZ5CP9D87		
10123463	Anakin Skywalker	n	NULL	y			12500	11999	03-06-2017		256DKEM74DOLF8521	Financing given	Financing given
10123464	Padme Amidala	y	NULL	NULL	3500			NULL	27-09-2017	81S2Q4JFMEWL54218			
10123465	Leia Organa	--	y	-		5500	11000	10100	01-01-2016	526DOEM78D9E124DL	71DE6E55R2F3Q4A1Z	Discount applied: repeat customer	Discount applied: repeat customer

Table 1: Transactions

TABLE: CUSTOMERS

TRANSACTION	CUSTOMER_ NAME	PHONE	STREET	CITY
10123456	Baggins, Frodo	202-555-0109	7405 Oak Meadow Road	Elk Grove Village
10123457	Gamgee, Samwise	701-555-0109	9372 Stillwater Ave.	Champaign
10123458	Took, Peregrin	202-555-0182	24 West Beechwood Drive	Urbana
10123459	Brandybuck, Meriadoc	202-555-0147	8 Hall Lane	Savoy
10123460	Wormtongue, Grima	701-555-0136	628 Center Rd.	Zionsville
10123461	Bolger, Fredegar	202-555-0179	9827 Morris Ave.	Bloomington
10123462	Goatleaf, Harry	701-555-0137	6 Blue Spring Court	Des Plaines
10123463	Willow, Old Man	701-555-0192	7186 Wintergreen St.	Champaign
10123464	Angmar, Witch-King of	701-555-0190	12 Rockaway Street	Urbana
10123465	Gandalf	701-555-0172	7390 E. Glenridge Rd.	Rantoul

Table 2: Customers

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

Table 3: Legend

III. ER Diagrams

We will follow a consistent conceptual Schema for Integration purpose. Let it be Crows Foot. My preference is Crow's foot because it is more like a tabular form. It doesn't involve different types of shapes. Denoting modality and cardinality is also convenient.

Here, we convert our existing Conceptual Schema is Crows Foot notation.

Let's draw an ER Diagram for Pre-Owned Cars Dealer. We will keep our notation consistent. Below is the conceptual schema for Pre-Owned car dealer.

ER diagrams will be orange for AutoDealer, Green for Pre-Owned Car Dealer and Pink for Integrated Database, throughout this paper.

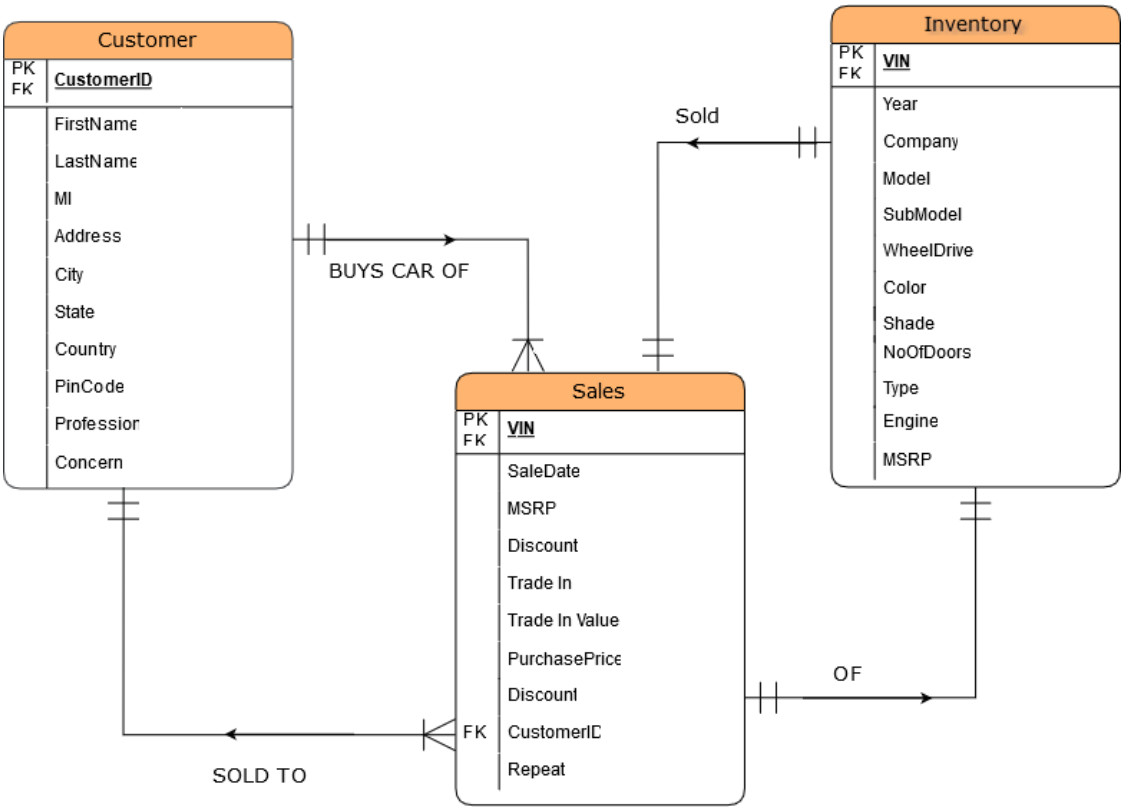


Image 3: ER Diagram-AutoDealer, Crow's Foot notation

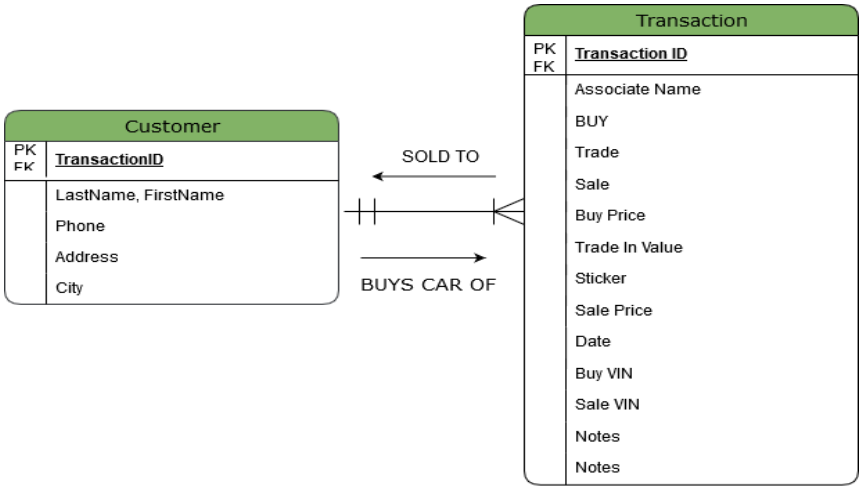


Image 4: ER Diagram-Pre-Owned Car Dealer, Crow's Foot notation

IV. Analysis

Analysis of AutoDealer Database has been done previously. The database is cleaned, removing all redundant values, attributes maintain consistency and accuracy among all the relations.

Analysis of Pre-Owned Car Dealer is required, as it is not properly maintained, contained redundant values and all the relations are not properly connected to each other in database.

IV. I File A (Transactions): It is an excel table. The excel table has 14 Columns. The columns are namely,

TRANSACTION

ASSOCIATE_NAME
BUY
TRADE (BUY AND SALE)
SALE
BUY_PRICE
TRADE_VALUE
STICKER
SALE_PRICE
DATE
BUY_VIN
SALE_VIN
NOTES
NOTES

- The transaction Id is unique to this table.
- Associate Name is the name of the employee/associate at this pre owned car dealer who helps customer to purchase the cars.
- There are three parameters which are mutually exclusive, Trade, Buy and Sell. These are Boolean Values, which determine the entries in other columns.
- Sticker is the MRP of the selling cars
- Buy VIN, SALE VIN are 2 different VINs.
- There are 2 different Notes Columns

Problem Areas:

- When the transaction type is YES, the values of BUY and SELL are inconsistent. At few places BUY is NO and SELL is YES, while at others both BUY and SELL are NO
- Row 3, BUY and SELL are no, while TRADE is NULL. We cannot infer any kind of transaction for this row
- In BUY, SELL and TRADE null values definitions are inconsistent. It is defined by NULL, NO or a dash for no value at different places.
- There are 2 Columns for NOTES, which looks redundant as at multiple entries both of them contains the same value.
- When there is a TRADE, Trade amount is written but no amount is given specially for the BUY VIN.
- When there is a discount mentioned in the NOTES, there is no detail of DISCOUNT AMOUNT.

Data Processing:

- To link this table to Customer table, we need to add customer ID as foreign key to Transaction Table.

- Associate Name, attribute, It is unclear from the format is what is the first and last name. But we assume the format is LastName, FirstName. We will need to maintain a separate table for Associates details. Therefore to link transaction Table with Associates table, we will add associates ID here.

- There are three attributes, BUY, SELL and TRADE. These are mutually exclusive to each other. Only one of the three can happen at a time. Instead of maintaining 3 attributes, let maintain one attribute. Attribute name will be Transaction Type. The values of this attribute should be one of the three, which is BUY, SELL or TRADE.

- In notes, there are 3 things given. The Reason of discount, Repeat Customer or if financing has been given or not. Let's remove the 2 notes attributes and replace them with 1. Discount Reason 2. Financing Given 3. Repeat Customer. Financing Given & Repeat Customer attributes can hold Boolean values, Yes, or NO.

- When we have given a discount to a customer & we know the reason of the discount. We should be able to capture it in our records. A discount Value attribute should be a must for transactions file. We will create one such attribute.

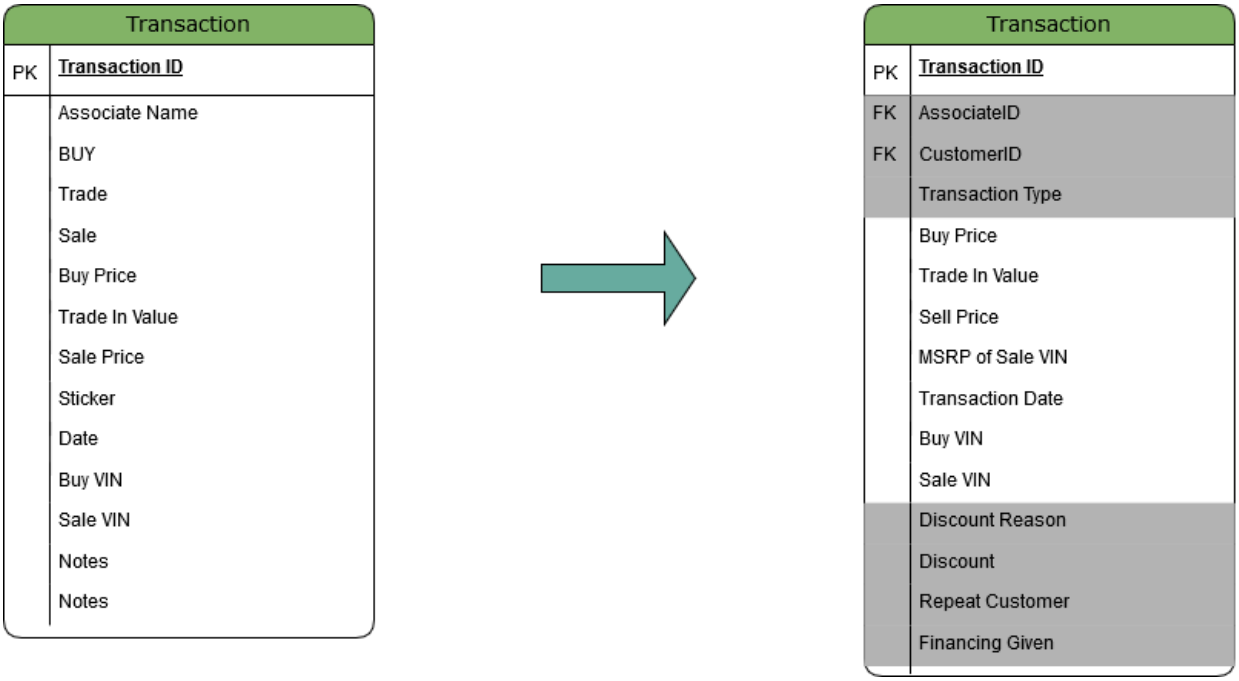


Image 5: Improvising Transaction Diagram-Pre-Owned Car Dealer

IV.II File B (Customers): This is an excel table. It contains 5 Columns.

TRANSACTION
CUSTOMER_NAME
PHONE
STREET
CITY

Transaction ID is unique in this table

Problem Areas:

- No state, Country and PIN is detail is given
- Customer first name and last name are entered in one column, separated by a comma.
- Other detail of customer not given, example profession or their financing details.

Data Processing:

- There is no unique identifier for relation Customer. We would need to generate Customer ID to access and extract required data. Generating Customer ID will also help in linking Customer table with other tables, maintain CustomerID as foreign key.
- FirstName & LastName of Customer are assumed to be in LastName, FirstName format. This format, firstly makes the values unusable because of compound values separated by comma. Moreover, we are not sure if which one is first and which is the last name. Therefore, we will make 2 columns, one for FirstName & other for LastName.
- In the address details, we notice that State, country and pin Code are missing. These details are necessary as, same address can be there in different States. Even the city names tend to be same sometimes in different States. Therefore, we will add these 3 columns.

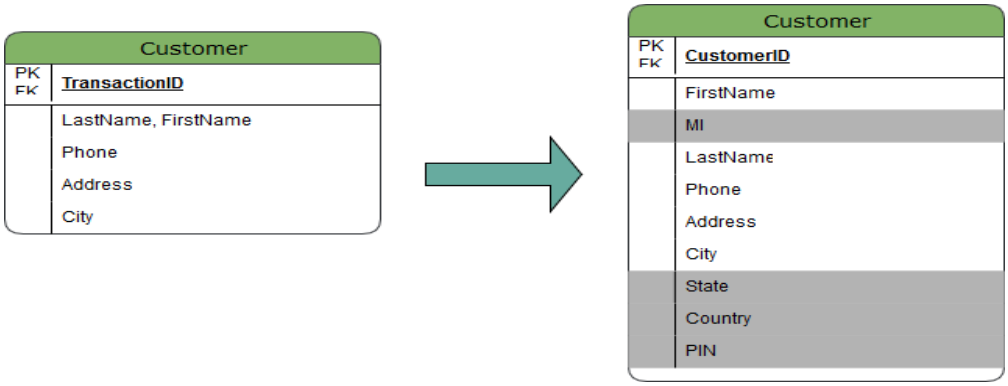


Image 6: Improvising Customers Diagram-Pre-Owned Car Dealer

IV.III Associates: As we see in transaction details, name of the associates/employees of pre-owned car dealer is given. We would need a proper relation to maintain details of the employees. Therefore, in Transaction schema, we have replaced the name of the associates, with their IDs. To maintain a record of the associates, we would need its relation. We would add all the necessary details needed for the associate records. We would leave more room to add more details about them. For now, improvising the Associate table, adding more attributes, other than just the names to this relation.

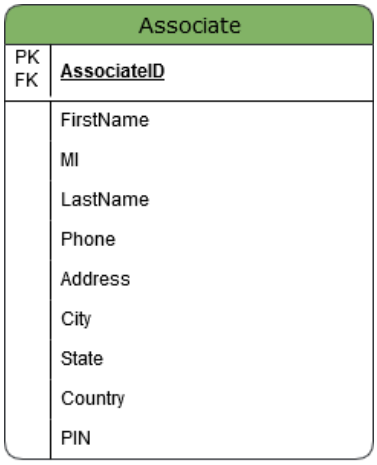


Image 7: Associate Diagram-Pre-Owned Car Dealer

V. Pre-Owned Car Dealer Schema

After analysis and implementing data curation activities on the Pre-owned Car Dealer Database, we have developed a new and better Conceptual Schema. Combining the ERs of each relations of the database, to form ER Diagram in Crow's Foot format.

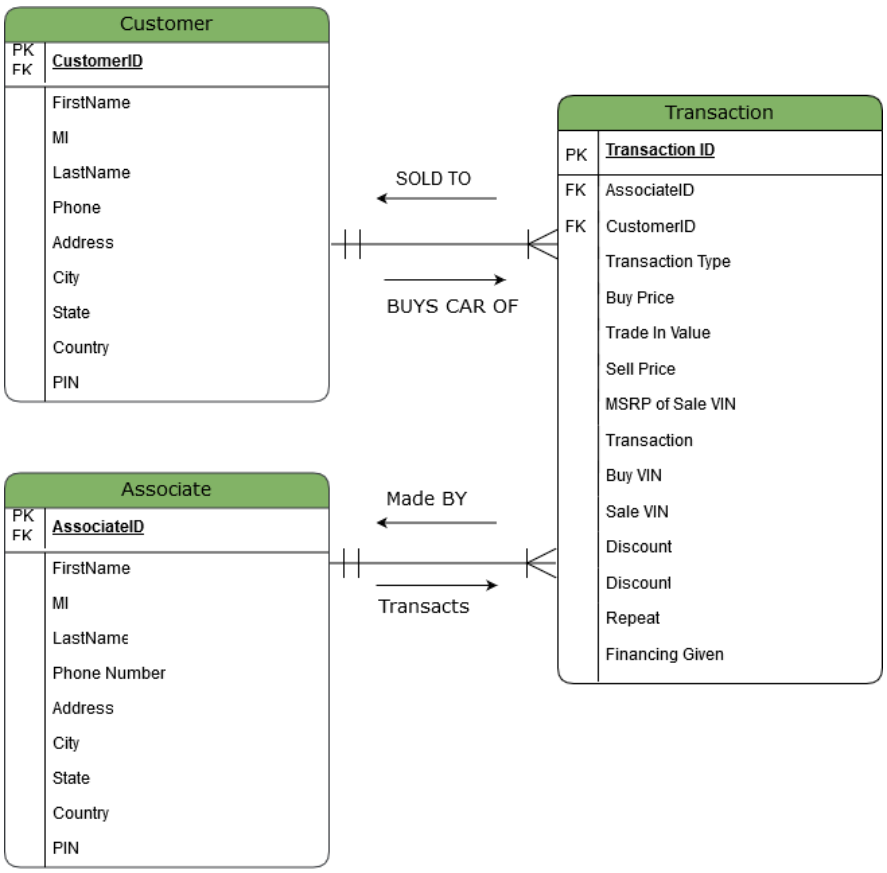


Image 8: ER Diagram-Modified Pre-Owned Car Dealer, Crow's notation

VI. INTEGRATION

We will approach both the database together, retain the same type of items in the merged database and make decisions about retaining of the individual columns/values from databases in the merged database.

After looking at both the Databases, we have arrived at few important observations:

1. Sales table (AUTO DEALER) is the subset of Transactions Table (PRE OWNED CAR Dealer).
2. CUSTOMER table in both the Database are comparably same.
3. INVENTORY Table is missing from PRE-OWNED CAR Dealer database, but present in AUTO DEALER Database.
4. Associates Table is present in PRE-OWNED CAR Dealer database, but missing in AUTO DEALER Database.

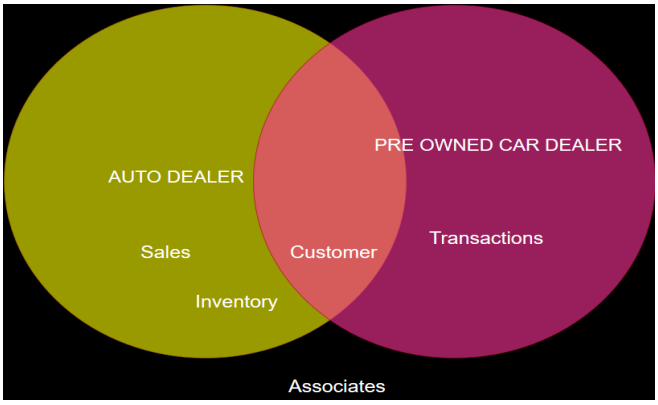


Image 9: Venn diagram of Databases before Data Curation

We can conclude from our observation that, we would need to

1. Merge transactions and sales table.
2. Merge Customer data from both the databases
3. Retain Inventory table from AUTO DEALER Database, and maintain data of PRE OWNED CAR Dealer in Inventory table of all the BUY VINs and SALE VINs

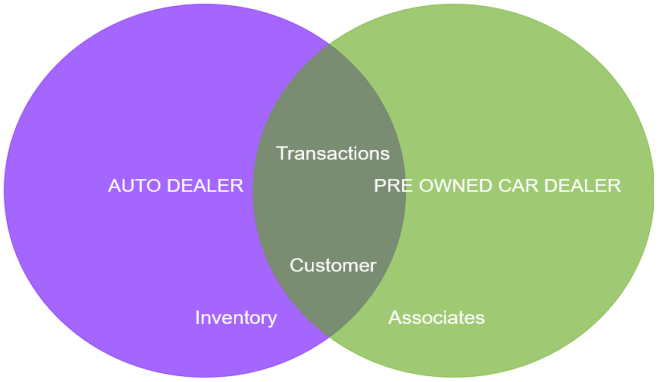


Image 10: Venn diagram of Databases after Data Curation

VII. INTEGRATION STEPS

For integration purposes we will use Chens like notation, but without attributes.
Once final schema is ready we will get back to Crow foot notation.
Attributes are not displayed to make the diagrams contact. Moreover, all finalised attributes in all the tables have been displayed previously and coming diagrams.

STEP 1:
Draw the ER Diagram for both the databases. We drew in Chens Format, without the attributes.

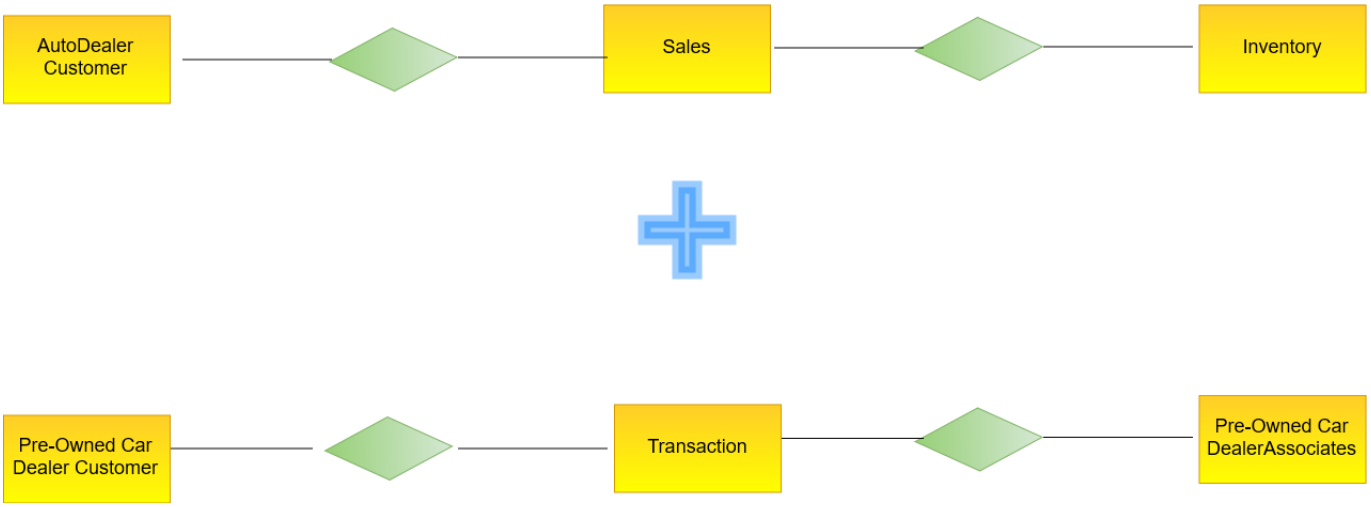


Image 11: Combine Schemas, Chens Notation without attributes

STEP 2: Merge the 2 databases. We merged the transaction and sales table, as both the tables share a close relationship.

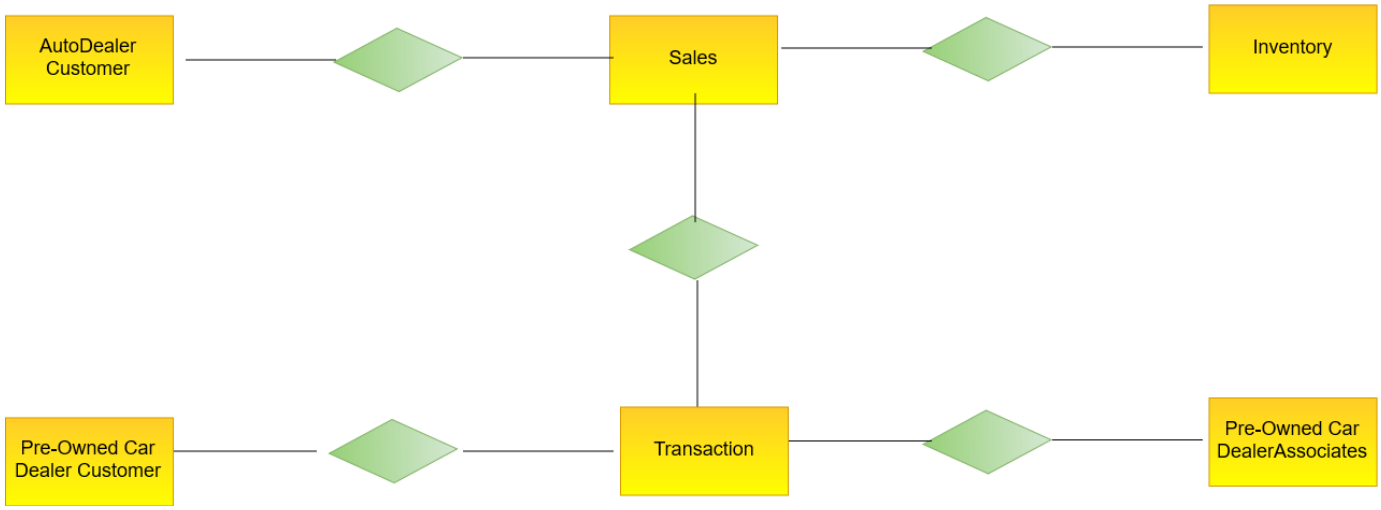


Image 12: Combine Schemas, Chens Notation without attributes

STEP 3: Sales table is a subset of Transaction Table. Sales Table contains all the attributes of Sales Table, and few extra attributes. Therefore, we will create a subset, and display the same in the ER Diagram.

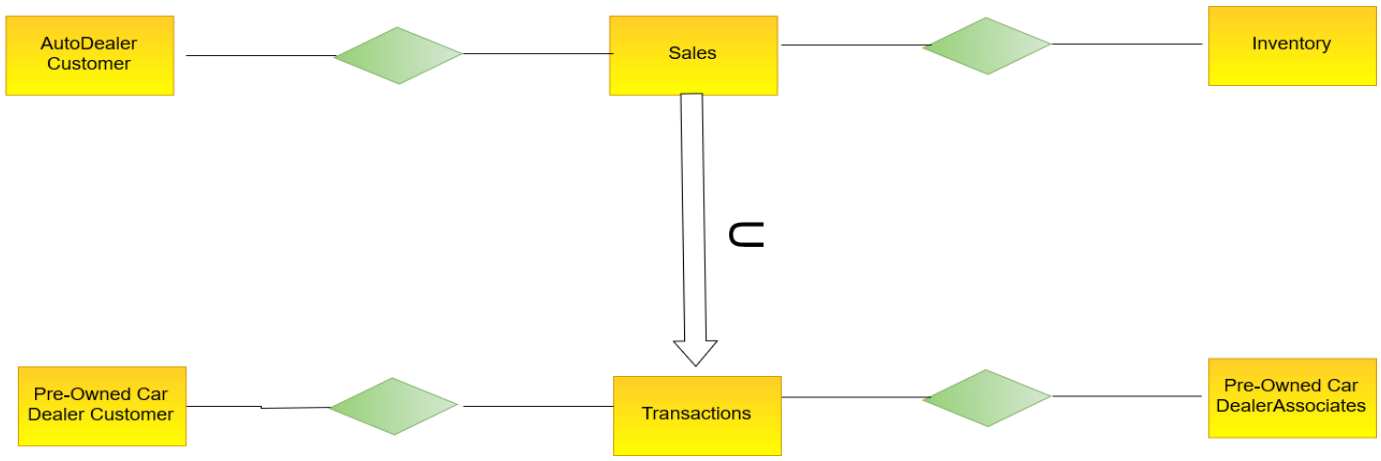


Image 13: Create Subset

STEP 4: We will merge the two Customer Tables. Here we will merge the same attributes of both the Customer's table into one single attribute of the merged table. The unique attributes in each Customer Table will be added separately. First we will show the attribute list in Crows format .

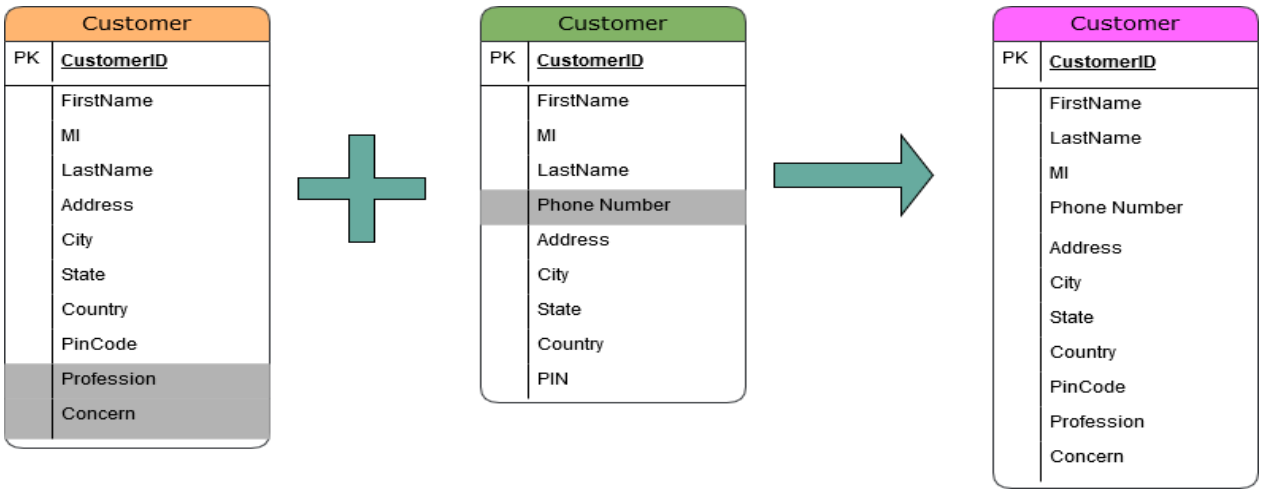


Image 14: Integration of Customer Relation

STEP 5: In ER Diagram, we remove the relationship of Customer Table and Sales Table, as we have already figured out that Sales table is the subset of transaction Table.

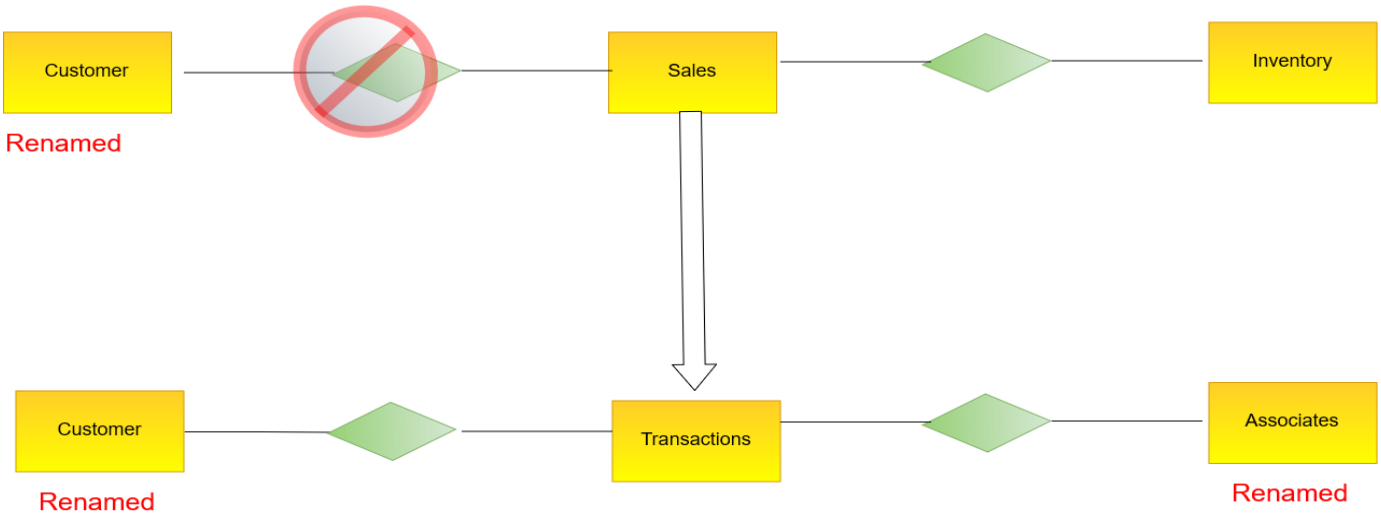


Image 15: Remove relationship

STEP 6: Customer table which was valid for the Sales Table should also be valid for the transaction Table.

We use the Chen format for integration. Customer Table is related with the transaction table.

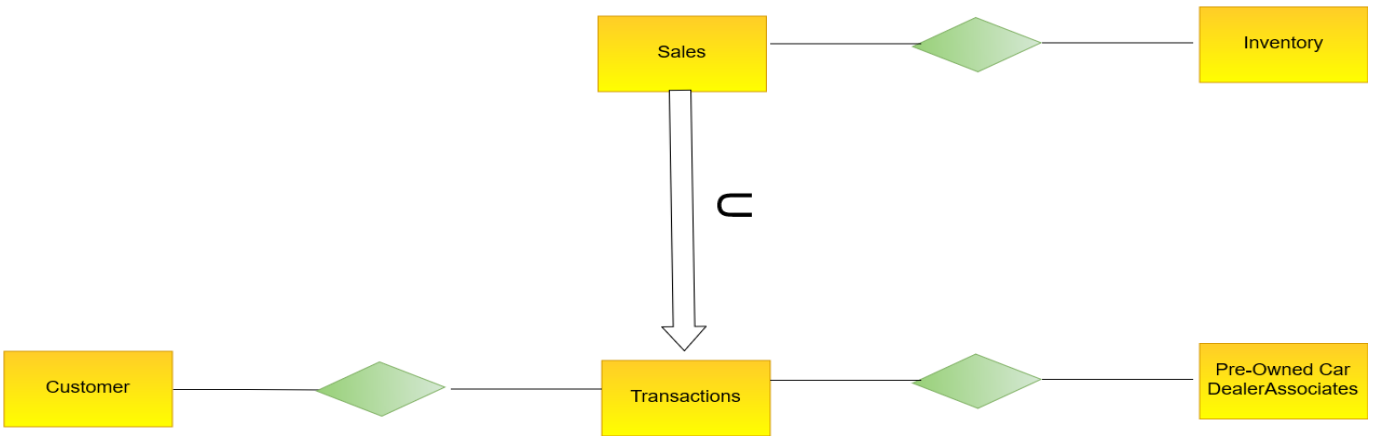


Image 16: Merge Customer Relations in one

STEP 7: Customer table which was valid for the Sales Table should also be valid for the transaction Table.

We use the Chen format for integration. Customer Table is related with the transaction table.

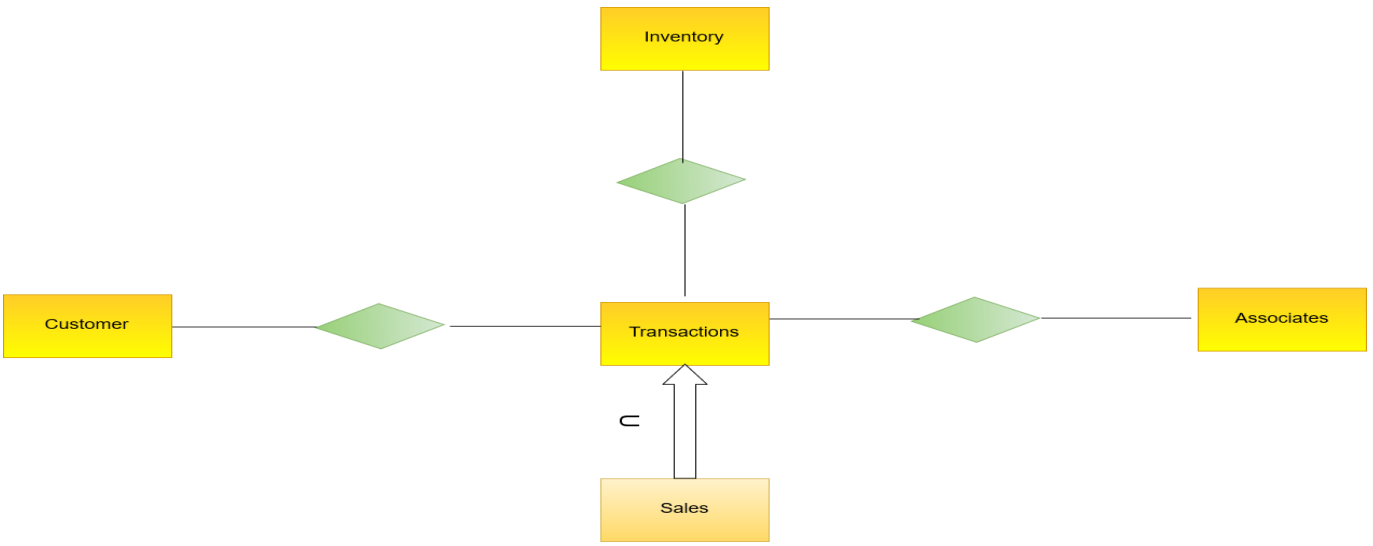


Image 17: Integrated Schema, Chens notation

STEP 8: The previous Schemas works well, in loosing no information. We will go one step ahead and make the final schema after removing the complete Sales table from Auto Dealer database and just keep the transaction table from Pre-owned car dealer.

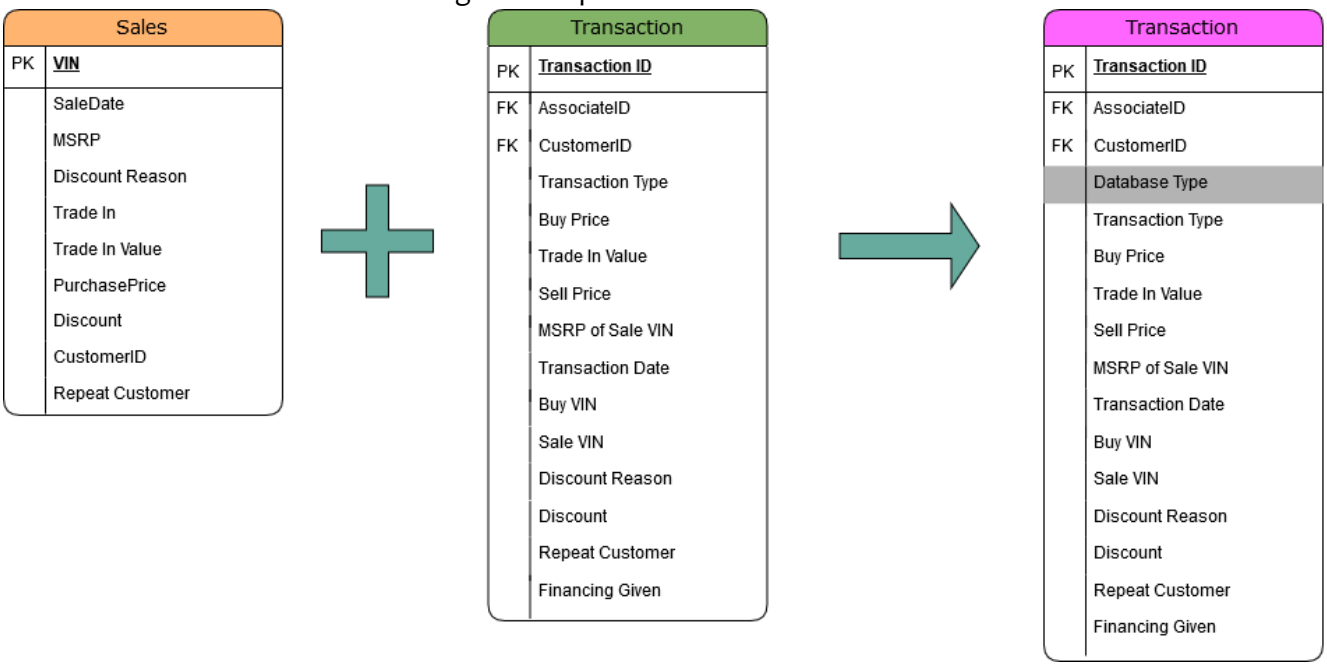


Image 18: Integration of Transaction Relation

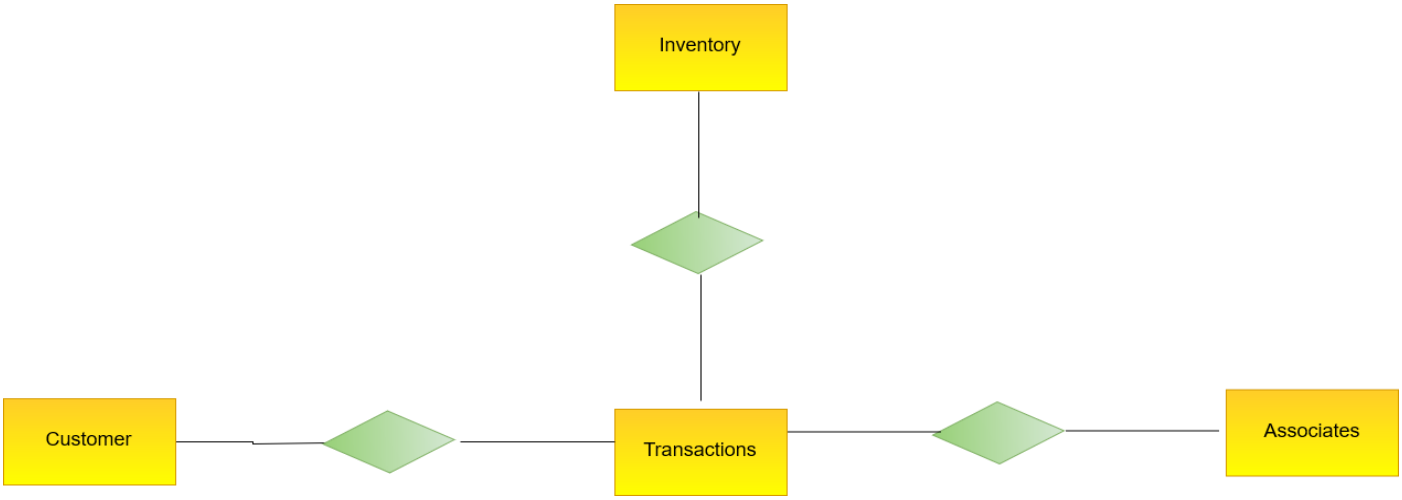


Image 19: Integrated Schema, Chens notation

VIII. Integrated ER Diagram

We will draw the final Schema in Crows Foot, which will depict all the attributes of all the relations, which were missing in our previous incomplete Chen's notation. Here, we convert our existing ER Diagram into Crow Foot notation.

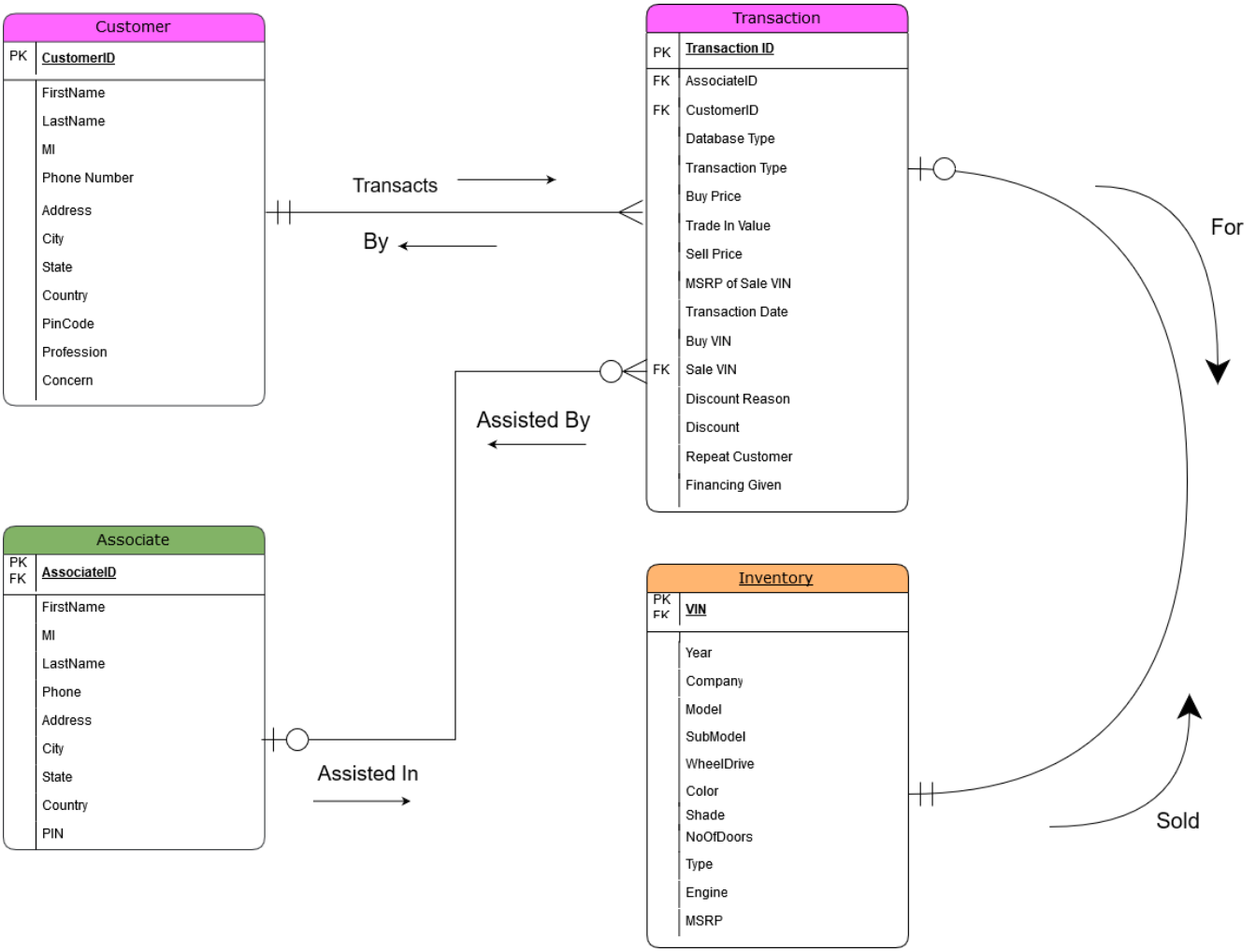


Image 20: Integrated ER Diagram, Crow's Foot notation

IX. Integrated Relational Schema

Logical Schema is represented below in Relational format for the same Integrated Database.

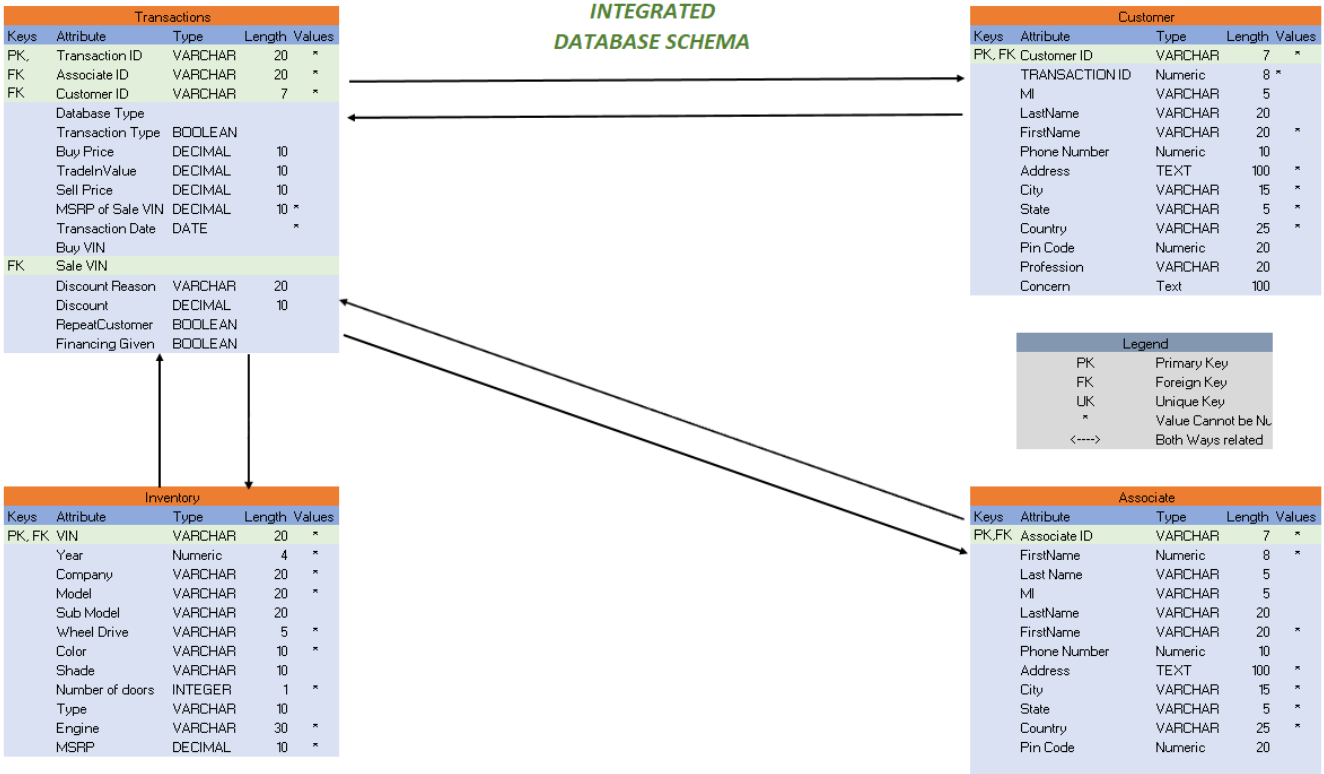


Image 21: Integrated Relational Schema

X. Integrated Database Schema

Transactions

Transaction ID	Associate Name	Customer ID	DataBase Type	Transaction Type	Buy Price	TradeInValue	Sell Price	MSRP of Sale VIN	Transaction Date	Buy VIN	Sale VIN	Discount Reason	Discount	RepeatCustomer	Financing Given
10000000	NA	CID008	Auto Dealer	Trade	NA	7,500.00	1,27,500.00	1,35,000.00	09-08-2017	NA	S7enznmKTrKsbm4ceC	NA	0.00	No	No
10000001	NA	CID002	Auto Dealer	Sell	NA	NA	19,500.00	23,475.00	10-09-2018	NA	UW7W4XUcxaMBL2PHqS	EndofYear	3,975.00	No	No
10000002	NA	CID006	Auto Dealer	Sell	NA	NA	60,615.00	60,615.00	08-08-2017	NA	vHxfKmtZ8bSd4JqP5y	NA	0.00	No	No
10000003	NA	CID003	Auto Dealer	Sell	NA	NA	62,000.00	68,000.00	08-09-2017	NA	ZdspCskTUsEMuA5xj4	EndofYear	6,000.00	No	No
10000004	NA	CID007	Auto Dealer	Trade	NA	1,250.00	26,512.00	30,740.00	20-10-2017	NA	Ab3F3AR5QX4jmxQGNX	EndofYear	2,978.00	No	No
10000005	NA	CID005	Auto Dealer	Sell	NA	NA	23,475.00	23,475.00	28-02-2018	NA	AQm44N9vhHn6DsWvsr	NA	0.00	No	No
10000006	NA	CID010	Auto Dealer	Trade	NA	2,500.00	20,975.00	23,475.00	15-06-2018	NA	amdRVQn8AVfrdP48CY	NA	0.00	No	No
10000007	NA	CID009	Auto Dealer	Sell	NA	NA	89,300.00	94,000.00	05-05-2018	NA	eLqdyxVVA2q5vRZNg5	First Time Driver	4,700.00	No	No
10000008	NA	CID004	Auto Dealer	Sell	NA	NA	25,232.25	29,685.00	04-03-2018	NA	3T3zsvzUp5Vm5r2SGm	Repeat Customer	4,452.75	Yes	No
10000009	NA	CID001	Auto Dealer	Trade	NA	5,000.00	57,685.00	74,500.00	21-01-2018	NA	QMsFeqUT38MFLV4NxW	Senior Citizen	11,815.00	No	No
10123456	Kylo Ren	CID011	Preowned	Buy	6,200.00	NA	NA	NA	04-02-2016	1BJ38LO45129JUT4I	NA	NA	0.00	No	No
10123457	Leia Organa	CID012	Preowned	Sell	NA	NA	9,500.00	9,700.00	01-08-2016	NA	5UD5LODK8W62DLKIEM	Autumn sales even	200.00	No	No
10123458	Anakin Skywalke	CID013	Preowned	Trade	5,595.00	1,205.00	6,800.00	7,000.00	06-07-2017	25D9MEI2NMDLPDK8	6S58W2S3F6G8G4D1D	Regular Discount	200.00	No	YES
10123459	R2-D2	CID014	Preowned	Trade	4,400.00	4,200.00	8,600.00	9,000.00	05-06-2017	6S58W2S3F6G8G4D1I	1BJ38LO45129JUT4I	Regular Discount	400.00	No	No
10123460	Padme Amidala	CID015	Preowned	Trade	6,975.00	1,025.00	8,000.00	8,500.00	05-01-2016	74EHF4F8YT56SMZA9	526DOEM78D9E124DL	Senior Citizen	500.00	No	No
10123461	Kylo Ren	CID016	Preowned	Buy	1,450.00	NA	NA	NA	06-09-2016	1E02D58GMZ5CP9D87	NA	NA	0.00	No	No
10123462	Anakin Skywalke	CID017	Preowned	Sell	NA	NA	9,995.00	11,000.00	02-05-2017	NA	1E02D58GMZ5CP9D87	Regular Discount	1,005.00	No	No
10123463	Anakin Skywalke	CID018	Preowned	Sell	NA	NA	11,999.00	12,500.00	03-06-2017	NA	256DKEM74DOLF8521	Regular Discount	501.00	No	YES
10123464	Padme Amidala	CID019	Preowned	Buy	3,500.00	NA	NA	NA	27-09-2017	81S2Q4JFMEWL54218	NA	NA	0.00	No	No
10123465	Leia Organa	CID020	Preowned	Trade	4,600.00	5,500.00	10,100.00	11,000.00	01-01-2016	526DOEM78D9E124DL	71DE6E55R2F3Q4A1Z	Repeat Customer	900.00	Yes	No

Table 4: Integrated Transaction Relation

Customer

Customer ID	TRANSACTION ID	LastName	FirstName	MI	Phone Number	Address	City	State	Country	Pin Code	Profession	Concern
CID011	10123456	Baggins	Frodo		202-555-0109	7405 Oak Meadow Road	Elk Grove Village	IL	USA	NA	NA	NA
CID012	10123457	Gamgee	Samwise		701-555-0109	9372 Stillwater Ave.	Champaign	IL	USA	NA	NA	NA
CID013	10123458	Took	Peregrin		202-555-0182	24 West Beechwood Drive	Urbana	IL	USA	NA	NA	NA
CID014	10123459	Brandybuck	Meriadoc		202-555-0147	8 Hall Lane	Savoy	IL	USA	NA	NA	NA
CID015	10123460	Wormtongue	Grima		701-555-0136	628 Center Rd.	Zionsville	IN	USA	NA	NA	NA
CID016	10123461	Bolger	Fredegar		202-555-0179	9827 Morris Ave.	Bloomington	IL	USA	NA	NA	NA
CID017	10123462	Goatleaf	Harry		701-555-0137	6 Blue Spring Court	Des Plaines	IL	USA	NA	NA	NA
CID018	10123463	Willow	Old Man		701-555-0192	7186 Wintergreen St.	Champaign	IL	USA	NA	NA	NA
CID019	10123464	Angmar	Witch-King of		701-555-0190	12 Rockaway Street	Urbana	IL	USA	NA	NA	NA
CID020	10123465	Gandalf			701-555-0172	7390 E. Glenridge Rd.	Rantoul	IL	USA	NA	NA	NA
CID001	10000009	Dumbledore	Albus	R	NA	557 Rodeo Trl	Rantoul	IL	USA	61866	Dean	NA
CID002	10000001	Granger	Hermione	S	NA	190 Clemton Ave	Champaign	IL	USA	61821	Archivist	Needs loan
CID003	10000003	Longbottom	Neville	R	NA	34 Lark Meadow Dr	Savoy	IL	USA	61874	Doctor	NA
CID004	10000008	Lovegood	Luna	D	NA	245-B Church St	Urbana	IL	USA	61802	Student	Needs loan
CID005	10000005	Lupin	Remus	W	NA	911 Megellan Ave	Bloomington	IL	USA	61701	Doctor - pediatrician	NA
CID006	10000002	Malfoy	Draco	M	NA	987 Withrop Lane	Urbana	IL	USA	61801	Unknown profession	NA
CID007	10000004	Pettigrew	Peter		NA	55 Shadow Canyon Trl	Indianapolis	IN	USA	46077	Librarian	Needs financing
CID008	10000000	Potter	Harry	D	NA	2008 Williams Dr	Chicago	IL	USA	60007	Professor, UIC	NA
CID009	10000007	Weasley	Ginny		NA	8890 Winston St	Champaign	IL	USA	61820	Stay at home mother	Inquiry into financing options
CID010	10000006	Weasley	Ronald	R	NA	54 Lane Ave	Chicago	IL	USA	60018	Research scientist	NA

Table 5: Integrated Customer Relation

Inventory

VIN		Year	Company	Model	Sub Model	Wheel Drive	Color	Shade	Number of doors	Type	Engine	MSRP
vHxfKmtZ8bSd4JqP5y	Auto Dealer	2017	Ford	Expedition	King Ranch	4WD	White	Pearl	4		Internal Combustion	60,615.00
Ab3F3AR5QX4jmxQGNX	Auto Dealer	2017	Ford	Fusion	Titanium	FWD	Gold		4		Hybrid	30,740.00
S7enznmKTrKsbm4ceC	Auto Dealer	2017	Tesla	Model S	P100D	AWD	White		4		Electric	1,35,000.0
ZdspCskTUsEMuA5xj4	Auto Dealer	2017	Tesla	Model S	60	AWD	Gray		4		Electric	68,000.00
QMsFeqUT38MFLV4NxW	Auto Dealer	2018	Tesla	Model S	75D	AWD	White		4		Electric	74,500.00
eLqdyxVVA2q5vRZNg5	Auto Dealer	2018	Tesla	Model S	100D	AWD	White		4		Electric	94,000.00
UW7W4XUcxaMBL2PHqS	Auto Dealer	2018	Toyota	Prius		FWD	Blue		4	Sedan	Hybrid	23,475.00
AQm44N9vhHn6DsWvsr	Auto Dealer	2018	Toyota	Prius		FWD	White		4	Sedan	Hybrid	23,475.00
amdRVQn8AVfrdP48CY	Auto Dealer	2018	Toyota	Prius		FWD	Silver		4	Sedan	Hybrid	23,475.00
3T3zsvzUp5Vm5r2SGm	Auto Dealer	2018	Toyota	Prius		FWD	Black		5	Hatchback	Hybrid	29,685.00
S7enznmKTrKsbm4ceC	Pre Owned											1,35,000.0
UW7W4XUcxaMBL2PHqS	Pre Owned											23,475.00
vHxfKmtZ8bSd4JqP5y	Pre Owned											60,615.00
ZdspCskTUsEMuA5xj4	Pre Owned											68,000.00
Ab3F3AR5QX4jmxQGNX	Pre Owned											30,740.00
AQm44N9vhHn6DsWvsr	Pre Owned											23,475.00
amdRVQn8AVfrdP48CY	Pre Owned											23,475.00
eLqdyxVVA2q5vRZNg5	Pre Owned											94,000.00
3T3zsvzUp5Vm5r2SGm	Pre Owned											29,685.00
QMsFeqUT38MFLV4NxW	Pre Owned											74,500.00
5UD5LODK8W62DLKIEM	Pre Owned											9,700.00
6S58W2S3F6G8G4D1D	Pre Owned											7,000.00
1BJ38LO45129JUT4I	Pre Owned											9,000.00
526DOEM78D9E124DL	Pre Owned											8,500.00
1E02D58GMZ5CP9D87	Pre Owned											11,000.00
256DKEM74DOLF8521	Pre Owned											12,500.00
71DE6E55R2F3Q4A1Z	Pre Owned											11,000.00
1BJ38LO45129JUT4I	Pre Owned											6,200.00
25D9MEI2NMDLPDK85	Pre Owned											5,595.00
6S58W2S3F6G8G4D1D	Pre Owned											4,400.00
74EHF4F8YT56SMZA9	Pre Owned											6,975.00
1E02D58GMZ5CP9D87	Pre Owned											1,450.00
81S2Q4JFMEWL54218	Pre Owned											3,500.00
526DOEM78D9E124DL	Pre Owned											4,600.00

Table 6: Inventory Relation

Employee Id	First Name	Last Name	MI	Phone Number	Address	City	State	Country	Pin Code
EMP001	Kylo	Ren							
EMP002	Leia	Organa							
EMP003	Anakin	Skywalker							
EMP004	R2	D2							
EMP005	Padme	Amidala							

Table 7: Associates Relation

XI. METADATA

Title	ER Integration
Databases	Auto Dealer & Pre-Owned Car Dealer
ER Format	Crow's Foot
Logical Schema Format	Tabular Format
Database Type	Microsoft Excel
URL	https://shrashtisinghal.github.io/Data-Curation/
Author	Shrashti Singhal
No. Of pages	14
Text Format	Candara

XII. Questions

1. How did you decide to represent the data in the way that you did?

The data is integrated using ER Diagrams in Chen's format. Final attributes along with complete ER Diagrams are displayed Crows Foot notation. After the integration a relational schema is developed for the integrated ER Diagram. Relational Schema is developed because it allows us to add constraint and data type to the attribute values. After the conceptual and logical Schema is completed, the entire database tables are shown with the appropriate data. These data tables depicts the actual database.

Here, in this paper all kind of schemas are used. ER diagrams in Chen's Format is used for the integration purpose of the tables. In this notation, we haven't shown the attributes to make the ERs neat and compact, as there are many attributes in each table. While the final integrated ER Diagram is shown in Crow's Foot format with all the attributes and relationships titles.

2. What were the hardest decisions you had to make in this design process?

Hardest Decision was to create a separate Associates table. We didn't have any information about the associates of the both the auto companies. We added more attributes for associates, which were not provided to us.

3. Did you leave out any information? If so, why?

No information has been left out. The purpose of creating Sales table as the subset of transaction table was to protect any information loss.

I leave it to analytics operation to filter the information needed by them. Here, I have captured all the details found out from the legacy dataset. Though, I have rearranged the information.

4. How does your design support data independence?

Data independence is supported as Crow's Foot ER Diagrams is used as a conceptual schema which allows interoperability between other ER Diagram notations and can be changed without impacting the data. Therefore it provides conceptual data independence.

It provides data independence from storage. If physical storage changes the end programs using the storage will continue to operate like before.

Also, new data constructs can be added without impacting the data, relationship and keys between the existing data.

5. How may your design support the overarching goals of data curation (revisit objectives and activities of Week 1)?

The overarching goal of Data Curation is concerned with all aspects of the data management, therefore, I have implemented most of the data Curation techniques/activities such as:

Collection: Collecting data from 2 different sources. Data was in different formats as well. Data from auto Dealer was in text, excel and word doc. Pre-owned database Dealer data was in excel format.

This data was collected. Of particular importance: recording information (as metadata) related to collection activity so that all relevant aspects of context are available later to support full understanding, authentication, and provenance. This recoding has been done at the beginning of the paper where screenshots of data has been taken. Later a Meta data is also generated.

Organization: Determination an appropriate data model and schema. Various kinds of standard have been adopted such as, Chens, Crow foot notation for ER Diagram. Tabular format for relational database as a logical Schema.

Storage: Appropriate mix of storage strategies have been used. Data was in different formats. Excel is finally used for storage after analysing it the best fit for data storage for the given data.

Preservation: This document is maintained as a preservation strategy. At the same time this document is readable and understandable. Moreover another copy of document is maintained over the internet in a safe repository, URL of which can be accessed from the metadata.

Discoverability: A metadata is developed to support searching for and finding relevant data in relevant formats. The ability to search for and locate relevant data is achieved by the help of primary and foreign keys provided in the conceptual and logical schema.

Integration: Integration of data from different sources using different data models has been done.

Use of schema alignment and cross-walking techniques to integrate data. Documentation of integration strategies in detail so that any conflation, data loss, etc. is noted.
Reproducibility: Our integration process support the ability to reproduce results, ensuring scientific validity and reliability. Data curation for reproducibility included documenting not only data collection and management, but also documenting processing and analysis.
Provenance: One data set (or view) is derived from another in case of Customer and Transaction tables, moreover few attributes are calculated from values of inventory table. These derivations and calculations are reliable use and understandable due to the use of primary and foreign key. This is also ensured by data constraints and kind of data types used for particular attributes.

This will allow efficient and reliable support to the analysis of data, and will also enable reuse over the time. How these activities have been incorporated and have enhanced the database design is mentioned above in this paper.

XIII. Conclusion

This integration of ER Schema, involved most of the Data Curation activities alongside as mentioned in the last question, question 5 in the questions heading.
On one hand ER integration was done while on other many of the Data curation activities were done to reach to the appropriate ER Diagrams. Moreover, Logical Database Schema & Database tables are added in addition.