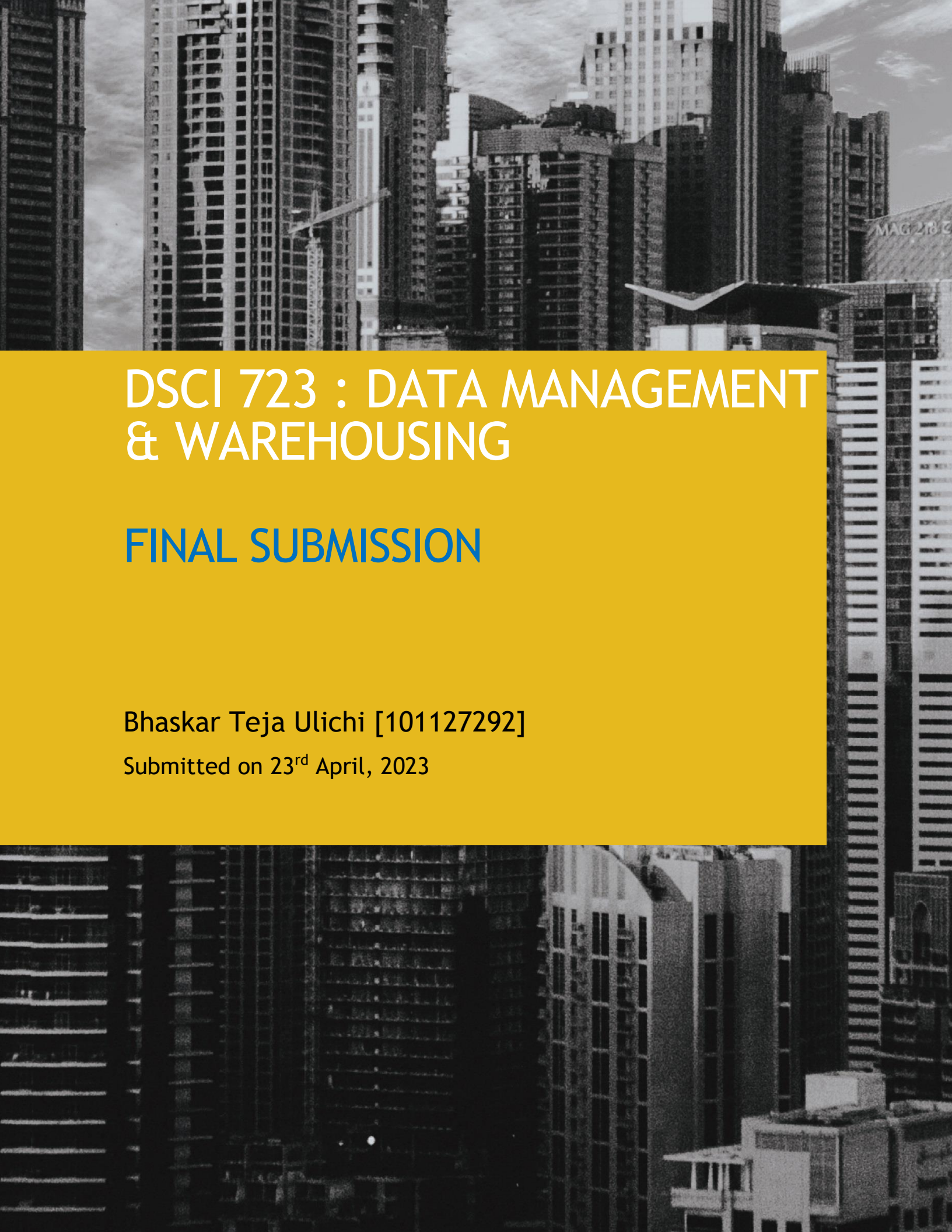


DSCI 723 : DATA MANAGEMENT & WAREHOUSING

FINAL SUBMISSION

Bhaskar Teja Ulichi [101127292]

Submitted on 23rd April, 2023

ABSTRACT

The report presented details on the development of a proposed database system for Custom Closet Contractors (C3), a company that specializes in custom home organization solutions. The current paper-based data management system has led to various issues like a miscommunication between departments, errors in the production process, and a lack of comprehensive data analysis. The proposed database aims to provide an electronic system for storing and managing customer data, job details, and production information. However, there are risks associated with implementing the database, including data breaches, technical glitches, employee resistance, and training costs.

The report outlines the necessary entities, requirements, relationships, and cardinalities for developing the database system. Eight entities have been identified: Customer, Company, Consultation, Designer, Quote, Unit, Installation, and Installer, along with their attributes. The report explains the relationships between these entities - Refers, Works_For, Gets, Involves, Receives, Includes, etc., along with their respective cardinalities and participations. The report presents an Entity Relation Diagram and the resultant Relational Schema along with full key functional dependencies. It also indicates the attributes and the composite primary keys they depend on to develop the database system.

Furthermore, the report provides a detailed database diagram and captures the first ten rows of data populated into each table. It also includes a data dictionary for each relation of the database, which defines each attribute and its data type. The report includes a warehousing process with the ER diagram and the star schema implemented to develop the data warehouse, providing information about each table in the data warehouse along with the first five rows of data that are populated via the ETL process. Lastly, the report emphasizes the importance of considering ethical and privacy concerns when collecting and storing data.

In conclusion, the report outlines the development of a proposed database system for C3 and presents details on entities, requirements, relationships, cardinalities, and necessary diagrams. The report also highlights the benefits and risks associated with implementing the proposed database system. By implementing, C3 can leverage the proposed database system to increase efficiency, streamline operations, and support data-driven decision-making, leading to improved profitability and scalability.

Table of Contents

ABSTRACT 1

TABLE OF FIGURES 3

LIST OF TABLES 3

INTRODUCTION 4

BUSINESS PROBLEM 4

EXPECTED BENEFITS 4

PROJECTED RISKS 5

DEVELOPING DATABASE 6

ENTITIES 6

RELATIONSHIPS & CARDINALITIES 7

FUNCTIONAL DEPENDENCIES 11

DATABASE - POPULATED 13

DATA DICTIONARY 17

DATA WAREHOUSING 21

DATAWAREHOUSE – POPULATED 23

ETHICAL & PRIVACY ISSUES 24

APPENDIX 26

TABLE OF FIGURES

Figure 1: The Entity-Relationship Diagram for the CRM System of C3 10

Figure 2: The Relational Schema for the CRM System of C3 11

Figure 3: Database diagram 12

Figure 4: ER Diagram for Datawarehouse Model 21

Figure 5: STAR SCHEMA FOR DATAWAREHOUSE MODEL 21

Figure 6: DATAWAREHOUSE DIAGRAM 22

LIST OF TABLES

Table 1: Database Tables - Populated 13

Table 2: Data Dictionary 17

Table 3: Populated Datawarehouse Tables 23

INTRODUCTION

BUSINESS PROBLEM

The business problem that the proposed database is supposed to address is the paper-based data management system used by Custom Closet Contractors (C3). C3 is a successful family business that designs, manufactures, and installs custom home organization solutions for individual clients and home builders. The company's operations involve customer acquisition, design consultation, production, and installation. C3's existing system of data management relies heavily on paper, with customer information and job details stored in blue binders. This system has led to miscommunication between the front office and production staff and has resulted in errors in the production process, causing wasted time and materials. The lack of an electronic database has also hindered the company's ability to analyze its operations comprehensively and strengthen profitability. The proposed database aims to address these issues by providing an electronic system for storing and managing customer data, job details, and production information, enabling the company to streamline its operations and improve efficiency.

EXPECTED BENEFITS

The proposed database is expected to improve data management by allowing Custom Closet Contractors (C3) to store all customer information, job orders, and other important data in a centralized, electronic format, thereby eliminating the need for paper-based records. This will lead to increased efficiency by streamlining C3's operations, from customer acquisition to final installation, providing real-time access to job orders and reducing miscommunication and errors.

The database will also facilitate data-driven decision-making by providing valuable insights into C3's operations, including pricing, inventory, and marketing. This will ultimately increase profitability by improving efficiency, reducing waste, providing better customer service, and optimizing production processes. Additionally, the database will enhance C3's scalability by providing real-time visibility into its operations, enabling the company to optimize its resources more effectively. Finally, the database will provide enhanced security measures to protect sensitive customer data, ensuring compliance with data privacy regulations and building customer trust.

PROJECTED RISKS

The implementation of the proposed database may face some risks, including data breaches due to the electronic database's vulnerability to cyberattacks or unauthorized access, potentially compromising sensitive customer data. Additionally, technical glitches such as system crashes or data corruption may occur, leading to data loss or disruptions in operations. The implementation of the new database may also be delayed due to technical or operational issues, potentially impacting the company's operations and profitability. Some employees may resist the implementation of the new database, leading to delays or disruptions in operations. Finally, the implementation of the new database may require significant training and support costs, which could impact the company's budget.

DEVELOPING DATABASE

The Database requirements are presented below.

ENTITIES

1. **Customer:** Attributes include CustomerID (Unique), CustomerName (Composite Attribute – FirstName, LastName), Phone (Multivalued attribute), Email, Address (Composite Attribute – Apartment, Street (Optional), City, ZipCode, State and Country). Customer's street address is optional, as street address becomes important during installation phase where it is collected madatorily.
2. **Company:** Attributes include CompanyID (Unique), CompanyName, Email, Phone (Multivalued), Website, AccountType.
3. **Consultation:** Attributes include ConsultationID (Unique), Date, Time, Location.
4. **Designer:** Attributes include DesignerID (Unique), DesignerName (Composite – FirstName, LastName).
5. **Quote:** Attributes include QuoteNo (Unique), TotalAmt (Derived), Taxes, DiscountRate (Optional), Note (Optional), Status
6. **Unit:** Attributes include UnitID (Unique), UnitName, Notes
7. **Installation:** Attributes include InstallationNo (Unique), Date, Time, Install_Address (Composite - Apartment, Street, City, ZipCode, State and Country)
8. **Installer:** Attributes include InstallerID (Unique), Name (Composite – FirstName, LastName), Phone (Multivalued).

RELATIONSHIPS & CARDINALITIES

1. **Refers:** *Unary Relationship involving Customer*
    - **Cardinalities:** An existing customer may refer more than one potential customer. A potential customer can only be referred by one existing customer (***One-to-Many***)
    - **Participation:** As there is no mandatory requirement for a customer to refer, the relationship is **optional on the 'Many' side**. Assuming that the information about potential customers is obtained not only from existing customers, but also from other sources, the relationship becomes **optional on the 'ONE' side** as well
2. **Works_For:** *One-to-Many relationship (1:M) between Customer and Company*
    - **Company to Customer:** We are keeping track of information about all companies whose employees are our customers, the relationship mandates company participation **(Participation: Mandatory)**. A given company could employ many customers. (**Max Cardinality: Many**)
    - **Customer to Company:** Not all customers need to work for a company. A given customer could be employed by only a single company—or no company at all **(Participation: Optional; Max Cardinality: 1)**
3. **Gets:** *One-to-Many relationship (1:M) between Customer and Consultation*
    - **Customer to Consultation:** It was possible for a customer to have multiple consultations, if needed. **(Participation: Optional; Max Cardinality: Many)**

- **Consultation to Customer:** We maintain records of all scheduled consultations and each consultation involves only one customer **(Participation: Mandatory; Max Cardinality: 1)**

4. **Involves:** *One-to-Many relationship (1:M) between Consultation and Designer*
    - **Consultation to Designer:** It is impossible to have a consultation without a designer. Each consultation involved only one designer **(Participation: Mandatory; Max Cardinality: 1)**
    - **Designer to Consultation:** Though it has been said that a designer was involved in many different consultations, we can reasonably assume that a newly recruited designer may not participate in a consultation, but his details need to be updated in the database **(Participation: Optional; Max Cardinality: Many)**
5. **Receives:** *One-to-Many relationship (1:M) between Customer and Quote*
    - **Customer to Quote:** A customer does not receive a quote if the consultation is not successful. A customer could have many quotes **(Participation: Optional; Max Cardinality: Many)**
    - **Quote to Customer:** A successful consultation leads to a quote. If a quote exists, it must be sent to a customer. Each quote was always specific to one customer **(Participation: Mandatory; Max Cardinality: 1)**
6. **Includes:** *Many-to-Many relationship (M:N) between Quote and Unit*
    - This relationship includes attributes – Quantity, Price
    - **Quote to Unit:** Each quote must have at least one unit described in it **(Participation: Mandatory; Max Cardinality: Many).**

- **Unit to Quote:** A unit could exist without being part of a quote (e.g. Some standard units that are bought & used for customers irrespective of their specifications). A given unit could appear on many different quotes **(Participation: Optional; Max Cardinality: Many)**.

7. **Leads_To:** *One-to-Many relationship (1:M) between Quote and Installation*
    - **Quote to Installation:** Installation can be scheduled only when a quote is accepted, and some jobs require multiple installations. **(Participation: Optional; Max Cardinality: Many).**
    - **Installation to Quote:** An installation cannot happen without a quote and each installation was always related to one quote. **(Participation: Mandatory; Max Cardinality: 1)**
8. **Is_Assigned_To:** *Many-to-Many relationship (M:N) between Installation and Installer*.
    - **Installation to Installer:** Installation must require an installer. Some large jobs may involve more than one installer. **(Participation: Mandatory; Max Cardinality: Many)**
    - **Installer to Installation:** Not all installers need to participate in installation (eg: newly hired installers). Only one installer is usually assigned for each installation. **(Participation: Optional; Max Cardinality: One)**

FIGURE 1: THE ENTITY-RELATIONSHIP DIAGRAM FOR THE CRM SYSTEM OF C3

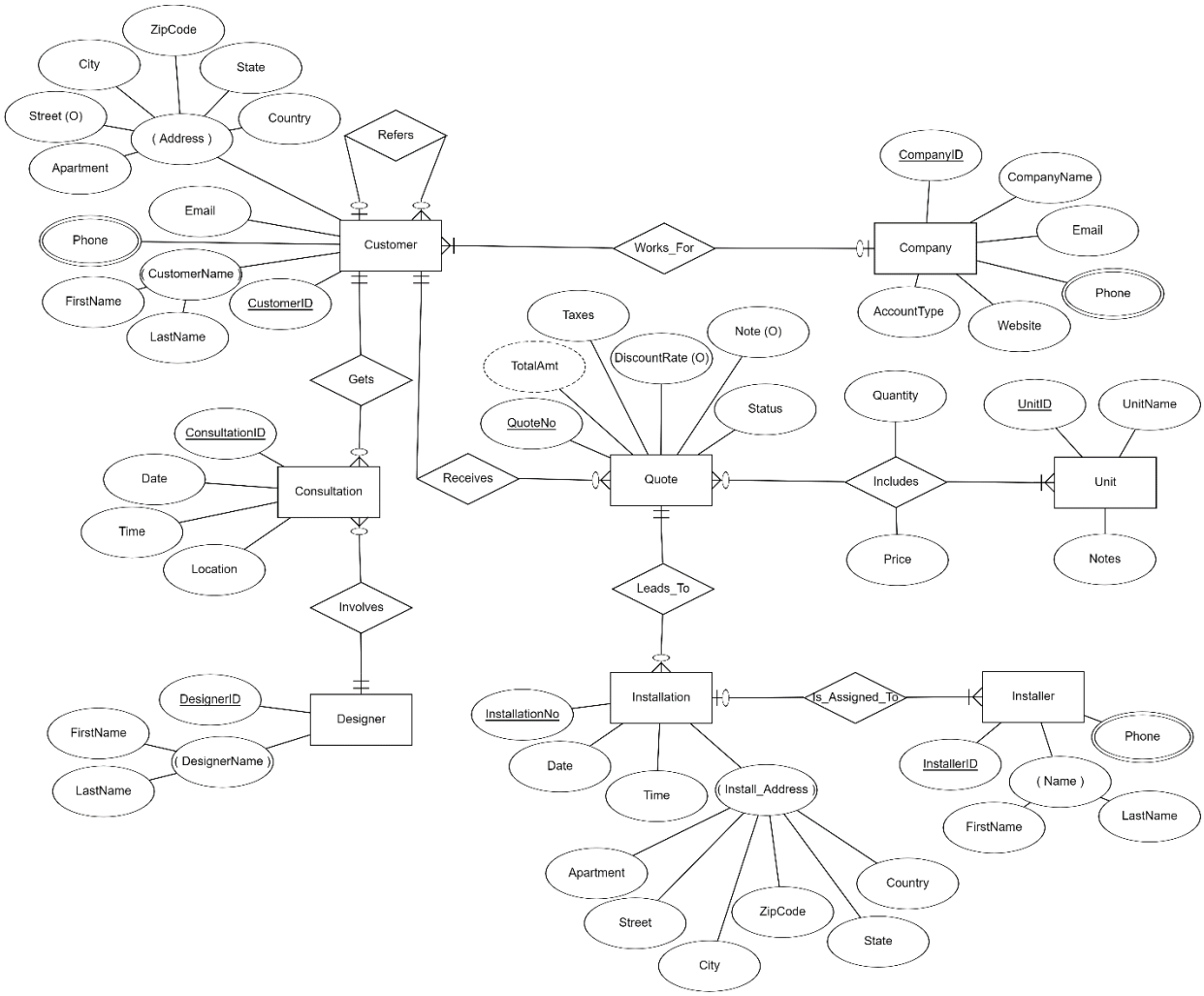
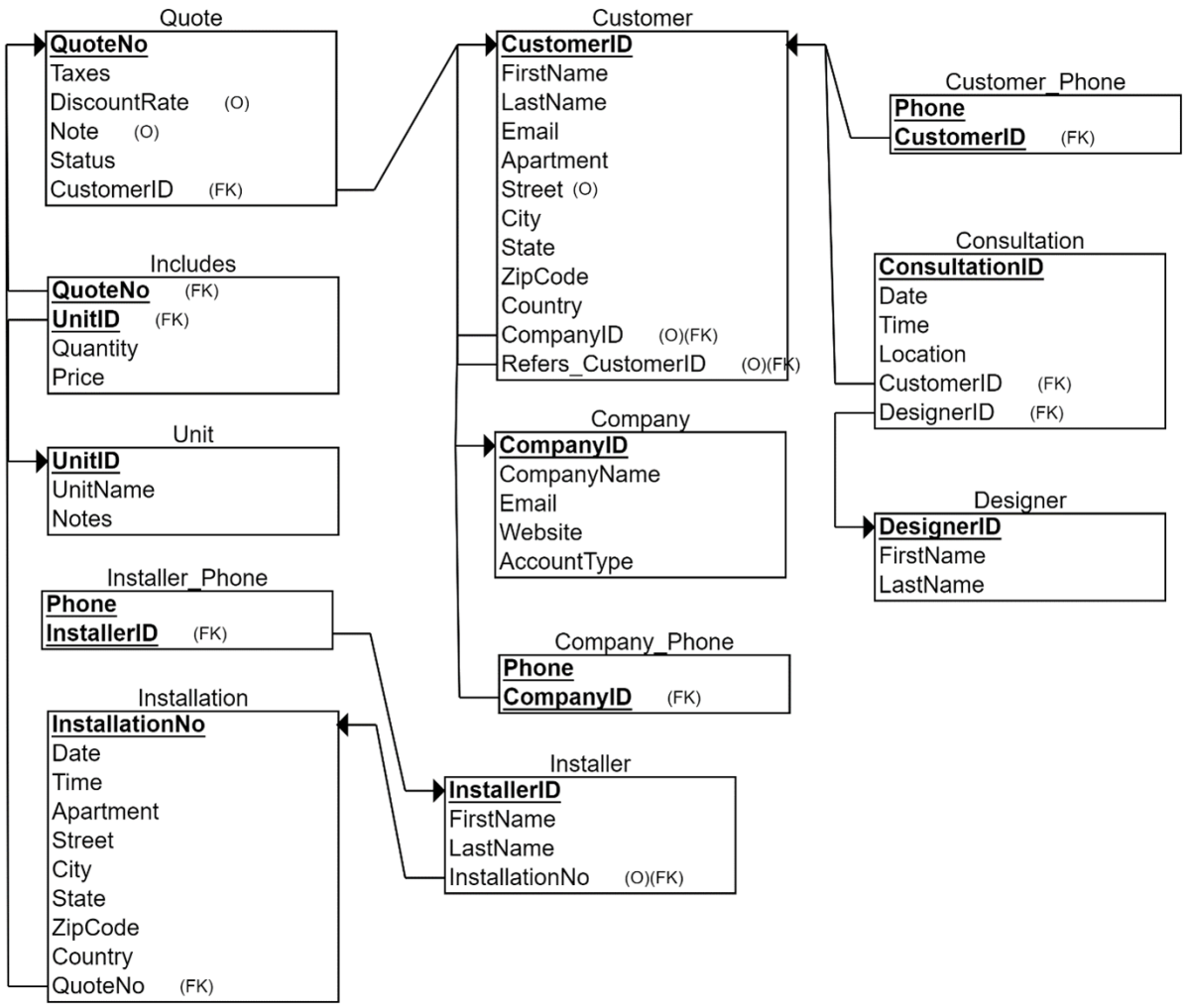


FIGURE 2: THE RELATIONAL SCHEMA FOR THE CRM SYSTEM OF C3

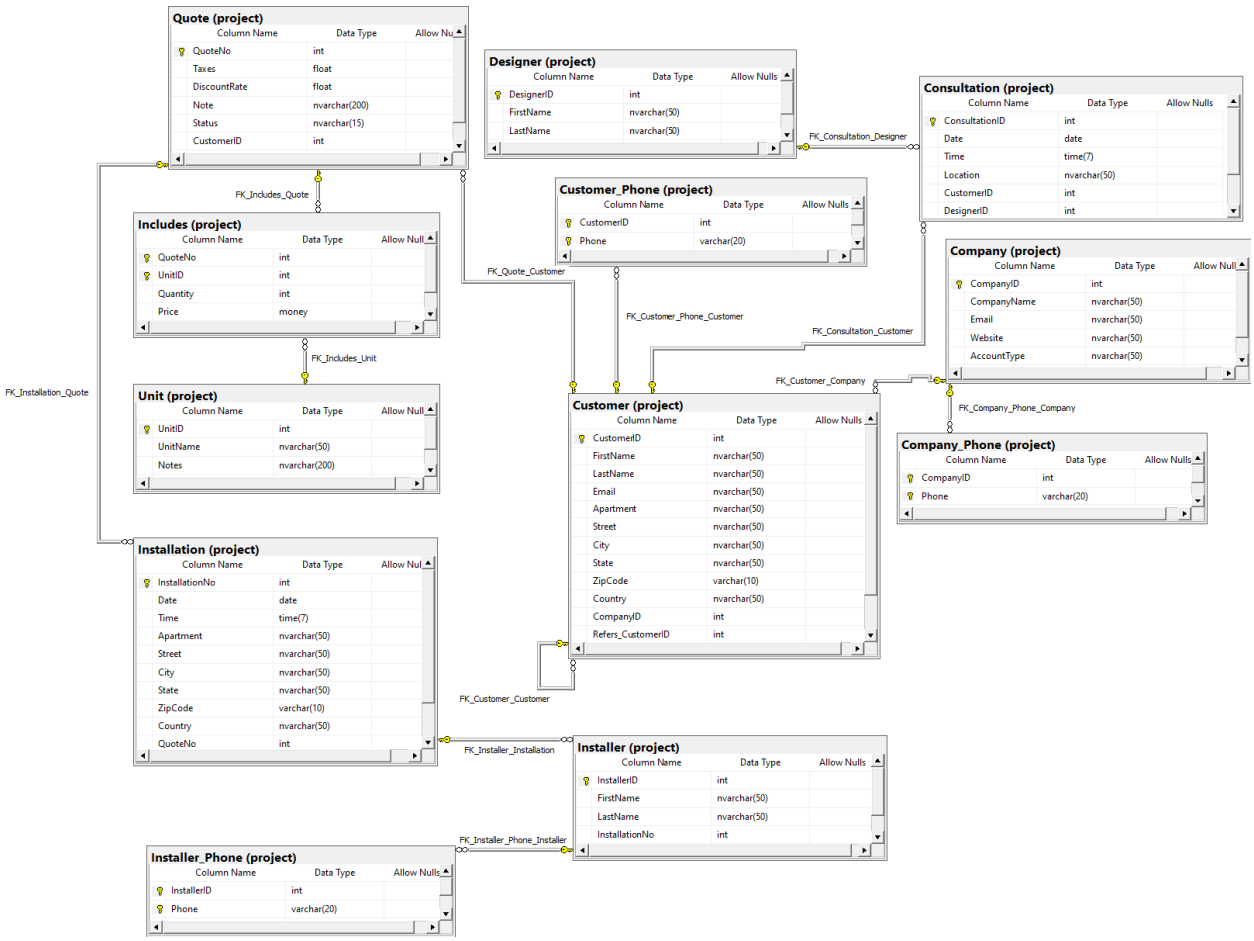


FUNCTIONAL DEPENDENCIES

1. **CustomerID** → FirstName, LastName, Email, Apartment, Street, City, State, ZipCode, Country, CompanyID, Refers_CustomerID
2. **CompanyID** → CompanyName, Email, Website, AccountType
3. **ConsultationID** → Date, Time, Location, CustomerID, DesignerID
4. **DesignerID** → FirstName, LastName
5. **QuoteNo** → TotalAmt, Taxes, DiscountRate, Note, Status, CustomerID

6. **UnitID** → UnitName, Notes
7. **QuoteNo, UnitID** → Quantity, Price
8. **InstallerID** → FirstName, LastName, InstallationNo
9. **InstallationNo** → Date, Time, Apartment, Street, City, State, ZipCode, Country, QuoteNo
10. **CustomerID, Phone** → Phone
11. **CompanyID, Phone** → Phone
12. **InstallerID, Phone** → Phone

FIGURE 3: DATABASE DIAGRAM



DATABASE - POPULATED

The following captures show the first 10 rows of each table in the database, created under the schema “project”.

TABLE 1: DATABASE TABLES - POPULATED

Company

	CompanyID	CompanyName	Email	Website	AccountType
1	1001	Ace Builders	acebuilders@example.com	www.acebuilders.com	Builder
2	1002	Dream Homes	dreamhomes@example.com	www.dreamhomes.com	Builder
3	1003	Cityscape Properties	cityscapeproperties@example.com	www.cityscapeproperties.com	Real Estate
4	1004	Sunset Interiors	sunsetinteriors@example.com	www.sunsetinteriors.com	Interior Designor
5	1005	Elite Builders	elitebuilders@example.com	www.elitebuilders.com	Builder
6	1006	Royal Interiors	royalinteriors@example.com	www.royalinteriors.com	Interior Designor
7	1007	Green Garden Landscaping	greengardenlandscaping@example.com	www.greengardenlandscaping.com	Landscaping
8	1008	Luxury Builders	luxurybuilders@example.com	www.luxurybuilders.com	Builder
9	1009	Artisan Interiors	artisaninteriors@example.com	www.artisaninteriors.com	Interior Designor
10	1010	Home Renovations	homerenovations@example.com	www.homerenovations.com	Remodeling and Renovation

Customer

	CustomerID	FirstName	LastName	Email	Apartment	Street	City	State	ZipCode	Country	CompanyID	Refers_CustomerID
1	2001	Tom	Cruise	cruise@sag.com	611	Beverly Park	Syracruse	New York	10012	USA	NULL	NULL
2	2002	Brad	Pitt	brad@sag.com	345	Rosewood Ave	Los Angeles	California	90004	USA	1001	NULL
3	2003	Cate	Blanchett	cate@sag.com	204	Hudson Street	Sydney	New South Wales	2011	Australia	NULL	2002
4	2004	Dwayne	Johnson	dwayne@sag.com	27	Clarendon St	Beverly Hills	California	90210	USA	NULL	2015
5	2005	Hugh	Jackman	hugh@sag.com	22	Hickson Rd	Sydney	New South Wales	2000	Australia	1002	NULL
6	2006	Leonardo	Dicaprio	leo@sag.com	148	S Carolwood Dr	Los Angeles	California	90024	USA	NULL	2050
7	2007	Priyanka	Chopra	priyanka@sag.com	402	Carter Road	Mumbai	Maharashtra	400050	India	NULL	2012
8	2008	Rachel	Weisz	rachel@sag.com	45	Lexington Ave	New York	New York	10010	USA	1003	NULL
9	2009	Steven	Spielberg	steven@dga.com	375	Crestwood Drive	Los Angeles	California	90049	USA	NULL	2010
10	2010	Tom	Hanks	tom@dga.com	301	N Carolwood Dr	Los Angeles	California	90077	USA	1004	NULL

Company_Phone

	CustomerID	Phone
1	2001	+1 (234) 567-8901
2	2002	+1 (345) 678-9012
3	2003	+1 (456) 789-0123
4	2004	+1 (567) 890-1234
5	2005	+1 (678) 901-2345
6	2006	+1 (789) 012-3456
7	2007	+1 (890) 123-4567
8	2008	+1 (901) 234-5678
9	2009	+1 (012) 345-6789
10	2010	+1 (123) 456-7890

Customer_Phone

	CompanyID	Phone
1	1001	+1 (605) 675-8231
2	1002	+1 (234) 567-8901
3	1003	+1 (987) 654-3210
4	1004	+1 (123) 456-7890
5	1005	+1 (555) 555-5555
6	1006	+1 (111) 222-3333
7	1007	+1 (444) 555-6666
8	1008	+1 (777) 888-9999
9	1009	+1 (222) 333-4444
10	1010	+1 (888) 999-0000

Designer

	DesignerID	FirstName	LastName
1	3001	Lily	Aldridge
2	3002	Helena	Christensen
3	3003	Gigi	Hadid
4	3004	Kendall	Jenner
5	3005	Bella	Hadid
6	3006	Karlie	Kloss
7	3007	Mario	Testino
8	3008	Patrick	Demarchelier
9	3009	Steven	Meisel
10	3010	Annie	Leibovitz

Consultation

	ConsultationID	Date	Time	Location	CustomerID	DesignerID
1	4001	2022-02-14	10:00:00.0000000	USA	2033	3003
2	4002	2022-03-20	13:30:00.0000000	United Kingdom	2011	3012
3	4003	2022-04-01	16:00:00.0000000	Online	2038	3004
4	4004	2022-04-05	09:00:00.0000000	USA	2028	3014
5	4005	2022-05-12	14:30:00.0000000	United Kingdom	2017	3013
6	4006	2022-05-20	11:00:00.0000000	USA	2031	3009
7	4007	2022-06-02	15:00:00.0000000	Online	2024	3007
8	4008	2022-07-08	10:30:00.0000000	United Kingdom	2019	3008
9	4009	2022-07-12	12:00:00.0000000	USA	2045	3015
10	4010	2022-08-15	14:00:00.0000000	United Kingdom	2009	3010

Quote

	QuoteNo	Taxes	DiscountRate	Note	Status	CustomerID
1	5001	0.07	NULL	Quote for a complete redesign of the living room wi...	Accepted	2005
2	5002	0.1	0.03	NULL	Rejected	2002
3	5003	0.07	0.02	Quote for a full kitchen remodel with new applianc...	Accepted	2040
4	5004	0.1	NULL	Quote for a complete bedroom redesign with a bo...	Rejected	2032
5	5005	0.07	0.05	Quote for a living room and dining room redesign ...	Accepted	2020
6	5006	0.1	0.04	Quote for a home office redesign with a minimalist ...	Accepted	2017
7	5007	0.07	NULL	Quote for a complete bathroom remodel with new ...	Rejected	2007
8	5008	0.1	0.01	Quote for a bedroom and bathroom redesign with ...	Accepted	2038
9	5009	0.07	0.05	NULL	Accepted	2049
10	5010	0.1	0.02	Quote for a complete kitchen and dining room rem...	Rejected	2024

Unit

	UnitID	UnitName	Notes
1	6001	Shoe Rack	A beautiful shoe rack with 4 shelves, made with pr...
2	6002	Kitchen Island	A custom kitchen island with granite countertop an...
3	6003	Lamp	A stylish lamp with a silver base and a white shade...
4	6004	TV Stand	A sleek and modern TV stand with a glossy black f...
5	6005	Bookcase	A sturdy and spacious bookcase with adjustable s...
6	6006	Dresser	A beautiful and functional dresser with 6 drawers, ...
7	6007	Nightstand	A simple and elegant nightstand with a single draw...
8	6008	End Table	A classic and versatile end table with a dark brown...
9	6009	Coffee Table	A stylish and functional coffee table with a glass to...
10	6010	Accent Chair	A comfortable and stylish accent chair with a bold ...

Includes

	QuoteNo	UnitID	Quantity	Price
1	5001	6009	2	75.00
2	5001	6010	3	100.00
3	5001	6076	5	300.00
4	5001	6087	1	50.00
5	5001	6105	4	200.00
6	5002	6012	1	25.00
7	5002	6033	3	150.00
8	5002	6056	4	200.00
9	5002	6081	5	350.00
10	5002	6111	2	100.00

Installation

	InstallationNo	Date	Time	Apartment	Street	City	State	ZipCode	Country	QuoteNo
1	7001	2022-01-05	09:30:00.0000000	Sky Palace	Rose Street	Los Angeles	California	90001	USA	5001
2	7002	2022-02-10	12:45:00.0000000	Green Acres	Meadow Lane	Chicago	Illinois	60007	USA	5003
3	7003	2022-03-15	16:15:00.0000000	Ocean Breeze	Marine Drive	Miami	Florida	33101	USA	5005
4	7004	2022-04-20	11:00:00.0000000	Fairy Tale	Wonderland Road	New York	New York	10001	USA	5006
5	7005	2022-05-25	14:30:00.0000000	Royal Garden	Palace Street	San Francisco	California	94101	USA	5008
6	7006	2022-06-30	10:00:00.0000000	Golden Nest	Sunrise Avenue	Houston	Texas	77001	USA	5009
7	7007	2022-07-05	15:45:00.0000000	Sunny Side	Brighton Lane	Miami	Florida	33101	USA	5011
8	7008	2022-08-10	08:15:00.0000000	Misty Mountains	Foggy Drive	Chicago	Illinois	60007	USA	5012
9	7009	2022-09-15	13:00:00.0000000	Crimson Villa	Redwood Street	San Francisco	California	94101	USA	5014
10	7010	2022-10-20	17:30:00.0000000	Windsor Castle	Queens Road	New York	New York	10001	USA	5016

Installer

	InstallerID	FirstName	LastName	InstallationNo
1	8001	Stephen	King	7001
2	8002	Neil	Gaiman	7002
3	8003	Haruki	Murakami	7003
4	8004	Alexander	Desplat	7004
5	8005	George	Orwell	7005
6	8006	Ernest	Hemingway	7005
7	8007	John	Williams	7005
8	8008	Sidney	Sheldon	7006
9	8009	Terry	Pratchett	7007
10	8010	George	Martin	7007

Installer_Phone

	InstallerID	Phone
1	8001	+1 (123) 456-7890
2	8002	+1 (234) 567-8901
3	8003	+1 (345) 678-9012
4	8004	+1 (456) 789-0123
5	8005	+1 (567) 890-1234
6	8006	+1 (678) 901-2345
7	8007	+1 (789) 012-3456
8	8008	+1 (890) 123-4567
9	8009	+1 (901) 234-5678
10	8010	+1 (012) 345-6789

DATA DICTIONARY

Data Dictionary containing details of each attribute for every relation of the database is provided below as an Excel spreadsheet.

TABLE 2: DATA DICTIONARY

Customer

schema_name	table_name	column_name	data_type	data_type_ext	nullable	default_value
project	Customer	CustomerID	int	int	N	
project	Customer	FirstName	nvarchar	nvarchar(50)	N	
project	Customer	LastName	nvarchar	nvarchar(50)	N	
project	Customer	Email	nvarchar	nvarchar(50)	N	
project	Customer	Apartment	nvarchar	nvarchar(50)	N	
project	Customer	Street	nvarchar	nvarchar(50)	Y	
project	Customer	City	nvarchar	nvarchar(50)	N	
project	Customer	ZipCode	varchar	varchar(10)	N	
project	Customer	State	nvarchar	nvarchar(50)	N	

Customer_Phone

schema_name	table_name	column_name	data_type	data_type_ext	nullable	default_value
project	Customer_Phone	CustomerID	int	int	N	
project	Customer_Phone	Phone	varchar	varchar(20)	N	

Company

schema_name	table_name	column_name	data_type	data_type_ext	nullable	default_value
project	Company	CompanyID	int	int	N	
project	Company	CompanyName	nvarchar	nvarchar(50)	N	
project	Company	Email	nvarchar	nvarchar(50)	N	
project	Company	Website	nvarchar	nvarchar(50)	N	
project	Company	AccountType	nvarchar	nvarchar(50)	N	

Company_Phone

schema_name	table_name	column_name	data_type	data_type_ext	nullable	default_value
project	Company_Phone	CompanyID	int	int	N	
project	Company_Phone	Phone	varchar	varchar(20)	N	

Designer

schema_name	table_name	column_name	data_type	data_type_ext	nullable	default_value
project	Designer	DesignerID	int	int	N	
project	Designer	FirstName	nvarchar	nvarchar(50)	N	
project	Designer	LastName	nvarchar	nvarchar(50)	N	

Consultation

schema_name	table_name	column_name	data_type	data_type_ext	nullable	default_value
project	Consultation	ConsultationID	int	int	N	
project	Consultation	Date	date	date	N	
project	Consultation	Time	time	time(7)	N	
project	Consultation	Location	nvarchar	nvarchar(50)	N	
project	Consultation	CustomerID	int	int	N	
project	Consultation	DesignerID	int	int	N	

Quote

schema_name	table_name	column_name	data_type	data_type_ext	nullable	default_value
project	Quote	QuoteNo	int	int	N	
project	Quote	Taxes	float	float	N	
project	Quote	DiscountRate	float	float	Y	
project	Quote	Note	nvarchar	nvarchar(200)	Y	
project	Quote	Status	nvarchar	nvarchar(15)	N	
project	Quote	CustomerID	int	int	N	

Unit

schema_name	table_name	column_name	data_type	data_type_ext	nullable	default_value
project	Unit	UnitID	int	int	N	
project	Unit	UnitName	nvarchar	nvarchar(50)	N	
project	Unit	Notes	nvarchar	nvarchar(200)	N	

Includes

schema_name	table_name	column_name	data_type	data_type_ext	nullable	default_value
project	Includes	QuoteNo	int	int	N	
project	Includes	UnitID	int	int	N	
project	Includes	Price	money	money	N	
project	Includes	Quantity	int	int	N	

Installation

schema_name	table_name	column_name	data_type	data_type_ext	nullable	default_value
project	Installation	InstallationNo	int	int	N	
project	Installation	Date	date	date	N	
project	Installation	Time	time	time(7)	N	
project	Installation	Apartment	nvarchar	nvarchar(50)	N	
project	Installation	Street	nvarchar	nvarchar(50)	N	
project	Installation	City	nvarchar	nvarchar(50)	N	
project	Installation	State	nvarchar	nvarchar(50)	N	
project	Installation	ZipCode	varchar	varchar(10)	N	
project	Installation	Country	nvarchar	nvarchar(50)	N	

Installer

schema_name	table_name	column_name	data_type	data_type_ext	nullable	default_value
project	Installer	InstallerID	int	int	N	
project	Installer	FirstName	nvarchar	nvarchar(50)	N	
project	Installer	LastName	nvarchar	nvarchar(50)	N	
project	Installer	InstallationNo	int	int	Y	

Installer_Phone

schema_n	table_nam	column_n	data_type	data_type	nullable	default_va
project	Installer_P	InstallerID	int	int	N	
project	Installer_P	Phone	varchar	varchar(20	N	

DATA WAREHOUSING

Entity Relationship Diagram showing all entities (3 Dimensions and 1 Fact) for the data warehouse is shown below along with the relationships and cardinalities.

FIGURE 4: ER DIAGRAM FOR DATAWAREHOUSE MODEL

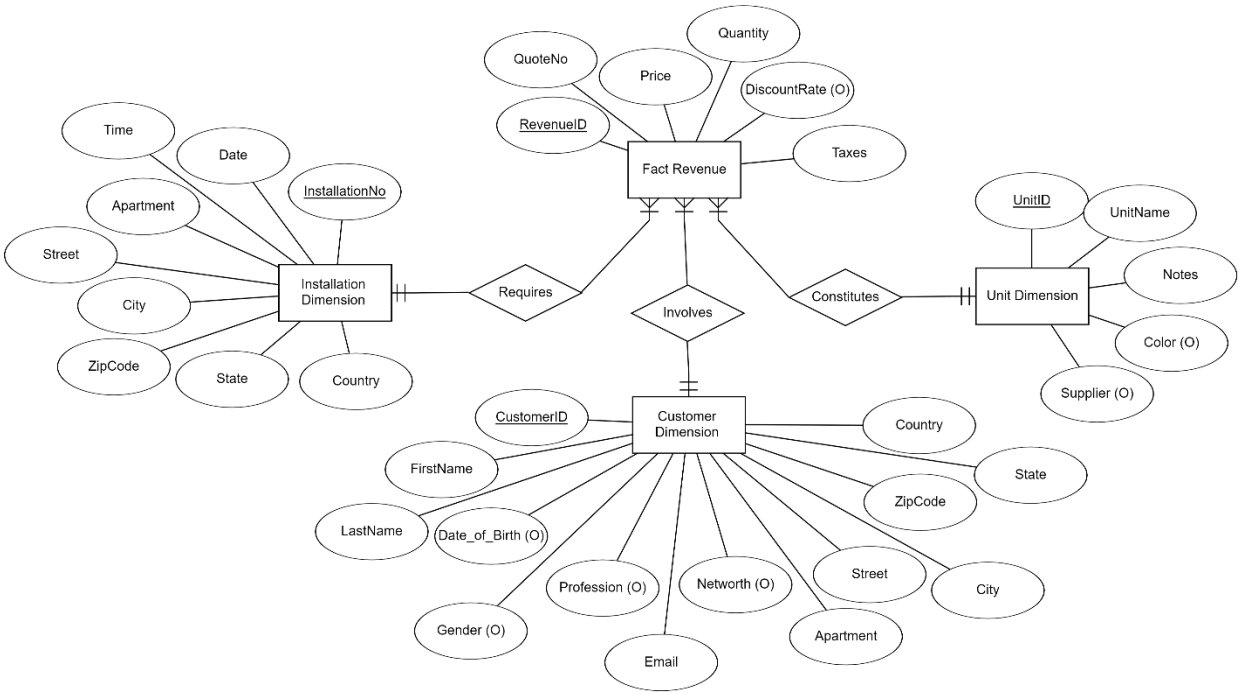


FIGURE 5: STAR SCHEMA FOR DATAWAREHOUSE MODEL

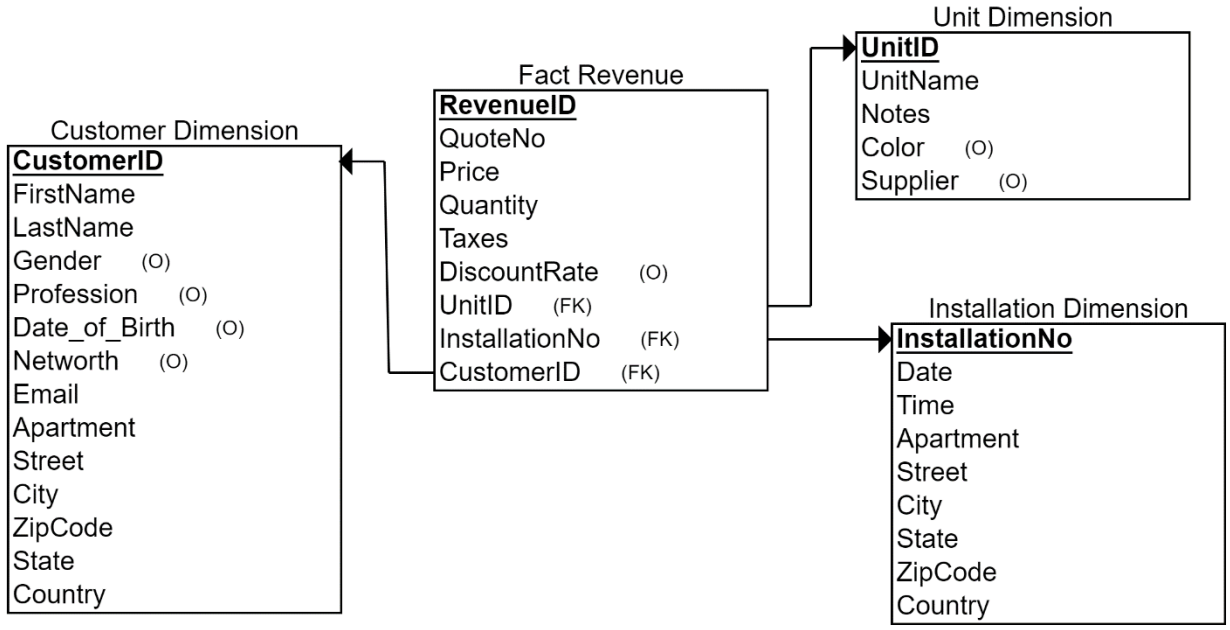
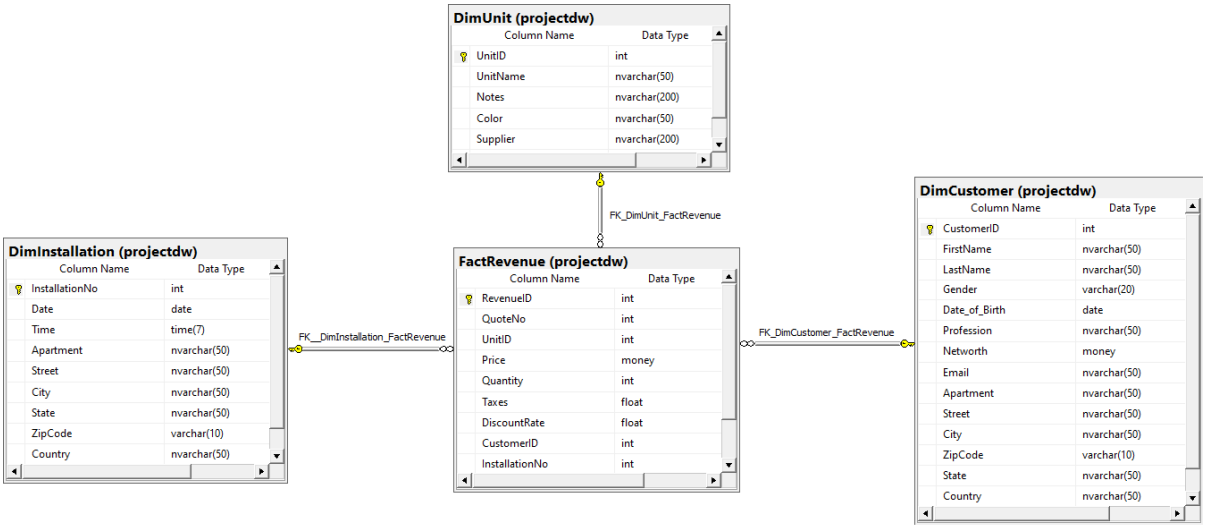


FIGURE 6: DATAWAREHOUSE DIAGRAM



The figures 5 & 6 depicted above, reflect the star schema designed and implemented in the Datawarehouse. Tables for the data warehouse are created under the schema "projectdw" and they have the following attributes:

1. **DimUnit:** UnitID (*Primary Key*), UnitName, Notes, Color (optional), Supplier (Optional)
2. **DimCustomer:** CustomerID (*Primary Key*), FirstName, LastName, Gender (Optional), Date_of_Birth (Optional), Profession (Optional), Networth (Optional), Email, Apartment, Street (Optional), City, ZipCode, State and Country.
3. **DimInstallation:** InstallationNo (*Primary Key*), Date, Time, Apartment, Street, City, State, ZipCode, Country
4. **FactRevenue:** RevenueID (*Primary Key* with Seed = 1 & Increment = 1), QuoteNo, UnitID, Price, Quantity, Taxes, DiscountRate (Optional), CustomerID, InstallationNo

DATAWAREHOUSE – POPULATED

It is evident that most columns in the tables within the data warehouse are sourced from the operational database within the "project" schema. The process of populating these tables is executed via the ETL process, and the code utilized for this purpose can be accessed in the Appendix. Furthermore, a snapshot of the first five rows of data that is loaded into each table within the data warehouse is presented below.

TABLE 3: POPULATED DATAWAREHOUSE TABLES

DimCustomer

CustomerID	FirstName	LastName	Gender	Date_of_Birth	Profession	Networth	Email	Apartment	Street	City	ZipCode	State	Country
2001	Tom	Cruise	NULL	NULL	NULL	NULL	cruise@sag.com	611	Beverly Park	Syracruse	10012	New York	USA
2002	Brad	Pitt	NULL	NULL	NULL	NULL	brad@sag.com	345	Rosewood Ave	Los Angeles	90004	California	USA
2003	Cate	Blanchett	NULL	NULL	NULL	NULL	cate@sag.com	204	Hudson Street	Sydney	2011	New South Wales	Australia
2004	Dwayne	Johnson	NULL	NULL	NULL	NULL	dwayne@sag.com	27	Clarendon St	Beverly Hills	90210	California	USA
2005	Hugh	Jackman	NULL	NULL	NULL	NULL	hugh@sag.com	22	Hickson Rd	Sydney	2000	New South Wales	Australia

DimInstallation

InstallationNo	Date	Time	Apartment	Street	City	State	ZipCode	Country
7001	2022-01-05	09:30:00.0000000	Sky Palace	Rose Street	Los Angeles	California	90001	USA
7002	2022-02-10	12:45:00.0000000	Green Acres	Meadow Lane	Chicago	Illinois	60007	USA
7003	2022-03-15	16:15:00.0000000	Ocean Breeze	Marine Drive	Miami	Florida	33101	USA
7004	2022-04-20	11:00:00.0000000	Fairy Tale	Wonderland Road	New York	New York	10001	USA
7005	2022-05-25	14:30:00.0000000	Royal Garden	Palace Street	San Francisco	California	94101	USA

DimUnit

UnitID	UnitName	Notes	Color	Supplier
6001	Shoe Rack	A beautiful shoe rack with 4 shelves, made with pr...	NULL	NULL
6002	Kitchen Island	A custom kitchen island with granite countertop an...	NULL	NULL
6003	Lamp	A stylish lamp with a silver base and a white shade...	NULL	NULL
6004	TV Stand	A sleek and modern TV stand with a glossy black f...	NULL	NULL
6005	Bookcase	A sturdy and spacious bookcase with adjustable s...	NULL	NULL

FactRevenue

RevenueID	QuoteNo	UnitID	Price	Quantity	Taxes	DiscountRate	CustomerID	InstallationNo
1	5001	6009	75.00	2	0.07	NULL	2005	7001
2	5001	6010	100.00	3	0.07	NULL	2005	7001
3	5001	6076	300.00	5	0.07	NULL	2005	7001
4	5001	6087	50.00	1	0.07	NULL	2005	7001
5	5001	6105	200.00	4	0.07	NULL	2005	7001

It can be observed from the above that certain columns like Net Worth in DimCustomer and Color in DimUnit still have null values. This is because the values for those columns are not available in the operational database. These columns can be pulled into the data warehouse from an external data source, if needed.

ETHICAL & PRIVACY ISSUES

It is crucial to consider potential ethical and privacy issues when collecting and storing data. In relation to the database created in this project, the following are some of the issues that should be considered:

Data Privacy: Personal information such as names, addresses, phone numbers, and email addresses are sensitive data that must be safeguarded. It is crucial to guarantee that the information is securely stored, and access is restricted to authorized personnel only.

Multivalued Attributes: The phone attribute in the Customer and Company tables is a multivalued attribute that could raise privacy concerns. Collecting multiple phone numbers for a customer or company without their explicit consent could be viewed as intrusive.

Composite Attributes: The address attribute in the Customer and Installation tables is a composite attribute made up of multiple components such as street, city, state, and zip code. Gathering detailed address information could result in potential privacy concerns if not handled appropriately.

DW Attributes: Certain attributes in the data warehouse are sensitive. For example, Net Worth in DimCustomer. Collecting such information without a valid reason or the individual's explicit consent could raise privacy concerns.

Data Security: Data breaches and cyber-attacks can result in the loss or theft of sensitive information. To prevent this, the database should be secured with strong passwords and encryption, and regular security audits should be conducted.

Data Ownership: It is essential to clarify who owns the data and how it can be used. This includes obtaining permission to use data collected by third parties.

Data Accuracy and Integrity: It is essential to ensure that the data collected is accurate and reliable. This can be achieved by implementing data validation rules, data cleaning procedures, and regular backups.

APPENDIX

The following code is used to populate Data Warehouse tables.

Populating DimUnit

```
INSERT INTO projectdw.DimUnit (UnitID, UnitName, Notes)

    SELECT UnitID, UnitName, Notes

    FROM project.Unit;
```

Populating DimCustomer

```
INSERT INTO projectdw.DimCustomer ( CustomerID, FirstName, LastName, Email, Apartment,
Street, City, ZipCode, State, Country )

    SELECT CustomerID, FirstName, LastName, Email, Apartment, Street, City, ZipCode,
    State, Country

    FROM project. Customer;
```

Populating DimInstallation

```
INSERT INTO projectdw.DimInstallation ( InstallationNo, Date, Time, Apartment, Street,
        City, State, ZipCode, Country )
       SELECT InstallationNo, Date, Time, Apartment, Street,
        City, State, ZipCode, Country
        FROM project. Installation;
```

Populating FactRevenue

```
INSERT INTO projectdw.FactRevenue ( QuoteNo, UnitID, Price, Quantity, Taxes, DiscountRate,
CustomerID, CompanyID, InstallerID, InstallationNo )
       SELECT inc.QuoteNo, inc.UnitID, inc.Price, inc.Quantity, q.Taxes, q.DiscountRate,
       q.CustomerID, cus.CompanyID, iler.InstallerID, ins.InstallationNo
       FROM project.Includes as inc JOIN project.Installation as ins
       ON ins.QuoteNo = inc.QuoteNo JOIN project.Quote as q
       ON q.QuoteNo = ins.QuoteNo JOIN project.Customer as cus
       ON q.CustomerID = cus.CustomerID JOIN project.Installer as iler
       ON ins.InstallationNo = iler.InstallationNo;
```