CONESTOGA
Connect Life and Learning

COURSE CODE: 1372

PROGRAM CODE- PROG8651

SECTION – 10

ASSIGMENT TITTLE: Task 5 – Data Warehouse (A)

STUDENT NAME: VAISHNAVI GOPAL WADHWA

DATE: 08 DEC 2023

Tasks:

Referring to narrative section 6.4.1.1. “Orders Database” (Case - CBR - 6.4.1.1. – Orders Database.pdf). The client Organization wishes to better understand shipping performance based on the observable variance in shipping estimation and the actual time of arrival. To perform analysis on this data, they require a data warehouse (data mart) that will help identify critical data regarding shipping data as gathered from the Orders Database).

1) Data Mart

From: Matt Kozi MKozi@cbr.org
To: You
Subject: New Data Mart
Hello.
I thought it would be helpful to send you a summary of what we discussed during the last meeting. The higher-ups want to kick-off a new initiative that will have our Analytics team evaluate shipping performance so that we can better understand which shipping methods are falling short on delivery Time expectations and to what regions/cities this is most commonly occurring. However, I do agree with what you stated during the meeting, that giving the Analyst’s direct access to the Orders database was an unnecessary risk, and that providing them with a cleansed data mart would be the best choice. We had some additional meetings, and everyone agrees that your data mart idea would ensure that we wouldn’t need to add more users to the Orders database, and would allow us to retain full control over what data we want the Analysts to access.

As it stands, I’ll need you to set up and document the new data mart for data export so that we can have some of our analyst’s start to evaluate shipping performance to date. You mentioned that you needed to first document a [data warehouse] star schema layout so you’d have a reliable understanding of what data you’d need to export and after discussing with the DBA, the following Information was identified for the design:

Fact table – ShippingEfficiency Table which will include all relevant connections from the associated dimensional tables along with a new fact table measure which will calculate the difference in date between estdateshipped and actualdateshipped

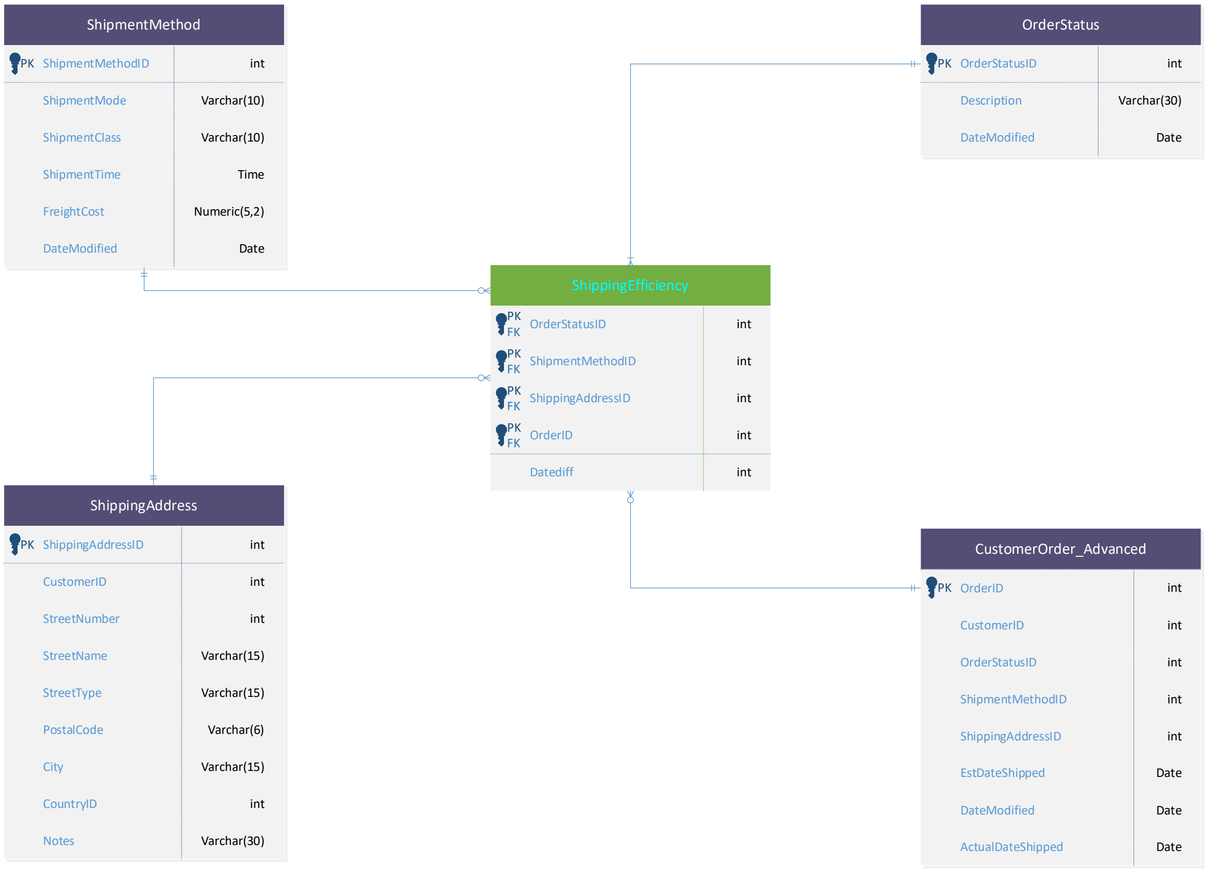
Dimensional table – CustomerOrder, ShipmentMethod, OrderStatus, ShippingAddress.

o I spoke to the legal department, they notified me that we should avoid including any cost or pricing data from the CustomerOrder table, so for the Dimensional table you’re setting up, skip those attributes.

o Also, let’s change the name of the table slightly so we don’t get it confused with the regular CustomerOrder table in the future. – CustomerOrderA

Can you build it in VISIO or something? I know it is a really basic diagram, just want it to look professional so we don’t have to go back and rework things…Send me the PDF when you’re done.

Matt Kozi
Solutions Architect, CBR
mKozi@cbr.org



ShipmentMethod	
PK ShipmentMethodID	int
ShipmentMode	Varchar(10)
ShipmentClass	Varchar(10)
ShipmentTime	Time
FreightCost	Numeric(5,2)
DateModified	Date

OrderStatus	
PK OrderStatusID	int
Description	Varchar(30)
DateModified	Date

ShippingEfficiency	
PK FK OrderStatusID	int
PK FK ShipmentMethodID	int
PK FK ShippingAddressID	int
PK FK OrderID	int
Datediff	int

ShippingAddress	
PK ShippingAddressID	int
CustomerID	int
StreetNumber	int
StreetName	Varchar(15)
StreetType	Varchar(15)
PostalCode	Varchar(6)
City	Varchar(15)
CountryID	int
Notes	Varchar(30)

CustomerOrder_Advanced	
PK OrderID	int
CustomerID	int
OrderStatusID	int
ShipmentMethodID	int
ShippingAddressID	int
EstDateShipped	Date
DateModified	Date
ActualDateShipped	Date

2) Table Scripts
From: Matt Kozi MKozi@cbr.org
To: You
Subject: URGENT: Need those table scripts
Now that you've gotten the model built, I need you to move onto writing out the script to build the tables for our new data mart.
P.S. Also, I really hate to bother you with this again, but don't forget to set the composite key for the fact table, last time you skipped this and it became a big mess… Additionally, while you did set the constraints last time you forgot to ensure you didn't allow updates or deletes through the connections.
Forward the .SQL file to me so I can pass it off to our DBA.
Matt Kozi
Solutions Architect, CBR
mKozi@cbr.org

- ➔ IN SQL

```
-- TASK-5 QUESTION-2 VAISHNAVI GOPAL WADHWA 8939083

USE [TASK-5]
GO
```

```
CREATE TABLE ShipmentMethod (
       ShipmentMethodID     Integer                   NOT NULL,
       ShipmentMode         Varchar(10)               NOT NULL,
       ShipmentClass        Varchar(10)               NOT NULL,
       ShipmentTime         Time                      NOT NULL,
       FreightCost          Numeric(5,2)              NOT NULL,
       DateModified         Date                      NOT NULL,
       CONSTRAINT           ShipmentMethod_PK         PRIMARY KEY (ShipmentMethodID)
       );


CREATE TABLE ShipmentAddress (
       ShipmentAddressID   Integer                    NOT NULL,
       CustomerID           Integer                   NOT NULL,
       StreetNumber         Integer                   NOT NULL,
       StreetName           Varchar(15)               NOT NULL,
       StreetType           Varchar(15)               NOT NULL,
       PostalCode           Varchar(6)                NOT NULL,
       City                 Varchar(15)               NOT NULL,
       CountryID            Integer                   NOT NULL,
       Notes                Varchar(30)               NOT NULL,
       CONSTRAINT           ShipmentAddress_PK        PRIMARY KEY (ShipmentAddressID)
       );


CREATE TABLE CustomerOrder_Advanced (
       OrderID              Integer                   NOT NULL,
       CustomerID           Integer                   NOT NULL,
       OrderStatusID        Integer                   NOT NULL,
       ShipmentMethodID     Integer                   NOT NULL,
       ShipmentAddressID    Integer                   NOT NULL,
       EstDateShipped       Date                      NOT NULL,
       DateModified         Date                      NOT NULL,
       ActualDateShipped    Date                      NOT NULL,
       CONSTRAINT           CustomerOrder_Advanced_PK PRIMARY KEY (OrderID)
       );


CREATE TABLE OrderStatus (
       OrderStatusID        Integer                   NOT NULL,
       [Description]        Varchar(30)               NOT NULL,
       DateModified         Date                      NOT NULL,
       CONSTRAINT           OrderStatus_PK            PRIMARY KEY (OrderStatusID)
       );

CREATE TABLE ShippingEfficiency (
       OrderStatusID        Integer                          NOT NULL,
       ShipmentMethodID     Integer                          NOT NULL,
       ShipmentAddressID    Integer                          NOT NULL,
       OrderID              Integer                          NOT NULL,
       [Datediff]           Date                             NOT NULL,
       CONSTRAINT           ShippingEfficiency_PK
       PRIMARY KEY(OrderStatusID, ShipmentMethodID, ShipmentAddressID, OrderID),
       CONSTRAINT           ShipmentMethod_ShippingEfficiency_FK
       FOREIGN KEY(ShipmentMethodID)
                                          REFERENCES  ShipmentMethod(ShipmentMethodID)
                                          ON UPDATE CASCADE
                                          ON DELETE CASCADE,
       CONSTRAINT           ShipmentAddress_ShippingEfficiency_FK
       FOREIGN KEY(ShipmentAddressID)
                                          REFERENCES ShipmentAddress(ShipmentAddressID)
                                          ON UPDATE CASCADE
```

```
                                         ON DELETE CASCADE,
       CONSTRAINT          CustomerOrder_Advanced_ShippingEfficiency_FK
       FOREIGN KEY(OrderID)
                                         REFERENCES  CustomerOrder_Advanced(OrderID)
                                         ON UPDATE CASCADE
                                         ON DELETE CASCADE,
       CONSTRAINT          OrderStatus_ShippingEfficiency_FK
       FOREIGN KEY(OrderStatusID)
                                         REFERENCES  OrderStatus(OrderStatusID)
                                         ON UPDATE CASCADE
                                         ON DELETE CASCADE
       );
```

3) Explain why referential integrity isn't a major focus of the FACT table within a data mart based on the data warehouse star schema? What issues would arise if update and delete were enforced within the key CONSTRAINT(s)?

This answer can be included on the PDF Visio file (PDF) of the dimensional table.

→ In a data warehouse star schema, the FACT table contains the quantitative data (usually numerical values or measures) and is surrounded by DIMENSION tables that provide context and descriptive information. The referential integrity between the FACT table and the DIMENSION tables is crucial for maintaining data consistency and accuracy. However, enforcing referential integrity within the FACT table itself is often not a major focus for several reasons:

1. Volume of Data:- FACT tables in data marts tend to be large and store a significant amount of historical data. Enforcing referential integrity on such large datasets can be resource-intensive and impact performance.

2. Performance Concerns:- Frequent updates and deletes on a FACT table can be performance-intensive, especially when dealing with large volumes of data. Enforcing referential integrity within the FACT table may lead to performance degradation, affecting the overall responsiveness of the data mart.

3. Analytical Nature:- FACT tables are primarily used for analytical queries and reporting, where read operations significantly outweigh write operations. Users are more interested in analysing historical data than in making updates or deletions to existing records

4. Aggregated Data:- FACT tables often store aggregated data, computed from various sources. In such cases, enforcing referential integrity at the detailed level might not be as relevant, as the aggregated values are more critical for analytical purposes.

If referential integrity constraints were enforced within the FACT table, issues could arise due to the nature of analytical data warehouses and the star schema:

1. Performance Impact:- Enforcing referential integrity through constraints like foreign keys in the FACT table could slow down data loading processes, especially when dealing with large volumes of data.

2. Complexity of Operations:- Updates and deletes on a FACT table with referential integrity constraints can be complex and may require cascading updates or deletes across multiple tables. This complexity can make maintenance more challenging.

3. Historical Data Challenges:- Enforcing referential integrity may become problematic when dealing with historical data that has been aggregated or modified over time. Updates or deletes in the FACT table could have cascading effects on historical reporting.

To conclude with, while referential integrity is crucial between the FACT table and DIMENSION tables, enforcing it within the FACT table itself may not be a primary focus due to performance considerations and the analytical nature of data warehousing. Instead, the emphasis is typically on maintaining consistency and relationships between the FACT table and associated DIMENSION tables.

4) Sample Data
From: Matt Kozi MKozi@cbr.org
To: You
Subject: Wrapping up
Can you provide a sample script for the data? Just 5-lines of data per table would suffice. That way we can test this out before we pass it along. I know, I know, not the most exciting work, but hey, look at it this way. You're almost done the project!
Btw, you can just use data from the existing orders database you originally built a few months back if that's easier, else, just make some new data by hand or through generatedata.com. Doesn't matter to me.
Just toss these commands to the end of the .SQL file from earlier, submit it and head home, thanks.
Matt Kozi
Solutions Architect, CBR
mKozi@cbr.org

➔ In SQL

```
-- TASK-5 QUESTION-4 VAISHNAVI GOPAL WADHWA 8939083

--Insert into ShipmentMethod
INSERT INTO
ShipmentMethod(ShipmentMethodID,ShipmentMode,ShipmentClass,ShipmentTime,FreightCost,DateModified)
VALUES
(89390831,'PickUp','Overnight','07:00:00',230.00,'03/12/2023'),
(89390832,'Air','Priority','09:30:00',786.00,'07/12/2023'),
(89390833,'Ground','Standard','10:30:00',243.00,'09/12/2023'),
(89390834,'PickUp','Priority','08:30:00',269.00,'12/12/2023'),
(89390835,'Air','Standard','12:30:00',289.00,'04/12/2023');

--Insert into Orderstatus
INSERT INTO OrderStatus(OrderStatusID,[Description],DateModified)
VALUES
(5001,'Delivered','11/11/2023'),
(5002,'In Transit','12/11/2023'),
(5003,'Processing Order','01/11/2023'),
(5004,'Delayed','12/11/2023'),
(5005,'Preparing to ship','12/11/2023');

--Insert into ShippingAddress
```

```
INSERT INTO
ShipmentAddress(ShipmentAddressID,CustomerID,StreetNumber,StreetName,StreetType,Postal
Code,City,CountryID,Notes)
VALUES
(8939083,8000040,220,'Ira Needles','Road','N2N0C6','Kitchener',548,'Fragile product'),
(8939084,8000041,221,'River Manchest','Road','N2N1K1','Waterloo',577,'Handle with
Care'),
(8939085,8000042,222,'Weber','Bay','N2B1K6','Kitechener',578,'Expensive Kids Toys'),
(8939086,8000043,223,'Public square','Street','N2C1N6','Cambridge',579,'Fragile
product'),
(8939087,8000044,222,'Highlandwest','Road','N2N3K6','Guelph',580,'Expensive Kids
Toys');

--Insert into CustomerOrder_Advanced
INSERT INTO
CustomerOrder_Advanced(OrderID,CustomerID,OrderStatusID,ShipmentMethodID,ShipmentAddre
ssID,EstDateShipped,ActualDateShipped,DateModified)
Values
(7000051,8000040,5001,89390831,8939083,'01/12/2023','11/12/2023','10/12/2023'),
(7000052,8000041,5002,89390832,8939084,'02/12/2023','12/12/2023','01/12/2023'),
(7000053,8000042,5003,89390833,8939085,'03/12/2023','03/12/2023','02/12/2023'),
(7000054,8000043,5004,89390834,8939086,'04/12/2023','04/12/2023','03/12/2023'),
(7000055,8000044,5005,89390835,8939087,'05/12/2023','05/12/2023','04/12/2023');

--Insert into ShippingEfficiency
INSERT INTO ShippingEfficiency
(OrderID,OrderStatusID,ShipmentMethodID,ShipmentAddressID,[Datediff])
Values
(7000051,5001,89390831,8939083,3),
(7000052,5002,89390832,8939084,4),
(7000053,5003,89390833,8939085,5),
(7000054,5004,89390834,8939086,6),
(7000055,5005,89390835,8939087,7);
```