**IST 722 Data Warehouse**

**Prof. Humayun Khan**

**A picture containing indoor

Description automatically generated**

**Company Introduction**

FudgeMart, Inc. is a conglomerate that currently manages two databases for its subsidiaries Fudgemart and FudgeFlix. FudgeFlix is an at-home movie service, Fudgemart provides a wide range of products to its customers.

**Project Description**

Perform ETL for external sources into the data warehouse to visualize the data to get insights.

**Business Case**

The Data Warehouse is being created to allow for the seamless integration of two databases into a single data warehouse that is interoperable with Business Intelligence Tools. This is done primarily to provide Subject-Orientation and Non-volatility Integration, hence satisfying the demand for a Data Warehouse for Customer Analysis.

**Business Requirements**

**Customer Analysis** – Having details about the Customer, Products and orders made by customers our team is trying to build reports of existing data and were able to get sales insights of our existing customers. Furthermore, the team was dedicated to generating a trend line and predict the sales of the company. The data for this purpose is to be integrated into the data warehouse and transformed according to the business requirements

**Functional Requirements**

External Sources to store the transactions of orders. A staging area to store the transformed data from external sources. A data warehouse to perform analytics on the transformed data and create views that help to visualize data using Power BI.

**Business Process for integration implementation**

We have selected **Customer Analysis** and **Sales Facts** for integration across both FudgeMart and Fudgeflix.

**Facts​:**

1. **Customer Product Ratings** – The details of reviews given by customer on the products they have bought or ordered

**Dimensions:**

1. **Customers** – Customer information such as name, city, state and zip
2. **Products** – Products available or sold in the organization with department details

**Design**

**Bus Matrix**

Screenshot

**Data Dictionary**

Below is the list of all attributes we used in our database.

Screenshot

**Dimensional Modeling and Schema**

Diagram

Description automatically generated

**ETL using SSIS**

According to the ETL process we are Extracting the data from the following sources

1. ExternalSources
   1. Date Dimension
2. FudgeMart
   1. Customers Dimension
   2. Products Dimension
   3. Customer Product Reviews Fact
3. FudgeFlix
   1. Accounts Dimension
   2. Titles Dimension
   3. AccountTitle Rating Fact

From the two databases i.e., FudgeMart and FudgeFlix we are integrating the data into one table in our Data warehouse i.e.

1. DimCustomers
2. DimProducts
3. FactCusomterProductReviews
4. DimDate

Below are the screenshots at each phase i.e., Extraction and Loading. Yes, transformation is done in between the Extraction and Loading.

**Staging**

1. **Staging FudgeMart**

Diagram

Description automatically generated

Graphical user interface

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

1. **Staging Data from FudgeFlix**

Graphical user interface, application

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidence

1. **Loading stage data to data warehouse from FudgeMart**

A picture containing timeline

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, chat or text message

Description automatically generated

1. **Loading stage data to data warehouse from Fudgeflix**

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, Word

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, text, application, Word

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

**Business Intelligence**

The various reports and dashboard we developed using the cube, facts and dimensions.

Screenshots

**SQL Scripts**

**Generated SQL**

/\*\*\*\*\*\* Object: Database ist722\_hhkhan\_ca2\_dw Script Date: 8/14/22 1:58:43 PM \*\*\*\*\*\*/

/\*

Kimball Group, The Microsoft Data Warehouse Toolkit

Generate a database from the datamodel worksheet, version: 4

You can use this Excel workbook as a data modeling tool during the logical design phase of your project.

As discussed in the book, it is in some ways preferable to a real data modeling tool during the initial design.

We expect you to move away from this spreadsheet and into a real modeling tool during the physical design phase.

The authors provide this macro so that the spreadsheet isn't dead-end. You can 'import' into your

data modeling tool by generating a database using this script, then reverse-engineering that database into

your tool.

Uncomment the next lines if you want to drop and create the database

\*/

/\*

DROP DATABASE ist722\_hhkhan\_ca2\_dw

GO

CREATE DATABASE ist722\_hhkhan\_ca2\_dw

GO

ALTER DATABASE ist722\_hhkhan\_ca2\_dw

SET RECOVERY SIMPLE

GO

\*/

/\*

USE ist722\_ksulfika\_dw

;

DROP SCHEMA fp

;

CREATE SCHEMA fp

;

\*/

IF EXISTS (SELECT Name from sys.extended\_properties where Name = 'Description')

EXEC sys.sp\_dropextendedproperty @name = 'Description'

EXEC sys.sp\_addextendedproperty @name = 'Description', @value = 'Default description - you should change this.'

;

/\* Drop table fp.DimDate \*/

IF EXISTS (SELECT \* FROM dbo.sysobjects WHERE id = OBJECT\_ID(N'fp.DimDate') AND OBJECTPROPERTY(id, N'IsUserTable') = 1)

DROP TABLE fp.DimDate

;

/\* Create table fp.DimDate \*/

CREATE TABLE fp.DimDate (

[DateKey] int NOT NULL

, [Date] date NULL

, [FullDateUSA] nchar(11) NOT NULL

, [DayOfWeek] tinyint NOT NULL

, [DayName] nchar(10) NOT NULL

, [DayOfMonth] tinyint NOT NULL

, [DayOfYear] smallint NOT NULL

, [WeekOfYear] tinyint NOT NULL

, [MonthName] nchar(10) NOT NULL

, [MonthOfYear] tinyint NOT NULL

, [Quarter] tinyint NOT NULL

, [QuarterName] nchar(10) NOT NULL

, [Year] smallint NOT NULL

, [IsWeekday] bit DEFAULT 0 NOT NULL

, CONSTRAINT [PK\_fp.DimDate] PRIMARY KEY CLUSTERED

( [DateKey] )

) ON [PRIMARY]

;

INSERT INTO fp.DimDate (DateKey, Date, FullDateUSA, DayOfWeek, DayName, DayOfMonth, DayOfYear, WeekOfYear, MonthName, MonthOfYear, Quarter, QuarterName, Year, IsWeekday)

VALUES (-1, '', 'Unk date', 0, 'Unk date', 0, 0, 0, 'Unk month', 0, 0, 'Unk qtr', 0, 0)

;

/\* Drop table fp.DimCustomers \*/

IF EXISTS (SELECT \* FROM dbo.sysobjects WHERE id = OBJECT\_ID(N'fp.DimCustomers') AND OBJECTPROPERTY(id, N'IsUserTable') = 1)

DROP TABLE fp.DimCustomers

;

/\* Create table fp.DimCustomers \*/

CREATE TABLE fp.DimCustomers (

[customerID] int IDENTITY NOT NULL

, [customer\_firstname] nvarchar(30) NOT NULL

, [customer\_lastname] nvarchar(30) NOT NULL

, [customer\_city] nvarchar(15) NOT NULL

, [cutomer\_state] nchar(15) NOT NULL

, [customer\_zip] nvarchar(15) NOT NULL

, [RowIsCurrent] nchar(1) NULL

, [RowStartDate] datetime NULL

, [RowEndDate] datetime DEFAULT '12/31/9999' NULL

, [RowChangeReason] nvarchar(200) NULL

, CONSTRAINT [PK\_fp.DimCustomers] PRIMARY KEY CLUSTERED

( [customerID] )

) ON [PRIMARY]

;

SET IDENTITY\_INSERT fp.DimCustomers ON

;

INSERT INTO fp.DimCustomers

(customerID,

customer\_firstname,

customer\_lastname,

customer\_city,

cutomer\_state,

customer\_zip,

RowIsCurrent, RowStartDate, RowEndDate, RowChangeReason)

VALUES ( -1,'Unk Attribute2', '', '', '', '', 'Y', '12/31/1899', '12/31/9999', 'N/A')

;

--SET IDENTITY\_INSERT fp.DimCustomers OFF

--;

/\* Drop table fp.DimProducts \*/

IF EXISTS (SELECT \* FROM dbo.sysobjects WHERE id = OBJECT\_ID(N'fp.DimProducts') AND OBJECTPROPERTY(id, N'IsUserTable') = 1)

DROP TABLE fp.DimProducts

;

/\* Create table fp.DimProducts \*/

CREATE TABLE fp.DimProducts (

[ProductID] int IDENTITY NOT NULL

, [product\_department] nvarchar(20) NOT NULL

, [product\_name] nvarchar(50) NOT NULL

, [product\_retail\_price] money NULL

, [product\_wholesale\_price] float(20) NULL

, [product\_is\_active] bit NULL

, [RowIsCurrent] nchar(1) NOT NULL

, [RowStartDate] datetime NOT NULL

, [RowEndDate] datetime DEFAULT '12/31/9999' NOT NULL

, [RowChangeReason] nvarchar(200) NOT NULL

, CONSTRAINT [PK\_fp.DimProducts] PRIMARY KEY CLUSTERED

( [ProductID] )

) ON [PRIMARY]

;

--SET IDENTITY\_INSERT fp.DimProducts ON

--;

INSERT INTO fp.DimProducts (ProductID, product\_department, product\_name, product\_retail\_price, product\_wholesale\_price, product\_is\_active, RowIsCurrent, RowStartDate, RowEndDate, RowChangeReason)

VALUES (-1, 'Unk Attribute1', 'Unk Attribute2', NULL, NULL, NULL, 'Y', '12/31/1899', '12/31/9999', 'N/A')

;

--SET IDENTITY\_INSERT fp.DimProducts OFF

--;

/\* Drop table fp.DimOrders \*/

IF EXISTS (SELECT \* FROM dbo.sysobjects WHERE id = OBJECT\_ID(N'fp.DimOrders') AND OBJECTPROPERTY(id, N'IsUserTable') = 1)

DROP TABLE fp.DimOrders

;

/\* Create table fp.DimOrders \*/

CREATE TABLE fp.DimOrders (

[OrderID] int IDENTITY NOT NULL

, [order\_date] datetime NOT NULL

, [shipped\_date] datetime NULL

, [customer\_id] int NOT NULL

, [RowIsCurrent] nchar(1) NOT NULL

, [RowStartDate] datetime NOT NULL

, [RowEndDate] datetime DEFAULT '12/31/9999' NOT NULL

, [RowChangeReason] nvarchar(200) NOT NULL

, CONSTRAINT [PK\_fp.DimOrders] PRIMARY KEY CLUSTERED

( [OrderID] )

) ON [PRIMARY]

;

---SET IDENTITY\_INSERT fp.DimOrders ON

--;

INSERT INTO fp.DimOrders (OrderID, order\_date, shipped\_date, customer\_id, RowIsCurrent, RowStartDate, RowEndDate, RowChangeReason)

VALUES (-1, '-1', 'Unk Attribute1', 'UnkAttribute2', 'Y', '2009-01-01 00:00:00.000', '2009-01-01 00:00:00.000', 'N/A')

;

--SET IDENTITY\_INSERT fp.DimOrders OFF

--;

/\* Drop table fp.DimProductReviews \*/

IF EXISTS (SELECT \* FROM dbo.sysobjects WHERE id = OBJECT\_ID(N'fp.DimProductReviews') AND OBJECTPROPERTY(id, N'IsUserTable') = 1)

DROP TABLE fp.DimProductReviews

;

/\* Create table fp.DimProductReviews \*/

CREATE TABLE fp.DimProductReviews (

[customer\_review\_ID] int IDENTITY NOT NULL

, [customer\_id] int NOT NULL

, [product\_id] int NOT NULL

, [review\_stars] int NOT NULL

, [RowIsCurrent] nchar(1) NULL

, [RowStartDate] datetime NULL

, [RowEndDate] datetime DEFAULT '12/31/9999' NULL

, [RowChangeReason] nvarchar(200) NULL

, CONSTRAINT [PK\_fp.DimProductReviews] PRIMARY KEY CLUSTERED

( [customer\_review\_ID] )

) ON [PRIMARY]

;

SET IDENTITY\_INSERT fp.DimProductReviews ON

;

INSERT INTO fp.DimProductReviews (customer\_review\_ID, customer\_id, product\_id, review\_stars, RowIsCurrent, RowStartDate, RowEndDate, RowChangeReason)

VALUES (-1, -1, 'Unk Attribute1', 'Unk Attribute2', 'Y', '12/31/1899', '12/31/9999', 'N/A')

;

SET IDENTITY\_INSERT fp.DimProductReviews OFF

;

ALTER TABLE fp.Order\_details ADD CONSTRAINT

FK\_fp\_Order\_details\_order\_qty FOREIGN KEY

(

order\_qty

) REFERENCES DimAudit

( AuditKey )

ON UPDATE NO ACTION

ON DELETE NO ACTION

;