CS 6400 Database Project:
Will-Mart Sales System (WiMSS)

Summer 2021

Project Overview

The purpose of this project is to analyze, specify, design, and implement a reporting system for Will-Mart, a US-based furniture store. The project will proceed in three phases as outlined in the methodology for database development: Analysis & Specification; Design; and Implementation & Testing. The system will be implemented using a Database Management System (DBMS) that supports standard SQL queries.

Will-Mart Sales System

Will-Mart is a furniture store that has stores throughout the United States. Your team has been asked to design and build a prototype sales reporting system for Will-Mart. This section describes in detail the requirements for the Will-Mart Sales System (WiMSS).

What makes WiMSS different is that it is meant to import and aggregate sales data. Unlike *transactional* databases which are generally designed to record repetitive day-to-day business transactions (e.g., point of sale, buy and sell stock orders, online shopping carts, etc.), WiMSS is meant for reporting and analysis over millions of records to support enterprise-wide decision making. As an example, a large online merchant like amazon.com or bestbuy.com relies on a transactional (also called *operational*) database system for recording customer orders and payments in real time. A data analyst tasked with generating a report that compares sales of a certain product among the different regions of the United States will typically query a reporting database for the report instead of accessing the transactional databases directly. There are several reasons for this: the system can store data from multiple transactional databases in a consolidated form, its schema is designed to support complex queries aggregating millions of rows, and queries do not impact the performance of the transactional database which must support high transaction throughput and availability.

For this project, you will design the database schema for WiMSS and attach it to a rudimentary user interface. ***For this project you should create a normalized schema with as little redundancy as possible***. You also need not be concerned with the details of the transactional databases that we assume exist to support the point-of-sale system at each Will-Mart store. Instead, you will design the schema to support a consolidated view of the products offered and sold in all Will-Mart stores across the country. What follows is a description of the requirements for the system in terms of what information must be stored to support a set of reports defined by the Will-Mart executive team.

Data Requirements

WiMSS maintains information about each *store*, including a unique *store number*, the store's *phone number*, and the store's *street address*. (You do not need to store individual components of the address, such as street number, prefix, etc., as the addresses will be unparsed.) Some stores are *"Will's Grand Showcase"* stores, which have in-store product demonstration and delivery/assembly services available.

WiMSS should also maintain information about each store's *city*, including the *city name*, the *state* in which the city is located, and the *population* of the city. City names are not duplicated within a state, so you can assume they are unique to each state. A city may have multiple stores.

WiMSS contains information about every *product* for sale at Will-Mart stores. Products have a numeric unique identifier (*PID*), similar to a UPC barcode, as well as the *name* of the product. Assume that all products are available and sold at all stores—that is, there is no need to specify that a certain product is only available at a certain store.

Each product is made by a *manufacturer,* which has a unique *name.* It is possible that multiple products are made by the same manufacturer. It is also possible for a manufacturer to have been entered into the system but currently offers no products for sale at Will-Mart.

A product must be assigned to one or more *categories*. Each category has a *name*, which we assume to be unique. A category may or may not have products associated with it.

Every product has a retail *price*. The retail price is in effect for all sales transactions when the product is not discounted. WiMSS maintains the *discount date* and *discount price* of any product that goes on discount. If a product is discounted for multiple days in a row, then a record is stored for each day. It is possible that the same product is discounted multiple times (i.e., different days) with different prices. **If a product is discounted, it is for the same price in all stores—i.e., stores are not allowed to discount items independently or have store-specific discount prices.**

WiMSS stores information about which products are *sold*, including the *store* where it is sold, the *date* of the sale, and the *quantity* of the product purchased on that day. The total amount of sales is not stored explicitly but can be derived based on the date purchased and the quantity of items sold, as individual item prices can be determined from either the retail price or the discount price if one is in effect. For reporting purposes sales tax values are ignored. The system is not required to record the details of individual sales transactions, such as which products were sold together.

Will-Mart also has *special savings days* where all sales on that day are given a *percent discount*. This might be due to a "no sales tax" promotion where the discount is intended to offset any sales tax, a holiday, or a marketing promotion. The special savings is applied to an entire purchase, regardless of any discount at the product level. You will need to store these dates and their corresponding discount amount in the database and ensure any revenue reports properly calculate the impact of any special savings days.

The Will-Mart executive team understands that some details provided here may not be utilized in the reports they initially request but would like to have that data stored in the event they decided to enhance the system with more reports, or for any ad-hoc querying after you have built the system.

An important item to consider is that despite a "date" being a relatively simple piece of data, it may be optimal to treat dates as a set of values distinct from discount information, sale information, and special savings day information, as this helps ensure better consistency in your database.

Will-Mart's DBAs are working on a data extract for the system, but it will be approximately one to two months before it can be made available. You will need to ensure that your schema design matches the data as described here so that any transformation prior to loading is kept to a minimum. Since this is a proof-of-concept for WiMSS, you do not need to worry about any additional data loads.

WiMSS User Interface

All of your reports will be accessible from a "dashboard" UI that must be developed. All access must be controlled via a login screen to the system, utilizing a *username* and *password*. As this is a prototype version of the system, it is acceptable to store passwords as plaintext in the database. For the first version of this system, all user maintenance will be done manually by the DBA in the system's database, and you will be given an initial listing of users, passwords, role assignments, and store assignments as part of the data load. Your login screen should handle any input errors (such as invalid username or password) with appropriate error messages to the user.

The system will initially have three kinds of users: *read-only corporate* users, who can only view reports for all stores, *store managers,* who can only view reporting data for the store or stores that they manage and have no other permissions, and *marketing* users, who can view reports for all stores and are allowed to change certain values in the system (which will be described further). It is possible that a store manager user may not have any stores assigned to them in the system, and it is possible that a store is not associated with any store manager users.

There should be a main menu screen which can be used to access all functionality of the system that has been or will be described in this specification. On this main menu, the following statistics should be displayed along with any buttons/links to reports or functionalities: the count of stores, count of Grand Showcase stores, if applicable, the count of stores the user can view (that means this should not be displayed for read-only or marketing users!), count of manufacturers, count of products, and count of special savings days. These statistics will be used to determine if the data loaded in WiMSS is generally accurate before viewing reports. For store manager users, the count of stores they can view should link to a second page/screen that lists the details of those stores: store number, street address, and phone number, and if it is or is not a Grand Showcase store.

Your UI must allow marketing users to update the population of any cities in the system, should a city's population change after data for it has been loaded, and must be accessible from the main menu. You must ensure only marketing users have access. The interface should require entering or choosing a city/state pair, displaying the current population value, and allowing the entry of a new value which can be submitted to update the database. If a city/state pair does not exist, the user should be presented with an appropriate error message.

WiMSS Reports

Will-Mart's management has put your team in charge of developing the queries necessary to produce the following reports. Many of the reports have derived and/or aggregate data. These reports will be accessed with the user interface that you will create.

Some of the report queries are expensive to run given the large number of rows in the WiMSS. Therefore, whenever possible you should include the filter conditions specified. For example, some reports ask for data from only a certain time period. If you leave off this filtering condition, the query will likely take a long time to return any results.

Report 1 – Manufacturer's Product Report

For each manufacturer, return the manufacturer's name, total number of products offered by the manufacturer, average retail price of all the manufacturer's products, minimum non-discounted retail price, and maximum non-discounted retail price. Sort the results by average price with the highest average price appearing first, for only the top 100 manufacturers based on average price.

This report should also have “drill-down” detail (in other words, each line in the master report should have a method for loading its detail, such as a hyperlink or a button) for the manufacturer, which shows in the report header the manufacturer’s name, the summary information from the parent report, and lists for each of the manufacturer’s products the product ID, name, category (or categories), and non-discounted price, ordered by price descending (high to low). If a product has multiple categories, it must not show up as multiple rows on the report, but as a single row with multiple categories concatenated together.

Report 2 – Category Report

For each category, including those without products, return the category name, total number of products in that category, the minimum non-discounted retail price, the average non-discounted retail price, and the maximum non-discounted retail price of all the products in that category, sorted by category name ascending.

Report 3 – Actual versus Predicted Revenue for Couches and Sofas

Will-Mart executives want to predict whether offering items at a discount actually helps to increase revenue by encouraging a higher volume of sales. This report compares how much revenue was actually generated from a product’s sales versus if the product were never discounted. After speaking with some marketing consultants, Will-Mart executives have learned that product discounts introduce on average a 25% increase in volume (quantity sold). Therefore, we assume that if an item that was offered at a discount were instead offered at the retail price, the quantity of items sold would be reduced by 25%. However, it is still possible that the predicted revenue would be higher since the reduced volume of products would be sold at a higher price per product. Initially, the executives are only interested in seeing the report for products in the “Couches and Sofas” category.

Here is a simple example:
Assume that Product Z has a retail price of $10. Assume that it was offered at a discount for on 6/1/2012 and 6/2/2012. Also assume the following transaction data for Product Z:

Date	Price	Quantity	Actual Revenue
5/1/2012	10.00	5	50.00
6/1/2012	8.00	10	80.00
6/2/2012	7.00	5	35.00
TOTALS		**20**	**$165.00**

Table 1 - Actual Revenue

The predicted revenue is calculated by assuming that the product is never offered at a discount and only 75% of the original quantity was actually sold on discounted days. Note that because this is just a predicted average, we assume that it is possible to sell a fraction of a product (e.g., 7.5 couches).

Date	Price	Quantity	Predicted Revenue
5/1/2012	10.00	5	50.00
6/1/2012	~~8.00~~ 10.00	10 * .75 = 7.5	75.00
6/2/2012	~~7.00~~ 10.00	5 * .75 = 3.75	37.50
TOTALS		**16.25**	**$162.50**

Table 2 - Predicted Revenue

In this example, the discounted prices resulted in slightly more revenue due to the higher volume of sales ($2.50 more).

Generate the following report: For each product in the Couches and Sofas category, return the product ID, the name of the product, the product's retail price, the total number of units ever sold, the total number of units sold at a discount (i.e., during discount days), the total number of units sold at retail price, the actual revenue collected from all the sales of the product, the predicted revenue had the product never been discounted (based on 75% volume selling at retail price), and the difference between the actual revenue and the predicted revenue. If the difference is a positive number, it means that the discounts worked in favor of Will-Mart because the predicted revenue is less than the actual revenue collected. If it is a negative number, it indicates that Will-Mart would have been better off not offering the product discounts. Only predicted revenue differences greater than $5000 (positive or negative) should be displayed and sorted in descending order. For this report, you do not need to include the impact of special discount days in the computation of prices and revenue.

Report 4 – Store Revenue by Year by State

This report shows the revenue collected by stores per state grouped by year. The states available for querying should be presented in a drop-down box. For example, the user would select "New York" and the system would show each store in New York state, show the store ID, store address, city name, sales year, and total revenue. Be sure the revenue calculation takes into account items that were sold at a discount. Sort the report first by year in ascending order and then by revenue in descending order.

Report 5 – Outdoor Furniture on Groundhog Day?

Some of the sales staff have noticed that outdoor furniture sales appear to spike on Groundhog Day (which falls on February 2 each year). They surmise that this is because customers begin thinking about the warm spring weather ahead. The Will-Mart marketing team would like to prove this, so they have requested the following report.

For each year, return the year, the total number of items sold that year in the outdoor furniture category, the average number of units sold per day (assume a year is exactly 365 days), and the total number of units sold on Groundhog Day (which is always February 2) of that year. Sort the report on the year in ascending order. The report will show if the total number of units sold on Groundhog Day each year is significantly higher than the average number of units sold per day.

Report 6 – State with Highest Volume for each Category

Will-Mart management is planning to recognize all stores in the states that sell the greatest number of units for each category. They want to view this monthly, so the user interface must allow choosing a year and month from the available dates in the database before running the report. The report will return for each category: the category name, the state that sold the highest number of units in that category (i.e., include items sold by all stores in the state), and the number of units that were sold by stores in that state. This output shall be sorted by category name ascending. Note that each category will only be listed once unless two or more states tied for selling the highest number of units in that category. You may exclude from this report products which do not have a category.
This report can take a significant time to run, which may require tuned indices for the final implementation, but do not focus on their creation until the final phase.

Report 7 – Revenue by Population

To help forecast expansions into other cities, Will-Mart management would like to see what the total revenue is for specific population categories, and to see if there is a trend for growth, the revenue should be broken down on an annual basis. The categories for city size are: Small (population <3,700,000), Medium (population >=3,700,000 and <6,700,000), Large (population >=6,700,000 and <9,000,000) and Extra Large (population >=9,000,000). (While it may make sense to set a city's size category as part of the city's data, you'll need to update the category when the population is updated in the system as well, or your data will be inconsistent.) The team would like this in a tabular form, with city sizes as previously ordered as the columns on the report. Ensure that both rows and columns are arranged in ascending order (oldest to newest for years, smallest to largest for city size categories) so that no matter how it is formatted it is properly organized and understandable. (It would be confusing to have city sizes in the order of Large, Small, Medium, etc. or years ordered as 1999, 2004, 2002, etc.)

Report 8 – Grand Showcase Store Revenue Comparison

There is a considerable expense in operating "Will's Grand Showcase" stores versus a normal Will-Mart store. The Will-Mart executive team would like to compare the performance of these stores to non-Showcase stores. The report header must show the count of Grand Showcase stores and the count of all non-Grand Showcase stores. The body should have a row for each year, and in the columns, the minimum, average, maximum, and total revenue of all Grand Showcase stores should be displayed, followed by the the minimum, average, maximum and total revenue of all non-Grand Showcase stores. (Average revenue can be determined by taking total revenue and dividing it by the count of stores of that particular type.) The rows should be ordered by year ascending.

Report 9 – Grand Showcase Store Category Comparison

The marketing team believes that the product mix at Grand Showcase stores could be different from non-Showcase stores, such as having more of certain categories in stock, or not offering some categories/products that are not purchased as frequently. This report will help identify these differences.

There are two parts to this report, which should utilize all sales data available. The first part is a summary of each category, with a row for each category. One column should show the quantity of products sold at Grand Showcase stores from that category, the next column the quantity of products sold at non-Grand Showcase stores from that category, and the last column should show the difference. The results should be ordered by difference, in descending order and category name in ascending order.

The second part to this report will be a "drill-down" for each category, showing the detail for each product in the category, and, similar to the parent report, with a row for each product's ID and name, should show the quantity sold at Grand Showcase stores in the first column, the quantity sold at non-Grand Showcase stores, and the difference. The results should again be ordered by difference, in descending order, and product ID in ascending order. However, only the top five and bottom five rows should be displayed. A sample, with the top 1 and bottom 1 result, is below.

Product ID	Product Name	Grand Showcase Qty	Regular Qty	Difference
3	Bokustinlar	5	2	3
252	Agleepoyo	6	25	-19

Document History

Version	Notes	Date
1.0	New version for Summer 2021	6/2/2021