Data Warehousing

Gym and Fitness Center

By

Dawit	**Chusetthagarn**	**5888032**
Sirinut	**Euaungkanakul**	**5888046**
Papichaya	**Quengdaeng**	**5888146**
Waran	**Taveekarn**	**5888200**

A Report Submitted in Partial Fulfillment of the Requirements for

ITCS 453 Data Warehousing and Data Mining

Faculty of Information and Communication Technology
Mahidol University

2017

An Overview of the Business Domain

Gym and Fitness Center is a place where people go in order to do exercises and build up their bodies. It offers wide ranges of sport facilities, starting from basic exercising facility like treadmill to more specific facility like barbell, to satisfy the needs of the customers. Gym and Fitness Center also allows the customers to have access to specialized trainers for supervised training or giving advice and consultation. To use the facilities, the customers can register their membership in different levels which will allow them access to more facilities and services.

Objectives and Scope of the Project

- To create an organized system where all the data are associated and linked together.
- To make it easy to manage all the transactions with the customers and stakeholders.
- To prevent the redundancy, inconsistency, and loss of data.
- To reduce the delay that appeared in the traditional Gym and Fitness Center Domain.
- To collect usage information and progresses. Collected information can be used for managing sales, reducing costs, and identifying regular customer and peak time.
- To facilitate the movement of data stored in the data warehouse

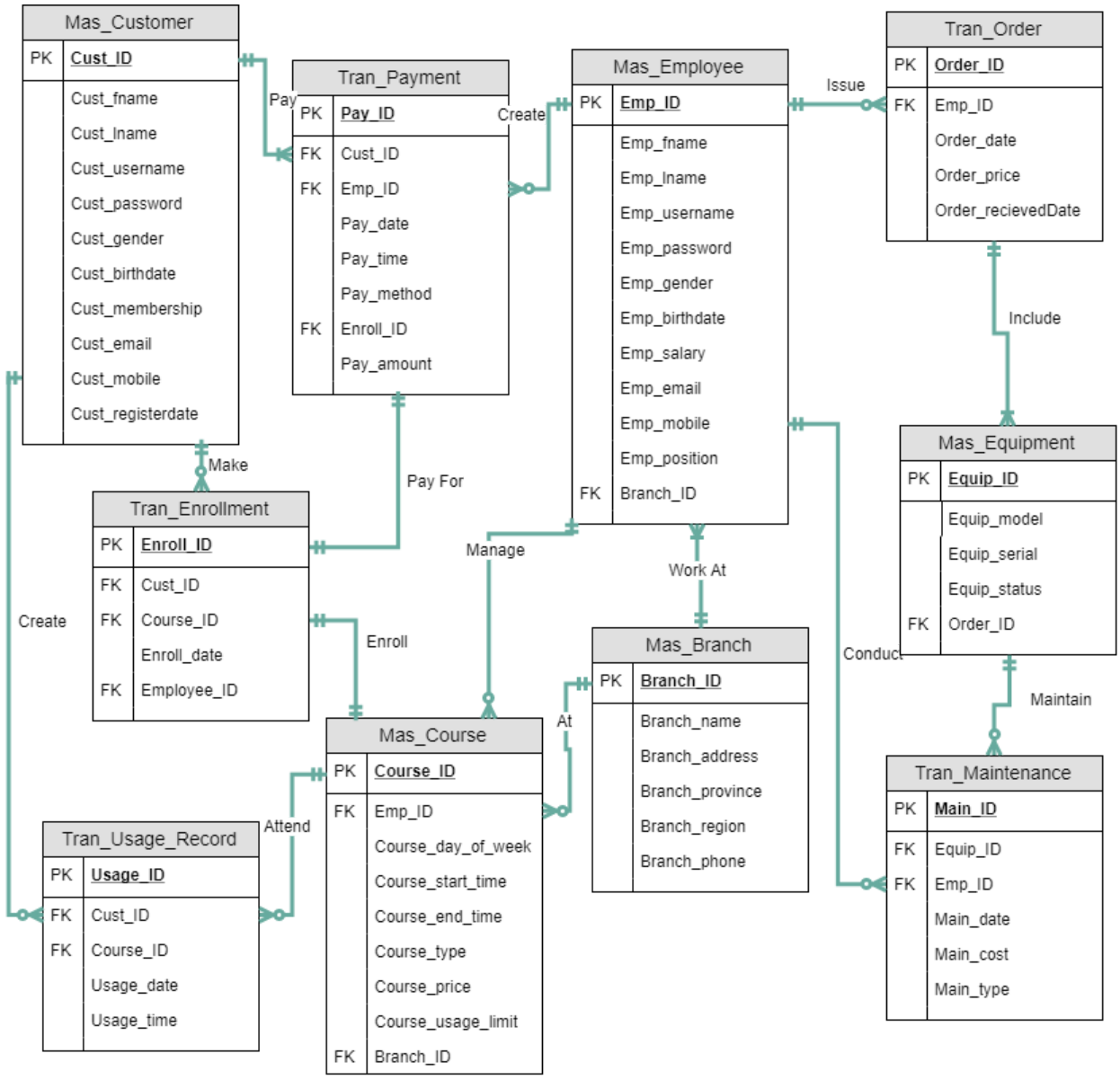
Business Requirements for Building a Data Warehouse

- Show the total number of course usage of the each type.
- Show the total number of equipment maintenance of each equipment type.
- Show the total cost of the maintenance in dollar of each equipment type.
- Show the total sales in dollar of each region, or each membership type.
- Show the number of course sold of each region, or each membership type.

Expected Benefits

- Gain insight on popular course type usage to provide more or less courses on each type.
- Compare durability of each equipment model, allowing better decision on future purchases of equipment in order to minimize maintenance cost.
- Get useful details on sales and source of revenue each branch, region, or different courses or membership type.
- Better decision on future course advertisement and target advertisement audiences on popular courses or membership type.

ER diagrams



This ER diagram shows the relationship between each entity for gym and fitness center, where data flows are, and how they are connected with each other.

Data Dictionary

TABLE NAME	ATTRIBUTE NAME	CONTENTS	TYPE	FORMAT	RANGE	REQUIRED	PK or FK	FK REFERENCED TABLE
Trans_Payment	Pay_ID	Payment ID	INT	99999999	00000000-99999999	Y	PK	
	Cust_ID	ID of customer making payment	INT	99999999	00000000-99999999	Y	FK	Mas_Customer
	Emp_ID	ID of employee responsible for the transaction	INT	99999999	00000000-99999999	Y	FK	Mas_Employee
	Pay_date	Date of payment	DATE	dd-mm-yyyy		Y		
	Pay_time	Time of payment	TIME	hh-mm		Y		
	Pay_method	Method of payment	VARCHAR	xxxxxxxxx		Y		
	Enroll_ID	ID of enrollment	INT	99999999	00000000-99999999	Y	FK	Tran_Enroll
	Pay_amount	Amount of money in the transaction	INT	999999999	0-999999999	Y		
Mas_Customer	Cust_ID	ID of customer	INT	99999999	00000000-99999999	Y	PK	
	Cust_fname	First name of customer	VARCHAR	xxxxxxxxx		Y		
	Cust_lname	Last name of customer	VARCHAR	xxxxxxxxx		Y		
	Cust_username	Username of customer	VARCHAR	xxxxxxxxx		Y		
	Cust_password	Password of customer	VARCHAR	xxxxxxxxx		Y		
	Cust_gender	Gender of customer	VARCHAR	X	(M,F)	Y		
	Cust_birthdate	Birth date of customer	DATE	dd-mm-yyyy		Y		
	Cust_membership	Level of membership	VARCHAR	xxxxxxxxx		Y		
	Cust_email	Email of customer	VARCHAR	xxxx@xxx.xxx		N		
	Cust_mobile	Mobile number of customer	VARCHAR	9999999999		N		
	Cust_registerdate	Date registered as a member	DATE	dd-mm-yyyy		Y		
Tran_Usage_Record	Usage_ID	ID of usage record	INT	99999999	00000000-99999999	Y	PK	
	Cust_ID	ID of customer	INT	99999999	00000000-99999999	Y	FK	Mas_Customer
	Course_ID	ID of course	INT	99999999	00000000-99999999	Y	FK	Mas_Course
	Usage_date	Date of usage	DATE	dd-mm-yyyy		Y		
	Usage_time	Time of usage	TIME	hh-mm-ss		Y		
Mas_Course	Course_ID	ID of course	INT	99999999	00000000-99999999	Y	PK	
	Emp_ID	ID of employee/trainer	INT	99999999	00000000-99999999	Y	FK	Mas_Employee
	Course_day_of_the_week	Day that the course takes place	VARCHAR	xxxxxxxxx		Y		
	Course_start_time	Start time of the course	TIME	hh-mm-ss		Y		
	Course_end_time	End time of the course	TIME	hh-mm-ss		Y		
	Course_type	Type of course	VARCHAR	xxxxxxxxx		Y		
	Course_price	Price of course	INT	999999999	0-999999999	Y		
	Course_usage_limit	Usage limit of the course	INT	99999999		Y		
	Branch_ID	ID of branch	INT	99999999	00000000-99999999	Y	FK	Mas_Branch
Mas_Employee	Emp_ID	ID of employee	INT	99999999	00000000-99999999	Y	PK	
	Emp_fname	First name of employee	VARCHAR	xxxxxxxxx		Y		
	Emp_lname	Last name of employee	VARCHAR	xxxxxxxxx		Y		
	Emp_username	Username of employee	VARCHAR	xxxxxxxxx		Y		
	Emp_password	Password of employee	VARCHAR	xxxxxxxxx		Y		
	Emp_gender	Gender of employee	VARCHAR	X	(M,F)	Y		
	Emp_birthdate	Birth date of employee	DATE	dd-mm-yyyy		Y		
	Emp_salary	Monthly Salary of employee	INT	999999999	0-999999999	Y		
	Emp_email	Email of employee	VARCHAR	xxxx@xxx.xxx		N		
	Emp_mobile	Mobile number of employee	VARCHAR	9999999999		N		
	Emp_position	Position of employee	VARCHAR	xxxxxxxxxxx		Y		
	Branch_ID	ID of branch	INT	99999999	00000000-99999999	Y	FK	Mas_Branch
Tran_Order	Order_ID	ID of order	INT	99999999	00000000-99999999	Y	PK	
	Emp_ID	ID of employee	INT	99999999	00000000-99999999	Y	FK	Mas_Employee
	Order_date	Date of order	DATE	dd-mm-yyyy		Y		
	Order_price	Total price for order	INT	999999999	0-999999999	Y		
	Order_receivedDate	Date received	DATE	dd-mm-yyyy		N		
Mas_Equipment	Equip_ID	ID of equipment ordered	INT	99999999	00000000-99999999	Y	PK	
	Equip_model	Model of equipment	VARCHAR	xxxxxxxxx		Y		
	Equip_serial	Serial number of the equipment	VARCHAR	xxxxxxxxx		Y		
	Equip_status	Status of equipment	VARCHAR	xxxxxxxxx		Y		
	Order_ID	ID of order which the equipment was ordered	INT	99999999	00000000-99999999	Y	FK	Tran_Order
Tran_Maintenance	Main_ID	ID of maintenance	INT	99999999	00000000-99999999	Y	PK	
	Equip_ID	ID of equipment in maintenance	INT	99999999	00000000-99999999	Y	FK	Mas_Equipment
	Emp_ID	ID of employee responsible	INT	99999999	00000000-99999999	Y	FK	Mas_Employee
	Main_date	Date sent to maintenance	DATE	dd-mm-yyyy		Y		
	Main_cost	Cost of maintenance	INT	999999999	0-999999999	Y		
	Main_type	Type of maintenance	VARCHAR	xxxxxxxxx		Y		
Mas_Branch	Branch_ID	ID of branch	INT	99999999	00000000-99999999	Y	PK	
	Branch_name	Name of branch	VARCHAR	xxxxxxxxx		Y		
	Branch_address	Address of branch	VARCHAR	xxxxxxxxx		Y		
	Branch_province	Province that the branch is located	VARCHAR	xxxxxxxxx		Y		
	Branch_region	Region that the branch is located	VARCHAR	xxxxxxxxx		Y		
	Branch_phone		VARCHAR	9999999999		Y		
Tran_Enrollment	Enroll_ID	ID of enrollment	INT	99999999	00000000-99999999	Y	PK	
	Cust_ID	ID of customer	INT	99999999	00000000-99999999	Y	FK	Mas_Customer
	Course_ID	ID of course	INT	99999999	00000000-99999999	Y	FK	Mas_Course
	Enroll_date	Date of enrollment	DATE	dd-mm-yyyy		Y		
	Emp_ID	ID of employee	INT	99999999	00000000-99999999	Y	FK	Mas_Employee

Star/Snowflake Schema Diagrams

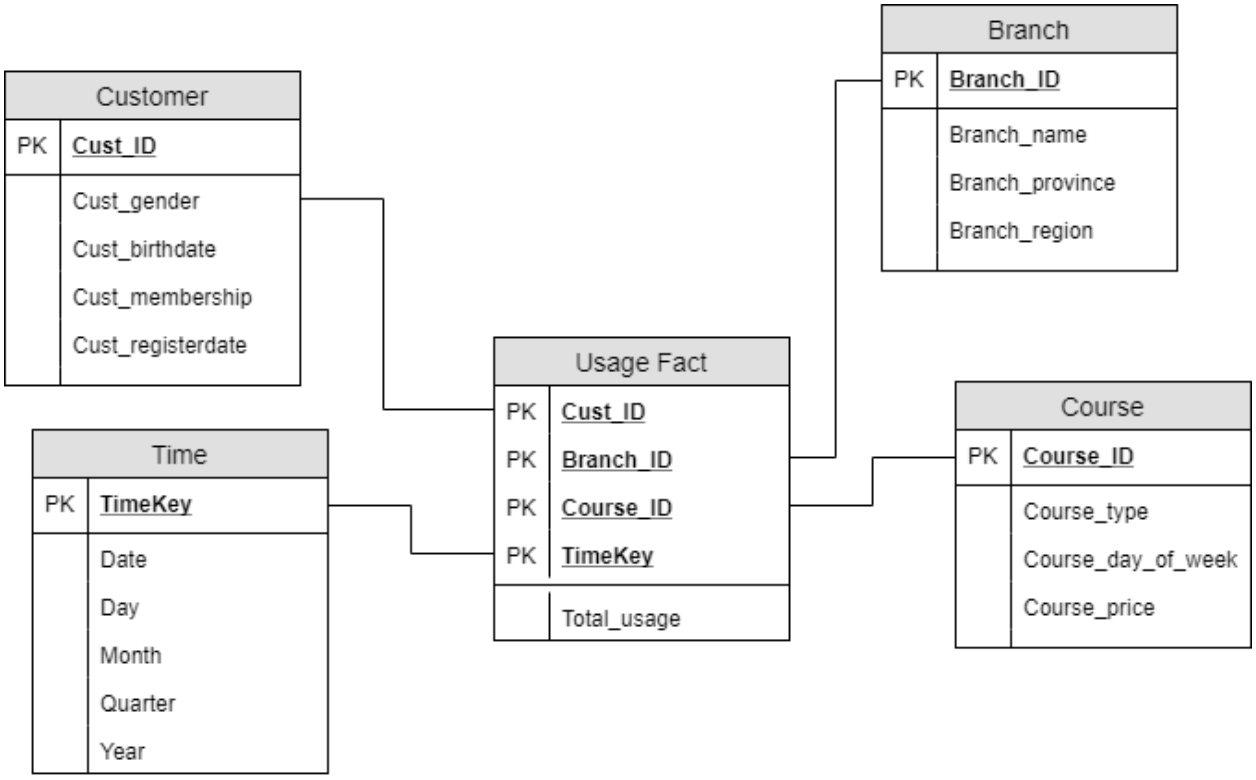
Usage Fact

Dimensions

- Customer - customer's demographic; gender, age, level of membership, membership age.
- Course - course's type and price
- Time - time of day, date, quarter, or day of week.

Facts

- Total_usage - aggregated total number of customer visitation and usage of the facility.



The total usage which derive from Customer, Course, and Time can be used for analysis the peak time for customer, so the gym can create promotion that suit for them.

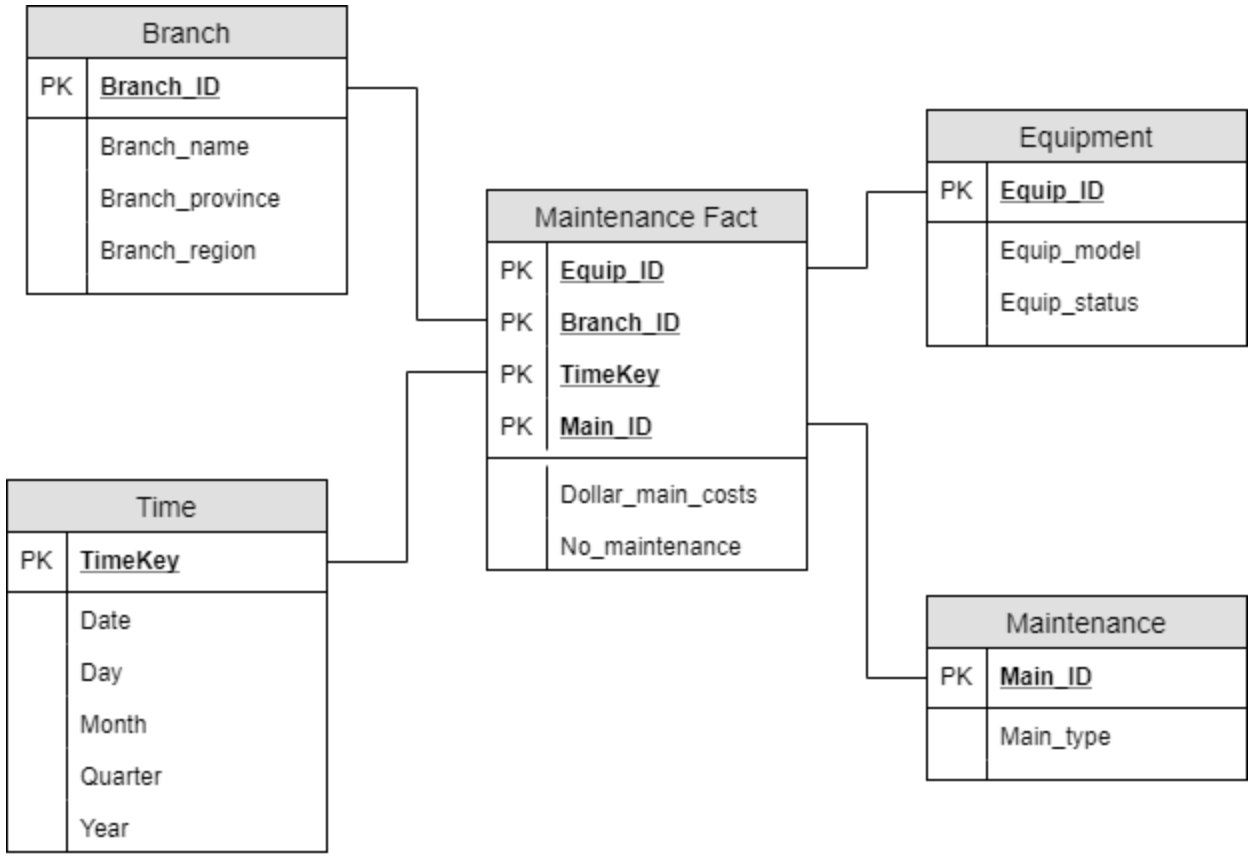
Maintenance Fact

Dimensions

- Branch - Branch's name, province, region.
- Time - date, quarter, or day of week.
- Equipment - equipment's model, status.
- Maintenance - Maintenance type

Facts

- No_maintenance - total number of equipment maintenance



The number of maintenance which derive from Branch, Equipment, Time and Maintenance can be used for analysis which branch need to be stock more equipment for customer or the lifetime of each types of equipment.

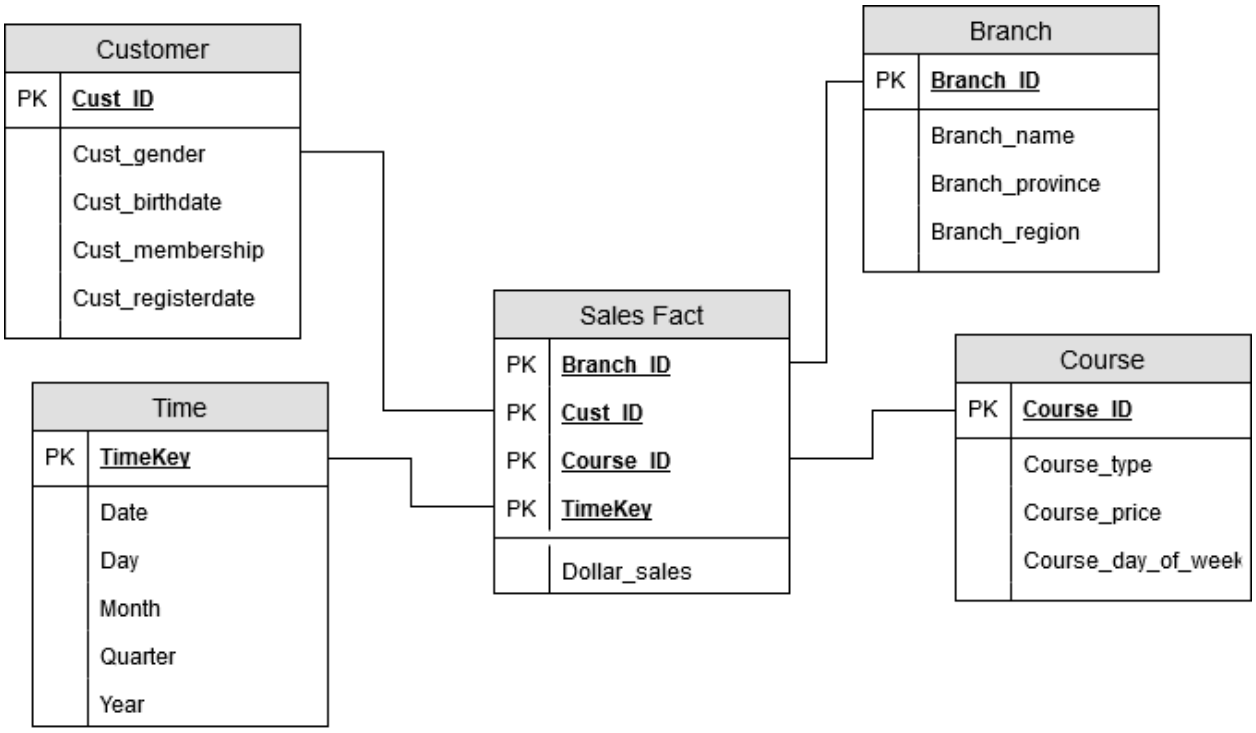
Sales Fact

Dimensions

- Customer - customer's demographic; gender, age, level of membership, membership age.
- Time - date, quarter, or day of week.
- Course - course's type and price

Facts

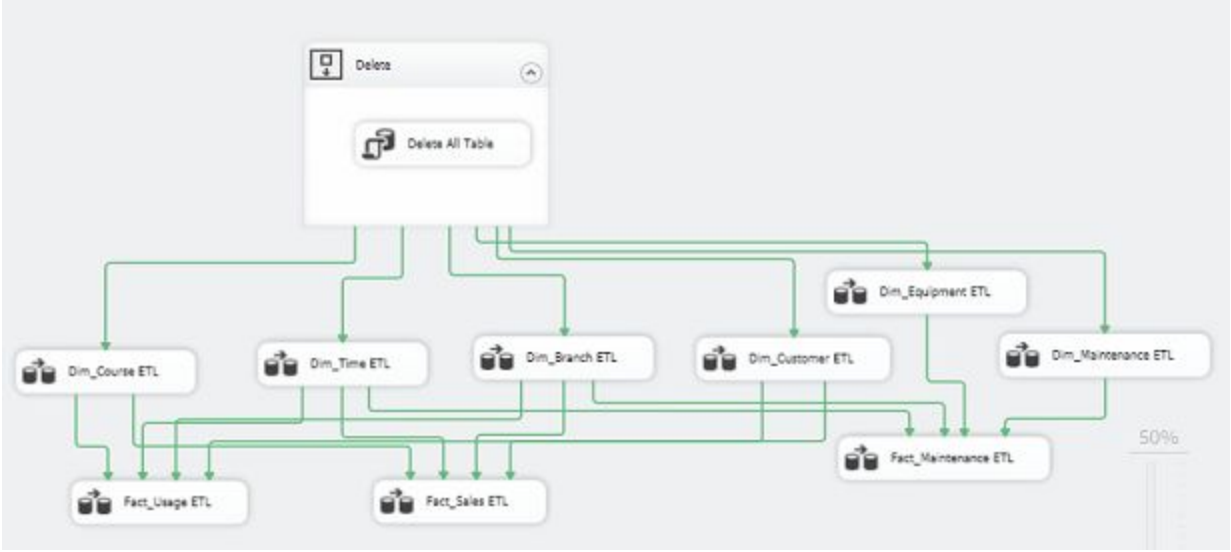
- Dollar_sales - aggregated total sales in dollar
- No_new_member - number of new member in the facility
- No_course_sales - number of course sales



The total sales, number of new member and number of course sales which derive from Customer, Branch, Time and Course can be used for analysis the profit of each branch or which course is the trend. Number of new customer also can be analysed for opening new branch.

ETL process description

Our ETL process of extracting data from the database to the data warehouse is done by using Microsoft Visual Studio. Each query selects the data from database and mapping with the ID of the data warehouse, so we can use those data for analysis. Below is the control flow of the ETL process.



Dim_Course ETL takes the ID, course type, price, and day of the week from Mas_Course.

Dim_Time ETL takes the date of all transactions in Payment, Order, Maintenance, and Usage_Record.

Dim_Branch ETL takes the ID, province, and region from Mas_Branch.

Dim_Customer ETL takes the ID, gender, birthdate, registerdate, and membership type from Mas_Customer.

Dim_Equipment ETL takes the ID, model, serial, and status from Mas_Equipment

Dim_Maintenance ETL takes the ID and maintenance type from Tran_Maintenance

Fact_Usage takes the unique key of customer, course, branch, and time from each respective table and initialize 1 as the Total_usage.

Fact_Sales takes the unique key of customer, time, branch, and course from each respective table, and pay amount and pay date from Tran_Payment.

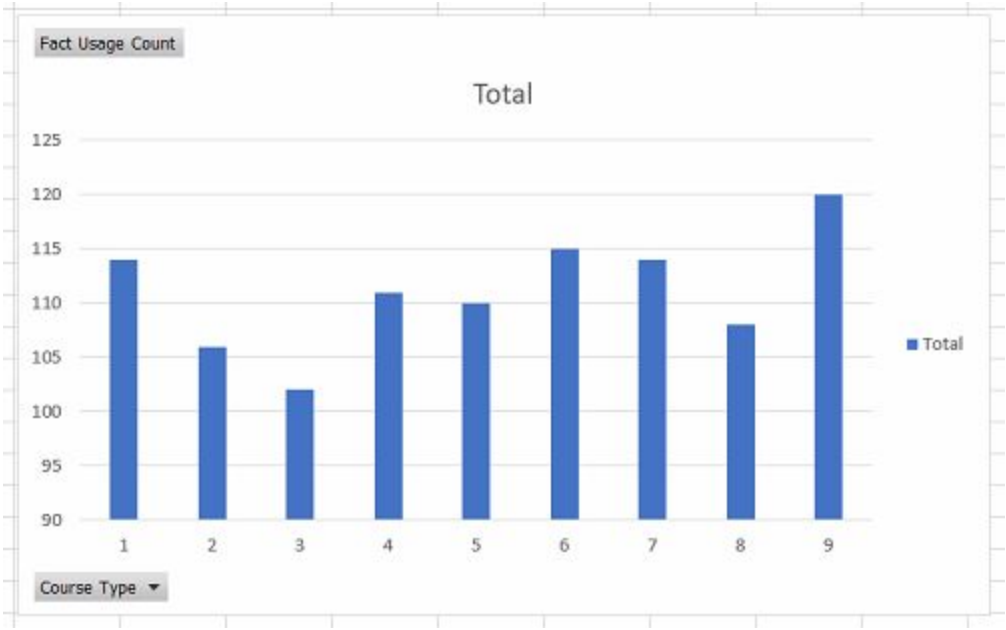
Fact_Maintenance takes the unique key of equipment, branch, time, and maintenance from each respective table and maintenance cost from Tran_Maintenance.

Analysis Reports

Each requirement is useful information that can be analysed deeper to improve the business. Each fact table will compute the information from data in the data warehouse. The fact tables are created based on the star schema diagram by deriving some tables with some attributes which are explained in data dictionary to aggregate specific information for analysis. This project has total three fact tables which are usage fact table, maintenance fact table, sales fact table.

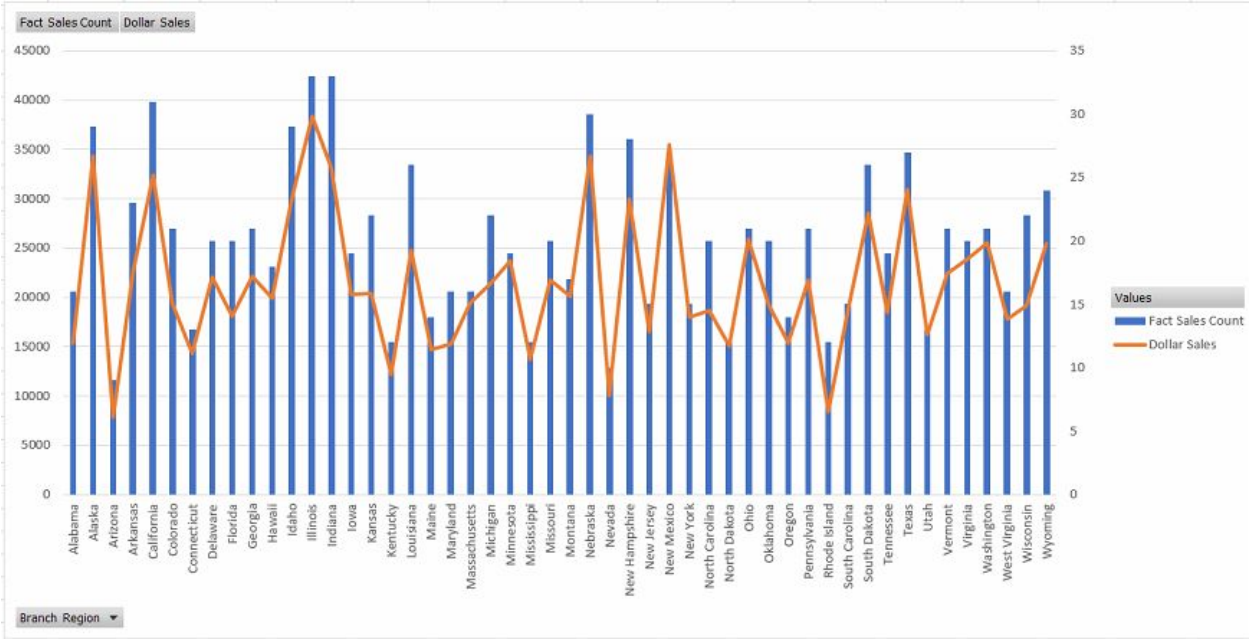
Usage fact table

This graph shows the total usage of the facility according to each course type. We can use this data to analyze which course is the most preferable course to the customers. The gym can adjust the price of the course to increase more profit and create new promotion for the course.

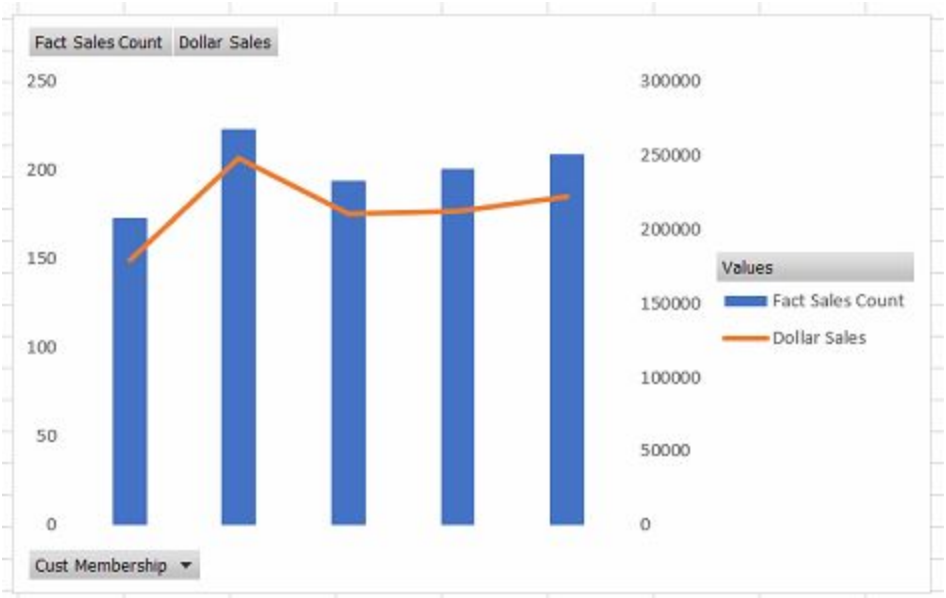


Sales fact table

This graph shows the total number of course sales count with the aggregated total sales in dollar of each region. This data can be used to find the potential regions or branches that bring the most profit to the gym and fitness center. The gym would be able to promote and invest more money in the regions that is highly profitable. The gym can also create promotions to attract more customers to the regions that are less profitable.



This graph shows the total number of course sales with the aggregated total sales in dollar according to each type of customer membership. This data can be used to analyze which type of customer membership bring in the most profit, so the gym would be able to create a new attracting promotion to attract customers to become that type of membership.



Maintenance fact table

This graph shows the total number of equipment maintenance and the total cost of the maintenance for each model of the equipment. This fact table can be used to analyze the rate of equipment deterioration and the cost efficiency of each model of the equipment. It can also be used to analyze the durability of each equipment model to know which model is the best decision to buy in the future.