TEAM MEMBERS:

Amruta Kulkarni

Ishani Desai

Lalita Gajbe

Sindhuja Gangadurai

**Data Warehousing and Business Intelligence**

**System for Retail Store**

**Prof. Dr. Jerry Fjermestad**

**MIS 620**

620-B

Contents

[Introduction 2](#_Toc418455758)

[Use Cases for the Report Generation tool 3](#_Toc418455759)

[Data Flow Diagrams 8](#_Toc418455760)

[Feasibility Analysis 13](#_Toc418455761)

[Economic Feasibility 13](#_Toc418455762)

[Technical Feasibility 15](#_Toc418455763)

[Organizational Feasibility 18](#_Toc418455764)

[Architecture of the DATA WAREHOUSE system 19](#_Toc418455765)

[Dimensional Modelling 22](#_Toc418455766)

[User Interface 26](#_Toc418455767)

[Conclusion 28](#_Toc418455768)

[References 29](#_Toc418455769)

## Introduction

Our client is a consumer product retail Company. It is a profitable organization but seeing its market share eroding. In order to analyze the reasons behind thinning market share and make the strategic changes, it needs a data warehousing and business intelligence system. This system is comprised of a data warehousing infrastructure, and a query analysis, and reporting environment. A physical repository where relational data are specially organized to provide enterprise-wide, cleansed data in a standardized format.

The Client Company has an existing point of sale system in place which simply records the transactions i.e. number of products sold, number of products added to inventory, customer records, purchase orders. Our DW system will help the client Company to analyse the reason of their eroding market share and set their strategy to improve their market share in next five years.

The new DW system will help to make decisions regarding

* Marketing the products according to their customer groups
* Placement of products based on sales
* Supply chain management
* Determining the best price for a product
* Improving customer satisfaction

Benefits of DW system:

* Integrates data from various sources (web portal, POS system, mobile app)
* Allows for analysis of data over time
* Provides faster and improved performance for complex analytical queries
* Reduces load on Operational database system
* Provides analysis necessary to make strategic decisions to improve market share
* Requires minimal investment in infrastructure

Below is the System Proposal stating functionalities, feasibility study and design of the new DW system

## Use Cases for the Report Generation tool

|  |  |
| --- | --- |
| **Use Case ID:** | 1 |
| **Use Case Name:** | Get list of the ‘products being purchased together’ |
| **Actors:** | Primary actor: Marketing staff |
| **Description:** | Marketing staff queries the data mart to check the pattern of products being purchased together |
| **Trigger:** | Marketing staff clicks Generate Report |
| **Preconditions:** | 1.   Segregated data is present in DataMart  2.   Marketing staff has passed the authentication for using the system |
| **Post conditions:** | The report of products being purchased together is generated |
| **Normal Flow:** | 1. User logs into his account   2. System displays the Generate Report menu   3. User selects ‘Generate Report for Product sales by orders’ from menu   4. System prompts user to enter the required data   5. User enters the month(s)/ year(s) for which he/she needs to attain information for the season sales.   6. User selects the stores for which he wants to collect the data   7. System matches the common products purchased in  various orders and generates a report   8. System displays the report in graphical form stating which products were sold together in how many transactions e.g. In 150 transactions, cartridges and printers were sold together   9. User reviews the report and instructs the Inventory management people to adjust the placement of products being purchased together in such a way that customers can easily locate the related products on shelf. For online shopping system, the Marketing staff asks admin user to display recommendation of related products when a customer selects a particular product. |
| **Alternative Flow:** | 6a. User requests the reports by store location  6b. Staff generates report based on city/ state  6c. User generates report based on the vendors  6d. User generates report based on the category (Electronics items, Accessories for electronic equipment) |
| **Includes:** | 1. System validates the user name and password entered by Marketing staff  2. System allows user to sort the result by week, month, year |
| **Frequency of Use:** | Once a week |
| **Special Requirements:** | The ELT processes are in place to capture, transform and load data from various data sources |
| **Assumptions:** | All special requirements are assumed to be fulfilled. |
| **Notes and Issues:** | Create separate use case for login, sort and filter functionalities |

|  |  |
| --- | --- |
| **Use Case ID:** | 2 |
| **Use Case Name:** | Record sales |
| **Actors:** | Primary Actor: Market Research staff  Secondary Actor: Management |
| **Description:** | Market Research staff queries the data mart to check the products that are sold and in which season people purchase it |
| **Trigger:** | Market Research staff clicks the generate report link |
| **Preconditions:** | 1.   Segregated data present in DataMart  2.   Market Research staff is authenticated to access the system. |
| **Post conditions:** | The report containing sales records and season of sales is generated |
| **Normal Flow:** | 1. Market Research staff logs into his/her account   2. System displays the Generate Report interface   3. Market Research staff enters the month(s) for which he/she needs to attain information for the season sales.   4. Staff selects the category( clothing, furniture, home essentials)   5. Staff selects the vendors as he/she needs to know the sales of products of particular vendor  6. Staff selects the price range to know how the price range reflects on sales  7. Market Research staff clicks on the generate report button  8. System shows that reports can be generated in various file types (excel, csv, text file) and the staff selects the preferred file type from the options available  9. Market Research staff reviews the report and passes his analysis information to the management team. His/her analysis report contains the details such as which products is sold frequently at all seasons, products getting sold only in particular months, the percentage of sales based on price and the sales of products of particular vendors. |
| **Alternative Flow:** | 3a. Market Research Staff selects the month/ months for getting the individual report on monthly sales.  4a. Market Research Staff selects the category (Clothing/ Furniture/Home essentials) and generates the individual report for category sales.  5a. Staff generates individual report based on the vendor sales  6a. Staff generates individual report based on the price range |
| **Includes:** | 1. System validates the user name and password entered by Market Research staff  2. System allows user to sort the result by date, price range and alphabetical order |
| **Frequency of Use:** | Once a week |
| **Special Requirements:** | The ETL processes are in place to capture, transform and load data from various data sources |
| **Assumptions:** | All special requirements are assumed to be fulfilled. |

|  |  |
| --- | --- |
| **Use Case ID:** | 3 |
| **Use Case Name:** | Record coupons |
| **Actors:** | Primary actor: Customer and Admin |
| **Description:** | To record the details of coupons used by customers. |
| **Trigger:** | Customer use discount coupons or member card to buy desired product. |
| **Preconditions:** | For placing order: Customer has access to the online shopping web portal |
| **Post conditions:** | The order is placed or cancelled and the action reflects in order history |
| **Normal Flow:** | 1. Customer logs into his account   2. Customer selects desired products and quantities and adds to cart   3. Customer enters the coupons code to avail discount offered on products by retailers.   4. Customer use membership cards details to gain points.   5. Customer proceeds to check out   6. Customer enters personal details as email id, shipping address   7. Customer enters payment details and completes payment   8. System sends confirmation email to the customer  9. System record the type of coupon and number of coupons used. |
| **Alternative Flows:** | 1a. Customer visit shop instead of online shopping  1a1. Customer selects product and got to the billing counter to products.  1a2. Admin scans the product and the coupons code along with membership card.  1a3. System will record the type of coupons used.  1a4. System will also able to find the number of coupons used from the data.  1a5. Customer enters his city name or zip code along with payment details.  3a. Customer enters gift card or discount coupon details  -10a1. System displays the discounted price of the products  -10a2. Customer completes payment |
| **Exceptions:** | 1a. Customer enters invalid credentials  -1a1. System prompts customer to enter correct credentials and try again  -1a2. Customer enters valid credentials  10a. Customer enters wrong card details or expired coupon code  -10a1. System gives error message  -10a2. Customer enters correct card details or another coupon code and      completes payment |
| **Includes:** | 1. System validates the user name and password entered by customer  2. System allows customer to filter the products by price, customer rating, brand or special offer  3. Payment processor validates the card details entered by the customer |
| **Frequency of Use:** | Several times a day |
| **Special Requirements:** | The ELT processes are in place to capture, transform and load data from various data sources |
| **Assumptions:** | All special requirements are assumed to be fulfilled. |
| **Notes and Issues:** | Create separate use case for login, search and filter functionalities |

|  |  |  |
| --- | --- | --- |
| **Use Case ID:** | 4 | |
| **Use Case Name:** | Generate feedback and perform analysis | |
| **Actors:** | | Primary actor: Administrator |
| **Description:** | | Administrator queries the data mart to determine which products have higher ratings and high product satisfaction |
| **Trigger:** | | Administrator clicks Generate Report |
| **Preconditions:** | | Administrator must authenticate credentials to access the system |
| **Post conditions:** | | The report of user feedback on products is generated |
| **Normal Flow:** | | 1. Administrator logs into the account  2. System displays the Generate Report menu  3. Administrator selects ‘Generate Report for Customer feedback’ from menu  4. System prompts administrator to populate fields for the query  5. Administrator selects the time period for which he wants to collect the feedback.  6. Administrator selects the products for which he wants to collect the data  7. System matches the all the feedback within the particular time frame with regards to the specified product and generates a report  8. System displays the report in graphical form stating the date of the feedback, the unique ID attached with the feedback, name of the person, the rating given to a particular product, the review that was attached with the product.  9. Administrator can further send this report to the marketing staff, which can determine which products are selling more and increase the inventory for that particular product or which products are selling less and whether to discontinue the product or find a way to better market the product to the customer. |
| **Alternative Flow:** | | 6a. User requests the reports by store location, city or state |
| **Includes:** | | 1. System validates the user name and password entered by Administrator  2. System allows administrator to sort the result by week, month, year |
| **Frequency of Use:** | | Once a week |
| **Special Requirements:** | | The ELT processes are in place to capture, transform and load data from various data sources |
| **Assumptions:** | | All special requirements are assumed to be fulfilled. |

## Data Flow Diagrams

**Context Diagram**

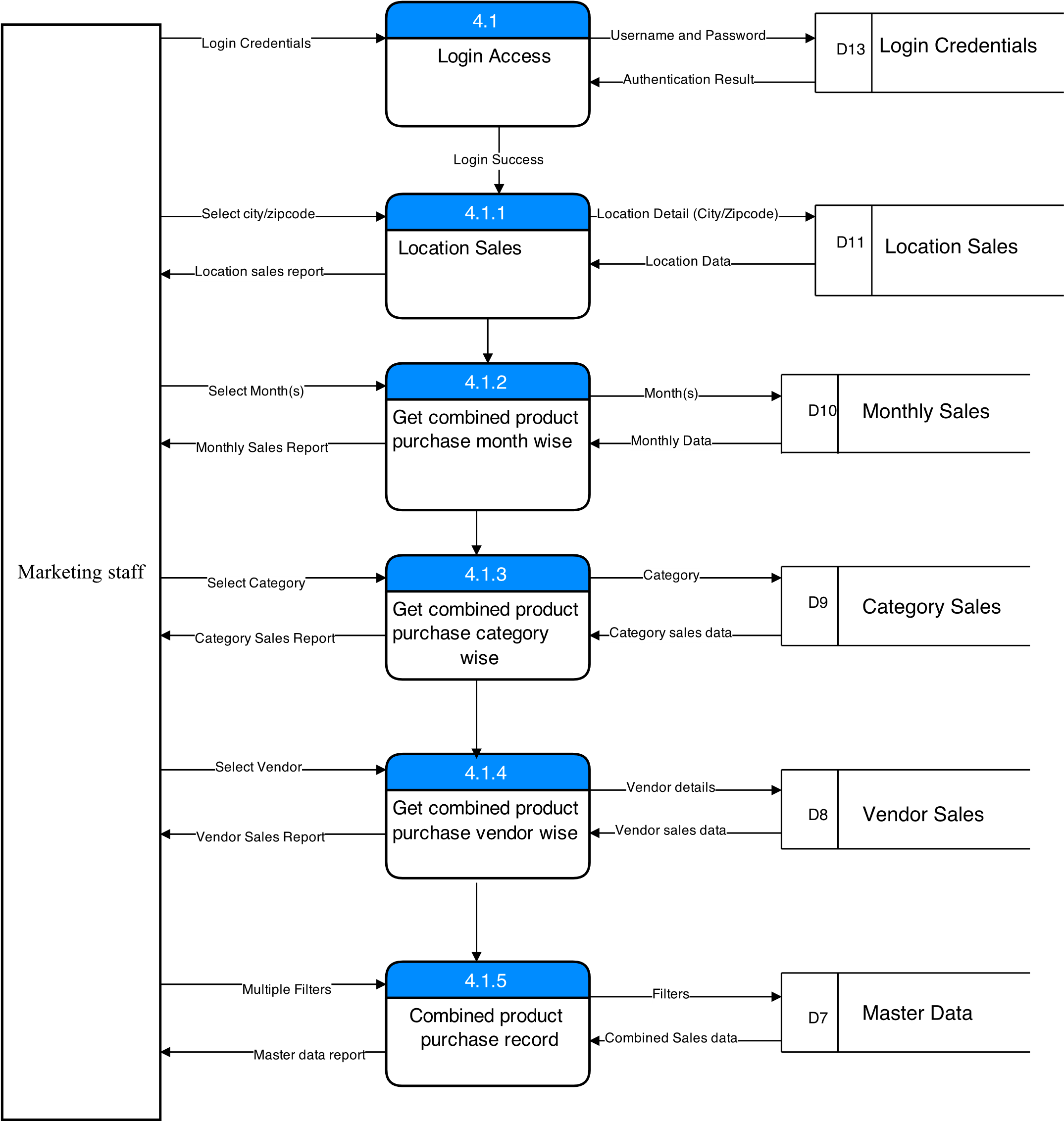


**Level Zero Diagram**

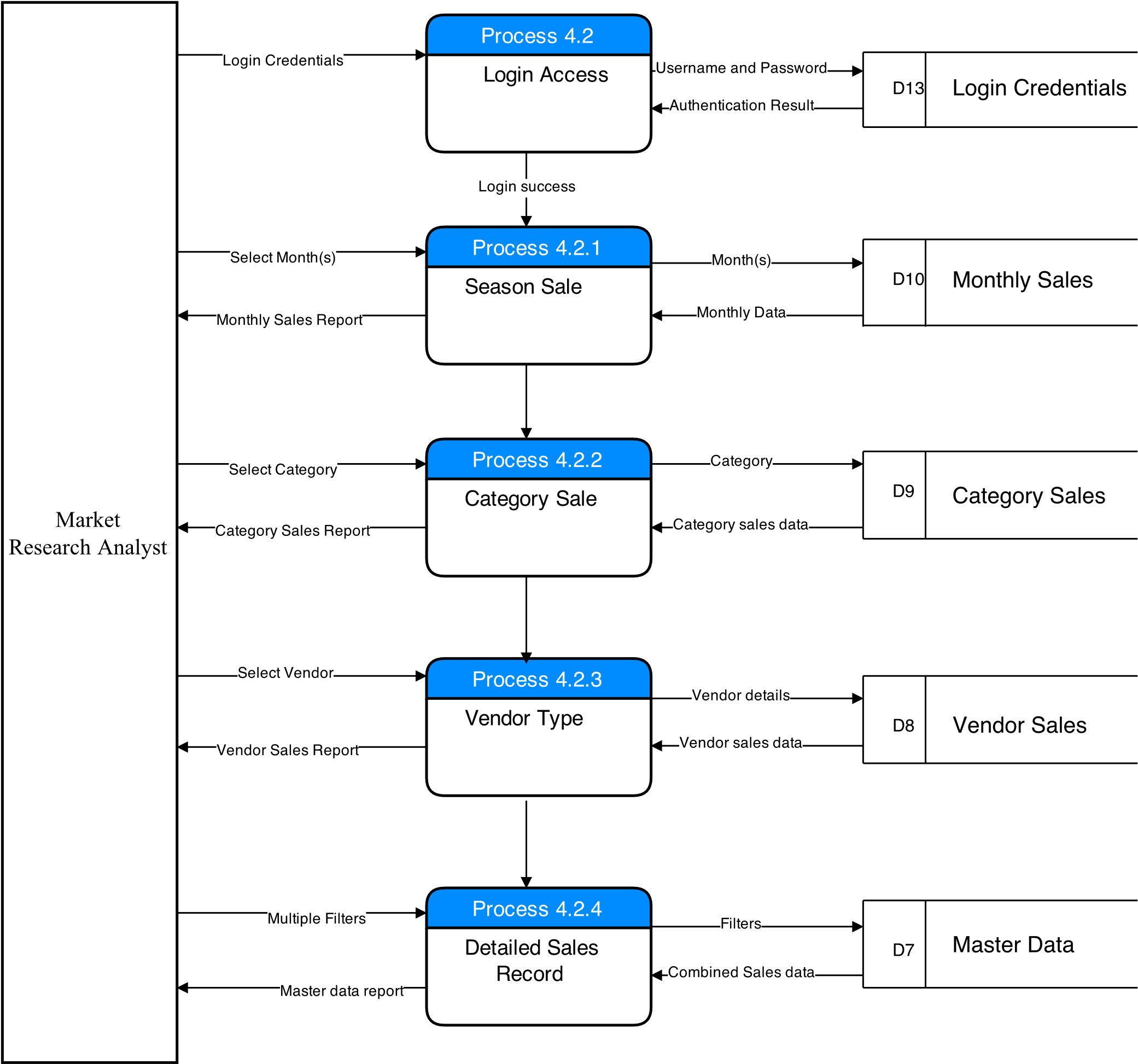


**Level 1 Diagrams**

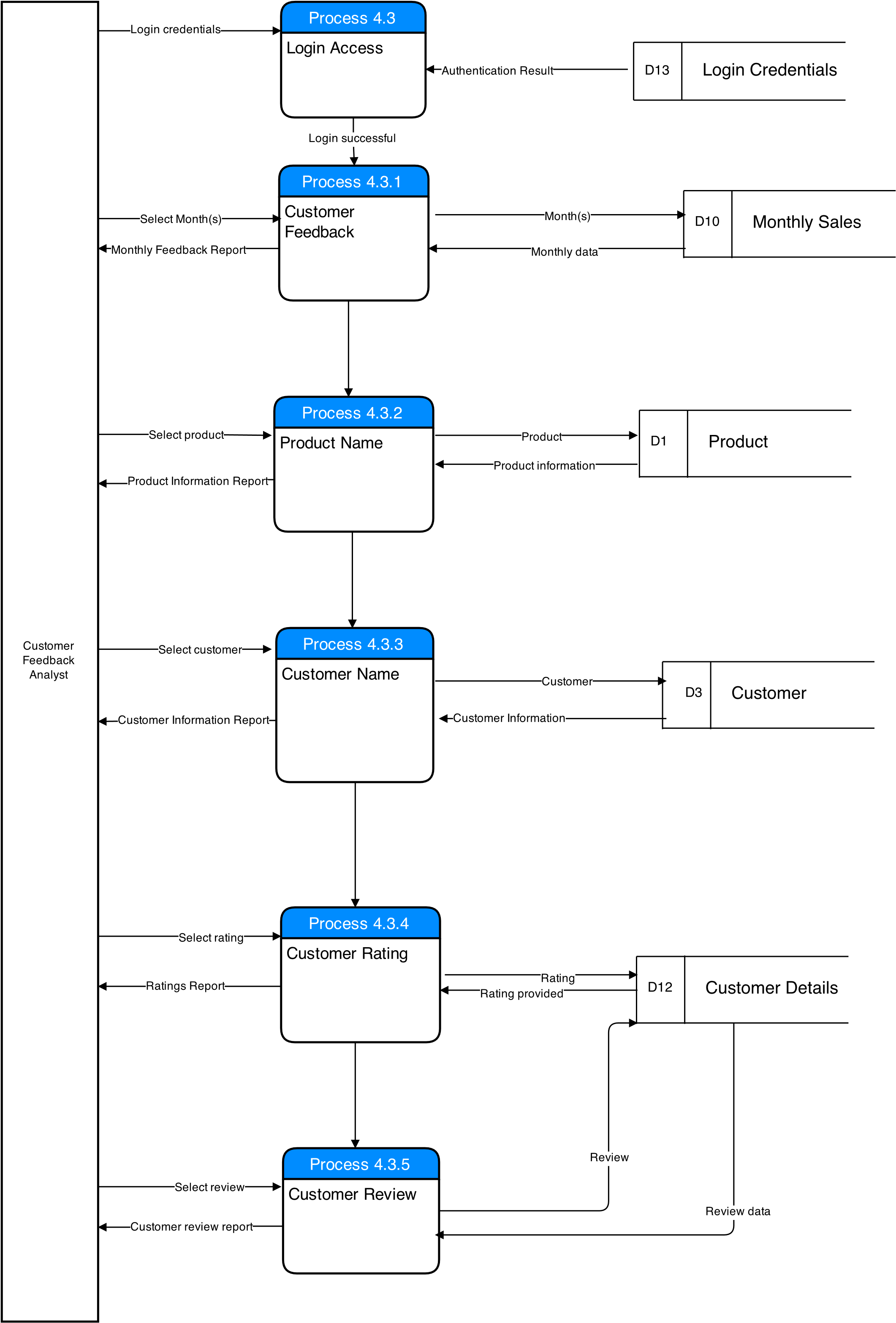
**Combined Product Sales Process**



**Record Sales Process**



**Customer Feedback Process**



## Feasibility Analysis

## Economic Feasibility

It deals with the cost benefit analysis of the system and measures the financial worth and risks associated. Mostly the system contains the initial implementation cost and then followed by the ongoing costs.

The Cost Analysis has to be done in the below areas while implementing the Data warehouse system.

Hardware Cost:

The Hardware configuration is most important to bring out the best performance of the data warehouse system. There is need of different servers like servers for ETL, Data warehouse database server, Reporting and Backup and recovery. Our company is planning to adopt the low cost blade solution provided by DELL. These blade servers serve as key elements in reducing the growing power costs and for implementation of environmental friendly IT initiatives. Below is the cost analysis of Dell blade configuration using an EMC storage array, larger drivers for each server and a gateway server and SAN hardware for backup purposes.

|  |  |  |  |
| --- | --- | --- | --- |
| ITEM | QUANTITY | PRICE | COST |
| Blade enclosure | 1 | 1200 | 1200 |
| Power Edge 1855 with 2 Intel Xeon dual core processors | 5 | 3200 | 16000 |
| Software, Accessories and maintenance | 1 | 22500 | 22500 |
| DELL EMC Storage Array | 1 | 160,000 | 160,000 |
| Gateway 840 Serial ATA SAN | 1 | 7000 | 7000 |
| TOTAL |  |  | 206,700 |

*Source: MTW Solutions Feasibility Study*

Software Cost:

Licenses needs to be purchased for the end user analytics tools and server software along with add-on packages. Infrastructure software costs include the Microsoft windows server and SQL Server.

|  |  |  |
| --- | --- | --- |
| ITEM | QUANTITY | COST |
| Microsoft Windows Server | 2 | 700 |
| SQL Server | 1 | 400 |
| ETL | 1 | 250,000 to 1,000,000 |
| OLAP | 1 | 250,000 to 1,000,000 |
| Reporting | 1 | 200,000 |

*Source: MTW Solutions Feasibility Study*

Development Cost:

The Cost associated with the development of the data warehouse system is higher when compared to the hardware and software costs. The development costs include the data modelling, design, mapping, ETL and the data population costs. ETL Costs are the software costs that include the ETL Software, metadata capture, transformation logic, source and target mapping etc. A lot of money will be spent on design, development and deployment of the data warehouse. The development costs occurs during the initial stages of building the system and slowly reduces as maintenance cost.

Storage Cost:

Storage costs include the cost of storage and the infrastructure. The cost of storage in the first few years of the data warehouse is not high. In the first few years there is relatively little stored data. In over time, the amount of data to be stored increases and there will be many subject areas introduced into the data warehouse.

Year

1

2 3 4

$

$

Short Term Cost of the Data warehouse system Long Term Cost of the Data warehouse system

Development

Storage

ETL

DBMS

*Source: Dataupia.com*

## Technical Feasibility

To implement Data warehouse, our management will increase their focus on defining data collection practices and communicating the data quality standards through improved training programs both inside and outside the company. This improvement in training and related policy development will facilitate the loading of high quality data into the data warehouse and have significant impact on the success and use of the data warehouse.

The success of a data warehouse as a strategic decision making tool is determined by three factors:

• The breadth, timeliness and quality of data contained within the data warehouse,

• The availability and usability (access) to the data warehouse by decision makers, and

• The power and scope of the data warehouse’s reporting and querying capabilities.

Since the architecture of the system can vary, a construct known as the operational data store can be used to create an additional source of information that is separate from the data warehouse. Data passes through the Extract, Transform, and Load (ETL) process into the data store and then passes an additional ETL process before being loaded into the data warehouse.

Because of the differences between systems and how systems define and store data, it is necessary to process the data so that the data from multiple source systems can come together in a new environment. This process is called **ETL**, an abbreviation for **Extract, Transform and Load**.

**Extract** is the process of reading the data from the source database. This can be done using batch, historic or real-time processes. Prior to the extraction process it may be necessary to clean the data. Data cleansing is the process of removing errors and inconsistencies with the source data.

**Transform** is the process of converting the extracted data from its previous form into the form it needs to be in so that it can be placed into another database. Transformation occurs by using rules or lookup tables or by combining the data with other data.

**Load** is the process of writing the data into the target database.

One approach to the architecture of a data warehouse is to use data marts to organize the data. A single **Data Mart** is data structure that is optimized for access. This approach supports the notion of simplicity by allowing the user to focus on one or just a few subject areas at a time. It is very important, that while designing data marts, that mart-spanning relationships be maintained globally.

After the data warehouse has data loaded, it is possible to begin to utilize the data for analysis. The multi-dimensional analysis process is completed using a variety of software products and provides the ability to manipulate the data by a variety of categories. The multi-dimensional analysis process is also referred to as “slicing and dicing” or “drilling-down”. Products that allow users to complete multi-dimensional analysis are referred to as OLAP tools. **OLAP** is an acronym for **On-Line Analytical Processing.**

The performance of data warehousing is largely dependent on query performance. Queries usually either select a very small set of data or perform aggregations on a fairly large data set. Although materialized views that stores pre-calculated aggregates are efficiently used to process queries with aggregations, this approach increases resource requirements in disk space and slows down updates because of the view maintenance problem. Multidimensional hierarchical clustering (MHC) of OLAP data overcomes these problems while offering more flexibility for aggregation paths. Clustering is introduced as a way to speed up aggregation queries without additional storage cost for materialization. End users can access OLAP Services by directly using the OLAP tools or they may access pre-defined reports via a Web browser.

**Benefits for the company in building the Data Warehouse:**

The integrated and homogeneous nature of a data warehouse, coupled with its innate reporting and querying strengths, should bring a broad spectrum of benefits:

* An integration of Company data from across many years
* Automation of the production and distribution of reports via email or the web
* Provision of an online analytical capability to authorized users at the state, regional, district and local levels
* Easier, quicker and more comprehensive reporting and querying capabilities
* More internally consistent and higher quality query and reporting
* Reduced staff time spent on data collection activities with more staff time available for data analysis.
* Retirement of redundant and/or obsolete data collection systems and a reduced data entry
* Improved reporting both from a timeliness and quality perspective with greatly enhanced capabilities for meeting Company’s increased reporting requirements

**Risks of Building the Data Warehouse:**

1. **Lack of strong project direction** - A data warehouse implementation is an incredibly daunting task. It will be resource intensive, expensive, and a complex undertaking that may last for several years. To assure project success, a strong project sponsor(s) must be attached to the effort. Not only must this individual have a clear and committed vision for the data warehouse, but must have the political and organizational clout to clear organizational and technical hurdles that may impede progress. Both the loss of a project sponsor or the replacement of a strong one by a weaker one can significantly increase the risk of project failure.

2. **Lack of, or misdirected, data warehouse missions and objectives** - A data warehouse implementation cannot be successful if it does not sufficiently meet the needs and expectations of its user community. This feasibility study attempts to identify and prioritize the missions and objectives of the data warehouse. Over the development and overall life of the data warehouse these missions and objectives should be regularly reviewed and revised based on evolving user community expectations.

3. H**elpdesk support, quality documentation and training -** The data warehouse is targeted for use by a variety of users with different levels of comfort and proficiency on computers. Insufficient support for users can critically undercut acceptance and use of a data warehouse and lead to criticisms of the tool that are not related to its development or implementation. Our company needs plans and budget support for all aspects of data warehouse operations including internal setup and maintenance (including ETL processes), internal staff training, help desk support, and product documentation

4. **Lack of a data steward function -** A data steward acts as a central policy and oversight body to assure the quality of data that is introduced into Company’s data collection systems, as well as providing a consistent policy for how collected information is used and distributed. Without this function, the data warehouse is highly susceptible to misuse of data, inconsistent policy decision-making, and an inability to evolve as the needs of Company change

5. **Inadequate preparation, distribution and training regarding privacy, security and confidentiality policies and procedures regulatory body is a driving force in determining appropriate use, access and distribution of information from the data warehouse.** To avoid conflicts, Company needs to have clearly define its privacy and confidentiality policies. The policies must be clear, universally applied, and understood throughout the agency. Even after the implementation of a data warehouse Company must continually communicate and train Company staff, as well as districts and schools, in the agency’s established policies

6. **Unrealistic expectations for the development timeline and the usability of the data warehouse.** To distribute the cost of building the data warehouse, Company will most likely use a phased implementation approach. This creates a risk in that users may not be able to immediately access the types of reports they expect from a data warehouse nor be able to complete the types of data queries they desire when the data warehouse is initially available. To avoid this risk, it will be important to communicate the project timeline and the phased implementation approach to internal and external users.

**Disaster Recovery**

An IT disaster recovery plan provides a process for resuming normal business and IT functions in the event of either short-term losses like temporary loss of electronic equipment or more significant disaster like flooding, long-term power outages, etc.

1. Identification of individuals who are responsible for data warehouse and compile a list of names and phone numbers who can access the data warehouse and carrying out recovery plans in case the data warehouse be affected by a disaster.

2. Identification of potential problems that will affect the operations and functionality of the data warehouse and development of recovery strategies and plans for how to address each problem. It is recommended to address both short-term and long-term disasters.

3. Determination of how Company Technical Support will respond when data source systems for the data warehouse are down or otherwise unavailable or when scheduled extractions of data are interrupted.  Detailed list of the servers used for the data warehouse with corresponding list of names and contact phone number of vendors where replacement parts for hardware can easily be obtained.

4. Define when scheduled data warehouse backups should occur, the format of the data backup and where the backup will be stored.

5. Review of Recovery plan on a regular basis.

**User Support** As with the introduction of any system to Company’s infrastructure, a data warehouse implementation will necessitate user assistance. Company’s existing help desk support infrastructure will need to be expanded to support a broad spectrum of data warehouse users. Because the user community will vary in both its level of sophistication, as well it access to the data warehouse, help desk staff must be able to offer assistance across a wide range of data consumers from casual users accessing the data warehouse the public interface to analysts needing assistance with reporting and querying tasks.

## Organizational Feasibility

Organizational feasibility analysis is conducted to determine whether a proposed business has sufficient management expertise, organizational competence, and resources to successfully launch its business. There are two primary issues to consider in this area: management prowess and resource sufficiency.

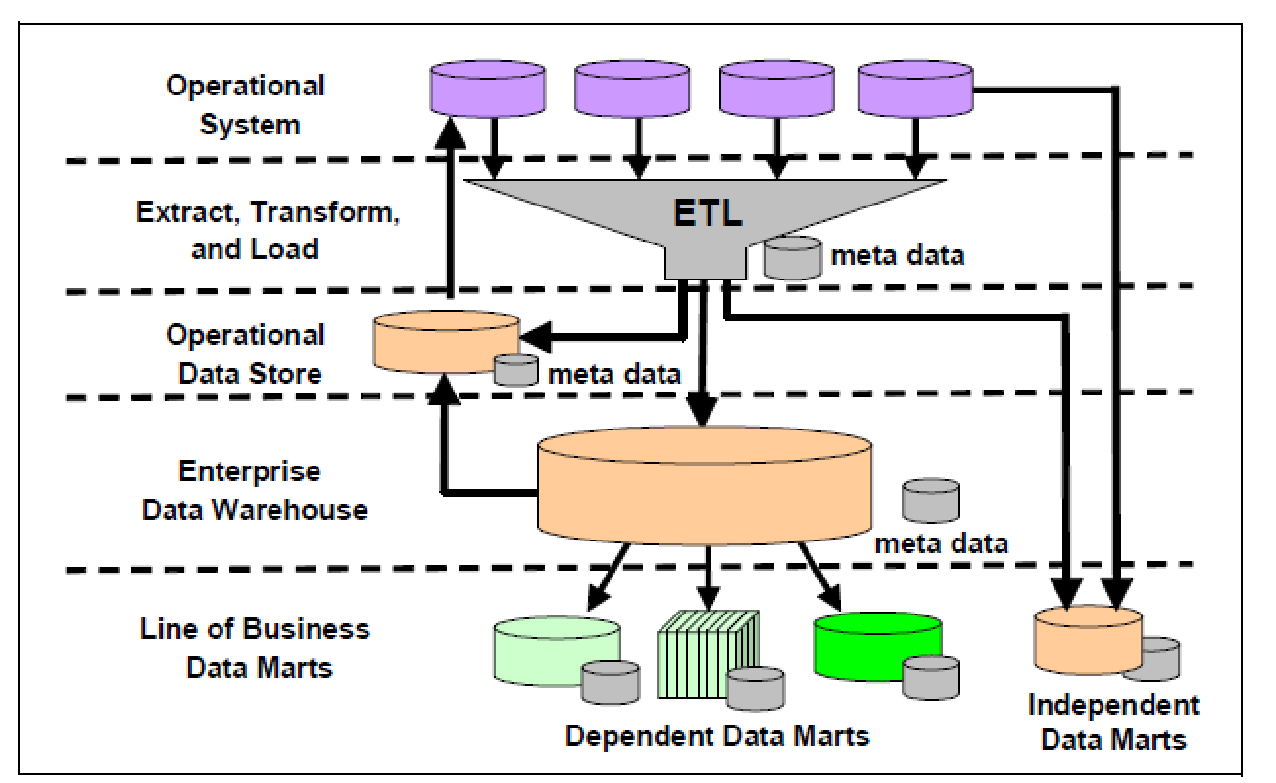
**Management Prowess:** A firm should candidly evaluate the prowess or ability of its management team. We believe that our management team has the proper understanding of the Data warehousing business and Retail Industry for which we are building this system and they can help us selecting the suitable and efficient technologies to build the system. Our president has built good relationship with the customer — that’s what it takes to win. Our management team with extensive professional and social networks has an advantage in that it is able to reach out to colleagues and friends to help them plug experience or knowledge gaps.

**Resource Sufficiency:** The second area of organizational feasibility analysis is to determine whether the potential new venture has sufficient resources to move forward to successfully develop a product or service idea. The focus in organizational feasibility analysis should be on nonfinancial resources, in that financial feasibility is considered separately. Several areas should be examined, including the availability of office space, the quality of the labor pool in the area where the business will be located, and the knowledge they possess about this new technology.

We have a team of 30 people having complementary skills required to build an information system. We have 5 system analysts who will gather requirements, create the use cases, data flow diagrams and help developers in understanding the business aspect of this system. We have 15 developers who have extensive experience in the area of data warehousing and BI. Our UI design team of 5 people will take care of the aesthetics and usability of the system. The design team has already created a sample User Interface for the report generation tool for the end user’s understanding. We also have a 5 person Project Management team who will coordinate the activities of project team and provide progress reports to the senior management. All of these resources are available for work for 40 hours per week.

**Organizational Assessment** includes an examination of the existing organizational structure and identification of the roles and responsibilities of both IT and the business community that need to be addressed. In conjunction with the Project Planning and Methodology Assessment (below), this review considers alignment of the project organization with the overall business and IT environments. Organizational readiness for warehousing is examined, including readiness to assume responsibility for ongoing technical and business support, hardware and software configuration management, continuing business requirements definition, and front end applications enhancement. Organizational assessment strongly focuses upon the organization’s ability to fulfill many warehousing roles. We believe that our organization is ready for planning and staging the overall implementation of the data warehouse.

## Architecture of the DATA WAREHOUSE system

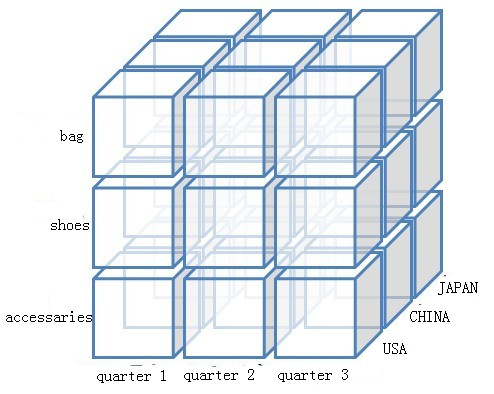


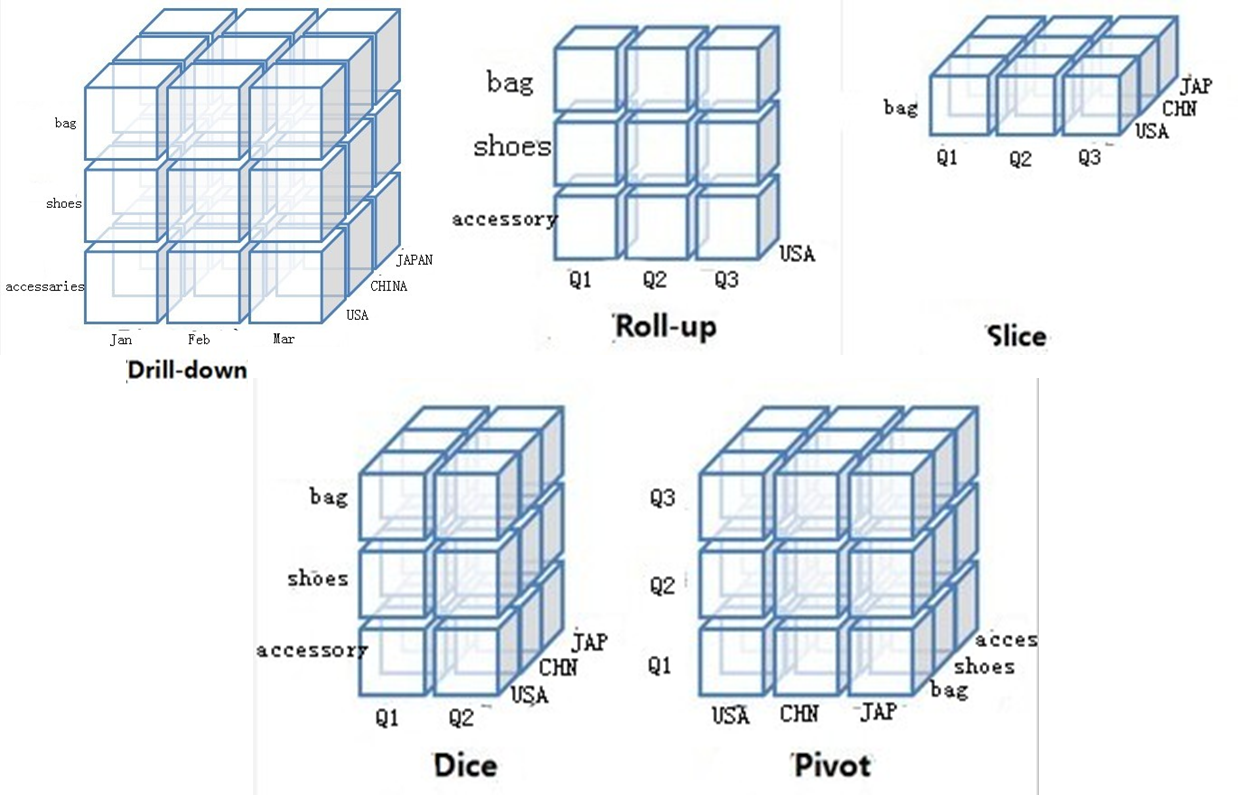
In our system, we are implementing data marts for marketing, market research and inventory departments.

DataMart approach: This approach extracts data from one or more databases in the operational system. This data is in pure form, ETL system transforms this data into cleansed, enhanced set and loads it into the primary Data Warehouse for DataMart application. The extraction has a special purpose. Hence, the repository of the data warehouse contains the subject of information, which is in DataMart format. The disadvantage of the architecture is that a data warehouse contains only limited knowledge. This approach cannot manage huge amounts of data. Use of the Datamart approach for the implementation of an enterprise’s data warehouse will develop many cubes and each cube is an independent data aggregation.

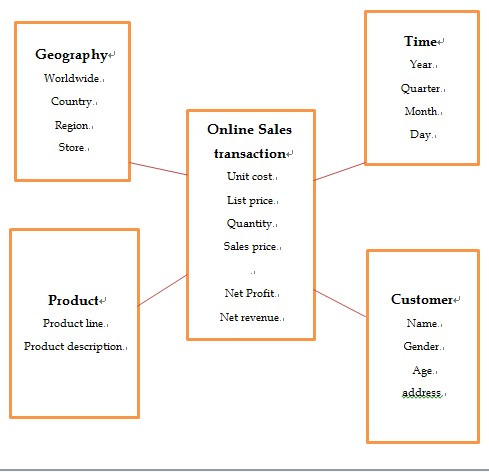
Users retrieve the knowledge from one single data mart and not from a global view; therefore, problems like data duplication, inconsistency and query integrity could occur.

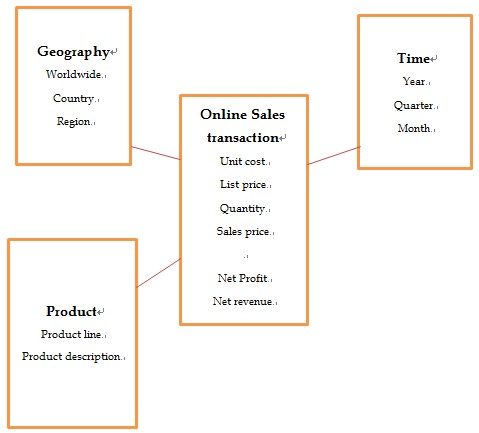
**E-Commerce cube**





Data marts might need updates if the requirement for the data mart changes over time, this update should be done during Data Warehouse update cycle to maintain consistency between DW and data marts.

**Aggregate tables:**



Aggregate table with atomic facts:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Online Sales Transaction | Product | | | | Geography | | | | Time | | | | | | | |
| Productline1 | | Productline2 | | Worldwide | | | | Year1 | | | | | Year2 | | |
| Product 1 | Product 2 | Product 1 | Product 2 | Country 1 | | Country 2 | | Quarter 1 | | Quarter 2 | | | Q1 | Q2 | |
|  |  |  |  | Region 1 | Region 2 | Region 1 | Region 2 | Month 1 | | Month 2 | | M1 | | M2 |
|  |  |  |  |  |  |  |  | D1 | D2 | D1 | D2 | D1 | | D2 |
| Unit cost |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |
| List price |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |
| Quantity |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |
| Sales price |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |
| Net profit |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |
| Net revenue |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |

## Dimensional Modelling

It’s commonly called the Star Schema. This model provides better query performance, especially on very large queries. It consists of large tables of facts known as fact table with number of other tables surrounding it that contain descriptive data termed as dimensions. This gives the resemblance of star. The Fact table has all the primary keys of dimension table and other numeric attributes which are changing frequently. Speed of data retrieval was good as all the data are got from the fact table. Strength of the dimensional modelling is symmetrical association i.e, predictable and standard framework, gracefully extensible to accommodate new data elements or new design decisions and common modelling situations in business environment. Gracefully extensible denotes that there is no need to take the copy of data, make changes and reload instead changes can be directly done and the application or query tool doesn’t require any changes. Existing query and report will continue and return the same answer set prior to changes.

**PRODUCT**

**CUSTOMER**

ProductID

ProductDesc

Quantity

CustomerID

CustomerContact

CustomerLocation

**FACT**

ProductID

CustomerID

CategoryID

YearID

MonthID

Sales

**TIME**

**CATEGORY**

Representation of Star Schema in Sales recording

YearID

MonthID

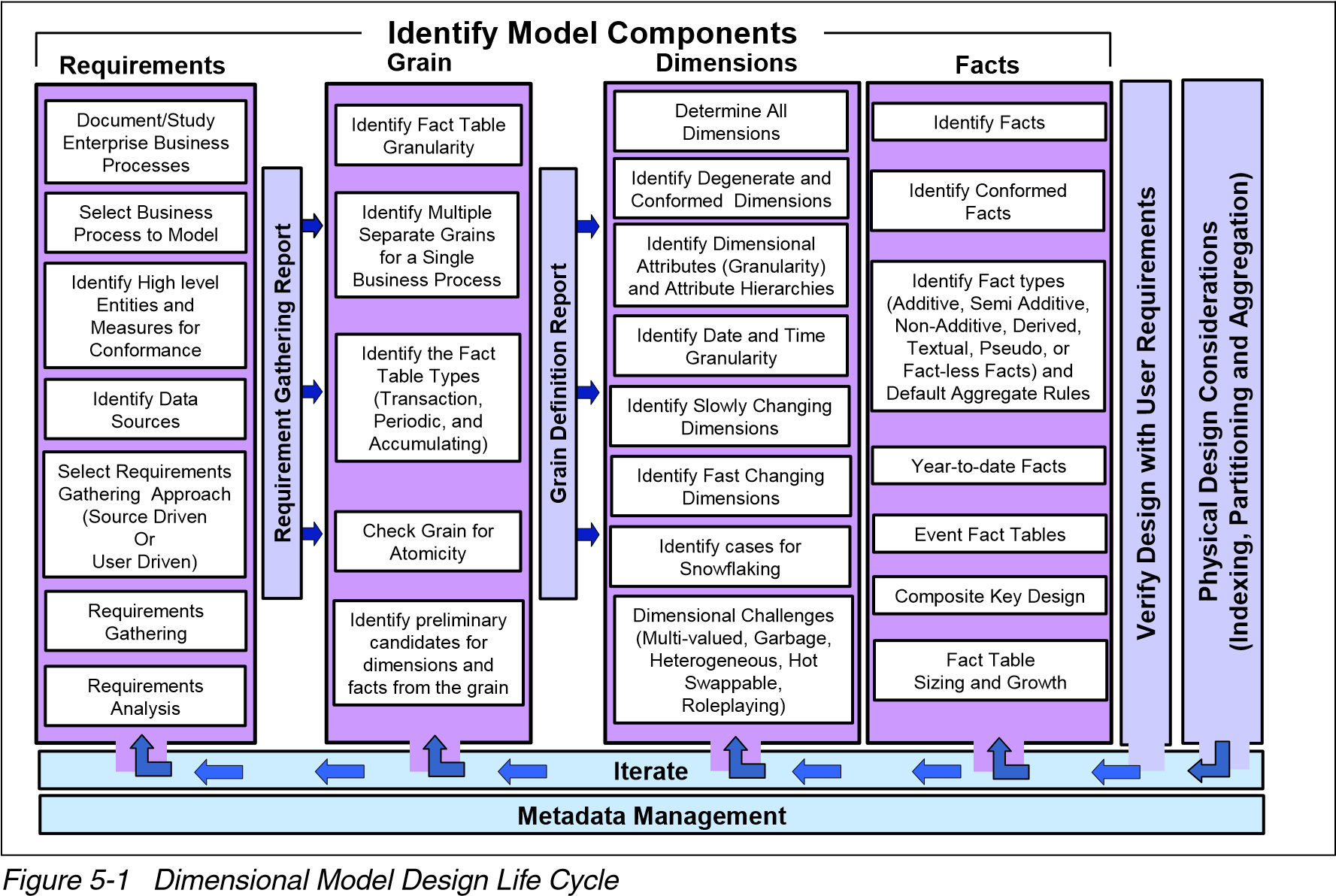
WeekID

CategoryID

CategoryName

CategoryDesc

The Dimensional model life cycle helps in identifying the phases and activities that are taken into account when designing the dimensional model. Below is the representation of Dimensional Model Life Cycle.



Dimensional

*Source: Dimensional Modelling: In a Business Intelligence Environment*

PHASES:

1. Requirements:

* Identify Business Process Requirements:

It involves selecting the business process based on which the dimensional model is being designed. Selection often considers criteria such as business process significance, quality of data in the source system, feasibility and complexity of the process.

The important business processes which are involved in the retail industry are:

* Retail Sales
* Finance
* Marketing
* Shipment

These business processes are assessed based on the assessment factors such as complexity, data availability, data quality, strategic business significance. The business processes are prioritized by giving points to the assessment factors. Next is finding out the high level entities. It involves the entities which are common among the business processes.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| High Level Entities  Business process | Product | Date of Sale | Customer | Vendor | Warehouse | Shipping |
| Retail Sales |  |  |  |  |  |  |
| Finance |  |  |  |  |  |  |
| Marketing |  |  |  |  |  |  |
| Shipment |  |  |  |  |  |  |

Next, we identify the different data sources which are involved with the business processes. We need to gather the requirements in order to get the data for designing the model. There are various methods for deriving the business requirements such as Source Driven and User Driven.

Next comes the requirement gathering which focuses on the business processes and the analysis activities in which the user is involved. It is the understanding of the problem domain for which modelling is done. Few business requirements for the sales division can be found below

|  |  |
| --- | --- |
| Business Requirement | Importance |
| Average sales quantity for this month in each product in each category | Medium |
| How much of each product did the customers order? | High |
| What are the top 10 products that are sold last month by total revenue? Who was the vendor for each product? | High |
| How many products are purchased together by customers? | Medium |
| How frequent customers are using the promotional offers provided? | Medium |
|  |  |

1. Identify the Grain:

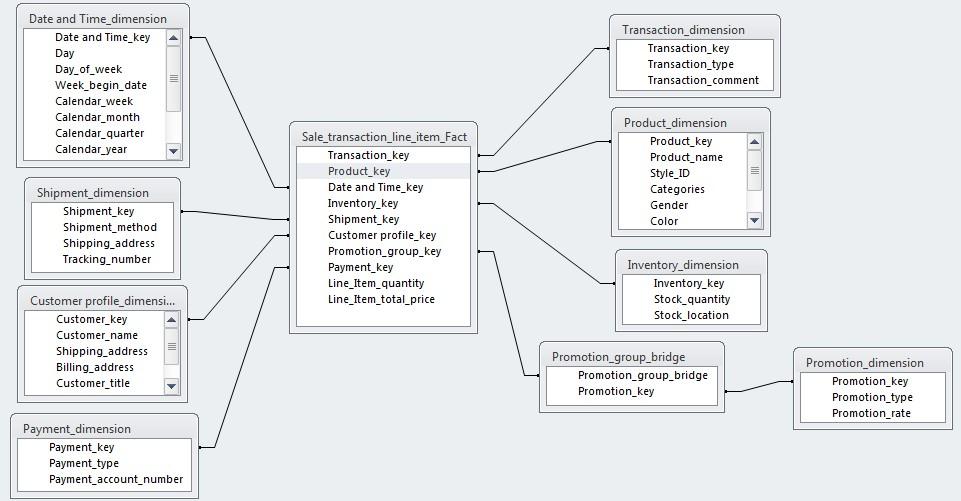
Grain is the level of detail available in the star schema .It refers to the level of detail associated with the fact table. The more the detail is, the lower the level of granularity. While choosing the granularity for the fact table, the following guidelines are taken into account.  
 During the requirement gathering phase, documents such as order form, invoice form are collected as they have transactional data associated with them, such as order number and invoice number. Documents often contain the lowest level of information that may be required. Date is an important consideration as we are storing the sales information at day, month and year level.

The Grain definition report is the final output report for this phase. It consists of multiple definitions of the grain for business process, the type of fact table (transaction, periodic) being used.

1. Identify the Dimensions:

Dimensions are identified that contain the lowest level of granularity of the fact table. Dimension tables contain the attributes that describe fact records in the fact table. Some of these attributes provide descriptive information. Dimension tables are relatively small, denormalized look up tables. It’s helpful to identify the restriction criteria’s.

Dimensions for the sales fact table is Time, Customer, Vendor, Product, Date, Order number.



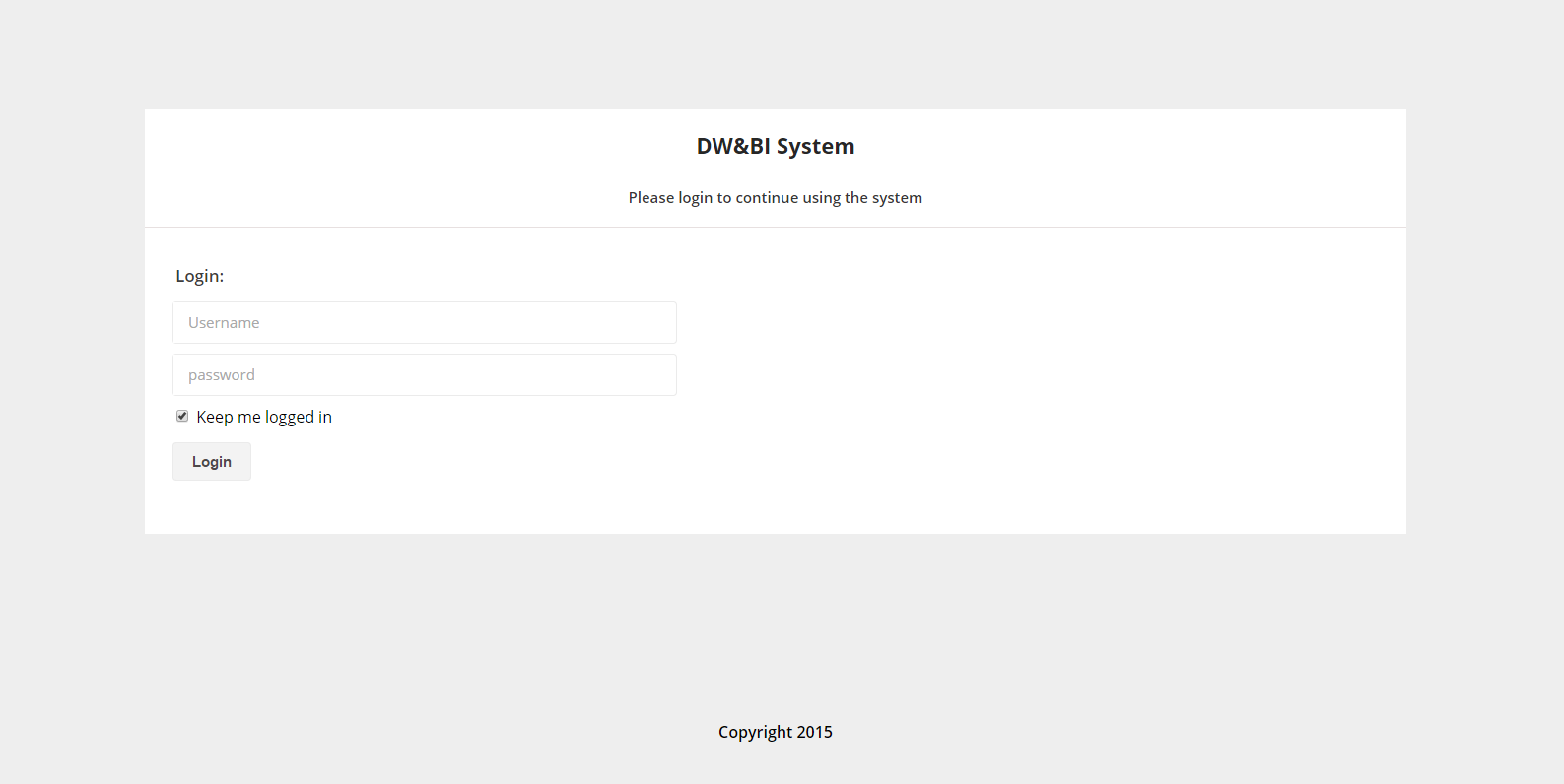
1. Identify the Facts:

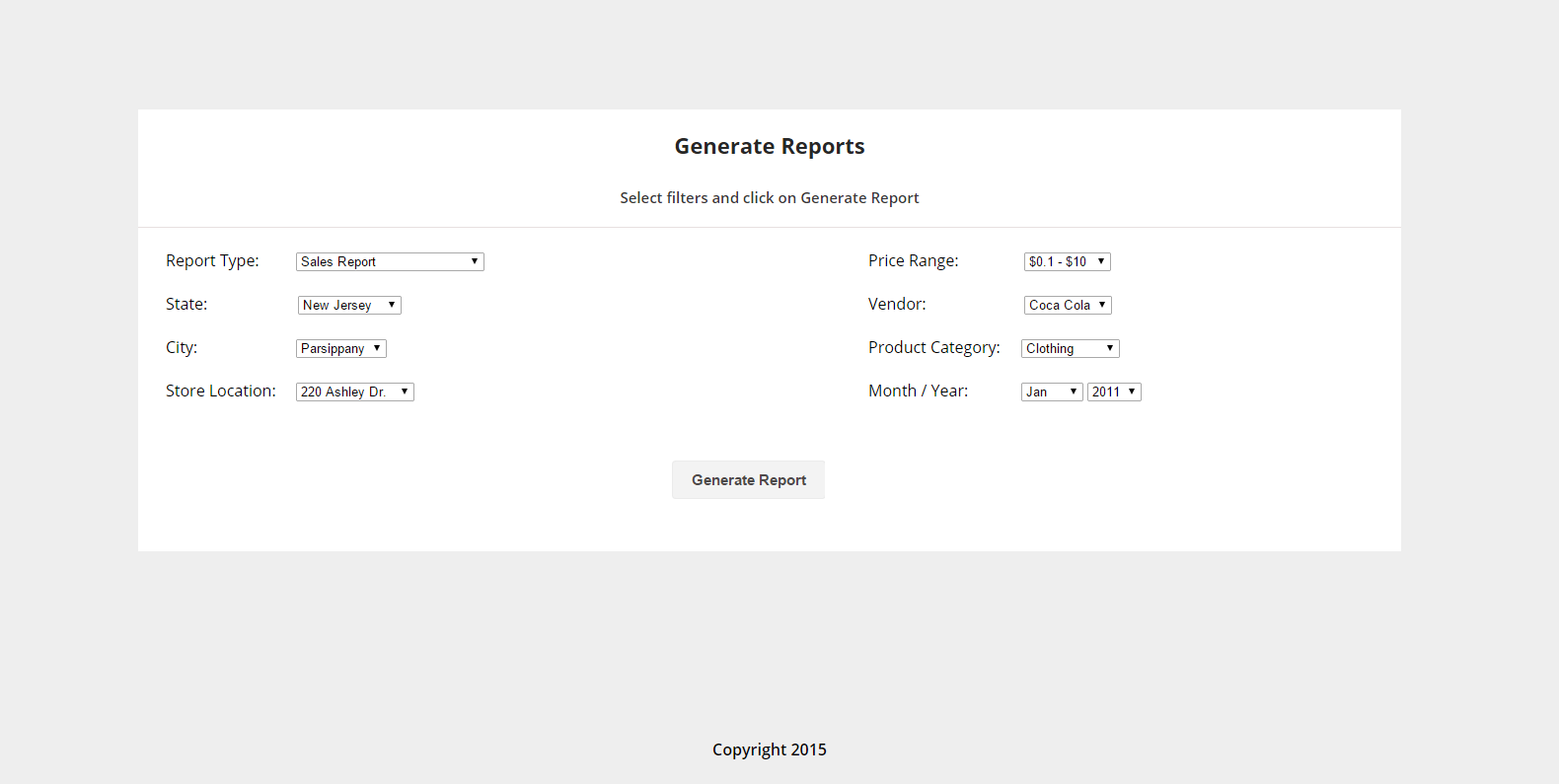
Facts true to the grain are identified. Some of the facts identified for our system is

|  |  |
| --- | --- |
| FACTS | FACT DESCRIPTION |
| Unit Sales price | Price of a single product |
| Quantity Sold | Quantity of each product that is sold |
| Amount | Number of units sold multiplied by Unit price |
| Discount | Discount given on single product |
| Promotional offers | Offers given on particular product |

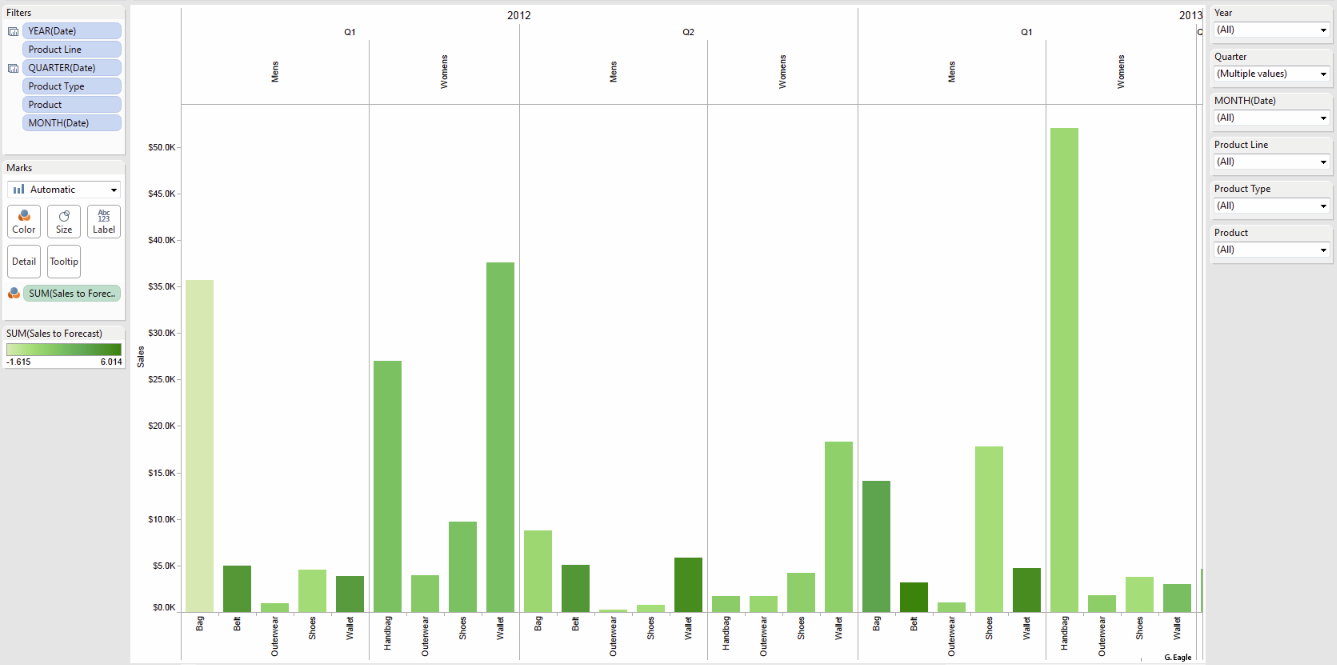
Thus the model is then verified with the user requirements collected. Then physical design is considered for managing the aggregation, indexing and portioning.

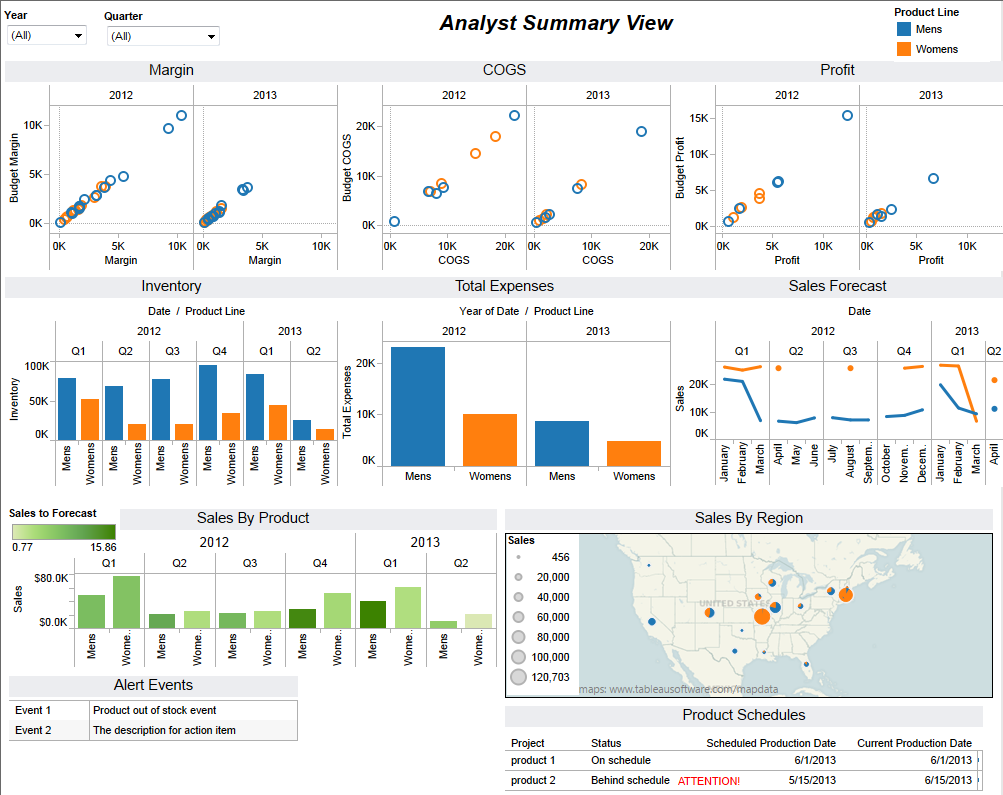
## User Interface











## Conclusion

The implementation of the data warehouse system coupled with Business Intelligence enhances the performance of the organization. It would articulate the following features in the organization

* The Users have access to the data they need to do the analysis is every aspect of sales.
* All data marts will share common identifiers and data definitions which supports the integrated reporting and analytics
* The multiple data marts are strong enough and facilitates the analysis that the user needs to perform in a specific functional area.

All these features enables to increase the efficiency in predictive analysis and reduces the user’s effort in retrieving the data. The future of the system is that the combination of DW with BI enables agility and maximum efficiency. The future developments include the management dashboards development, creating advanced financial reporting formats such as income and sales statements, integrating the planning and forecasting data with actual activity in a specific time period. Our users will expect to have easy access to reliable data and the trends in the data analytics will continue to drive the change.

## References

1.IEEE paper: Applications of DW in retail industry: <http://ezproxy.stevens.edu:2109/xpl/articleDetails.jsp?tp=&arnumber=1500153&queryText%3Ddata+warehousing+retail>

2. [http://www.Company.net/ILDS/pdf/feasibility\_study.pdf](http://www.isbe.net/ILDS/pdf/feasibility_study.pdf)

3. Dimensional Modelling: In a Business Intelligence Environment: <http://www.redbooks.ibm.com/redbooks/pdfs/sg247138.pdf>

4. ETL tool: <http://www.quora.com/Business-Intelligence/ETL-which-is-the-best-open-source-ETL-tool-to-start-working-specially-if-we-are-kick-starting-a-BI-project>

5. Science Direct paper: Data warehouse enhancement: A semantic cube model approach <http://www.sciencedirect.com/science/article/pii/S0020025506003719>

6. Stevens Library: The Data Warehouse Lifecycle Toolkit: Practical Techniques for Building Data Warehouse and Business Intelligence Systems. <http://web.a.ebscohost.com/ehost/ebookviewer/ebook/ZTAwMHhuYV9fNDEzMzUyX19BTg2?sid=3200e1c7-7568-4636-85ae-4a80d0e8293b@sessionmgr4005&vid=0&lpid=lp_Cover-2&format=EB>