

CHAPTER – 3

MATERIALS & METHODS

In this chapter, Research problems, research model, proposed assumption, techniques used to extract and analyze the data, and findings from the analysis are explained and discussed.

3.1 Objective and research Problem

According to various researcher research can be classified into three perspectives:

1. The application of the research study
2. The objectives in undertaking the research
3. The type of information sought

Due to above reasons it is very important in any research, clear demonstration of the problem that is problem definition which have a particular format or we can say some structured rules. According to prescribe rule of designing problem I find out my objective of research that is main problem of data ware house.

The adoption of data warehouse technology is costly and time-consuming with high probability of failure, compared with other information technology initiatives. Therefore, it is important to have a deeper understanding

of the factors which affect the adoption of data warehouse technologies. The research problem of this thesis can be portrayed as “what are the performance tuning strategies how we can improve our data ware house by different strategies”.

3.2 Outline of Research Design

Researcher have to state conceptual structure within which research could be conducted to achieve this research purpose it may be grouped into four categories

1. Exploration
2. Description
3. Diagnosis
4. Experimentation

There are several research design such as experimental or non experimental hypothesis testing but the preparation of research design, appropriate for particular research problem, involve usually the consideration of the following

1. The mean of obtaining the information
2. The availability and skill of the researcher and his staff
3. Explanation of the way in which selected means of obtaining information be organized and the reasoning leading to the selection
4. The time availability of research
5. The cost factor relating to research, i.e. the finance availability for the purpose.

3.3 Research Procedure

In dealing with any real life problem it is often found that data at hand are inadequate, and hence, it becomes necessary to collect data that are appropriate. There are several ways of collecting the appropriate data which differ considerably in context of money costs, time and other resources at the disposal of the researcher.

After deciding the most suited research strategy I have to decide how to collect data for my purpose. Because my research needs an in-depth data, and also considering the nature of qualitative research, I use documentation as my main source of data collection. In addition the use of interviews helps us to gather valid and reliable data that are relevant to my research questions and objectives.

I tried to arrange interviews with professionals working in companies which are already working with data warehouses. I mainly use semi-structured or unstructured interviews in this regard because I don't have any formal interview requirements. All the correspondence through email or on phone as the companies are not situated in the city. I asked them questions like how they are carrying out management of their data warehouses, what they think about data warehouse management and how it can be improved. To complement the interviews and gain as much information as possible, I also use the documents from the company's website if available.

In addition I gathered data from the internet including articles, research papers etc. The data from internet also help us a lot because I can find the latest developments in this field on the internet.

For the purpose of case study I have selected BSNL Bikaner, Sriganagnagar, Nagaur, Churu branches. BSNL has been using the data warehouse for more than five years and according to their staff their data warehouse is working properly and giving them the desired results. The data warehouse has become a very popular information source among all the divisions of BSNL and nearly all the business users are using the data warehouse in some way or the other, which shows that the data warehouse is properly maintained and given due support to function properly.

I arranged interviews with data warehouse project manager, data warehouse staff and the data warehouse users at BSNL involved in the day to day operations and the management of data warehouse. Interviews with these technical staff enable us to rightly analyze the best management policies in practice from a performance perspective. Keeping in view the location of the company and the cost of telephonic interviews I have decided to use some person as an intermediate interview. For this purpose I need a person who can communicate easily with us and the case site. I have identified the Nodal Officer of the project and General Manager, Asst. General Manager, Divisional Engineer, Sub Divisional Engineer, Junior Technical Officer, Chief Account Officers, Account Officers, Asst. Account Officers and clerical staff etc. also at BSNL. I send my interview questions to him and he sends us the interview response afterwards.

3.4 Research Methodology:-

3.4.1 Tools

Before presenting our approach let us briefly review .net platform, SQL Server, the store procedures. The Cluster index and Non cluster indexing we have already discuss.

.NET Framework

The **.NET Framework** is a software framework developed by Microsoft that runs primarily on Microsoft Windows. It includes a large library and provides language interoperability (each language can use code written in other languages) across several programming languages. Programs written for the .NET Framework execute in a software environment (as contrasted to hardware environment), known as the Common Language Runtime (CLR), an application virtual machine that provides important services such as security, memory management, and exception handling. The class library and the CLR together constitute the .NET Framework.

The .NET Framework's Base Class Library provides user interface, data access, database connectivity, cryptography, web application development, numeric algorithms, and network communications. Programmers produce software by combining their own source code with the .NET Framework and other libraries. The .NET Framework is intended to be used by most new applications created for the Windows platform. Microsoft also produces a popular integrated development environment largely for .NET software called Visual Studio.



Figure – 3.1 .Net Framework

In the .net OS manages the resources, the processes and the users of the machine Provides to the applications some services (threads, I/O, GDI+, DirectX, COM, COM+, MSMQ, IIS, WMI, ...) CLR is a separate process in the OS. CLR manages the execution of the.NET code and manages the memory, concurrency, security etc. It has rich object-oriented library with fundamental classes, input-output, collections, text processing, networking, security, multi-threading etc.

.NET is an integral part of many applications running on Windows and provides common functionality for those applications to run. It Unified technology for development of almost any kind of applications like GUI, Web, RIA, mobile, server and cloud etc .NET Framework is an environment for developing and executing .NET applications Unified programming model, set of languages, class libraries, infrastructure, components and tools for application development Environment for controlled execution of managed code. It provide environment for developing and executing .NET applications, Unified programming model, set of languages, class libraries, infrastructure, components and tools for application development, Environment for controlled execution of managed code.

Microsoft SQL Server

It is a relational database management system developed by Microsoft. As a database, it is a software product whose primary function is to store and retrieve data as requested by other software applications, be it those on the same computer or those running on another computer across a network (including the Internet). There are at least a dozen different editions of Microsoft SQL Server aimed at different audiences and for different workloads (ranging from small applications that store and retrieve data on the same computer, to millions of

users and computers that access huge amounts of data from the Internet at the same time).

Microsoft SQL Server 2005 is a database platform for large-scale online transaction processing (OLTP), data warehousing, and e-commerce applications; it is also a business intelligence platform for data integration, analysis, and reporting solutions.

It allows designing, creating, and managing multidimensional structures that contain detail and aggregated data from multiple data sources, such as relational databases, in a single unified logical model supported by built-in calculations. Analysis Services provides fast, intuitive, top-down analysis of large quantities of data built on this unified data model, which can be delivered to users in multiple languages and currencies. Analysis Services works with data warehouses, data marts, production databases and operational data stores, supporting analysis of both historical and real time data. We can use a combination of these features and tools to discover trends and patterns that exist in your data, and then use the trends and patterns to make intelligent decisions about difficult business problems.

3.4.2 Interviews And Survey Design

Both interviews and a survey were utilized to gather relevant data. Two formal interviews, each lasting about two hours, were conducted as a form of group discussion. Managers and business customers from different functional departments provided input. In addition, an informal meeting was held on a weekly basis with a senior manager who was in charge of the data warehouse system. The informal discussions were highly informative and were held for about two months while the author was serving as a faculty intern at the company during a summer period. Through the interviews, information such as

the data warehouse architecture, system specifications, financial effects, system usage, and future plans was obtained. Interviews also revealed detailed values of the data warehouse as well as roadblocks to its effective use.

The survey was designed to gather relevant information from existing data warehouse users. The first part of the survey obtained demographic information about data warehouse users including job title, position and General information regarding their organization and Questions on system quality, information quality, service quality, and user satisfaction with the system and with the information produced were included in the final section of the survey.

Most question items of the survey were developed to reflect the data warehouse environment. The database manager in charge of the data warehouse system also reviewed the survey questions from a practitioner's perspective to ensure they made sense to end users. The survey went through this iterative validation processes before it was distributed to the target users of the data warehouse.

3.4.3 Survey Data Collection

Data were collected from randomly selected data warehouse users from the company's 14 functional departments including finance, marketing, operations, and information technology (IT). There were Nodal Officer of the project and General Manager, Asst. General Manager, Divisional Engineer, Sub Divisional Engineer, Junior Technical Officer, Chief Account Officers, Account Officers, Asst. Account Officers and clerical staff etc at the company. This was in line with the survey result from Watson and Wixom [1997], which showed that senior executive officers were not directly exploiting the data warehouse as end-users. Senior managers, although not system users themselves, were the

main consumers of the reports produced from the data warehouse. Many respondents represented first-line and middle management. In addition, a significant number of contracted consultants were included in the selected sample. They were utilizing the data warehouse to undertake necessary services to the company. Heavy usage of the system by first-line and middle managers became the initial indication of the importance of the data warehouse in supporting managerial tasks.

An additional solicitation was sent to non-respondents to increase the sample size and to avoid a potential non-response bias. The response rate was 35 percent with 82 respondents. The IT manager attributed the relatively low response rate to the on-going system integration project as a result of recent acquisition of a large transportation company. BSNL have renewed their data warehouse technology recently. The reasons behind the replacement might be related to efficiency matters, or upgradeability to newer versions or overcoming problems experienced in the previous system. The following are the reasons behind changing the previous data warehouse, as stated by the respondents:

- Moving from one vendor to another for more flexibility, efficiency and automation of data storage, analysis and reporting.
- Moving from department-level to enterprise-level data warehouses.
- Upgradeability to newer versions.

The newly acquired company depended heavily on outsourcing for the maintenance and upgrade of its IT/IS infrastructure. Naturally, the lack of internal know-how became a major problem in undertaking the large scale system migration and integration.

3.5 Data analysis and discussion of research results

3.5.1 Experiment setup

The experimental setup was made in a PC environment where five computers had the same characteristics (Pentium 4, 2.8 GHz, window XP, 4GB RAM). I performed the tests using five computers. One of them was server and other four were clients. They were linked together on an Ethernet network. This section presents some performance results on a set of data taken from experiment. This dataset contains about 900000 records and the tests are conducted on searching queried attributes.

My experimental results show that grouping records search algorithm improves query response time of typical data ware house and improves scalability compared cluster indexing, non cluster indexing and full text indexing. We present the performance measurements in time to answer the retrieval queries.

3.5.2 Designation of post of respondent

To secure validity, relevance and reliability of data for analysis, it is important to ensure the relevance of respondents' backgrounds (educational and work-related background). The respondent should be an IT- and data warehouse-shrewd person and have regular interaction with a data warehouse. Such a person can be in the following positions Nodal Officer of the project and General Manager, Asst. General Manager, Divisional Engineer, Sub Divisional Engineer, Junior Technical Officer, Chief Account Officers, Account Officers, Asst. Account Officers and clerical staff etc. Divide these persons in top management, middle management and lower management. The figure below shows the distribution of respondent's title. The axis (x) represents the title of the post of the respondent and the axis (y) represents the percentage.

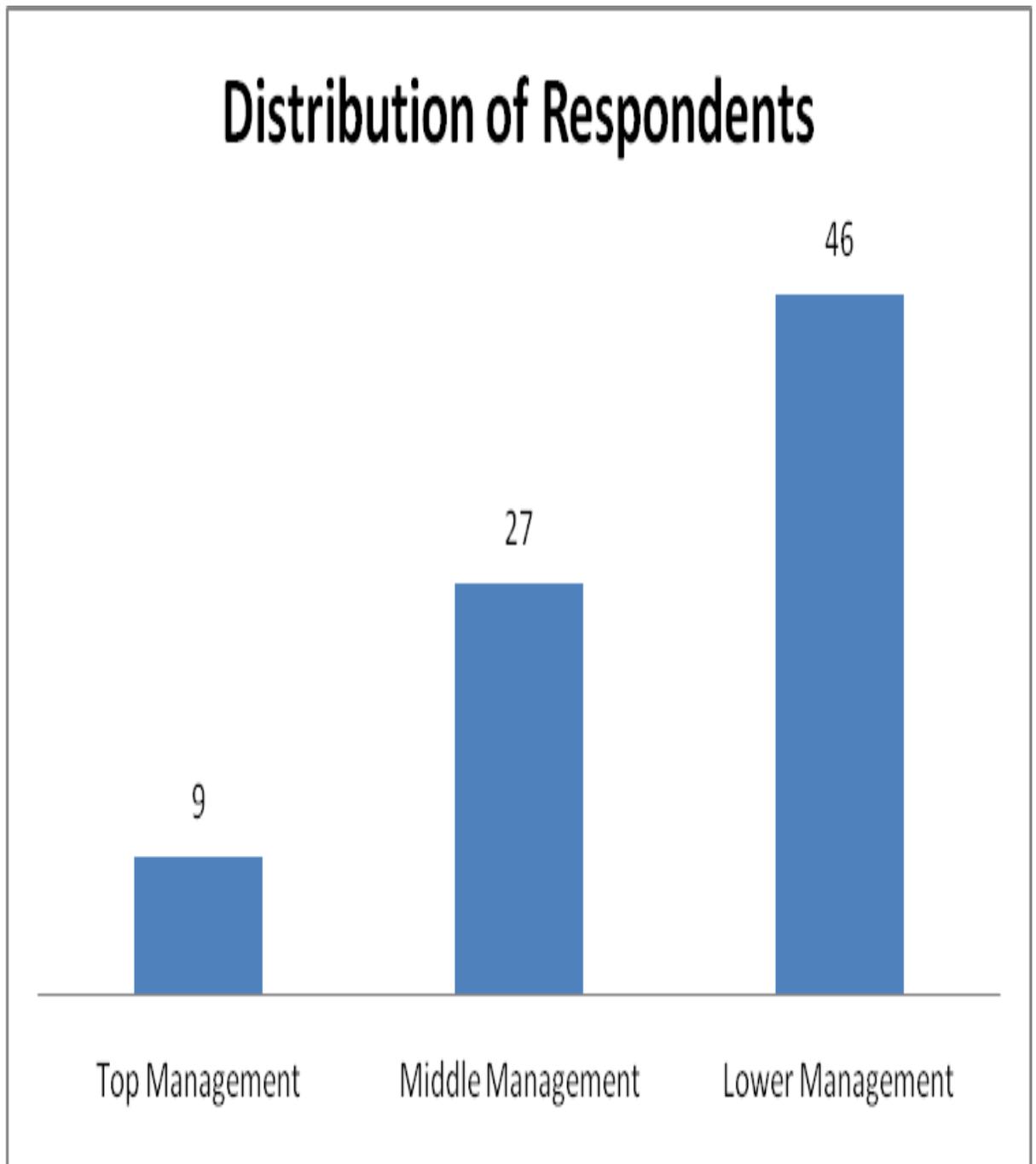


Figure 3.2 - Distribution of Respondent survey report chart

As noticed from the above figure, 09 respondents were top management 27 respondents were middle management and 46 respondents were lower management. These were playing different roles in strategic and managerial levels at their companies.

3.5.3 Analysis of the Data Quality

This section in the questionnaire was designed to collect basic data on the Data Quality; this section contains a mix of multiple choices. This questionnaire was design to get information regarding data that is current enough to meet their work needs and the data warehouse is accurate and reliable or not for their employees. The data warehouse maintains data at an appropriate level or not.

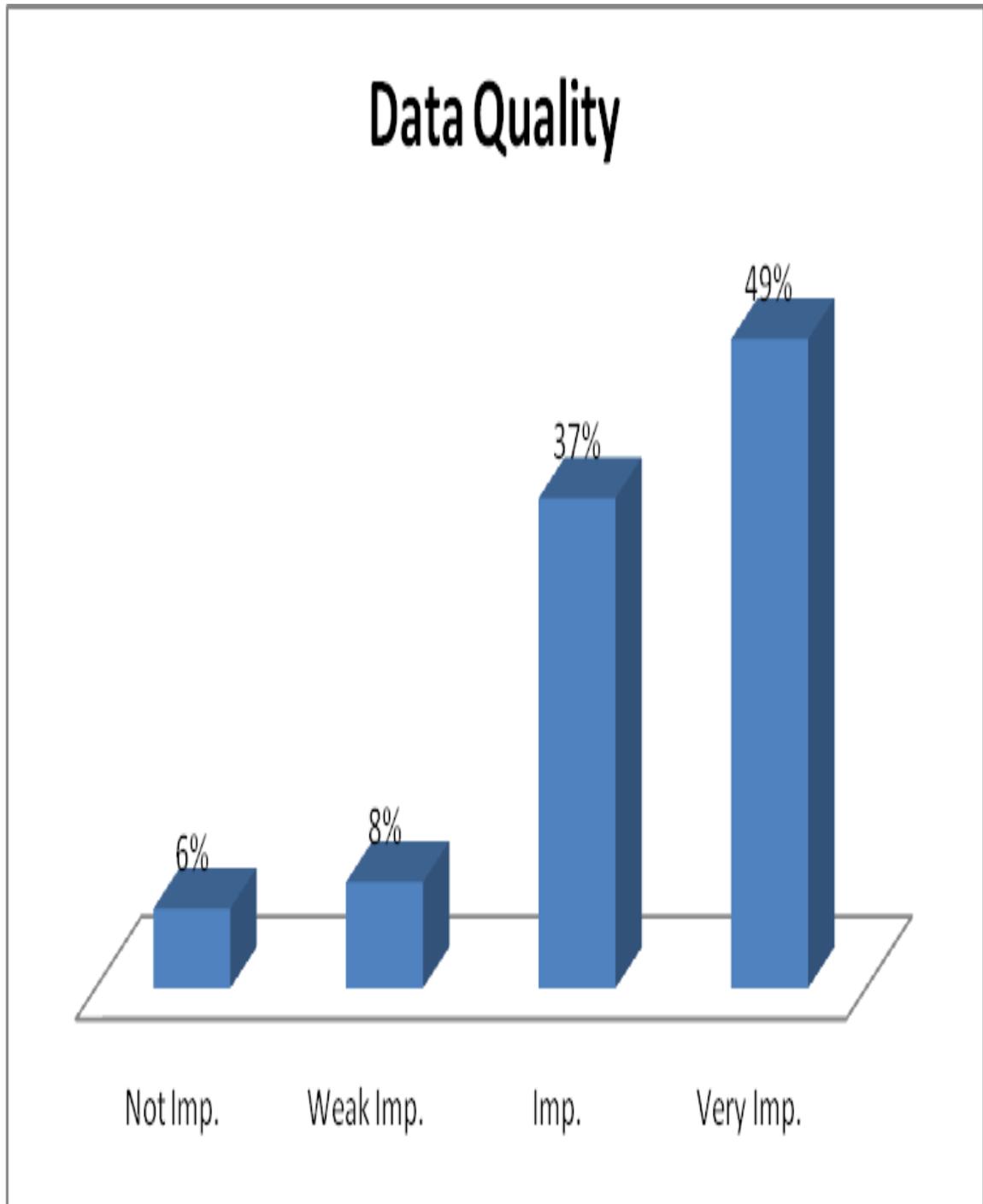


Figure 3.3 - Data Quality survey report chart survey report chart

As noticed from the above figure, 6% of the responses received from companies are Not Important, 8% of the responses received from companies are Weak Important, 37% of the responses received from companies are Important, and 49% of the responses received from companies are Very Important. By the help of this, it's come to know that data quality is very important for any organization. It must be reliable and accurate that fulfill requirement of the management.

3.5.4 Analysis of the Capability to Place data

This section in the questionnaire was designed to collect basic data on the capability to place data; this questionnaire was design to get information regarding data on a particular issue in the data warehouse, even if haven't used that data before and the data warehouse provides the level of detail that satisfies managements requirements or not

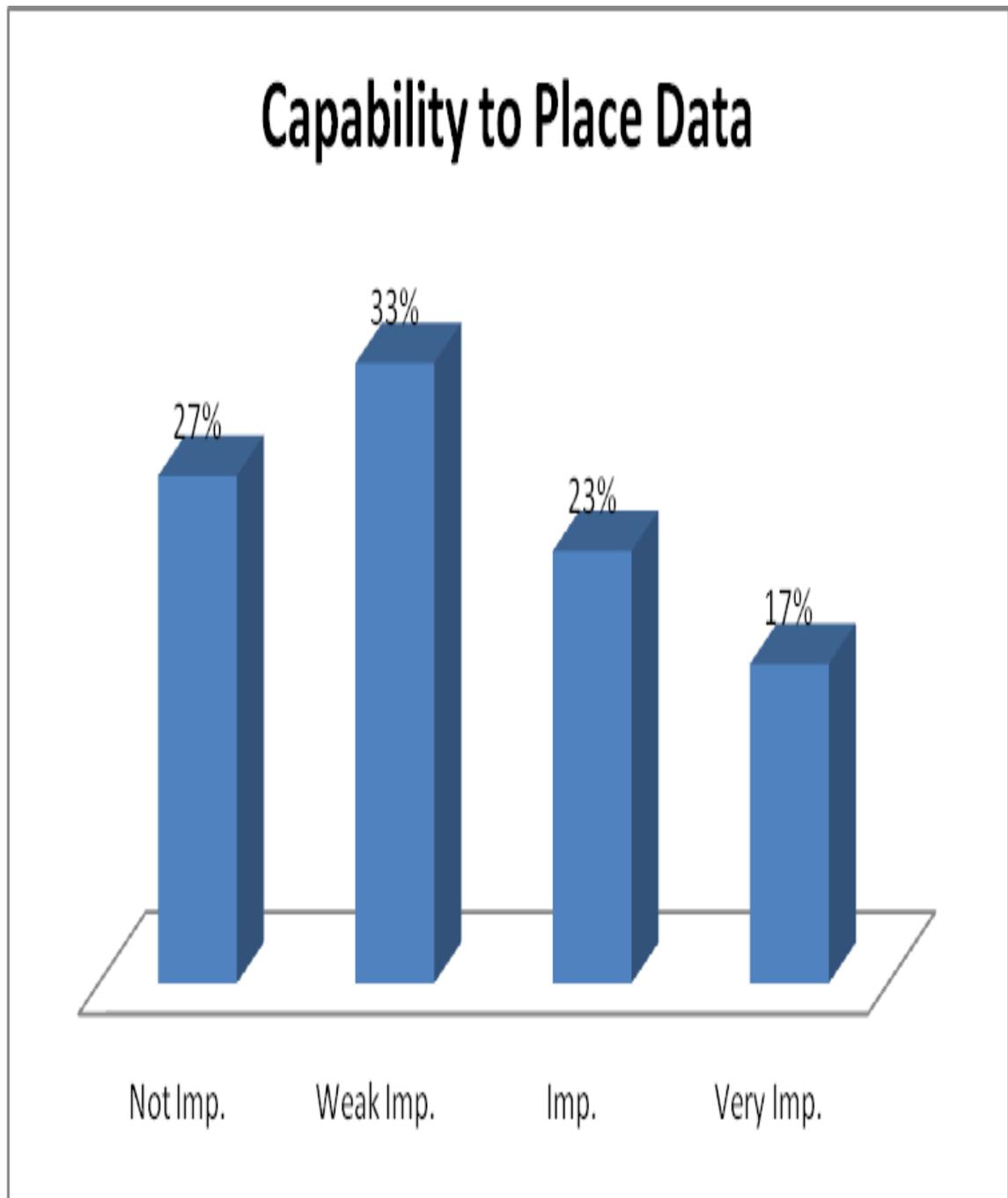


Figure 3.4 - Capability of Place Data survey report chart

As noticed from the above figure, 27% of the responses received from companies Not Important 33% of the responses received from companies are Weak Important 23% of the responses received from companies Important 17% of the responses received from companies Very Important By the help of this analysis it's come to know that current scenario data ware house not completing managements requirement. They not feel so good on some particular issue.

3.5.5 Analysis of the Access Authorization

This section in the questionnaire was designed to collect basic information of authorization. It is play an important role in security purpose. In this questionnaire the question was they want necessary authorization to access all the data useful for their task.

Its result was so good 57% of the management opinion was very important. Because it's important for their work responsibility and data quality and data speed because of that they were not depend to other.

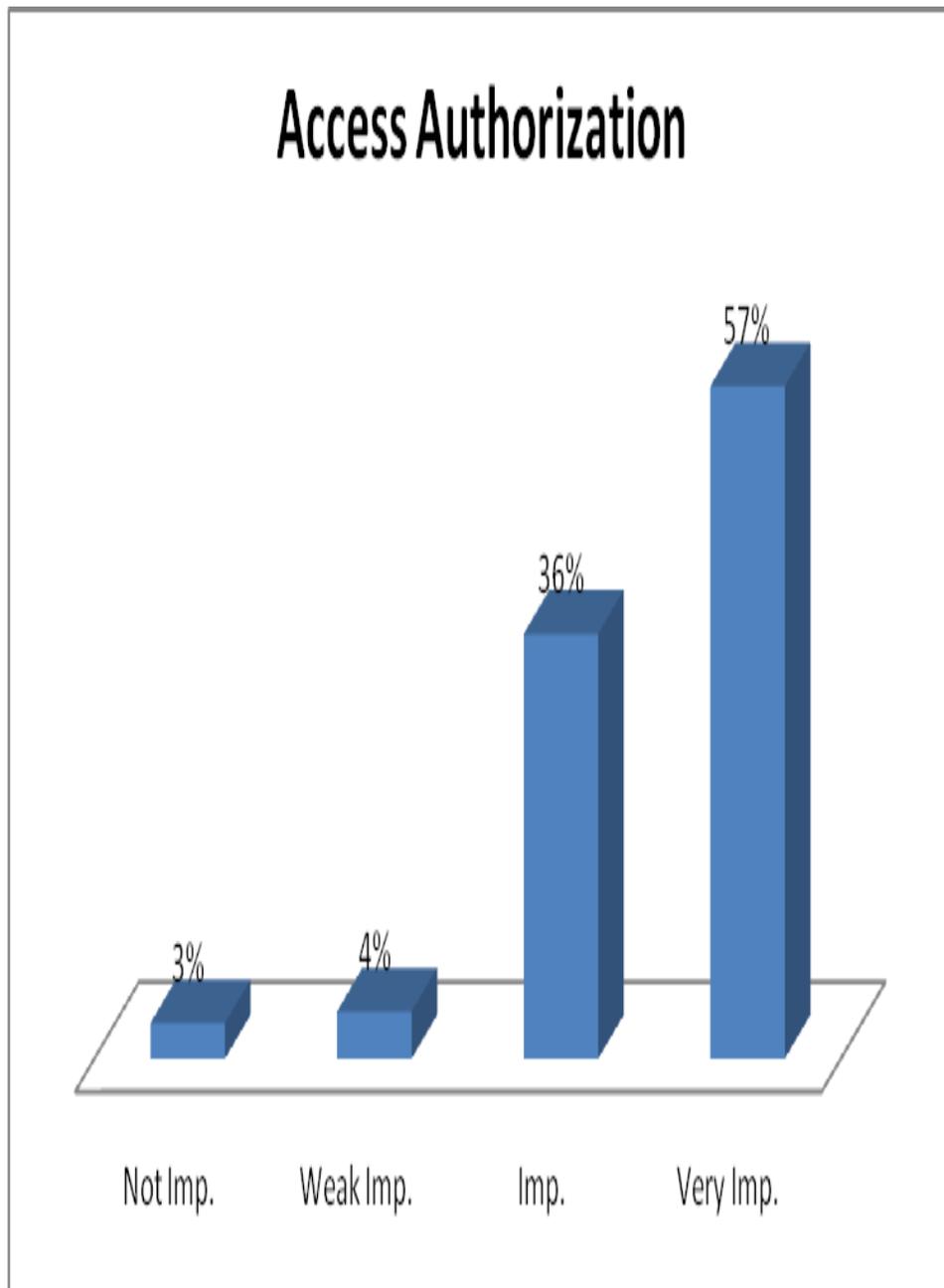


Figure 3.5 - Access Authorization survey report chart

As observe from the above figure, 3% of the responses received from companies Not Important 4% of the responses received from companies are Weak Important 36% of the responses received from companies Important 57% of the responses received from companies Very Important

3.5.6 Analysis of the Satisfaction

Employees and management satisfaction play an important role for any organization success. In this questionnair I ask the question that management satisfied with the overall quality of information generated from the data warehouse and the overall performance of the data warehouse system in its functionality, flexibility, and processing speed..in this analysis its show some employees and management is satisfy of the performance of the data ware house and other are not satisfy.because positivily and negetivily has not mojour diffrenence show to their respose as absolve the following chart.

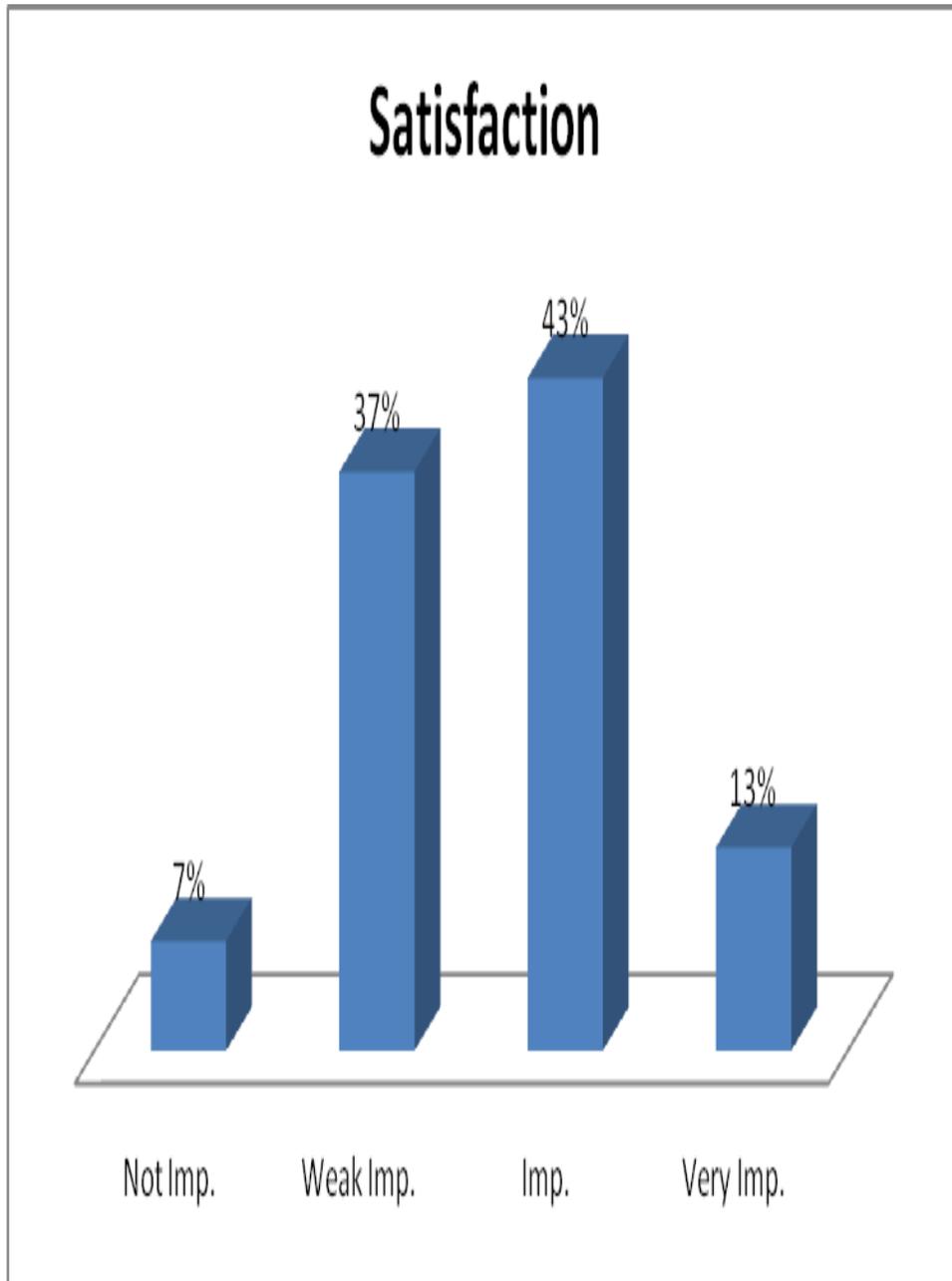


Figure 3.6 - User Satisfaction survey report chart

In this chart its observe that 7% of the responses received from companies Not Important 37% of the responses received from companies are Weak Important 43% of the responses received from companies Important 13% of the responses received from companies Very Important. Its shows weak important and Important views difference is only 6%.that may be some reason like data quality, authorization etc.

3.5.7 Analysis of Simplicity of Use

Simplicity of use mean The data warehouse system is convenient and easy to use for employees and

Simplicity of Use

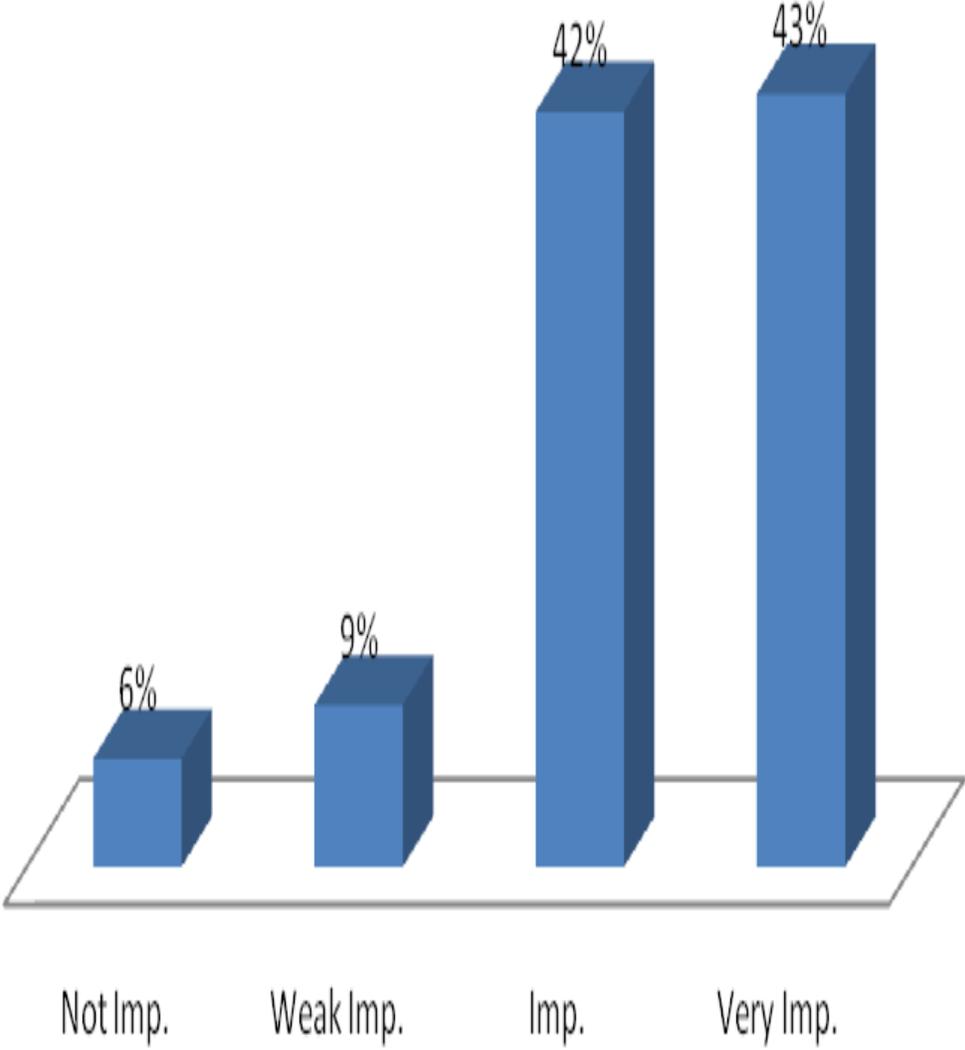


Figure 3.7 - Simplicity survey report chart

management of the organization. Its depend on application of the dataware house.By the review of this figure its shows that management are very esay to use this dataware house..

in this chart its examine that 6% of the responses received from companies Not Important 9% of the responses received from companies are Weak Important 42% of the responses received from companies Important 43% of the responses received from companies Very Important.

3.5.8 Analysis of User Training and Education

The quality of employees and their development through training and education are major factors in determining long-term profitability of a small business. Training needs to be focused on data warehouse concepts and terminologies, introduction to the organizational data, where is that located in the warehouse and how it is related to the reports or systems user already is using, the mechanics of using the tool. It is important for people to understand basic navigation within the tool.

User Training and Education

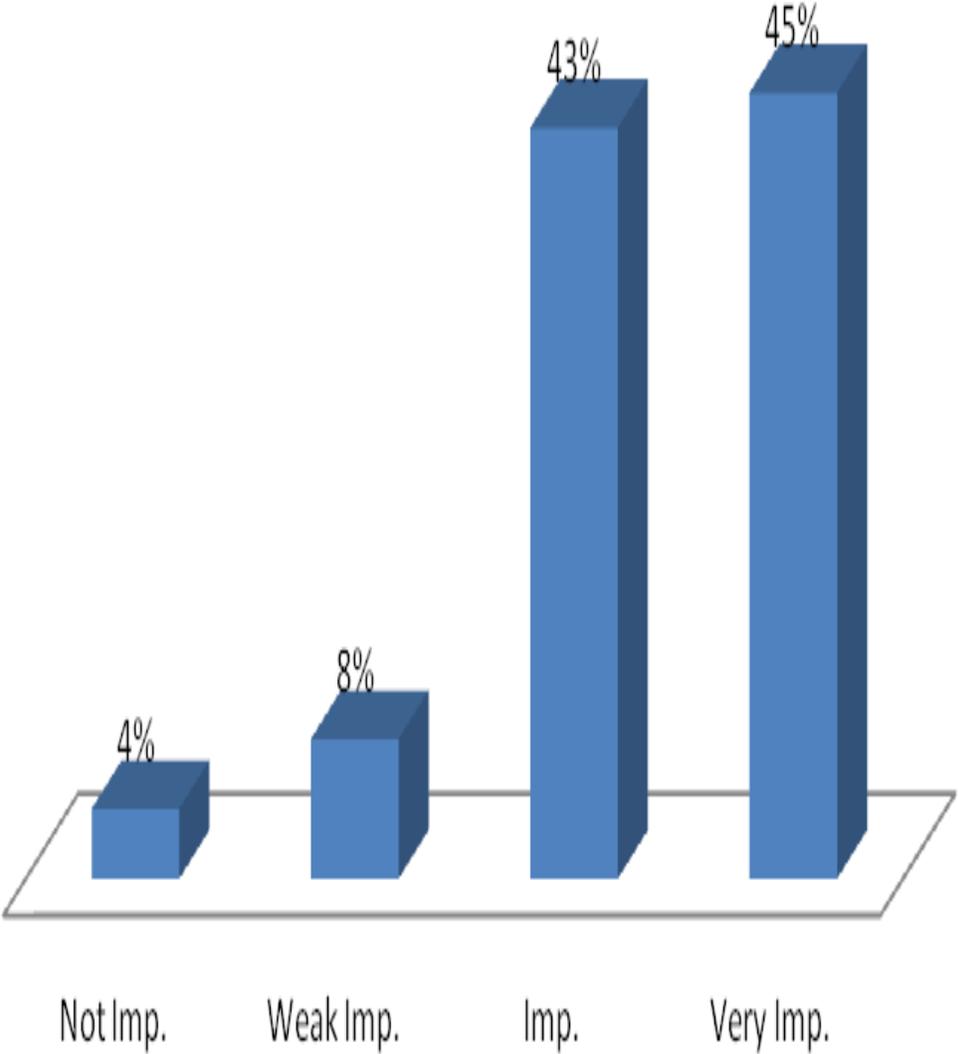


Figure 3.8 - User Training and education survey report chart

In this chart its observe that 4% of the responses received from companies Not Important 8% of the responses received from companies are Weak Important 43% of the responses received from companies Important 45% of the responses received from companies Very Important. Training of data warehouse users is significant and provides the desired output so management gives higher priority to user training and education.

3.5.9 Analysis of System Performance

This analysis is useful because its check the throughput of the system performance. Its check current scenario that management feels trouble to accessing the data warehouse and am forced to wait or not.

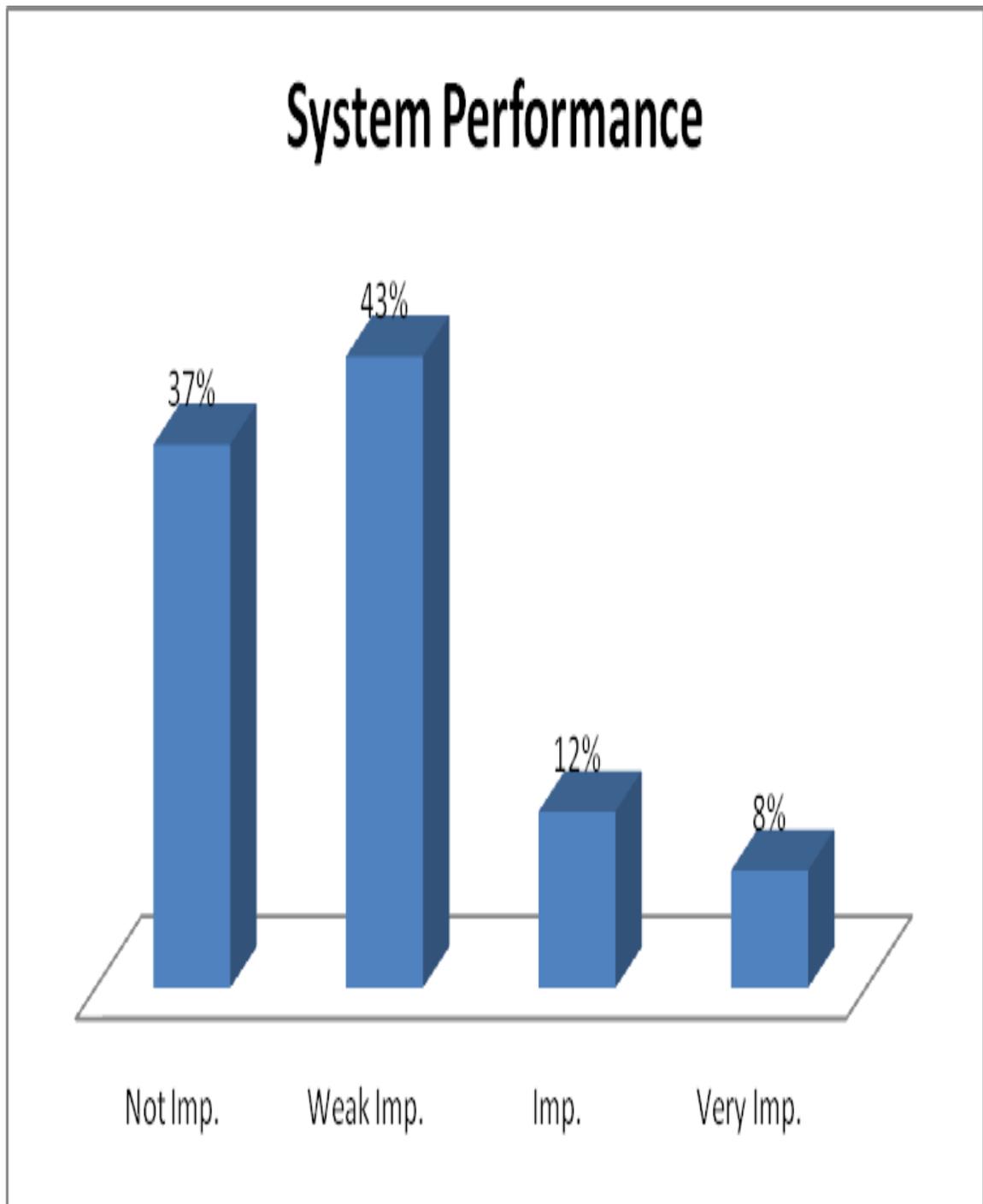


Figure 3.9 - System Performance survey report chart

In this chart its detect that 37% of the responses received from companies Not Important 43% of the responses received from companies are Weak Important 12% of the responses received from companies Important 8% of the responses received from companies Very Important. Its show that current data ware house performance is not so good because its response to not important and weak important is very high. Here not important and weak important mean system performance of present system output is not satisfactory.

3.5.10 Analysis of Information Usefulness

This analysis is valuable because it's about current data which is available in data ware house is useful and productive or not and the information from the data warehouse is important to work or not. Without accurate and reliable data there is no use data. Because it helps in decision making for future plans.

Information Usefulness

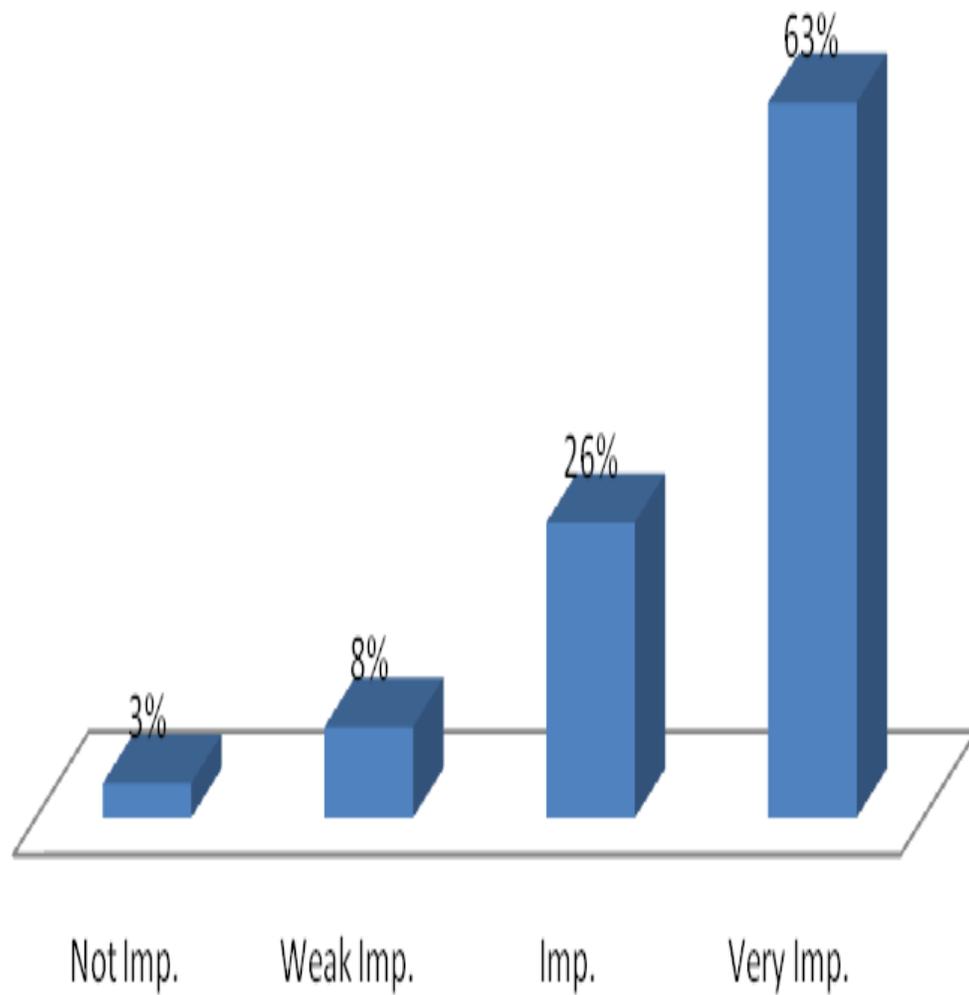


Figure 3.10 - Analysis of Information Usefulness survey report chart

In this chart its identify that 3% of the responses received from companies Not Important 8% of the responses received from companies are Weak Important 26% of the responses received from companies Important. 63% of the responses received from companies Very Important.

Its shows management is well known value of reliable and accurate information. Any organizations can success if organizations have useful and latest trend information of the market. it can have accurate if all the performance tuning strategies work properly.

3.6 Data Analysis

modes of data analysis provide ways of discerning, examining, comparing and contrasting, and interpreting meaningful patterns or themes. Meaningfulness is determined by the particular goals and objectives of the project at hand: the same data can be analyzed and synthesized from multiple angles depending on the particular research or evaluation questions being addressed. The varieties of approaches including ethnography, narrative analysis, discourse analysis, and textual analysis correspond to different types of data, disciplinary traditions, objectives, and philosophical orientations. HoIver, all share several common characteristics that distinguish them from quantitative analytic approaches. In quantitative analysis, numbers and what they stand for are the material of analysis. By contrast, qualitative analysis deals in words and is guided by fewer universal rules and standardized

procedures than statistical analysis.

They further say that data analysis consists of three concurrent flows of activities. These are reducing data, displaying data and drawing conclusion and verifying the conclusion. Sometimes it is assumed that data reduction is not part of data analysis but it's not true. Data reduction helps in making data sharp, sorted, focused, discarded, and organized in order to be able to draw and verify conclusion. Miles and Huberman conclude by saying 'I have few agreed-on canons for qualitative data analysis, in the sense of shared ground rules for drawing conclusions and verifying their sturdiness.

As I have selected qualitative research so I attempt to gather data from several sources to aid in the validation of the data collection The analysis of qualitative data is not nearly as straightforward as quantitative data and requires a great deal more thought and effort to do well. Qualitative data analysis is not as easy as statistical data analysis.

I use inductive reasoning for this purpose. First I tried to make some observations on data warehouse management on a small scale and then I draw inferences about data warehouse management on a large scale. I also use comparisons to compare the theoretical findings with the empirical findings. This help in understanding which things are common and which are not between the two and why. This help us in validating my research findings as well.

3.7 Validity and Reliability

The concept of validity refers to quality and can be applied to any aspect of the research process. With respect to measurement procedures it relates to whether a research instrument is measuring what it is used to measure. There are two approaches used to establish the validity of a research: the establishment of a logical link between the objectives of a study and the questions used in an instrument, and the use of statistical analysis to demonstrate this link. There are three types of validity which are: face and content, concurrent and predictive, and construct validity.

Reliability is defined as the extent to which a questionnaire, test, observation or any measurement procedure produces the same results on repeated trials. In short, it is the stability or consistency of scores over time or across raters. Keep in mind that reliability pertains to scores not people. The reliability of a research method refers to its ability to produce consistent measurements each time. When I perform any research under same or similar condition to the same or similar population and obtain similar results I say that the research is reliable.

I get validity and reliability in my research by consulting with the documentation as well as with the industry professionals working with data warehouses. Then I compare the results found in the documents to those what the professionals found while working with data warehouses. In this way I could say that I have achieved some form of validity and reliability in my research work.

To further validate the research findings I collect data from reliable sources, such as data warehouse administrators who are in charge of data warehouse management and interview questions I made based on

literature review to ensure validity of research. If required I may be able to also use triangulation where I gathered information from multiple resources and use that information to support my own findings.

3.8 Research Limitations

This research has limitations in the research methodology used. First, this study has strong attributes of a case study because the data were gathered from employees of a single company. Accordingly, certain system or non-system issues discussed here might be phenomena local to the studied organization.

In that sense, this is an intensive rather than an extensive study in which the generalize ability of analysis results may be lacking.

Meanwhile, the particular data warehouse system was recognized among industry watchers as a showcase for data warehousing success. Its system design and implementation and its successful adoption by prospective users were, therefore, expected to offer a reliable setting from which research questions could be rigorously pursued. Also, that the survey data represented data warehouse users from 14 different functional departments might have a positive effect on the generalize ability of the study.

A reduced number of indicator items in certain constructs might have affected their measurement reliability and accordingly the power of the regression analysis. For instance, two indicators were used for user satisfaction, system throughput and information utility. Three constructs (accessibility, ease of use, and user training) were also represented by a single item. Limited sample size (82) might have affected the integrity of the statistical inference.

Given that the survey was distributed within an organization, a 35 percent response rate could be regarded as rather low. Besides, although the survey instrument was reviewed and validated by five people from both academia and industry, more content validation could have been done through pilot testing the questionnaire items.