Ganapathy Balakrishnan

Business Intelligence Solution

Metropolia University of Applied Sciences

Master of Engineering

Information Technology

Master’s Thesis

26 April 2022

PREFACE

I was more excited and exhausted to bring this thesis to its final shape. I was approached by a friend of mine, who is a director in a construction company, about his organization's current issues, especially after their recent acquisition. I was not sure about the size of the work at that time, but this attracted my interest and I decided to take this forward. When I started learning about their problems in detail, I understood that it requires a deep dive into their issues to come up with an appropriate solution. It ended up as a huge excavation work, where I spent much time understanding the organization's processes, systems, data etc. I came across many resistances and barriers such as technical challenges, process complexities, security restrictions, people & their mind barriers, attitude towards change etc. It took much longer that planned. It was quite a long and steep learning curve and a very good brainstorming experience. I should thank my colleague Balaji Srinivasan, working as a principal consultant for a leading IT firm, for his timely support and his knowledge that added lot of value to this thesis.

Espoo, April 26, 2022
Ganapathy Balakrishnan

Abstract

Author:	Ganapathy Balakrishnan
Title:	Business Intelligence Solution
Number of Pages:	45 pages + 2 appendices
Date:	26 April 2022
Degree:	Master of Engineering
Degree Programme:	Information Technology
Professional Major:	Information and Communication Technology
Supervisors:	Ville Jääskeläinen, Principal Lecturer, Head of IT Master's program

---

A leading construction company which plays a major role in soil load capability improvement and management had a problem in monthly financial reporting, especially after their recent acquisition & merger. It took more than two weeks to generate the reports every month and when more details were requested, it took another couple of weeks. This strongly impacted the decision making. By the time the decisions were made they were almost in the mid of the next month, and it was too late for corrective actions. It was critical to solve this problem as they were not able to make timely corrective actions resulting in financial loss. A deep study was made to find the root cause of the problem. The entire landscape was analyzed, and the reporting requirements were refined. Based on these findings, a new solution is proposed. The scope of this thesis is limited to the analysis of the existing issues and IT landscape.

As a result of the analysis, a dedicated Business Intelligence Solution is suggested for reporting needs.

Keywords: IT Analysis, System analysis, requirement analysis, Business Intelligence Solution, Data Warehouse, BI Consulting, BI Reporting, Reporting Requirements, Gap Analysis, DWH/BI Solution, Solution Approach, Solution Architecture, Data Transformation

Contents

List of Abbreviations

1NF	First Normal Form
2NF	Second Normal Form
3NF	Third Normal Form
4NF	Fourth Normal Form
BI	Business Intelligence
CSV	Character Separated Value
DB	Database
DQ	Data Quality
DWH	Data Warehouse
ELT	Extract Load & Transform
ESA	Enterprise Solution Architecture
ETL	Extract Transform & Load
HR	Human Resource
IOS	iPhone Operating System
IT	Information Technology
KPI	Key Performance Indicator
MOLAP	Multidimensional Data Cube
OS	Operating System
PMO	Project Management Office
R&D	Research and Development
ROI	Return of Investment
ROLAP	Relational Data Cube
SA	Solution Architecture
SLA	Service Level Agreement

List of Figures

List of Tables

1 Introduction

The purpose of this thesis is to find a way to reduce the delay in generating monthly finance reporting for the case company. The case company has relatively modest IT infrastructure and does not have much supporting structure in place, including people and processes to guide the everyday data need. All the existing resources are focused on core business case and working at customer sites.

This has left the company IT development into sidetrack, but to prepare for future growth and drive the organization decision making based on data, it is important to improve the data platform by creating a central data repository.

The target is to investigate the current systems, data readiness and identify the best possible solution to meet the data need of the case company.

1.1 Business Context

The case company is a leading construction industry company which includes the soil load management and improvement tasks when it needs to withstand high loads. Its roots can be extended way back to 1980s and it has been operating in more than 50 countries. This company has carried out 100,000 plus projects and got merged with another company recently with a new logo.

1.2 Business Challenge, Objective and Outcome

The challenge for the case company was the prolonged reporting time, inefficiency, and inaccuracy of the data in the monthly finance reports.

The objective of this thesis is to find out the reasons for the inefficiency and propose a solution for timely reporting and to improve data accuracy.

The scope of this thesis is limited to the study of the problem analysis, source systems analysis, reporting requirement analysis, gap analysis, solution

approach, system architecture, and data transformation strategy. The following items are considered as out of scope of this thesis: IT tools selection, vendor selection, operational model, and team structure.

The outcome is an industry standard Business Intelligence (BI) reporting platform for efficient reporting.

1.3 Thesis Outline Challenge, Objective and Outcome

This thesis consists for seven sections.

Section 1 is the introduction, background to the thesis, case company, business problem and the outcome of the study. Section 2 describes the study methods and consolidate phases with quantitative analysis, qualitative analysis, implementation feasibility and approach.

Section 3 describes the business problem analysis, where a deep look on the problem was made and the problem is segregated and grouped for easier resolution. Section 4 describes the current state analysis of the IT landscape, describing in detail the source system, Operating System (OS) analysis, Database (DB) analysis, and Data Quality (DQ) analysis.

Section 5 describes the analysis of the reporting requirements detailing with number of reports, number of reporting users, reporting technology/ tool and data retention period. Prioritization of reporting needs using quadrant analysis is also covered in this section. Section 6 describes the four-step gap analysis process, spider chart analysis and the proposed resolution for the identified gaps.

Section 7 describes the business intelligence solution, solution approach, development approach, high-level system architecture and data transformation strategy. Section 8 summarizes the thesis work and the recommendation for the implementation of the proposed solution and a brief roadmap for the potential future actions.

2 Study Method and Material

This section introduces the method used in the study. The overall work is divided into two categories, study method and analyse & consolidate

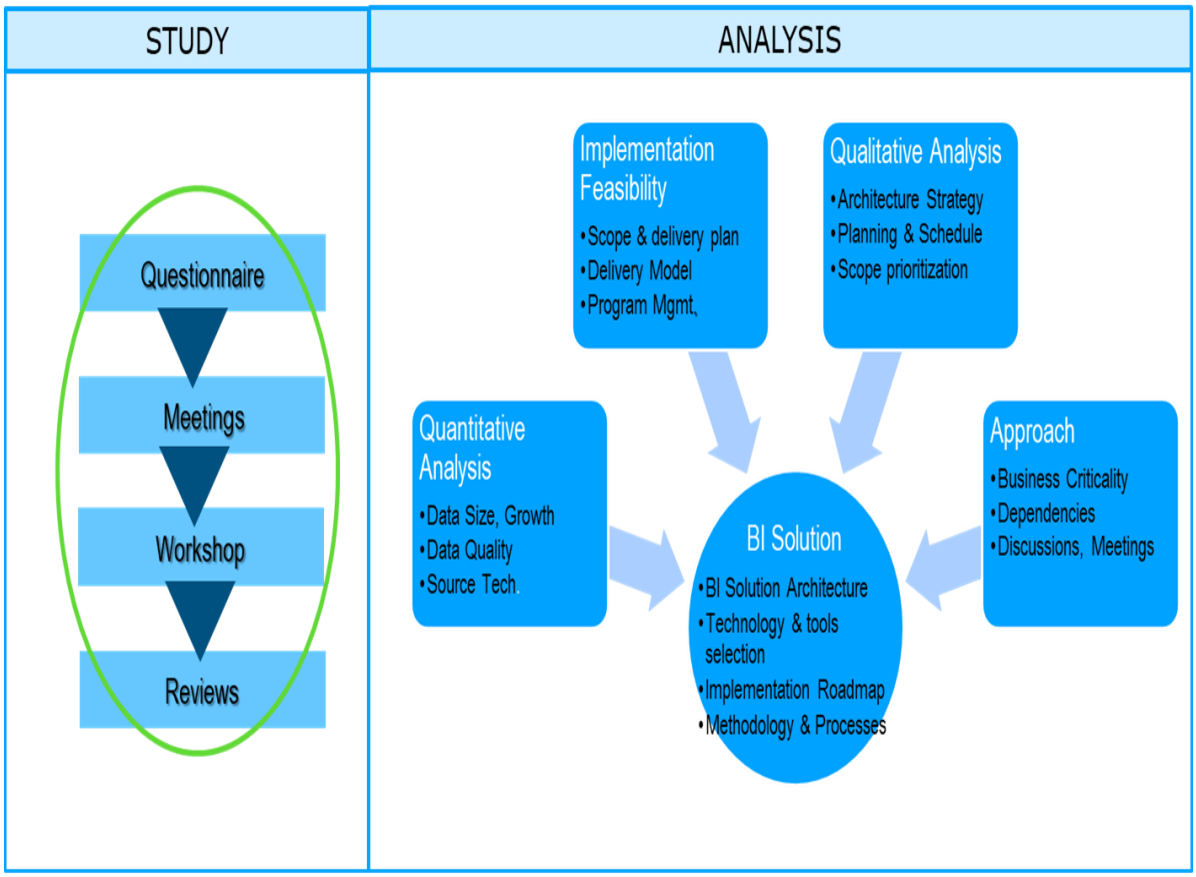

Figure 1: Study Methodology

2.1 Study method

This study started with a questionnaire preparation. List of questions, which enables the user to provide details were prepared. These questions were formed with more open-ended questions and with less close-ended question. The idea of this questionnaire was to make the users to provide as much details as possible and avoid Boolean answers. The questionnaires were sent in advance over emails, so that the users can take their own time, think about the topic, and respond more accurately. Understanding the current state of the systems,

process, security etc are the key activities. To do this, meetings were organized with users and stakeholders to discuss. Questionnaires which were sent earlier over emails were discussed in detail on those meeting. Unusual answers such as ‘maybe’, ‘don’t know’, Yes-no’ etc were also captured as a part of the meeting notes. Below is a screen shot of the sample questions. The full list of questionnaires is attached in the Appendix 1.

	System Analysis Sheet	Answers
2	**System Informations**	
3	What is the Name of the System Name	
4	what type of System is this? (Payment/ Gateway/ etc)	
5	Name of the System Owner	
6	Name of the System IT Contact Person1	
7	Name of the System IT Contact Person2	
8	Nameof the System Bussiness Contact Person	
9	**Infrastructure and Systems**	
10	Details of the Network Landscape in which the application works (LAN/ SAN/ WAN/ etc)	
11	What is the Hardware Environment for this application (Server type)	
12	what is the Software: Operating system	
13	What is the Software Languages used for this application	
14	Client Server Technology: Web Server / Web Application Server	
15	Is Any other S/W (Please give Version Numbers) used in this application	
16	What is the Existing System Landscape? Or How many instances of the applications are running? (Dev, TEST, PRE-PROD, PROD)	
17	What is the Mailing System: Notes/MS Exchange for this application	
18	**Application Informations**	
19	Please provide some General Description of the Application	
20	What are the List of Businesses units supported through this application	
21	What are the List of Legal Entities supported by this application	
22	Is any other Geographical Regions supported other than Finland?	
23	Is ther any other Countries/ Cities supported by this application?	
24	Is any Business Processes supported by the Application? (If yes, please explain)	
25	Any major business process handled outside the Application?	
26	Any Major Enhancements done/planned near future?	
27	Any Minor Enhancements done/planned near future?	
28	Any Major Interfaces from other applications?	
29	List the Critical Areas in the application?	
30	How is the application hosted?	
31	(Is the server hosting the application exclusive or shared by other pplications	
32	What is the Solution Architecture? Single or Distributed?	
33	Is it a Custom built Application ?	
34	What is the Stability of the application (how often it troubles)	
35	What is the Volatility of the application	
36	What is the Frequency of updating Development /Test/ Acceptance Environment	
37	Applications in production since<No of years>	
38	What is the numner of Named Users in the application?	
39	What is the numner of Concurrent Users in the application?	
40	What is the Usage of the system? And the expected Usage time of the application	
41	What is the Peak usage time of the application?	
42	What is the current average and peak work load of the application?	
43	What is the expected / projected Workload for next year?	
44	Are there any retirement Plans for this application? (If yes, please specify when and what.)	
45	What is the front end Technologies used?	
46	Please explain process dependencies, if any, between applications	

Figure 2: Questionnaire for Meetings/ Workshop

Next to the meetings, a workshop was conducted to understand the missing piece of information. The key users were invited for the workshop, the main idea of the workshop was to bring all the users to single understanding. This being a one-day workshop, the day was divided into multiple sections: general discussion, vision & business objective, BI functional requirements gathering and closing words.

General discussion was about the current problem, role of each system in the overall landscape of the company, and interaction and dependency between the systems. Importance of each system and data flow between the system were explained to the users. This brought light to the hidden corners and made everyone understand the problem in a better way. The same workshop was used to make the users understand the vision and the business objective of the company. Potential changes for the strategy, the reason for those changes and why such changes are essential at this point were discussed. The functional requirements were collected and were discussed in the workshop. The functional requirements were consolidated and prioritized for further processing.

2.2 Analyse and Consolidate

In this section different type of analysis were made and how this analysis helped to derive the solution was described.

Quantitative Analysis

Quantitative analysis is the use of mathematical and statistical methods in finance and investment management [2]. Quantitative analysis provides analysts with tools to examine and analyze past, current, and anticipated future events [3].In this analysis, the data size and the growth of the data in the IT systems were analyzed. The data quality of these systems is analyzed using random sampling and anonymization process. These samples were further used to determine the as-is state in the Gap analysis. Source technology of the systems and their age in terms of productization and the maturity in production were analyzed.

Qualitative Analysis

Qualitative analysis aims to increase the overall understanding of the quality, characteristics and meanings of the researched object or topic [4]. Qualitative research is a vast and complex area of methodology that can easily take up whole textbooks on its own [5]. Under this section, architectural strategy, planning & schedule, and prioritization of scope were analyzed.

Architecture strategy translates business strategy into objectives for building, enhancing, or replacing business and system capabilities together with an implementation roadmap, all-the-while maintaining resilience to change as a key architectural objective [6]. Architectural options like top-down approach, bottom-up approach and hybrid approach were analyzed in this phase. The suitable approach and the reason to adapt an approach was considered carefully. The schedule and the available resources like software, hardware, technology was analyzed and the best possible plan for the future solution was though about in this phase.

Prioritization of the subjects were analyzed during this stage. The subject with high financial implication and the subject with high architectural impact were chosen at high priority for Implementation

In this section the feasibility to implement a full stake IT/business solution was analysed. Pre-requisite and other geographical, environmental aspects were also analysed at this stage. The feasibility for the organization to take a project or course of action is an important aspect. A tentative plan, feasible resources needed, ability to manage a project, available resources, people, roles, and responsibilities were analysed. The more the number of ‘Yes’ received or the smaller number of ‘No’ received is considered to take the discussion to the next level. A suitable delivery model was discussed at this stage. Three delivery models were taken into consideration and their pros and cons were analysed. Based on all the above criteria decision about the ‘go’ or ‘no-go’ was made.

Table 1: Delivery Model

Delivery Model	Pros	Cons
Fully sourced model – in this model all the needed resources will be acquired, and all the necessary people will be hired to the company	Good strong technical team would be formed with all necessarily resources in house	Acquiring all the resources and people would be expensive. Making change is future would be difficult.
Fully outsourced model - in this model all the necessary resources and the people will be hired as a service from a third-party company (outsourced company)	Making change is easy Low cost than the above model	No control on the IT asserts and less control or limited control on the data.
Hybrid model – in this model some of the resources will be procured and few will be hired as service. Small core team will be hired, and the rest (extended team) will be outsourced to third parties	Good balance between resources and people	Core team spent much of the time in managing the extended team and may not be able to architectural development. More time spent on people management.

In the process of analyzing the delivery model, a project management team is also considered to find out the amount of management effort needed to execute such projects. A share project management team was considered as a suitable option.

Approach

In this section the feasible approach for the business problem has been analysed. Business criticality of the application was considered as a key element in this analysis. The more business critical a problem is, the higher its rank in the approach backlog list. In few cases the dependencies between the system in terms of operations and data was considered as important in the approach and been given higher rank in the backlog list. Discussions with the stakeholders and the key users more important at this stage to finalize the list before proceeding any further. Based on all the factors above a clear approach with its priority list was made.

BI Solution

In this section Business Intelligence Solution was derived based on the above analysis. The output from quantitative analysis, qualitative analysis, Implementation feasibility and approach were the input for this phase. Based on the input the best feasible solution was analyzed. First thing to be looked into is the solution architecture.

Solution Architecture

A solution architecture (SA) is an architectural description of a specific solution. SAs combine guidance from different enterprise architecture viewpoints (business, information and technical), as well as from the enterprise solution architecture (ESA) [7]. A multilayered solution architecture was made for here, which contains highly movable components to enable scalability of the platform.

BI Tools

Tools play a key role in BI solution. Procuring a matured tool makes the construction of the BI solution easier and quicker. Right from the database to the transformation and to the reporting usage of tools is unavoidable. Database tools like Netezza, Oracle, Hadoop are few examples. Transformation tools are usually

called as Extraction Transformation Load (ETL) tools. This is an important tool in BI solution. It is the process in which the data is extracted from any data sources and transformed into a proper format for storing and future reference purposes [8]. Data consolidated from various system need a big transformation and cleansing before loaded in the Data Warehouse (DWH). Some of tools call them self of Extract, Load, Transfer (ELT) tools. Every tool has its own advantages and disadvantages. The output from the implementation feasibility and approach plays a vital role in selecting the tool. Informatica, Xplenty, Dataddo, Talend are some of the leading ETL tools. Reporting tools are usually chosen based on the primary front-end application of the company. The report tools capability is more important to bring the value of the data to the decision makers or end-users. The flexible tools would have the capability to integrate dynamic data from multiple sources, these tools usually have good adaptability to modern handheld devices like IOS, android mobiles, tablets, and Java platform devices

3 Business Problem Analysis

Business problem analysis is a set of techniques that identify and evaluate areas in which stakeholders are unhappy with an existing situation [1]. Understanding the business problem is an important and the first step. Reading the problem statement and discussing with the stakeholders about the problem provides the best possible view of the problem. Looking into the real time reports and the other evidence of the problem provides more clarity to the problem. Discussing with the end-users provides another view to the problem.

When doing this type of analysis, the problem should not only be analysed from the top-down, but also from the bottom-up. The depth of the problem, where it starts and how that develops into the current form of the problem needs to be carefully analysed

With careful evaluation, the problem was classified under four categories.

1. Corporate financial reports are not generated on time
    a. Group reporting takes more than half a month to produce. This delays the compliance and the on-time corrective actions.
2. Lack of details (‘drill-down’) in corporate reporting
    a. Not able to find the reasons behind the numbers in the reports.
    b. Manual reporting and poor data mapping take more time to drill down to the details.
3. Lack of data clarity (single version of truth)
    a. Different system has different data quality. Not able to trust the data and hence the reliability is lost.
    b. Reporting done manually in silos.
4. Multi-directional flow of data
    a. Data flowing from one system to another and vice versa. This mess up the data.

4 Current State Analysis

The current IT system architecture was scattered across 14 individual systems. Each system was operating in silos and been developed over time. These systems were mostly older systems and had very less or no warranty support left. The most recent system was the sales system, where the customer relation management (CRM) tool was integrated.

4.1 Source System

Logistics

This system is used to mobilize the work tools, machineries etc between the work locations. Most of the time in a year these machineries stay at customer locations and been moved from one customer location to another. This system maintains the current location of such machineries and other work items. Machineries are classified into small, medium, and large based on the size and the functionality of the machines. The next big category under logistics is the tools. These tools are large in numbers and been used mostly as hand-held tools. Tracking these power tools was difficult in real time and hence the stock of the tools is updated once in a quarter. Logistics unit also includes the electronics devices used in the construction site and the real time tracking is not enabled for various reasons.

Procurement

Procurement unit is responsible for all the procurement activities. Procurement of machineries, power tools, electronic tools are within the scope of this unit. This unit also maintain the warranty details of the procured machineries and tools. Procurement of other essential raw materials, work items like chemical products etc are handled by the procurement team. Office supplies and other office essentials are also within the scope of this unit

Retail Sales

This unit is a subunit under sales; however, it operates as a separate unit. This unit handles the sales related to end customers. This unit is recently renovated with Sales force cloud solution. This unit have multiple frontend user interface and been used by sales team, managers, and the sales director. Customer details, work details and the work status are stored in the system. Call centres use this data for customer support. Potential customer identification and marketing are made in this unit. New campaigns are made and been sent to the potential customers as offers. Customer journey workflows is enabled in this unit which helps to the customer service and improves customer satisfaction.

Partner Sales

This unit handles the sales related to company customers. This unit came from the recent merger and been maintained as a separate unit focusing only on company customers. These sales are usually bigger in money value and been handled a bit different from the retail sales. Opportunities on this unit are being monitored directly by the sales director and been always handled with more confidence. Call centres use this data for support however this system provide only details of the completed projects to the call centre.

Operational Risk

This system holds the geological data, work details and the potential risks along with the mitigation plan. The data in this system is more secure and only few people in the organization have access to this system. The geological data is obtained from the local municipality and been periodically updated. This system also sends the information back to the authorities about their completed projects and their finding from the under groundwork. This system plays a vital role in rolling out offer to new customers. Based on the type of land and the potential risk, the offer to the customers varies.

Warranty Support

This unit is responsible to authorize a customer under warranty support. This system contains information about the invoiced work, work completion details, customer details along with customer contacts. Call centres use this data for warranty claims. Due to the 10-year warranty period, the data retention period for this system in 12 years. Due to this long retention period this system is highly under pressure and dedicated team is organized to keep this system live.

Corporate Reporting & CXO Reporting

This unit is called as Group Reporting or Corporate reporting unit. Reports produced by this unit are used for decision making. There is no separate system doing this reporting and the reports are handled manually in excel sheets and google sheets. Due to the manual nature of this unit, the data in these reports are prone to errors and typo issues. These reports don't provide insight of the data. This data is also used by the audit unit and Finance unit.

PMO unit

Project Management Office (PMO) is a work force management system. It is used for calculating the employees work timing and the compensation. This unit also contains Human Resource (HR) people. This unit maintain the employees and their work location details and these details are updated every day. This unit works as a point of contact for any HR queries from outside the organization.

Audit

Audit is a unit which handles all the legal entities in the company. The unit handles and ensures all the safety measure to be in a workplace. This unit acts as a primary point of contact for legal queries. Both internal and external audits are handled by this unit. The data from this system in considered as secure and only the audit team have access to the data.

Operations

This unit handles the real work in the construction site. This is the biggest unit with more than 90 percentage of people working for it. This unit use couple of systems, they connect with the PMO system for working time check-in and check-out. This system contains the site workers, site engineers, supervisors, and the designers. This unit work from various locations and the instruction for the work force is provided using this unit. The status of the work, situation in the worksite and other high-level and low-level details of the work are available with this unit. Current projects and machineries are handled by this unit. This unit also holds a separate Research & Development (R&D) unit, which is involved fulltime in developing the new solution for the customers

Finance

This unit handles the finance related activities. Recently this unit is renovated with Salesforce cloud solution. This unit is responsible for company finance, general ledger, and all finance related activities. This system holds all the finance details and have limited access to others outside the unit.

Compliance

This unit is responsible for the compliance from the local bodies. This unit also involves in the approval process from the local authorities for the work in the construction site. This unit is part of the Audit unit; however, this unit is always seen as a separate unit to hold the legal standards.

Online (Invoice & Payment)

This unit is responsible for the Invoice generation and the payment process. This unit receives details from Sales and operations systems every day. Based on the data received from these systems, invoices are generated and been despatched to the customers. This system contains the payment data, invoice data and the customer data in one place.

Warehouse

This unit is responsible for maintaining the stock of all the movable asserts. The data from this system includes the availability raw materials, machineries, and tools. This unit also have the full control of warehouse. This unit work closely with the procurement and logistics team. Logistics team receives the daily goods from the warehouse and deliver it to the worksite.

4.2 Source System Analysis

Operating System Analysis

An operating system is the most important software that runs on a computer [10]. It manages the computer's memory and processes, as well as all its software and hardware [10]. The operating system of all the source system was analysed at this phase. The following were the list of operating system under use.

- Windows
- Aix
- AS400
- Cloud
- Linux
- Sun Solaris.

Out of which Window is the most used operating system with 39 percentage of system usage and the least used operating system is Sun Solaris at 5 percentage. With this finding, windows can be the proposed operating system for the BI system, which will save cost in Operating system management.

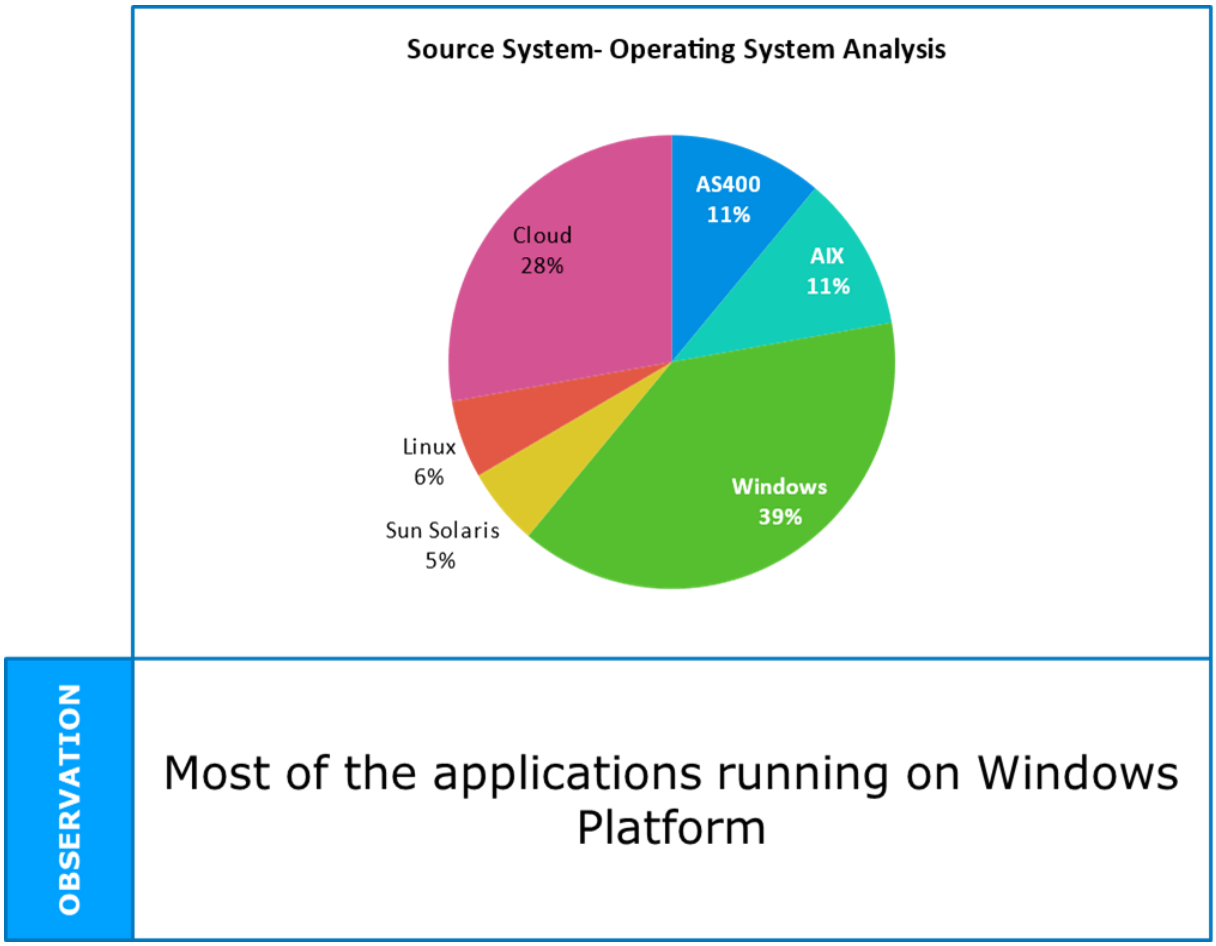

Figure 3: Source System- Operating System Analysis

Database Analysis

Database store for each of the source system is identified and analysed. The following were the list of databases under use.

- Oracle
- MYSQL
- DB2
- Salesforce Cloud
- Informix

Cloud storage and MYSQL were considered as the widely used database both scoring at 18 percentage. However, the mostly used data sources is Excel sheet, which is not considered under database list. Cloud storage and MYSQL can be the proposed database for the BI solution, however other leading database cannot be ruled out of scope for the database options. A detailed study along with the tool selection process needs to be made for database selection.

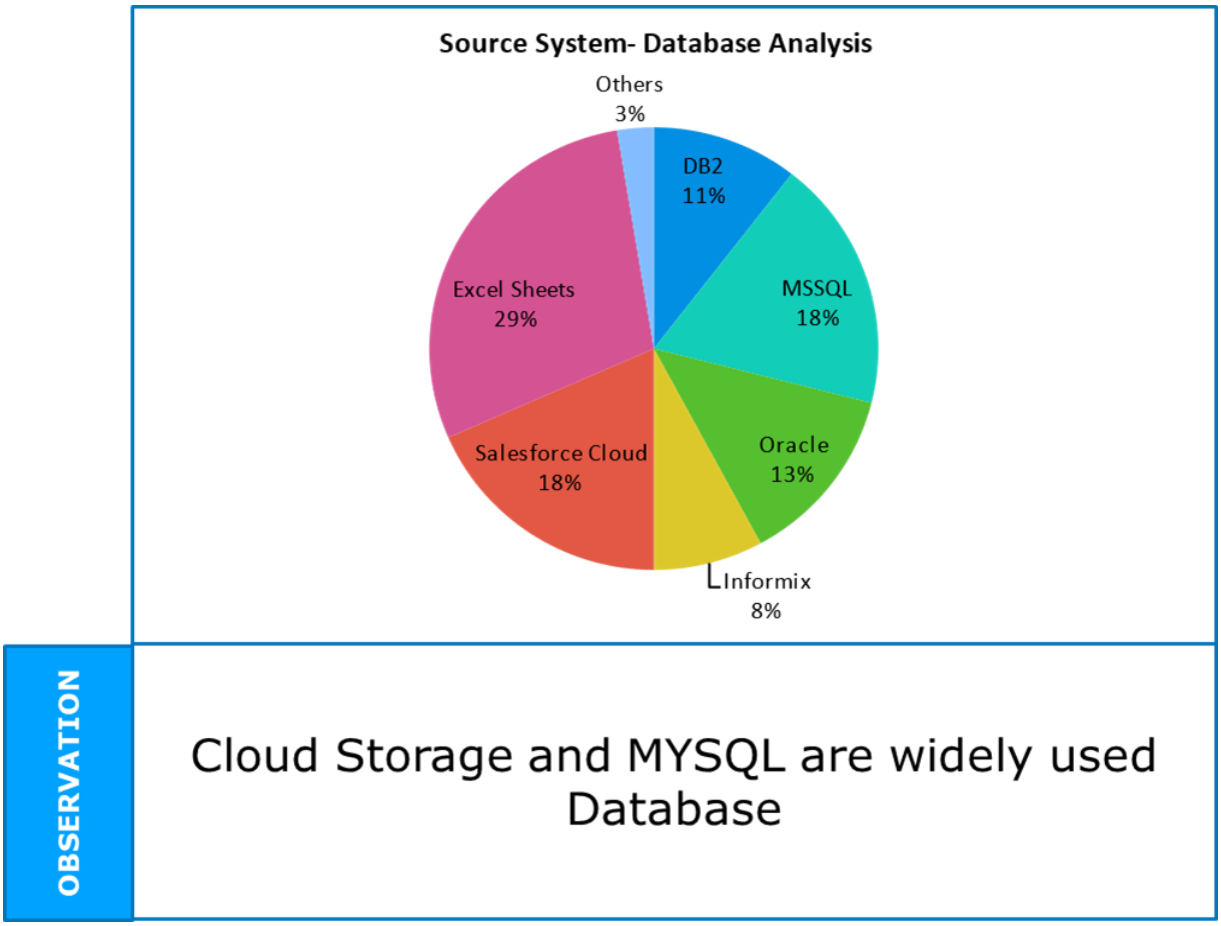

Figure 4: Source System - Database Analysis

Data Quality Analysis

Data quality is a measure of the condition of data based on factors such as accuracy, completeness, consistency, reliability and whether it's up to date [11]. Measuring data quality levels can help organizations identify data errors that need

to be resolved and assess whether the data in their IT systems is fit to serve its intended purpose [11].

Sample data from all the system been collected and was analysed using the data standard analysis technique.

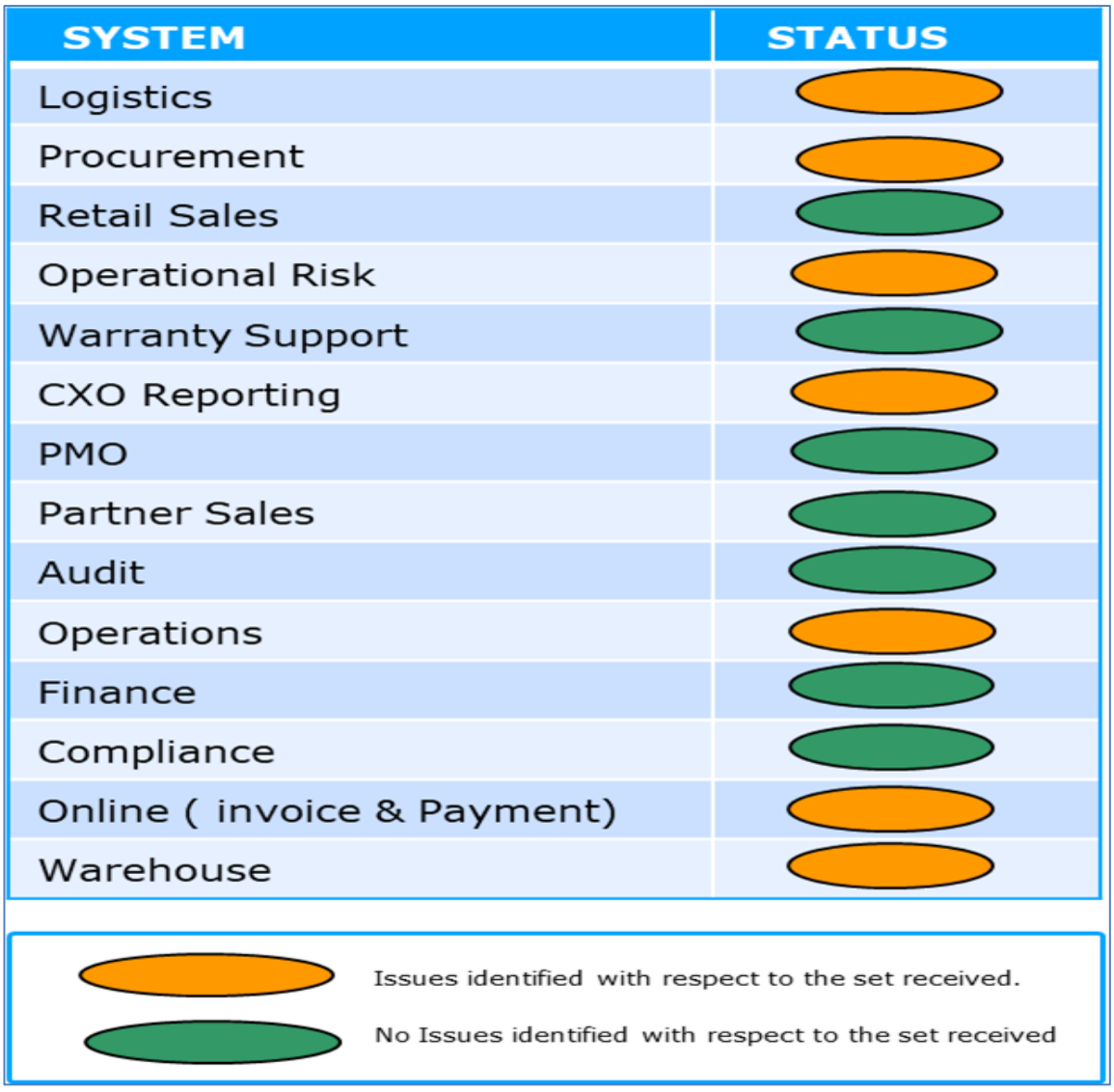

Figure 5: Source System - Data Quality Analysis

In this technique the data from each system was compared with the other system and the consistency of the data was analysed. The data for an event or date was compared and analysed across all the system for the same date.

For example, the work-start-date & work-end-date was compared between the operations unit, warehouse, online (invoice & payment) unit and warranty support.

Every discrepancy was closely analysed, and the root cause was identified. In the similar way the customer details, product details, work status were used as a use case for this analysis.

Based on the sample analyzed, the below issues were identified.

- The data between the applications had considerable number of discrepancies.
- Data duplications was identified in both high level and low level of data.
- Every system identify customer with a different unique code.
- Customers were identified by generic terms like invoice numbers, contract number etc.
- Both Finnish data and English data observed in the same column of the application.
- Unique code standard not used in all the applications.
- Date format mismatch identified between the system. Few systems use local date format and few systems using other country date format.
- Blanks, NULL and junk characters are identified in few tables.

5 Reporting Requirement and Requirement Analysis

Reporting requirements collected during the business meeting and the workshop were analysed, classified, and prioritized.

5.1 Based on Number of Reports

In this analysis the total number of existing reports in each unit was analysed against the total number of reports expected in the future from each unit. The increase or decrease in number between the existing reports and the expected reports is an important parameter for the load stability of the proposed reporting platform. In this analysis, the number of expected reports were higher than the existing reports, meaning that every unit needs more reports from the common reporting platform. This also indicates the need for the cross functional reporting capability.

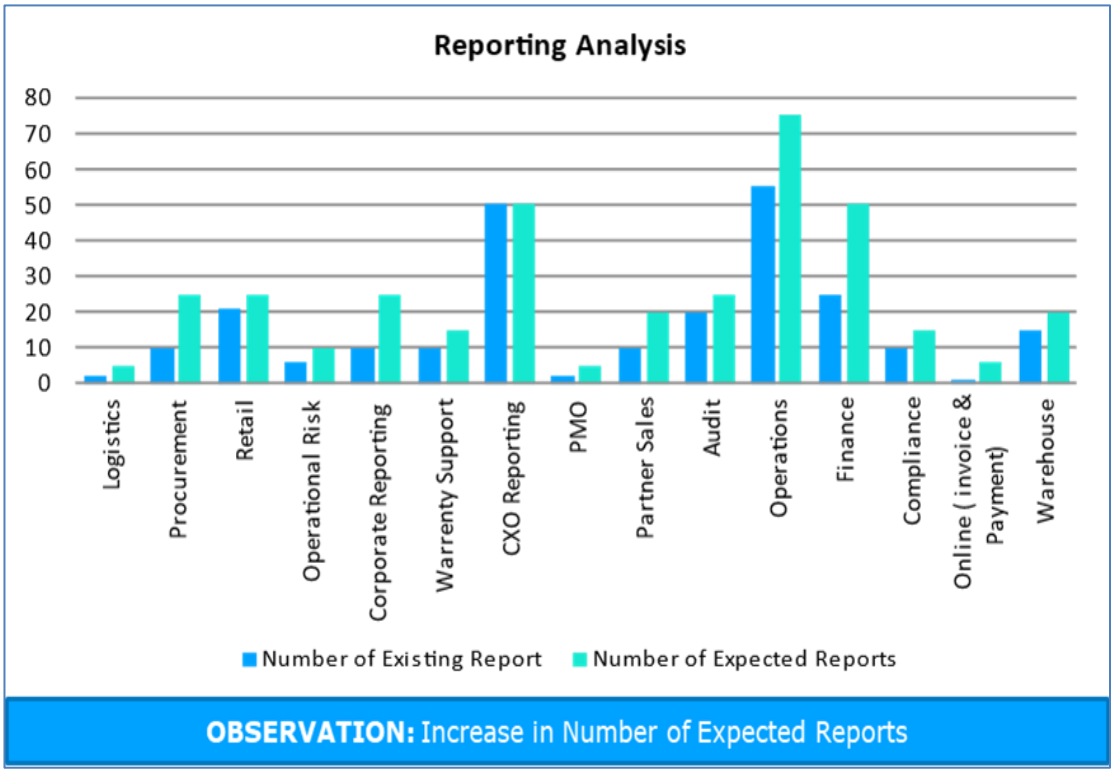

Figure 6: Requirement Analysis - Number of reports

5.2 Based on Number of Users

In this analysis the total number of existing reporting users in each unit was analyzed against the total number of expected reporting users. The increase or decrease in number between the existing number of users and the expected number of users is an important parameter for the capacity planning of the proposed reporting platform. In this analysis, the number of expected reporting users were higher than the existing number of reporting users, meaning that every unit is going to add more users in the common reporting platform. The expected number of reporting users also includes the tentative new joiners of the company in the coming years.

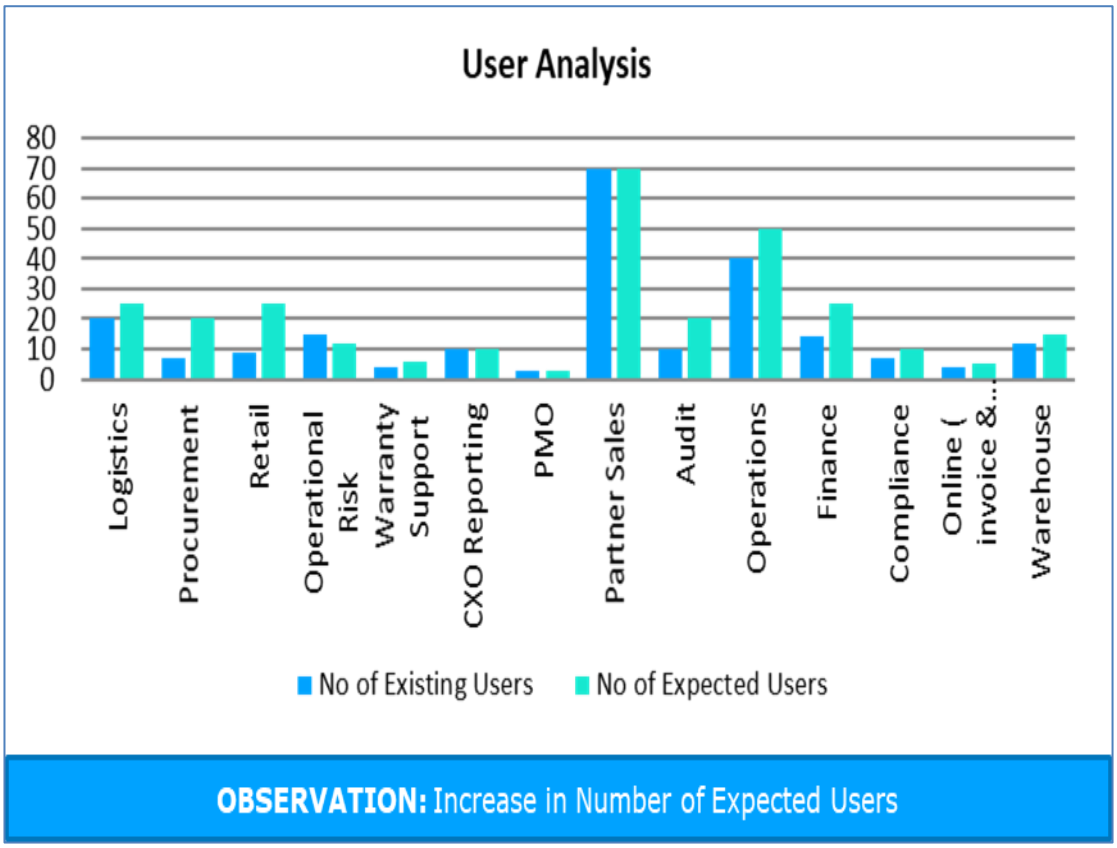

Figure 7: Requirement Analysis - Number of Users

5.3 Reporting Tools and Technology Analysis

In this analysis the reporting tool available in the company been analyzed. This analysis helps to understand the money spent on various reporting tools, their license model, and the capacity to extend the license were taken into consideration. The tool capability and the maturity of the tools were analyzed. In this analysis, Excel sheet & Google sheets are found to be the widely used reporting tool summing up to 40 percentage of the usage. Salesforce reporting as the second widely used reporting tool with 27 percentage of usage. The least used tool is Cognos at 3 percentage of usage.

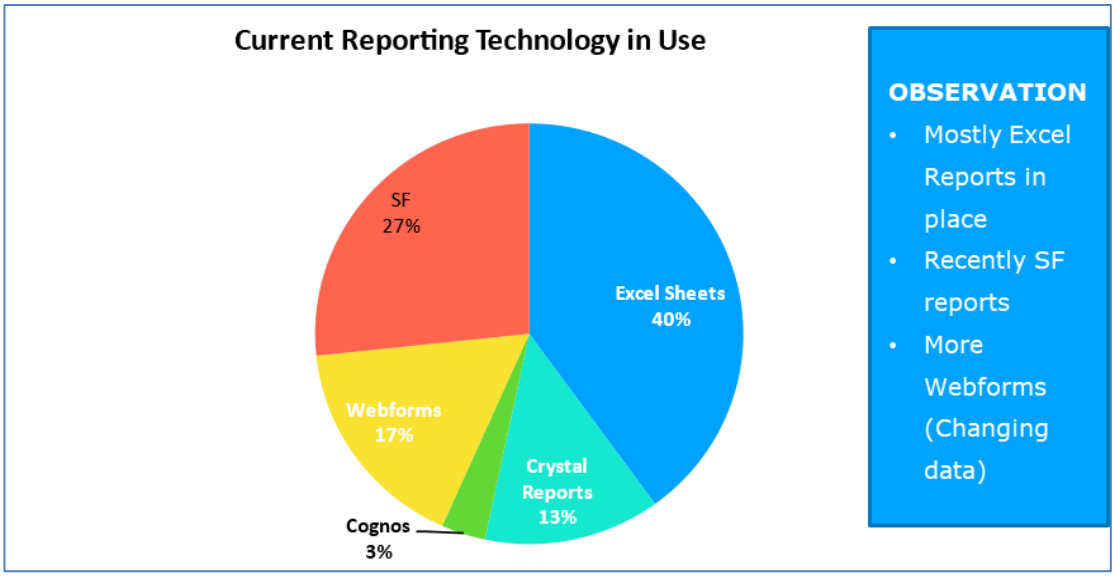

Figure 8: Requirement Analysis - Reporting Tools

5.4 Data Retention Requirement

In this analysis the data retention period for each business unit been carefully analyzed. The data retention period changes widely between the units for various factors. Data retention period is an important parameter for capacity planning especially for database volume capacity. Warranty support and compliance unit have the data retention period as 10 years. Audit and PMO unit have the data retention period as 3 years. The data retention period is based on the legal requirements in cases like audit and compliance and in other case like Retail and

Operations by company's policy. In this analysis the data retention periods are identified as longer. Below is the picture about the retention period.

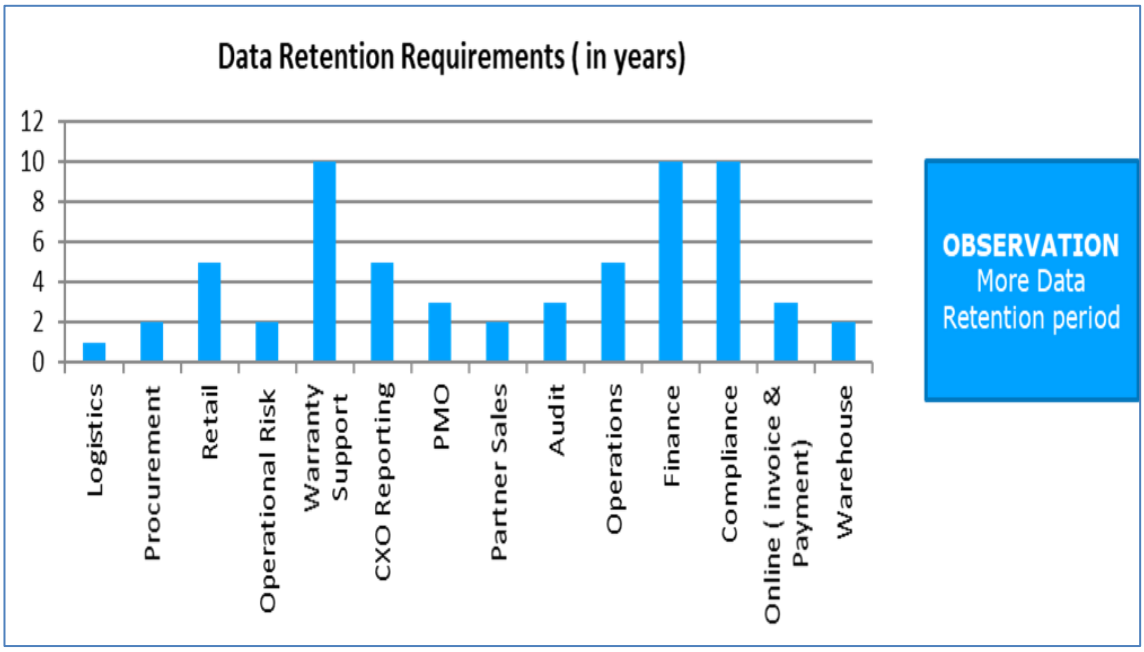

Figure 9: Requirement Analysis - Data Retention Requirement

5.5 Prioritization

In this exercise the overall reporting requirements are classified into three categories using quadrant analysis technique. The quadrant is built with financial benefit in one axis and Architectural benefit / Data readiness in another axis. The quadrant units are marked between low, medium, and high. The higher the value in financial business, the important it is for implementation; hence it is taken under Phase 1. The higher the architectural benefit, the data was identified to be important. The higher the data readiness, it adds more value to the architecture. As an output of this analysis, the reporting requirements are grouped as Phase 1, Phase 2, and Phase 3. Below is the picture where the requirements are classified using quadrant analysis technique.

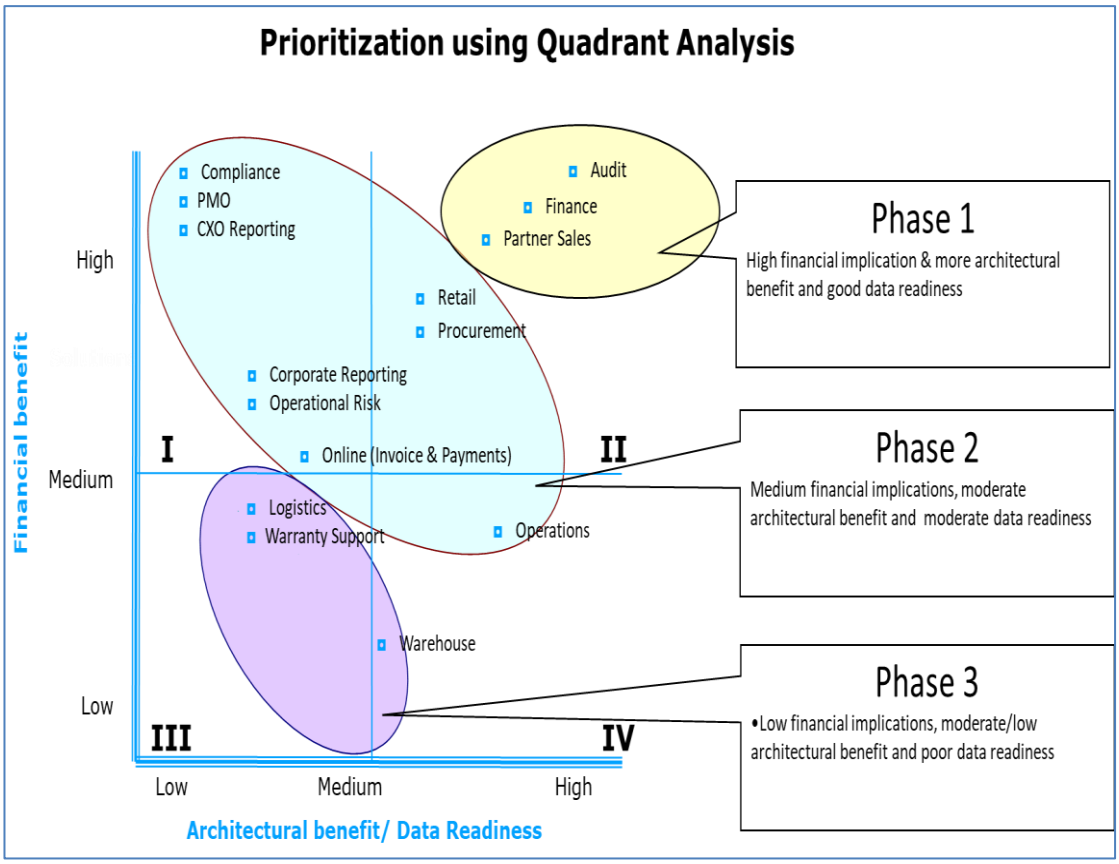

Figure 10: Prioritization using Quadrant Analysis

6 Gap Analysis

Gap Analysis is an evaluation of the current state for the purpose of identifying the difference between the current state of the business units and where they like to be. In this process the current state is analyzed using key parameters and been compared with the expected or desired state. Gap Analysis is a four-step linear process.

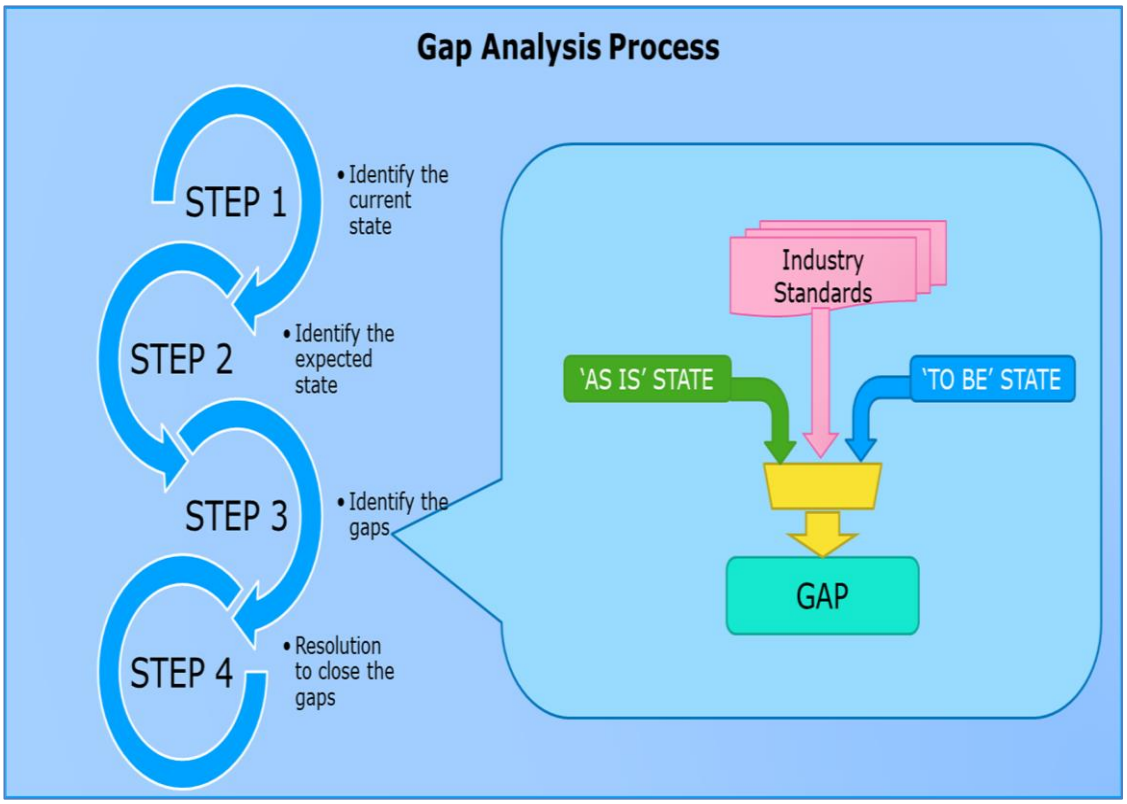

Figure 11: Gap Analysis Process

6.1 STEP 1: Identify the current state

In this step the current state of the all the systems was analyzed. This analysis is done based on the industry standard parameters. These parameters are chosen keeping in mind the industry standard of expected solution and the basic reporting requirements. The details about these parameters were gathered along with the workshop and later with multiple meetings with the system maintenance team. The collected information was further verified with the leads from each unit to

ensure the accuracy of the information been gathered. The following are the list of identified parameters

- Data Management
    - Data Availability
    - Data Quality
    - Metadata
    - Information Access
    - Data Integrity
    - Data Security
    - Data Governance
- Reporting
    - Report Generation Process
    - Reporting (Tools)
    - Planning (Tools)
- Infrastructure
- Competency
- Change Management/ Version Control
- Collaboration
- Architecture

6.2 STEP 2: Identify the expected state

Identification of the expected state was quite a wide area. This was narrowed down using the vision & Mission of the company, where the company's leadership team what to position their company in short tern (like next few years) or in the long term (like next 5 years).

When it comes to the expected state, two things played vital role. The first thing was derived from requirements from the users. These requirements include functional requirements, security requirement, data requirement (timely need), data accuracy (completeness) and data integration. The second thing was the industry standards, these are the standard that needs to be considered when

building a new system. These requirements were mostly like technical standards, design logs, connectivity need, platform stability, architectural scalability, accessibility of the system and data security.

6.3 STEP 3: Identify the Gaps

In this phase the parameters analyzed in the earlier steps were normalized using the standard five-point index. Considering the highly important parameter and its required state as the highest scale of 5, the rest of the parameters are normalized and been charted in the spider chart.

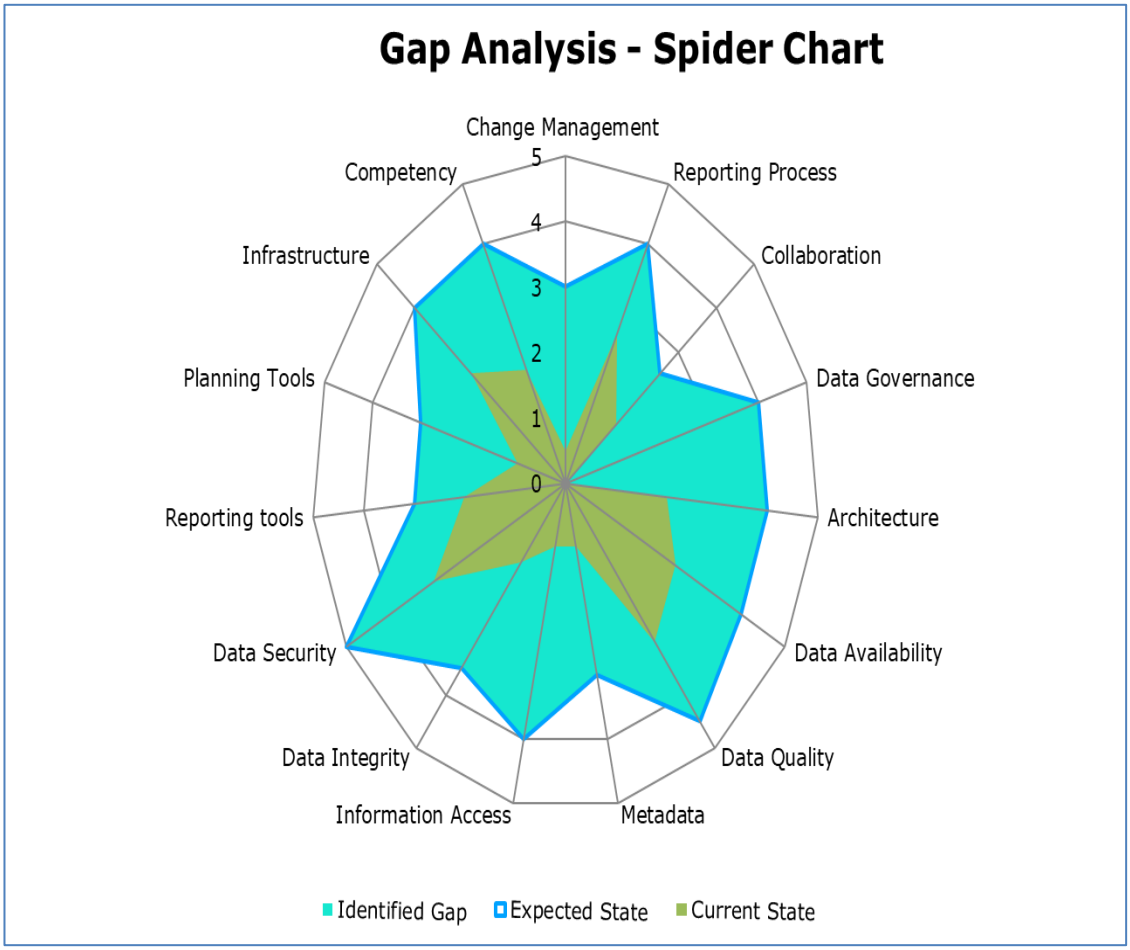

Figure 12: Gap Analysis - Spider Chart

In the above chart the current state was marked with green color and the expected state is marked with blue line (or border). The space between the

current state and the expected state is the identified Gap, which need resolution. In the above picture the identified gap is marked in turquoise blue.

6.4 STEP 4: Resolution to close the gaps

In this section, the identified gaps were carefully studied in details and the best possible recommendations were made. These recommendations are mostly based on the comparison between the industry standards and the current state of the systems against the parameters used for gap analysis.

Table 2: Gaps and resolution

Parameters	**Gaps**	**Resolution**
Data Management		
Data Availability	Not able to visualize informational trend as the source data is not integrated.	Data to be extracted, cleansed, transformed, and loaded into a common area and to be available for reporting and other needs
Data Quality	Data needs to be standardized	Data standardization needs to be done before loading to data Warehouse and Data must be restricted at the entry point
Metadata	Metadata not available in all levels	Enterprise-wide metadata management to be implemented
Information Access	Minimal Information access process/ procedures in place	Proper User access needs to be defined and implemented
Data Integrity	No data integrity between systems	Data Integrity needs to be maintained

Data Security	Not all sensitive data encrypted	Need to implement enterprise-wide data level security
Data Governance	No Data Governance in place	Data Governance model needs to be implemented
Reporting		
Report Generation Process	No Automated process - High manual intervention giving rise to typo errors and more human effort	With the BI Program, entire reporting process to be converted as tool-driven and automated
Reporting (Tools)	More number of reporting tools increasing the complexity	Single reporting tool/ platform needs to be implemented
Planning (Tools)	Minimal planning tools in place	Need to implement standard planning tool
Infrastructure		
	Enterprise-wide tools and standards are not available	Enterprise-wide tools, standards, and capabilities to be established
Competency		
	Lack of knowledge and technical capability in reporting area	With the BI program an exclusive BI competency needs to be developed through training.
Change Management/ Version Control		
	No Established Change management and Version control procedures	Proper Change management needs to be defined and version control tools to be implemented
Collaboration		

	Inputs given by IT don't meet the Business need as no consistent link or strategy between Business Unit's needs and IT.	With the BI program business can create their own reports without much IT interaction.
Architecture		
	All systems are independent, henceforth creating bottlenecks in generating reports.	Need to design the data model specifically for requirements which will consolidate the data in a single platform.

7 Business Intelligence Solution

This section describes the Business Intelligence solution for the case company. The Business Intelligence solution is derived based on the below approach.

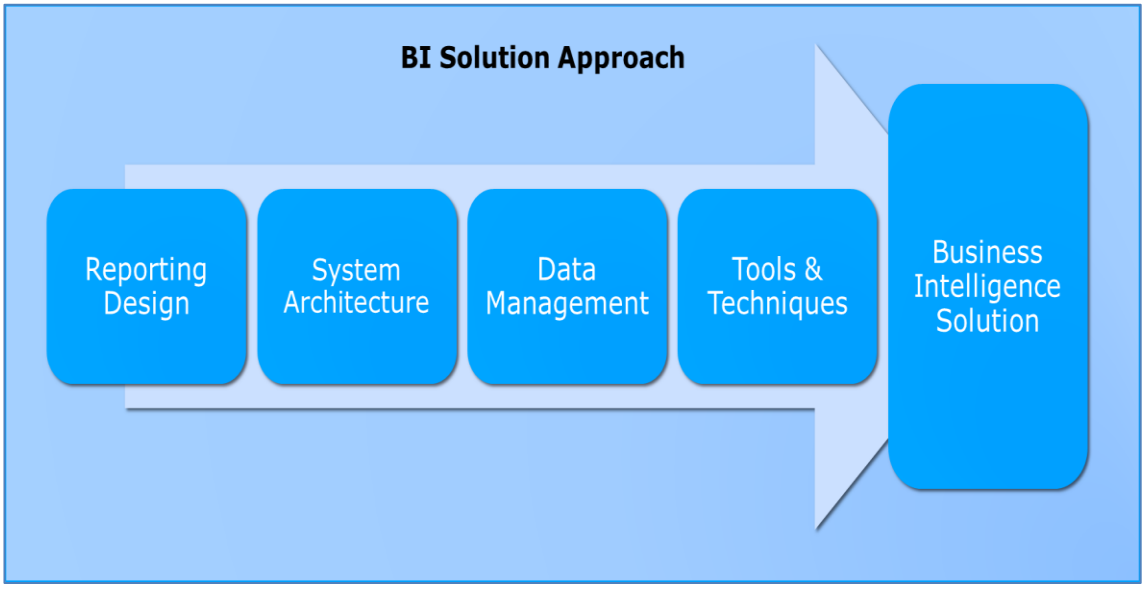

Figure 13: BI Solution Approach

This is a linear approach; each step needs to be completed before proceeding to the next step.

7.1 Reporting Design

In this step, the high-level design for the reports been analysed and the prototype was made. The design is highly depending on the requirement; however, the feasibility of the data is more analysed at this step. Based on the best possible information available the prototypes was made in simple design tools (like Excel sheets) and been discussed with the end-users. In this step, the prototype was made only for specific reports. These specific reports are chosen based on the discussion between the technical team and the end-users. The data feasibility plays a vital role in selecting the report for prototype.

7.2 System Architecture

In this step a high-level architecture was made. This architecture provides the information about the data flow, data store and the data display part of the solution. This is a multi-tier architecture, which enables the scalability and platform stability. This multi-tier architecture also makes the channel open for future change in tools and technology. The data flows always from left to right, which address one of the key business problems ‘multi-directional flow of data’.

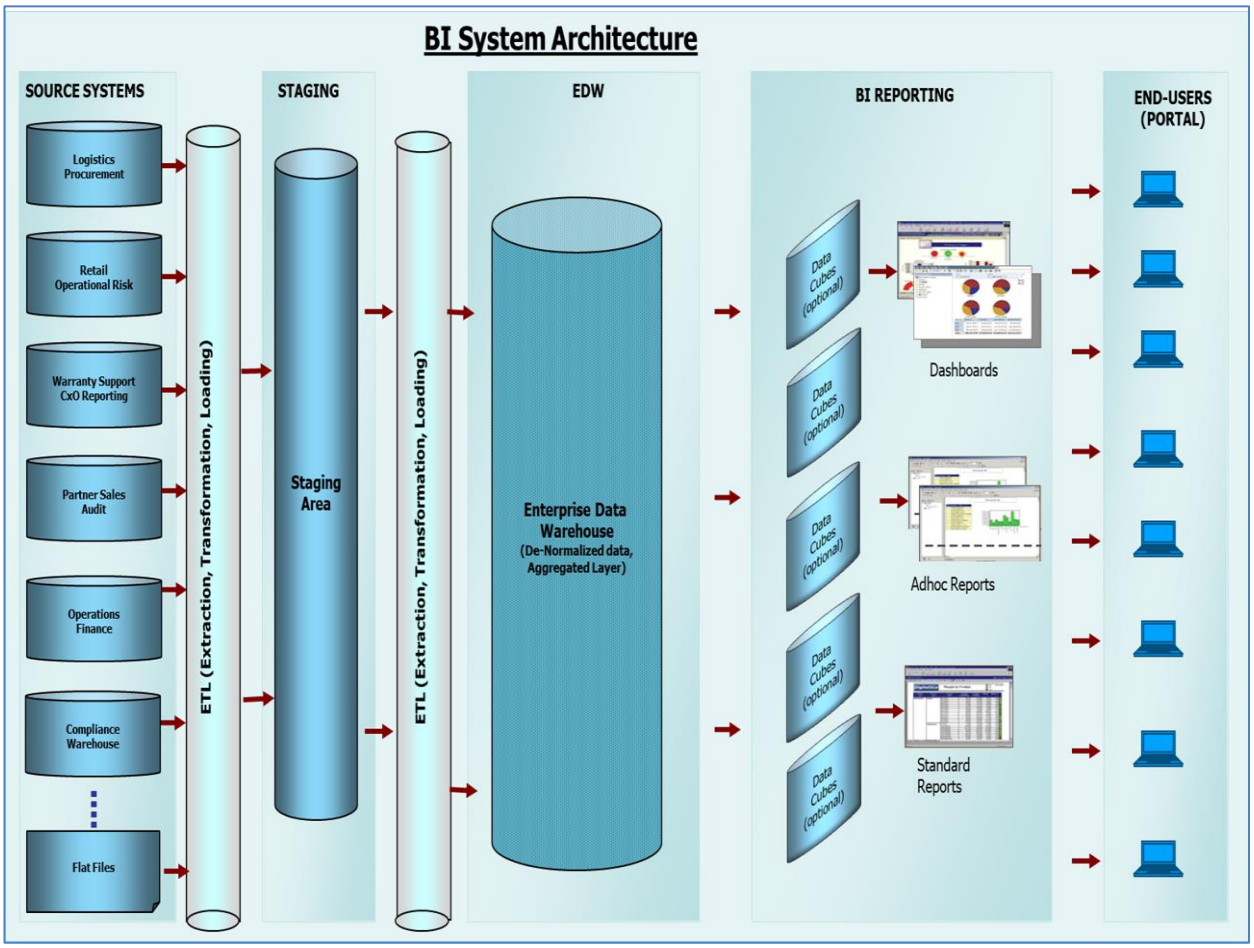

Figure 14: BI System Architecture

Source System

The first layer from the left is called as the source systems. This is a consolidation of all the available and necessary systems/ units. This is the place where the data

starts its journey in the BI System. The main purpose of this layer is to identify the list of system that needs to be integrated with the BI system. At this layer each system would have their own technical landscape like Operating system, databased, front-end tools etc

Staging

Staging is the second layer in the Business Intelligence system architecture. Data from all the source systems are moved to this layer using ETL tools. This layer is built on the same technology of the Enterprise Data Warehouse (EDW). All the necessary data from the source systems are moved to the staging layer using ETL tools. These tools are recommended to be configured in the same system as the EDW layer. These tools fetch data in two strategies, Push-type and Pull-type. Based on how data comes from the source system to the staging layer, these types are determined.

In Push-type the source system push data to the Staging layer on regular intervals. The frequency of the push is determined based on the need of the data in the BI system. The source system generates files and send the files to the staging layer. The responsibility of generating the files and delivering the files to the staging is with the source system. Due to the process of generating a file and transferring in stipulated frequency, there is always a delay in the data readiness between the source system and the staging

In Pull-type the ETL server connects directly with the database of the source system and pulls the required data whenever needed. Here no file generation needed, and the source system does not have any role to play other than providing database level access to the ETL tool. Here the ETL is configured in such a way that every change in the source system will reflect in the staging layer instantly, hence no delay between the source system and staging.

As the name specifies, staging layer is a temporary layer, the data reside in this layer until it gets loaded in to the DWH layer. The database tables in this layer have the same structure as the source tables and are truncated daily.

Enterprise Data Warehouse (EDW)

This the main layer in the BI system architecture where the data resides. The data from the staging layer is moved to this layer using ETL tools. The database tables on this layer are built in snowflake schema. The data on this layer is de-normalized and aggregated.

Normalization:

Normalization is the process of organizing the data in the database, to minimize the redundancy from a relation or set of relations. It is used to eliminate the undesirable characteristics like insertion, update & deletion anomalies [12]. Normalization divides the larger table into the smaller table and link the using relationship and it is used to reduce redundancy from the database tables [12].

There are four types of normal forms available.

Table 3: Normalized forms

Normal Forms	**Description**
First Normal Form (1NF)	A relation is in 1NF if it contains an atomic value.
Second Normal Form (2NF)	A relation will be in 2NF if it is in 1NF, and all non-key attributes are fully functional dependent on the primary key

Third Normal Form (3NF)	A relation will be in 3NF if it is in 2NF, and no transition dependency exists.
Fourth Normal Form (4NF)	A relation will be in 4NF if it is in Boyce Codd normal form and has no multi-valued dependency.
Fifth Normal Form (5NF)	A relation is in 5NF if it is in 4NF and not contains any join dependency and joining should be lossless.

Denormalization:

Denormalization is a strategy used on a previously normalized database to increase performance [13]. In computing, denormalization is the process of trying to improve the read performance of a database, at the expense of losing some write performance, by adding redundant copies of data or by grouping data [13]. One of the main purposes of denormalization is to reduce the number of physical tables that must be accessed to retrieve the desired data by reducing the number of joins needed to derive a query answer [14].

Star Schema:

A star schema is a data model that stores information in multiple table types: a single fact table and multiple dimensional tables [15]. In contrast to the classical database design of normalizing tables, star schemas connect dimensional data with fact data in a shape resembling a star [15].

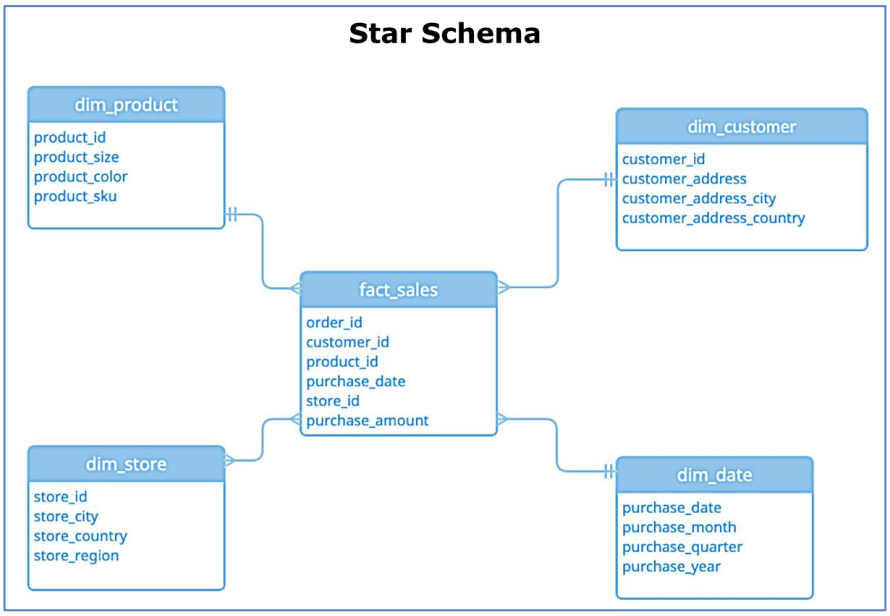

Figure 15: Star Schema [15]

In the above diagram the central table is the fact table, and the four other surrounding tables are the dimension table. The fact table holds the facts of the sales. The dimension table holds the information about the fact like customer details, purchase date, store where the purchase happened, and product purchased.

Snowflake Schema:

A snowflake schema is very similar to star schema, but it splits dimensional tables into further lookup tables [15]. The snowflake schema can have any number of dimensions and each dimension can have any number of levels [16]. Each dimension is split until it is normalized [16].

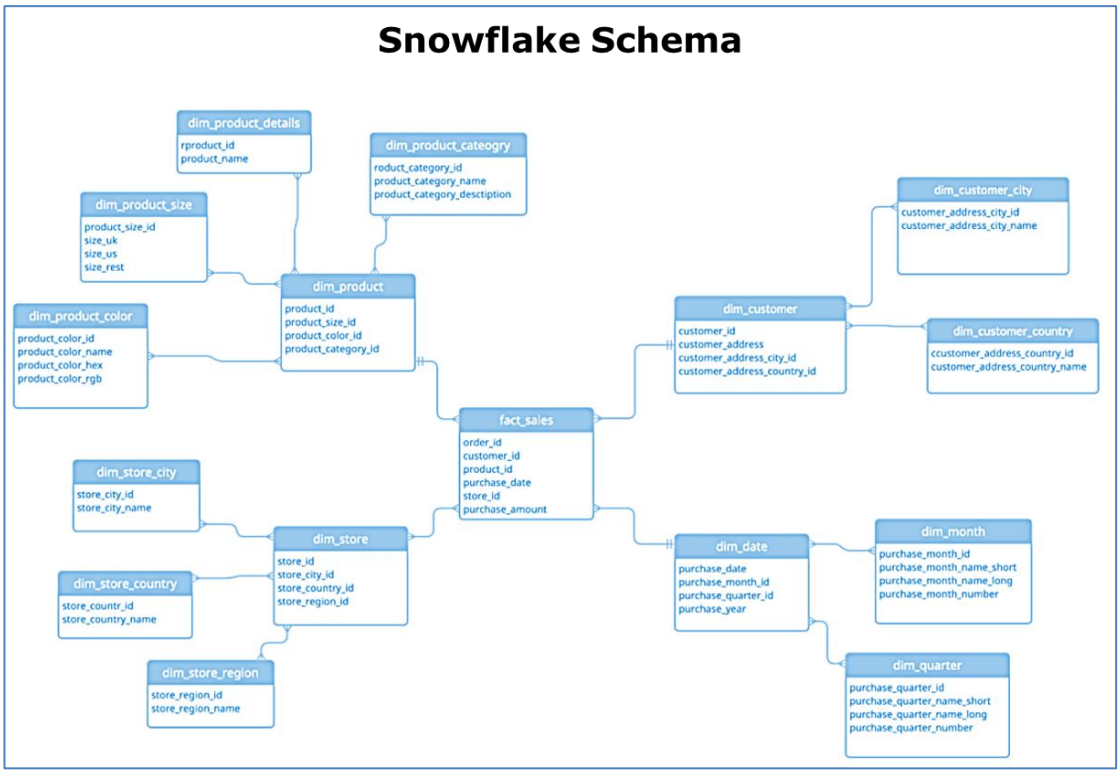

Figure 16: Snowflake Schema [15]

In the above diagram the customer country field is further lookup tables. The redundancy is fully avoided in the lookup tables.

The below are the seven important differences between star and snowflake schema

Table 4: Star schema and snowflake schema

Parameters	**Star Schema**	**Snowflake Schema**
Normalization	Denormalized	Normalized
Data redundancy	Redundant	Not redundant
Query complexity	Simple query	Complex query
Query performance	Faster	Slower
Disk Space	More storage space needed	Less storage space needed

Data integrity	More at risk due to duplications	Less or No risk of duplications
Setup and Maintenance	Easy to maintain	Hard to maintain

BI Reporting

In this layer of the architecture the actual reports are available. This layer is split into two sub layer, Data cubes and Reports. The reports are further classified under three categories, Dashboards, Ad-hoc reports, and standard reports.

Data cubes

A data cube in a data warehouse is a multidimensional structure used to store data, it is used to represents the aggregated data in terms of dimensions and facts [17]. The two main kinds of cubes are multidimensional cubes and relational cubes. Multidimensional Data Cube (MOLAP) are used to store data that represent the multidimensional view of the data. They help in storing large volume of data. Using indexing technique the multidimensional data cubes improves the accessing, retrieving, and storing of data. The Relational Data cube (ROLAP) are used to store the relational data, these are used as an extension of the relational database. Relational Data Cube used relational table to store data and they represent the dimension of the data cube.

Reports

Dashboards, Ad-hoc reports, and the standard reports are together called as reports. Dashboards provide very high-level information about the subject, these dashboards are mostly used by the decision makers, and they are updated periodically like weekly, monthly etc. Dashboards are used to measure the Key Performance Indicator (KPI) and Service Level Agreement (SLA). The Ad-hoc reports are reports generated only when needed. These reports are usually generated when there is a need to look at some specific case or incident. Incident analysis and the root cause analysis reports are generated under Ad-hoc reports. The report structure keeps changing ever time for the Ad-hoc reports. The

standard reports are the usual reports the company uses for the day-to-day business. These reports include the operational reports, financial reports and the CXO reports. The structure of the report does not change in the standard reports, and this helps to generate the reports in the stipulated time.

End-users

This is the last layer in the architecture, in this layer the end-users of the reports are provided access through the portal. The reporting tools provide these portals, and the users are restricted using their own credentials to limit the data access. This layer also provides report delivery via email and file delivery protocols. The format conversion of reports like Excel sheets, Character separated values (CSV), Text file, PowerPoint etc are done at this layer. This layer is often seen as the front-end layer for the overall BI solution. This layer also provides adaptability of report access to devices like mobiles, tablets, and other hand-held devices.

7.3 Data Transformation Strategy

Data Transformation Strategy is also called as ETL Strategy. In the overall BI solution, ETL is used in two places. The first one is between the source system and the staging layer and the second one is between the staging and the EDW layer.

Between the source system and staging, the ETL tools are used mostly for one-to-one data load. There is not much transformation involve at this point. The capability of the ETL tool to connect with multiple data sources is highly used here. The ability of the ETL tool to read files and unstructured data is much used here. Very basic verifications are done on the data at this level. The checks about empty file or file with zero kb, junk characters are some examples of the checks.

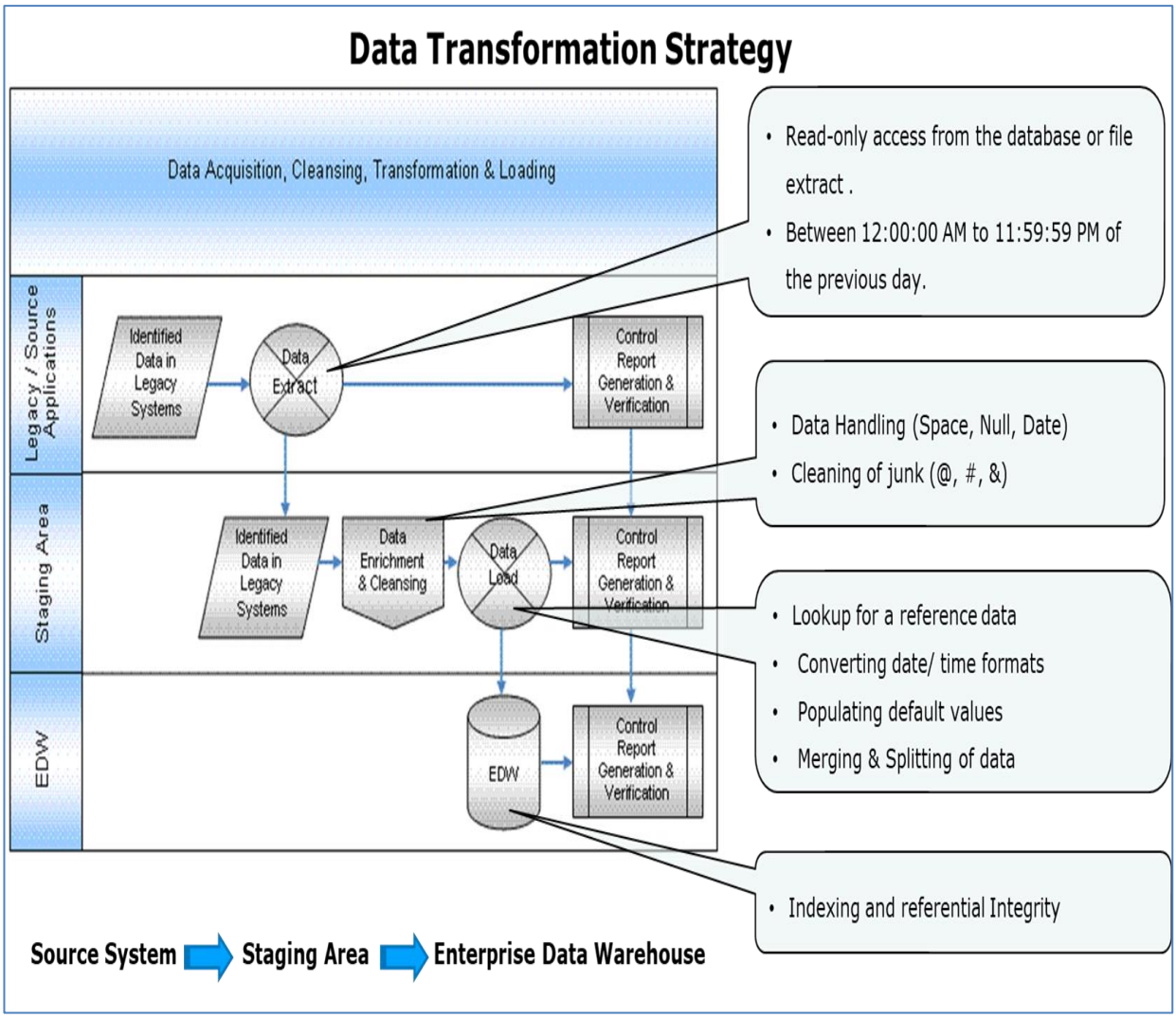

Figure 17: Data transformation strategy

The ETL tools are provided read-only access in the source systems. The data from the source system are read by the ETL tool or using file extract the data is moved to the staging. The data is filtered for the previous days records and been pushed to the staging tables.

Between the staging and EDW, the transformation ability of the ETL tools is fully used. Here the transformation of data from normal form to denormal is done by the ETL tool. Data enrichment and cleansing is an important activity done at this level. Increasing the readability of data and removing the unnecessary piece from the data is part of this activity.

The load from the staging to the EDW handles the data issues like null, blank space etc. The cleansing of data by removing symbols like @, &, # etc are handled. Next the lookup reference tables are made to ensure the data integrity. Conversion of date and time format to single format is handled. Data rules are applied at this point to enrich the data. The data rules like populating the default values and creating look up values to the data are made. Aggregation of data and splitting of data is also handled. Summary tables are created as a part of the aggregation activity.

8 Conclusions

This section summarizes the thesis and evaluates the success as well by comparing the business problem to the solution this thesis offers for the case company.

8.1 Executive Summary

The objective of this thesis was 1) to find out the reason for the delay in generating the monthly finance reporting every month and the reasons behind the inaccuracy of data in the report and 2) to come up with a solution to solve the issue.

The research for this thesis was done by studying the business problem analysis, analyzing the source system, reporting requirements and gap analysis along with literature research for applicable area.

The business problem analysis grouped the problem under four categories:

1. Corporate reports are not generated on time.
2. Lack of details in corporate reports.
3. Lack of data clarity.
4. Multi-directional flow of data.

With the source system analysis, the existing landscape of the case company was studied. The reporting need for the case company was further analyzed and been prioritized based on the Return of Investment (ROI) and the data availability. Gap Analysis brought light into the existing gaps in the current systems and the resolution to fill the gaps.

Based on the analysis a Business Intelligence Solution was proposed for the case company. This solution was designed more specific to address the business problems incorporating the resolutions from gap analysis.

The below table provides the details about how this thesis addresses all the identified business problem

Business Problem	**Problem Detail**	**Solution**
1. Corporate financial reports are not generated on time	Group reporting takes more than half a month to produce. This delays the compliance and the on-time corrective actions	The proposed BI Solution makes the data ready every day and stores in the EDW, hence the reports would be generated in couple of days, hence this issue addressed.
2. Lack of details (‘drill-down’) in corporate reporting	Not able to find the reasons behind the numbers in the reports	The proposed BI System architecture provided EDW layer with denormalized data which help to provide details to the report number, hence this issue addressed.
	Manual reporting and poor data mapping take more time to drill down to the details	The proposed ETL strategy along with the BI system architecture resolves the mapping and drill down problems, hence this issue addressed.
3. Lack of data clarity (single version of truth)	Different system has different data quality. Not able to trust the data and hence the reliability is lost	The proposed Data Transformation Strategy ensure the data quality in the DWH layer, hence this issue addressed

	Reporting done manually in silos	The proposed BI system architecture provides single reporting platform, hence this issue addressed
4. Multi-directional flow of data	Data flowing from one system to another and vice versa. This mess up the data	The proposed BI system architecture is designed with single data flow direction; hence this issue is addressed

8.2 Next Steps and Recommendations

The final proposal for the BI Solution is very useful for the case company, it offers clarity on the current issues, understanding of current landscape, details about the reporting requirements and the details of proposed solution.

The first step to implement the solution is to procure the necessary tools. This includes the ETL tool, database tool and the reporting tool. The next step is to choose the delivery model, based on the advantages and the disadvantages listed earlier in this thesis. The case company need to choose the best model of their choice. The next step is to roll-out a Request for Proposal (RFP) for the implementation vendor, the vendor selection plays a vital role in the success of the implementation, hence the company need to choose the vendor based on their capability. The last step is the execute the implementation along with the vendor. The case company needs to work closely with the vendor company throughout the implementation. Organizing monthly status meetings, monthly architectural forum discussion etc. are some of the recommendations to run the implementation project effectively.

References

1 A Basic Business Problem Analysis Technique (businessanalysisexperts.com) - Accessed 25 Mar 2022

2 Quantitative analysis (finance) - Wikipedia – accessed 26 mar 2022

3 Quantitative Analysis (QA) Definition (investopedia.com) – accessed 26 mar 2022

4 Qualitative Analysis — Jyväskylän yliopiston Koppa (jyu.fi) - accessed 25 Mar 2022

5 Trochim, William M., 2006. Qualitative Measures. The Research Methods Knowledge Base, 2nd Edition.

6 Architecture Strategy (bredemeyer.com) - accessed 25 Mar 2022

7 Definition of Solution Architecture - IT Glossary | Gartner - accessed 25 Mar 2022

8 15+ Best ETL Tools Available in the Market in 2022 (softwaretestinghelp.com) – accessed 25 Mar 2022

9 IT Roadmaps Explained with Examples – BMC Software | Blogs – accessed 25 Mar 2022

10 Computer Basics: Understanding Operating Systems (gcfglobal.org) - accessed 25 Mar 2022

11 What is Data Quality and Why is it Important? (techtarget.com) - accessed 26 Mar 2022

12 DBMS Normalization: 1NF, 2NF, 3NF and BCNF with Examples - javatpoint – accessed 28 Mar 2022

13 Denormalization - Wikipedia – accessed 28 Mar 2022

14 G. L. Sanders and S. K. Shin. Denormalization effects on performance of RDBMS - accessed 28 mar 2022

15 The 7 critical differences between a star schema and a snowflake schema (keboola.com) - accessed 29 Mar 2022

16 Snowflake schemas - IBM Documentation – accessed 29 Mar 2022

17 What is Data Cube in Data Warehouse? Definition, Types, Operations - Binary Terms – accessed 29 Mar 2022

Appendices

Appendix 1: List of Interview Questionnaire

1. Section: System Information
    1.1. What is the Name of the System Name?
    1.2. what type of System is this? (Payment/ Gateway/ etc.)"
    1.3. Name of the System Owner
    1.4. Name of the System IT Contact Person1
    1.5. Name of the System IT Contact Person2
    1.6. Name of the System Business Contact Person

2. Section: Infrastructure and Systems
    2.1. Details of the Network Landscape in which the application works (LAN/ SAN/ WAN/ etc.)"
    2.2. What is the Hardware Environment for this application (Server type)
    2.3. what is the Software: Operating system
    2.4. What are the Software Languages used for this application?
    2.5. Client Server Technology: Web Server / Web Application Server
    2.6. Is Any other S/W (Please give Version Numbers) used in this application?
    2.7. What is the Existing System Landscape? Or how many instances of the applications are running? (Dev, TEST, PRE-PROD, PROD)
    2.8. What is the Mailing System: Notes/MS Exchange for this application?

3. Section 3: Application Information

3.1. Please provide some General Description of the Application

3.2. What is the List of Businesses units supported through this application?

3.3. What is the List of Legal Entities supported by this application?

3.4. Is any other Geographical Regions supported other than Finland?

3.5. Is there any other Countries/ Cities supported by this application?

3.6. Is any Business Processes supported by the Application? (If yes, please explain)

3.7. Any major business process handled outside the Application?

3.8. Any Major Enhancements done/planned near future?

3.9. Any Minor Enhancements done/planned near future?

3.10. Any Major Interfaces from other applications?

3.11. List the Critical Areas in the application?

3.12. How is the application hosted? (Is the server hosting the application exclusive or shared by other applications

3.13. What is the Solution Architecture? Single or distributed?

3.14. Is there a Custom-built Application?

3.15. What is the Stability of the application? how often it troubles

3.16. What is the Volatility of the application?

3.17. What is the Frequency of updating Development /Test/ Acceptance Environment?

3.18. Applications in production since when? (No of years)

3.19. What is the number of Named Users in the application?

3.20. What is the number of Concurrent Users in the application?

3.21. What is the Usage of the system? And the expected Usage time of the application

3.22. What is the Peak usage time of the application?

3.23. What is the current average and peak workload of the application?

3.24. What is the expected / projected Workload for next year?

3.25. Are there any retirement Plans for this application? (If yes, please specify when and what.)

3.26. What are the front-end Technologies used?

3.27. Please explain process dependencies, if any, between applications

3.28. Any third-party tools (including interface tools) used during development/maintenance? If yes, please specify the details.

3.29. Does vendor of these tools support them now?

3.30. Are there any other dependencies of this applications?

3.31. Please explain about the Archiving process, if any, in detail.

3.32. Are there any processes and check points in place to ensure the accuracy of data input and output?

3.33. Is Disaster recovery plan adequate, well documented, and well tested? Details please.

3.34. What Industry Solutions, Country Versions/ Languages / Currencies are currently implemented? Are there any immediate additional requirements? (Local only / International)

3.35. What is the number of Forms used in this application?

3.36. Is any Data Conversion Programs in place in this application?

3.37. Does this system receive any other feed from External Vendors/ systems? (In terms of Data)

4. Technical Information - general
    4.1. What is the Database Technology used in this application? (With the version of the Database)
    4.2. How is the customer data maintained with in single application cluster/ group (Golden data)
    4.3. Mechanism - In what way the data can be delivered to EDW (MIS) system (Direct DB access/ File dump/ API/ Middleware/ Replication etc.)
    4.4. Latency - Is the data available is real time or near real time or batch
    4.5. Heap - Is data needs to be Incremental or Full refresh (in the daily operation)
    4.6. Mode - Availability of Data / Application for Push/ Pull (ETL)
    4.7. Topology - Is the system Centralized, Distributed or Federal

5. Technical Information - Data Volume and Governance
    5.1. What is the Number of Tables in this application?
    5.2. What is the Volume of Data available in Each tables/ entire application?
    5.3. What is the Type of Data (English or Arabic or Both or any other) in the application?
    5.4. what is the format of the dates in the application?
    5.5. From which date/ month/ year the Historical data is available (in any form)
    5.6. What is the Quality of the data (% of pure data)?
    5.7. Any Data Security Encryption Implemented / security (role based or user based)

5.8. Data Growth per day/ month/ year with units in terms of rows/ transactions and size

6. Reporting and Measurements
   6.1. What is the Present reporting Technology used (with Version)
   6.2. What is the number of Reports running in this application (with the frequency)
   6.3. Type of Reports (standard/ ad-hoc/ Dashboards)
   6.4. What is the required frequency of metrics reporting for production support issues?
   6.5. Any statistics-reports on the support level?
   6.6. Percentage of adherence to SLAs, issues resolved

7. General Information
   7.1. Any Major Overhaul in near future

8. Application Documentation Status
   8.1. Last updated date and Contact person?
   8.2. Is Business Requirements in place
   8.3. Is User Requirement Document in Place
   8.4. Is Design Document with architecture in place
   8.5. Is Workflow Diagrams in Place?
   8.6. Is Naming Conventions/ Standards document in place?
   8.7. Is Web-based front-end Style Guides UI design principles document in place?

8.8. Is Change/Configuration management Information document in place?

.

.

Appendix 2: Workshop brainstorming questionnaire

1. What is the Preferred DB for EDW?
2. What is the Preferred ETL for EDW?
3. What is the Preferred Reporting tool for EDW?
4. What is the Preferred OS (Windows, Unix, Linux)
5. What amount of Historical data to be Migrated to EDW? (Size in GB)"
6. Is History Data warehouse and data mining required in the new system?
7. What is the Data Standard to be followed in EDW?
8. What is the Date standard to be followed in EDW?
9. What is the Expected number of Standard reports from EDW?
10. What is the Expected number of Ad-hoc reports from EDW?
11. What is the Expected number of Dashboards reports from EDW?
12. What is the Data loading frequency to EDW?
13. What is the Expected Landscape (DEV, TEST, PRE-PROD, PROD)
14. Is Customer data Golden copy is expected from EDW?
15. Is Meta data reporting required from EDW?
16. Is Notification services through email and/ or SMS required in EDW?
17. Is Disaster recovery plan required for EDW?
18. Is Single sign on required for EDW?
19. Integrate dashboard/ reports with corporate portal is required?
20. What is the Number of Expected numbers of Named Users for EDW from your application?

21. What is the Number of Expected numbers of concurrent Users for EDW from your application?
22. What should be the Usage and availability of EDW?