Chapter 3

3. Technology Adapted

3.1. Introduction

Development of software to the end user satisfaction is not a simple task. What makes the whole process so complex is that there are many requirements, that has to be catered from a single piece of software, if one of which might not be fulfilled would lead to non compliance to the purpose for which the software was initially targeted. This will ultimately affect the end user un-satisfaction and hence leads to the failure in the whole project incurring major financial losses.

To overcome this situation, professionals have introduced, through their knowledge and experience, processors and methodologies which would streamline the software development process and improves communication, between the end user and the development team bi-directionally, which will make sure that all the requirements of the end user are exactly met.

It has been identified that there are fundamental activities which are common to all the software development processors as follows.

Software Specification

The functionality of the software and the constraints must be defined

Software design and implementation

The software to meet the specification must be produced

Software Validation

The software must be validated to ensure that it does what the customers want.

Software Evolution

The software must evolve to meet changing customer needs

3.2. Software Process Models

A software process model is an abstract representation of a process. It presents a description of a process from some particular perspective [4]

The waterfall model

Separate and distinct phases of specification and development

Evolutionary development

Specification and development are interleaved

Reuse/component-based development

The system is assembled from existing components

3.3. Data warehouse

Many consider Bill Inmon to be the father of data warehousing as he has contributed a great deal to build up the concept of data warehousing. According to the following paper 'The Customer-Centric Data Warehouse' [5], it refers to a book published by Bill Inmon in which he had stated the following definition for a data warehouse

"A Data Warehouse is a subject oriented, integrated, time variant, non volatile collection of data in support of management's decision making process" [5]

If we were to scrutinize this definition, further by analyzing each component

Subject Oriented – As opposed to Online Transactions Processing systems (OLTP) which deal with the functionality of a certain department (example Payroll system) Data warehousing deals a subject area (example Customer profitability in a Bank)

Integrated – integration process requires that data extracted from different (sometimes incompatible) systems are transformed into globally acceptable (uniform) data. For instance the same customer might have been given different identifiers in two different systems, needs to be assigned a global identifier.

Time Variant – Like OLTP systems, not restricted to current data, historical data is sometimes more important than current data when used for analytical purposes.

Non-volatile - Most of the time data is read-only in a data warehouse. With each loading, data is mostly added to the data warehouse, not updated.

3.4. Extraction, Transformation and Loading

One of the most time consuming and costly exercises in a data warehousing project would be the extraction and transformation of data with the intension of consolidating data. This scenario is quite common due to the fact the systems to be integrated most of the time would have been developed using different data structures and technologies, by probably different vendors and maybe at different point of time.

Extraction of data in this context is to identify where the required data (defined by the business requirements) lie's in the source systems, and to extract it to a staging area. Things become more complicated when a vendor has provided the source system and the internal data structures of the system are not known.

Transformation can mean many things such as cleansing the data in order to remove garbage, categorizing it using lookup tables, applying matching rules using known characteristics of the data and much more. This is to ensure that quality information is loaded into the data warehouse.

Loading data is the process of loading transformed data from the staging area to the data warehouse. Time duration taken for the load and management of disk space are things to be managed as time progresses.

3.5. Data Warehouse and Business Intelligence tools

Business Intelligence can be termed as the process that links bits and pieces of information within the data warehouse, turning it into meaningful information or knowledge, which can help us in making decisions. Business intelligence tools help us gain knowledge from the data warehouse using different techniques. These business intelligent tools can be categorized by the techniques they use. Given below are some of the main categories and tools associated with them. [8]

Reporting tools – They allow users to produce sophisticated, graphic intensive reports based on data warehouse data, which can be even deployed in the web. Oracle Reports is a good example for this category.

Query Tools – they enable users to submit ad hoc queries to pull data from the data warehouse and then display restricted data using parameters. It also allows slicing, drilling, pivoting on data retrieved. Tools such as Oracle Discoverer and Business Objects are very good examples of this category.

On-Line analytical Possessing (OLAP) tools – They provide multidimensional view of the data (cube) from the warehouse, where users can easily navigate through multiple dimensions and the hierarchies within each dimension. Oracle Express can be quoted as an example for this category. Modeling, forecasting and What-if analysis is some of advanced analytical tasks that can be performed by Oracle Express.

Data mining tools – help identify patterns and relationships in data. Also useful for building models that aid decision-making or predict behavior. In the banking industry, data mining can be used to detect trends in credit card frauds, Credit usage patterns, profiling customers who are most likely to take new loans and even forecast bad loans. The techniques used by these mining tools include Artificial Neural Networks, Decision Trees, Nearest Neighbor and Rule Induction. Examples of tools in this category would be Oracle10g Data Mining, Knowledge Seeker IV and Data Mind.