CHAPTER 1

1.1 INTRODUCTION

The exploration and production sector of hydrocarbon is not only highly competitive but also requires an enormous amount of investment (according to reports by HIS-[Information Handling Services] the capital expenditure was $641 billion in 2012) [154]. The world market today is facing shortage of oil and gas supplies leading to constant increase of prices, thus the requirement for further new discoveries as well as prompt decision making has become more important than ever. Most of the old oil fields are depleting and the new reserves are in difficult and sometimes almost in inaccessible location for easy exploration and production. Improper or untimely decision can lead to huge financial losses and may cause irreparable damage to the formation. For maintaining a competitive edge in this cut throat market accurate and efficient information and knowledge is required in real time or near real-time environment. For extraction from existing reservoirs high speed and intelligent simulation solution becomes a critical factor. Under this highly complicated scenario Exploration and Production (E&P) companies requires to utilize and implement information technologies to streamline their processes as well as for efficient, accurate timely decision making.
To be able to support speedy, intelligent analysis for optimum utilization of upstream assets, high performance computers as well as large data storage infrastructure is required. Oil companies are the companies which were the first to deploy clusters and grids. The concept of data warehousing was a comparatively recent technology which came into the market with the advent of more powerful tools like “ETL” (Extract. Transform. Load), and high level programming languages makes it possible today.

But in E&P sector the amount of data to be handled is in the range of 100s of “TB” (Terabytes) [153], which creates a challenge for the IT technologies being implemented. The technology has to be lot more sophisticated and the approach quite different from the conventional technology available in the market today.

1.2. Motivation of Work.

Separate independent studies [36] by “International oil companies” (IOC) corroborate the overall finding that “engineers spend half their time unproductively chasing data.” Consistency of data and information management customs has been acknowledged by one IOC as the basis of integration and automation of the whole enterprise [36]. It has ventured into building a software package called “production surveillance and optimization” (PS&O) solution, and under the circumstances it has a high rationality to go for it to reduce the unproductive time spent by engineers just preparing data for analysis. According to an internal survey conducted the company a staggering 44% of time is spent by
a technician or an engineer just for accessing and processing data for analysis purpose [36].

Further assessment of the same survey discovered that paltry 9% [36] of the operators were capable of retrieving data from the real time systems automatically, to feed into their geo-science or engineering routines. Whereas, 91% [36] of the operators conveyed that they on an average spend 50% of their time in identification and preprocessing of data, to enable it for different analysis tools. The results further showed that the time available for analysis and decision making was less than 25% for 55% [36] of the respondents. The above mentioned survey pointed out that a substantial amount of workforce is caught up in unproductive activities rather than utilizing their highly skilled potentials in productive and ground breaking work [36].

1.2.1. Review of Literature.

The primary literature which states the above given problem was published in SPE journal September 2009[36]. Which gave a clear identification of the problem in E&P data integration, identification and analysis? This states that almost 44% of productive working time is lost in unproductive and repetitive work. If this problem could be circumvented with the present technologies a huge amount of revenue (tangible & intangible) could be saved.[36]

The primary application of ETL techniques for the purpose of solving the problem of identification and cleaning of data could be easily used. Literature contains
many ideas put forward by many researchers to this end. “A conceptual modeling of ETL processes[2]” was published by Alkis Simitsis[2] and his partner. Another paper Spatial ETL Tools: The Bottom Line of an Enterprise GIS[155] was discussed by S.Raghavendran. It discussed the ETL tools which are being used for the purpose of data integration in GIS.

The ETL taxonomy was discussed in a paper written by Panos Vassiliadis et.al.[101]. In this paper the writer and his associates presented the method for identification of generic properties that characterize ETL activities. The concept of this paper is identification of generic properties that characterize ETL activities. For it they follow a black-box approach and provide a taxonomy that characterizes ETL activities in terms of the relationship of their input to their output and provide a normal form that is based on interpreted semantics for the black box activities. The proposed taxonomy can be used in the construction of larger modules, i.e., ETL archetype patterns, which can be used for the composition and optimization of ETL workflows.

The next important way of handling the problem is the use of Information Lifecycle Management which was nicely put forward by Mr.Paresh Chatterjee [29] in PCQUEST February. In this paper the “Storage Tiering” is discussed. Here, the capacity to be provisioned is divided into separate pools of storage space with various cost/ performance attributes. At the top resides the Tier 1 pool, which is the most expensive but high performing nonetheless. The bottom tier is occupied by more cost-effective storage arrays. The concept to devise a
sophisticated software layer that intelligently places data into the different tiers according to their value is also discussed. This concept is variously known as data classification or “Information Lifecycle Management” (ILM) which is highly required in my research work.

Another enlightening article was published in PCQUEST February edition 2010 by Rahul Sah, in which the author discussed the “advanced analytics”. [103]. Advanced analytics incorporates methodologies for answering future-oriented, proactive and predictive questions, as well as streaming data; so that consequent business decision is based on real-time questions about what's going on now. Advanced analytics leverages the same core features of typical analytics solution, i.e. reporting, dashboards, visualizations etc. but takes the analytics power to several steps further. Such powerful analytics capabilities would be required by enterprises where they generate bulk of data and there is a need to analyze that in real-time.

The above stated preliminary review of literature hints at the different technologies which could prove to be the avenue ahead to solve the problem at hand. The ETL technology could be the key to the solution of the problem if it could be modified and manipulated according to the specific requirement. This basically leads to the identification of methodology of the research work. Further in the second chapter a thorough review work of few related technologies is given in detail.
1.3. **Objective.**

The objective of this work would be as follows:

- To define and design an effective and efficient architecture.
- To develop an intelligent algorithm.

To facilitate a seamless real time data integration environment with ultimate aim of reducing the amount of time lost in information handling by E&P workforce in searching for data, integrating it from multiple sources, and preparing it for analysis in applications.

1.4. **Methodology**

1) Identify the different types of data generated in the following operations.(see chapter 2.6.2 for details)
   1. Geological Data
   2. Seismic Data/geophysical data.
   3. Drilling Data
   4. Well Log Data
   5. Production Data.

2) Specify the components for the required architecture.

3) Design efficient data integration architecture.

4) Develop the required data integration algorithm.

5) Identify the areas of further improvement.

6) Give suggestions for future improvement.
1.5. Structure of Thesis

The thesis is divided into four chapters. The first chapter discusses the problem and some review literature which authenticates the problem existing in the market. It also looks into the present technologies and its pros and cons for the specific market of Exploration and Production sector.

The second chapter reviews the different technological researches and development which has been done till date concerned towards the problem at hand. It also looks into the different aspects of Exploitation and Production Data Organization in attendance.

The third chapter contains the design and the components of the required architecture for the real time data integration in the E&P sector with the concerned discussions and arguments. The later part of the chapter states and explains the algorithm and the software developed for the testing of the algorithm.

The fourth chapter is the conclusion of the thesis, which looks into the problem; the solution formulated its limitations and scope of future work.