EXECUTIVE SUMMARY

The exploration and production sector of hydrocarbon is not only highly competitive but also requires an enormous amount of investment. 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 ruthless market, accurate and efficient information and knowledge is required in real time or in 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.

Separate independent studies by International oil companies (IOC) corroborate the overall finding that “engineers spend half their time unproductively chasing data.” One IOC has identified standardization of data and information management practices as a foundation for enterprise wide work-process integration and automation. The company is developing production surveillance and optimization (PS&O) solution, which was justified on its
potential to minimize the amount of time required by engineers and technicians to access data and prepare it for analysis. Based on an internal survey of current PS&O processes across all production units around the world, the IOC identified that engineers and technicians spend an average of 44% of their time accessing information and preparing it for analysis.

The survey further revealed that currently a meager 9% of the operator respondents are able to get data automatically from their real-time systems into their engineering or geo-science analysis routines. The other 91% of operator respondents had to consume more than 50% of their time and human resources to identify, format, and prepare the data for the analysis tools. A staggering 55% of the respondents had less than 25% of their professional time available for analysis, decision, and action. These statistics show that there is a significant potential workforce that is sequestered in the data commute, unavailable for value-adding activities that utilize their engineering and geo-science education and experience.

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. These 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, and high level programming languages. But in E&P sector the amount of data to be handled is in the range of 100s of TB, 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.

There are a lot of opportunity and space for development in this area. With the evolution of Information Technology (IT) there are a lot of technologies available today which could be incorporated for the streamlining of the information identification and integration.

The biggest problem of the E&P sector is not only the amount of data it has to handle but also the diverse computer system platforms which are used. The IT infrastructure of the E&P sector normally contains an assortment of operating systems ranging from windows, Linux as well as may contain Macintosh. The jumble does not end here; diversity is present in databases maintained by an E&P company. Data could reside in a proprietary database like Oracle, MS-SQL server as well as open-source databases like MySQL. This heterogeneity in case of Data Base Management System (DBMS) as well as operating system is a problem when communication is required between them for any kind of data transfer. There are solutions present in the market like Open Database Connectivity (ODBC), or VM Ware (Virtual Machine) to handle the above stated problem of communication but each one of it has its own pros and cons. An exhaustive study on organization of E&P data revealed further snags in its total data organization architecture leading to major predicament in data identification and process of integration. The whole architecture of IT infrastructure is in a chaotic state at a very best of time.

If the architecture of IT infrastructure could be streamlined appropriately most of the complications could be resolved. To solve the problem of
communication between all the assortments of databases is to put them onto a common platform like a cloud structure for convenient communication between them. The proposed cloud platform described in this thesis Eucalyptus Enterprise Edition (EE) 2.0 is an open-source, Linux-based software architecture which has the ability to realize a scalable, proficiency-augmenting private as well as hybrid clouds, utilizing the available IT infrastructure. It being an Infrastructure as a Service (IaaS) provider the company can utilize and implement its own assortment of resources (hardware, storage and network) all over its distributed assets. The cloud could be deployed on the On-Premise Data Center and can be accessed over the companies’ VPN (Virtual Private Network). As a consequence of which it enables a higher security to the confidential data of an E&P company.

The next step towards homogenization of the architecture is paramount for proper linking and removal of disorganized components in the IT Infrastructure. The propose architecture provides a perfect organized structure in the current scenario of an E&P company. The architecture streamlines all the different components of databases for the proposed real-time data integration process. Once the reorganization is done it becomes simple for an efficient algorithm for the extraction and integration of data from the source data marts to the target data warehouse.

The algorithm developed and described in this thesis looks into the different databases at located in different asset locations in a round robin basis within a scheduled interval of time. To be able to extract and integrate data at high speed for enabling real time data integration the algorithm needs to be simple,
light as well as agile. The algorithm designed here is very simple in its approach thus it is light. It generates a search command at a regular interval of time and follows all the enlisted distributed databases each at a time, whenever it encounters a new entry in it, it copies that entry into the central data warehouse with a time and date stamp. The next time the query visit the same database it looks for the last update date and time stamp and any entry later than that is again updated.

This enables a continuous real time data extraction and integration of E&P data into the central data warehouse enabling a faster and efficient process of Business analysis.

Through this process the basic problem identified in the beginning of the research of manually locating and extracting data for different analysis would become much more convenient and highly accurate.