CHAPTER 4

CONCLUSION

The main objective of this research work has been to identify and relegate the problems of data handling in the field of E&P sector. The varied type of data being handled in this field as well as the humongous quantity of data is an enormous problem in itself. But that is the challenge we are required to overcome. This research has tried to enumerate the different problems which are required to be detached from the data handling system, to pave the way for a real time architecture.

A lot of discussion has been done on the current market scenario of E&P sector in chapter 2 of this thesis. Other than that some even recent development worth mentioning are by Saudi Aramco[156] which is working on a conventional Data Warehouse project for handling its E&P data. The other one is by Microsoft known as Microsoft Upstream Reference Architecture Framework (MURA) [146], which is implementing the concept of Big-Data in the system. The basic drawback with a conventional Data Warehouse is, it has a high time lag for extraction of data from transactional systems, and the huge volume of data also augments the problem even further, so a safe conclusion could be drawn that the architecture does not support real time concepts. The Microsoft product is proprietary software which is based on the concept of big data so, it could be
safely presumed that it would be able to handle the enormous volume of data an E&P company is likely to generate. On the other hand high cost could be one of the major issues with it, not only for the software but also the hardware migration could lead to huge capital expenditure for a company.

Whereas the architecture developed in this research looks into the entire above mentioned potential problem. The architecture is far different from the conventional Data Warehousing architecture, fine tuned to make it as near as possible to Real-Time, it works with the existing infrastructure and there is no vendor locking. It also makes it a very cheap and viable option with the integration of open-source software. The SAP HANA also gives it an edge over any of the existing competitors by enabling higher throughput at an unbelievable speed of processing.

The technology present today is highly efficient as well as effective. The only requirement is to formulate a specific solution for the required problem and modify accordingly.

The technologies like ETL, Cloud Computing are perfect to create a solution for the specific problems faced by an E&P company in its IT Infrastructure. All it needs a streamlined architecture with the present efficient technologies and some innovating thinking.

Real-Time doesn’t have a definite or a single definition so it entirely depends on the realization of the requirement of the business an organization is handling.
Almost every chief E&P Company follows analogous workflow process for transmitting production data en route for office. Additionally there are homogeneous array of infrastructure and applications which support are aligned with projected workflows to acquire and supply production data into enterprise systems like SAP or JD Edwards. The main objective is to augment automation and diminish time and labor costs from work processes while obtaining timely production information from the field.

The best solution for any problem is the simplest one. In the above depicted architecture the base platform of all real time data generation like SCADA system is based over a cloud platform, which in turn will provide a highly virtualized environment to enable unimpeded communication among the different heterogeneous data sources. With cloud storage platform the most inconvenient problem which has dogged illustrate E&P industry till date, of heterogeneous source databases for integration would invariably reduce the time and effort for preprocessing of data required before diverse database integration.

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.

After implementation of the architecture with the developed algorithm the test results has shown to be highly effective in addressing the initial problem stated at the beginning. The tests have been conducted in a controlled and limited environment with emphasis on the speedy and flawless execution of the developed algorithm.

**SCOPE OF FUTURE WORK**

In this research the emphasis was on the identification and integration of data, but there is another aspect of quality which needs to be looked into. Any data which is to be put into a system for analytical processing needs to be “cleaned” and checked for any kind of discrepancies. Thus a lot of scope is present in this sphere for future work. The major concern in this research has been the different aspects of cloud computing as a whole, which has not been able to provide a completely reliable environment to the user as a system. A lot of concerns and apprehensions are present today with all the products available today in the market.

One of the biggest concerns in present day cloud system is the ownership of the data and its security. This is more the managerial aspect than the technicality of the problem. There is a lot of research opportunity present today in this area of cloud computing.
Another big problem the cloud computing is being ploughed by is vulnerability of
the client of being over dependent on a single provider and the products they
market. It limits the different options which could be profitably availed in the
market.