9th Biennial International Conference
& Exposition on Petroleum Geophysics

"HYDERABAD 2012"

P-136

Developing a seismic data information center: a discussion

Santanu Mitra*, Kuldeep Prakash, *ONGC*

Summary

Developing a robust seismic database, important enough to cater to the requirements of all geo scientists at every level, is a great challenge. All hardware and software used across different spheres of exploration activities widely differ in constitution and usage. Present discussion is on bringing a compatible solution for seamless data flow through a comprehensive but complex data base with a GIS based graphic user interface for both online and near line data accessibility to different forms of seismic data and related information without compromising with state security.

Introduction

Since independence, Oil exploration & exploitation in India was dominated by two giant national oil companies namely Oil & Natural Corporation Ltd. (ONGC) and Oil India Ltd. (OIL). The scenario started to change in the mid-nineties, after the implementation of "New exploration & licensing policy" (NELP) through Directorate General of Hydrocarbon. This policy was meant to encourage private oil companies to explore oil & natural gas on Indian shores. Going back to the late sixties, digital seismic data acquisition in India was in its infancy and so were its entire storage systems. With the passage of time, round reels (½ inch, 9-track recording) of the early seventies gave way to 10/20/30 GB capacity 3590-IBM cartridges in the early nineties. Although devices like LTO-1 & DLT were also available in the market, they could not emulate the reliability and simplicity of IBM designed '3xxx' series cartridges. By the turn of 21st century, it is widely thought that disks will eventually replace the utility of data cartridges only if it can emulate the reliability and linearity of the latter. Meanwhile, with the advent of newer cutting edge technologies, there was an explosion in data volume across the oil industry as a whole. Managing these diversified and huge data volumes required complex scientific database systems, at different geographic locations, with cutting edge technology. NELP has also warranted all oil companies to start building robust data centers' at their end for their own survival.

Problems of time overflow

It has been a standard practice for the oil companies in India to acquire land seismic data, in a given seismic project, over a period of six months or more. After completing field data acquisition, the data is then sent to the company's nearby **R**egional **C**omputer **C**enter (**RCC**) for processing (As chalked out in annual work program commitments). Normally the processing program spans from five to eleven months depending on the complicacies of geology, data volume and availability of hardware resources. But for re-processing of the same data (if required) after a gap of some years, a considerable period of time is spent in finding, loading and unloading of the same data and other related information's. Frequently the project time table spills over due to problems encountered during data loading. In the recent times, API companies across the globe are slowly embarking on a comprehensive data base solutions whereby all form of data volumes are available on a common pool with GIS based graphic user interface. These modern databases keep all type of trace data & related information separately in a '*Device independent'* and '*Format free*' environment. This eliminates the need for format conversions and data archival of old vintages periodically. The client has the liberty to select from the entire volume of data available online or near line through a GUI interface. It cuts down project time by a fourth and utilizes the hardware resources to the maximum. The initial investment on this database may be staggering but the benefits are enormous.

GEOPIC, Oil & Natural Gas Corporation Ltd, Dehradun
E-mail: santanu_ongc @ yahoo.com

Problems of plenty

Three decades ago, companies like ONGC was acquiring data with 96 channel DFS-V/MDS-10 units both in offshore and onshore with the data volume hardly in few hundreds of gigabytes. The offshore data from three to four prospects were acquired and sent to RCC's for processing in one financial year.

By todays standard, offshore 3-D seismic data are acquired mostly by foreign vendors which have multiple streamer capacity with array of source air guns and more than three thousand seismic channel recording system. A typical offshore seismic field data volume is received in few thousands of Gigabytes. Managing such huge data volume of data is a big challenge. For example, this year alone Geo Data Processing & Interpretation Centre (GEOPIC) at Dehradun received twelve offshore seismic field data volumes from twelve different offshore investigation numbers. Similarly, on land 3-D projects has also undergone a sea change. All in all, the annual intake of bulk data volume managed by the data managers has increased hundred fold.

Problems of Hardware compatibility

In present scenario, oil companies across India are trying hard to modernize and update the entire range of computer systems (field recording units in operation and processing /interpretation computer systems) spread across the country. Every new computer system brings with it a new set of operating system, application software's and compatible storage system. For example, the most commonly used field recording system in seismic field parties, SN-408UL operates on SUN-SOLARIS platform, SUN-SPARC hardware & 'NAS' (Network attached storage) type output devices. On the contrary, newly acquired "SCORPION" field recording unit still relied on IBM DLT-3 type of cartridges for output but operates on ultra modern DELL hardware coupled with LINUX & Window operating systems. SN-408UL delivers field data in SEGD format where as "SCORPION" field recording unit delivers data in SEGY format. Processing software's like "EPOS-4" of Paradigm Geophysical requires SMP (Symmetrical Multi-Processor) computing machines whereas "GeoCluster" of CGG Veritas and "OMEGA" of Western-Geco operates on multi-core PC Cluster systems.

Previously, 100 Kbps Ethernet networking system was enough to run a 32 CPU Origin2000 machine in full potential. Today's requirement are multi channel fiber optical cable network system with hundreds gigabytes per second capacity. The power requirement has also gone up exponentially. To get a perfect match of computer solution is difficult and rare.

The problem in different types of devices

In the early eighties, a 6250 round reel or a 3480E cartridge could hold kilobytes of data. The packing densities of these devices were low and therefore not prone to frequent breakdowns. As an example, field data of a typical 2-D seismic line are accommodated in more than ten (10) 3480/3490 cartridges. The same is not true for high density devices like 3592 cartridges, which have an intrinsic capacity of 500 Gigabytes. As a result, data of an entire 3-D project can be accommodated in one cartridge, leaving considerable space to spare. So if one data cartridge goes bad, the entire volume of data is lost forever. These devices are very sophisticated, but very allergic to dust. The breakdown rates are also on the higher side. As a result, the executives of the company should keep at least three sets of same data at three different geographic locations in secured databases.

The difference in data types coupled with compatible storage devices, demands a high level of expertise to handle them properly. It is also difficult to provide at least eighty (80) percent data volume for these high density cartridges for meaningful utilization.

A solution by using Storage Manager Device utilities

To solve this problem the most economical and viable option available in the market is to use a ***Storage Manager*** for storing of bulk seismic data instead of manual libraries. As an example, GEOPIC adopted **T**ivoli **S**torage **M**anager (**TSM** in short) for storing dataset backups and bulk seismic data. In this ***Storage Manager***, data is stored in form of files or folders (in data encapsulated form) in a device independent environment. In this environment alone, disks and cartridges are complementary in nature under a defined storage policy and do not require any kind of human intervention at any point in time while running jobs. Moreover, more than one copy of the stored data can be created by making alteration in the defined storage policy. The essence of this type of storage managers is its space management. For example, if the primary storage device is defined as cartridges and if the data volume within some of these cartridges goes below a certain "***Water Mark***", the data from all these cartridges are migrated to a new scratch cartridge and the space of the earlier cartridges are reclaimed for future use. This is explained in Figure-1, 2 & 3. In case of a bad sector within any cartridge or disk, the same process is repeated. This is a unique way of utilizing the full potential of all storage devices. Also, these **Storage Managers** are autonomous by definition (i.e. does not require any human intervention at any point except when the stock of input cartridges/disks are exhausted).

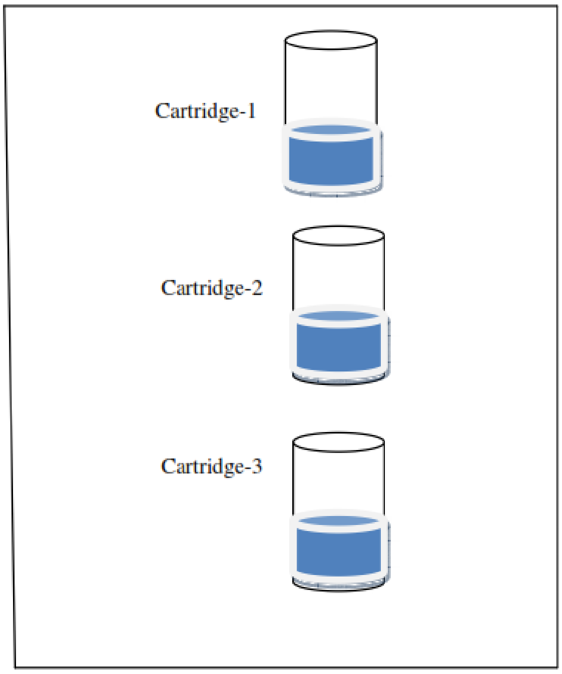

Figure-1: Data in cartridge-1, 2 & 3 are less than 33%

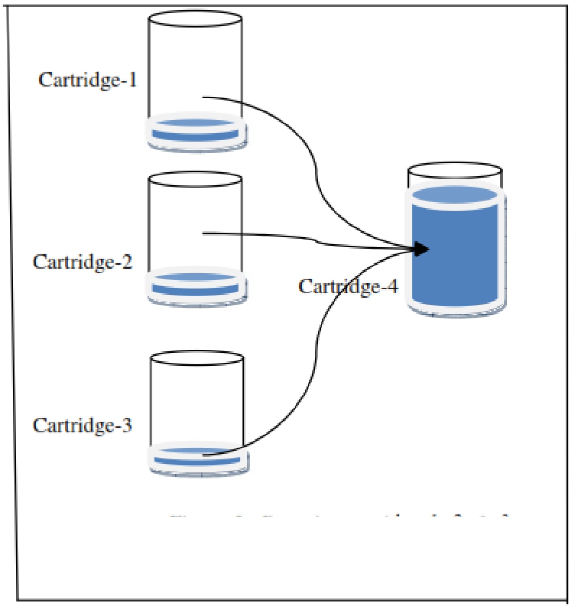

Figure-2: Data in cartridge-1, 2 & 3 are migrated to Cartridge-4

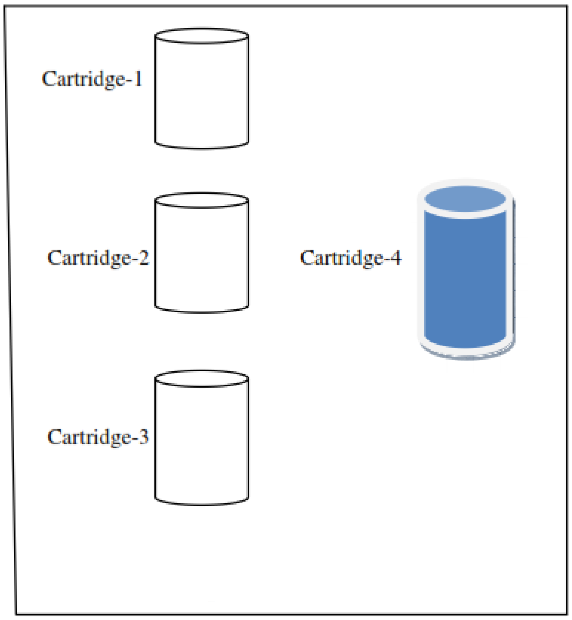

Figure-3: Cartridge-4 retains all the data of cartridge- 1, 2 & 3. Cartridge-1, 2 & 3 are reclaimed as scratch

The effective solutions

There are two road maps to solve this problem. One road map is to have three sets of all bulk & voluminous seismic gather data, stored in devices like cartridges, at three different geographical locations (including the location of the primary user and the owner of the data) with the help of a robust near line database connectivity so that locating them physically is very easy and are readily available at any point in time (if not issued).

The other road map is to have a regional GIS based database. In this database, all seismic stack data are decomposed into trace and trace header data. They are then stored separately in a device independent environment (i.e. trace data are stored in disks or cartridges through a **Storage Manager** & the trace header data in tables using database software's like '**Oracle®**') which can be accessed globally through a graphic user interface (using graphic software's like **'JAVA®'**). This is based on the assumption that stack data is the most compact form of presentable seismic information note. All seismic stack data is normally coupled with velocity information and some essentials like the highlights of field data acquisition and processing. One can have a bird's eye view of the history of seismic activity being carried out in the area of interest on a geographical scale. This type of database is more focused towards the senior executives of the company for taking administrative decisions (Figure-4a). Micro informative trace data volumes like shot gathers and CMP/CDP gathers are required by clients engaged in processing and other similar activities at the ground level. The data, they handle, are bigger in volume (Thumb rule for volume calculation of gather data is to multiply stack data with foldage) and therefore difficult to handle. As a result, online loading of these bigger versions of data in an interactive database is practically ruled out. The datasets are physically stored in devices (i.e. cartridges) & in racks with bar code labels. This information is then exported to a relational database which is commonly used by the end user (Figure-4b).

Also, for an effective database solution, precise information (in tabular form) on data type and its contents of information (**Metadata**) available within the company along with its exact geographical location is of utmost importance. Executives of the company have to decide on how best to publish this **Metadata.** It may be published nline or through hard copies. As an example, ONGC has published the seismic **Metadata** in hard copies with the name '**Atlas**'. This **Atlas** is exclusively for internal circulation.

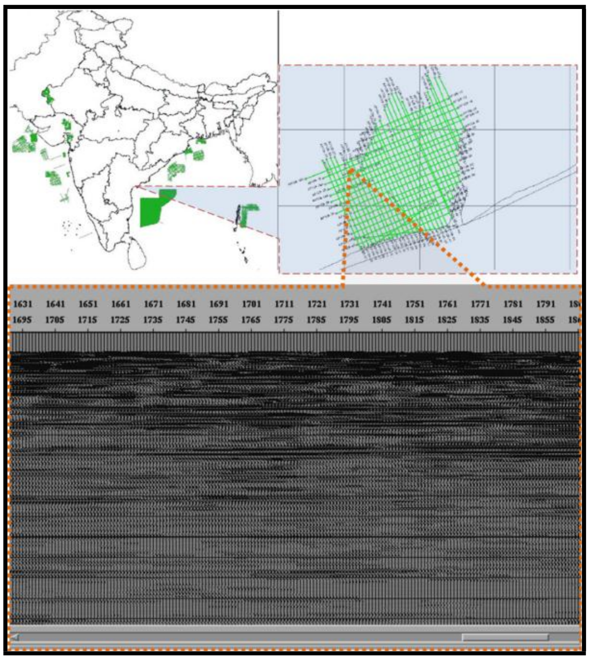

Figure-4a showing a typical GIS based database whereby a user with a valid user license can select seismic stack data (and other related information) through a graphic user interface.

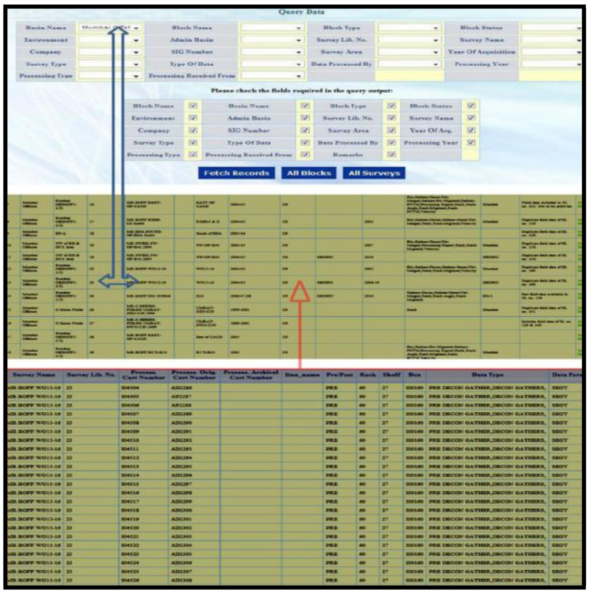

Figure-4b showing a near line database for geoscientists to find the physical location of devices/cartridges containing data of their choice.

However, a database structure of this magnitude (which is highly complicated structurally both from software's' and hardware's point of view) requires huge investment and a large skilled manpower to operate. Therefore oil companies with sound financial and technical base can only operate such databases.

Quality Control (QC)

QC of the available datasets is considered as one of the cornerstones for any database. The format and the type data, to be kept in the database, should fulfill certain criteria laid down in advance. As for a normal routine, input data should be in standard format, which unfortunately had been the bone of contention between a storage manager and the client for years now. It should be kept in mind that by keeping data in a standard format, the longevity and dependability of the data in question is increased by multiples of decades. For seismic trace data, the standard formats are either SEGY or SEGD (with a mention of the SEG version i.e. 8036, 8042, 8058 or 8068 etc.). Navigation data are normally in flat ASCII file with a standard format like UKOOA P-1/90, UKOOA P-2/94 or SPS format. Velocity information does not have any global standard format. Therefore they can be in non-standard ASCII flat file formats. There is also no universal standard format for loading interpreted seismic projects into the database.

Available database soft wares in the market

PetroBank® database software from Halliburton and FINDER® database software from Schlumberger are the only software's which fits into this database framework. These database soft wares are very complex in nature and contain more than one third party software (like Oracle®, JAVA® etc. in runtime version) which needs to be acquired and maintained separately. Some petroleum E & P companies have designed their own GIS based database system (as per the internal requirements) in the same line as discussed. E&P and Saudi Aramco, for example, has used GIS to integrate exploration and producing (E&P) spatial and tabular information and its E&P users have applied appropriate geographic analysis tools to thoroughly analyze potential for petroleum at a location.

Conclusion

The world is moving toward a device independent and format free storage solutions for storing the huge volume of input data along with their processed and interpreted derivatives. In today's scenario, oil companies require such complex, secured but easily accessible data bases whereby '*field_acquired_shot_point_addressed* sequential gathers' or '*CMP/CDP_addressed* sequential gathers' are stored in files or folders within a Storage system and a part of these files are kept online for display purpose to authorized users through a GIS based graphic user interface. More over, these databases should be customized heavily to meet the requirement of specific clients and countries. For example, the databases like **PetroBank®** was initially designed keeping in view the 'E & P' requirements of the Norwegian government. In an environment of cut throat competition, proper utilization of data for yielding better results is any geoscientists dream. Therefore a database which is not dependant on specific devices and stores data in the smallest possible fragments (so that the data can be supplied to clients in any format and in any type of devices) has become a necessity. A data storage centre having such a database in the centre stage, coupled with a facility to archive old vintage data of different formats, would be an institution in itself.

References

Tivoli Storage Manager V6.1 Technical Guide Redpapers, published 1 Sep 2011

PetroBank® Master Data Store™ E&P Database Software-Published overview by Halliburton.

E&P GIS: Integrating E&P Data and Applications by Imtiaz Ahmed & Adil Al- Marzooq; Petroleum Engineering Applications Services Department, Saudi Aramco, Dhahran, K.S.A.

Acknowledgement

The authors are thankful to ONGC for providing an opportunity to allow this discussion to be presented in a forum like SPG. Authors place on record their sincere thanks to Director (Exploration), ONGC, for his kind permission to publish this discussion. The authors are deeply thankful to Shri S. K. Das, ED & Head-GEOPIC, ONGC, Dehradun, for providing opportunity, guidance and note of encouragement at every stage of work. The authors extend thanks to all the members of GMS division for providing the necessary inputs to prepare the manuscript. Thanks are also extended to associated programmers of GMS for educating all about the technical know-how's of 'Robotic Tape Library', 'Tivoli Storage Manager' and 'Networking systems'.

NB: the views, expressed here, are solely of the authors and do not necessarily reflect the views of ONGC.