PaperID **AU209**
Author **Pradeep kumar Kukreja , ONGC , India**
Co-Authors **Akhil Puri**

Seismic Data Compression in E & P data storage A case study

Keywords: Seismic Data Compression(SDC), Roaming, E & P, SEG-Y.

Abstract

In E & P industry, seismic data consumes the maximum data storage, accounting for up to 80 - 85% of enterprise storage requirements. Size and complexity of data storage capacity are growing exponentially posing a variety of challenges with the introduction of rich, wide and full-azimuth seismic data. Adding to the challenges is the advanced processing, imaging and interpreting technologies that create datasets exceeding 40-50 terabytes and petabytes of data volumes that are fast becoming the new normal. In fact there are currently seismic surveys being planned with a volume of around 120 terabytes. Until today, the answer to growing datasets and projects, ensuring access to original SEG-Y files from selected petro-technical solution, and avoiding costly data duplication, has been more hardware upgrade oriented. Now IT departments are forced to decide between short-term “instantaneous need-based” hardware growth and more strategic approach to benefit both geoscientists’ immediate data requirements and IT’s long-term operations.

What if the geoscientists could handle large seismic volumes interactively from the desktop? What if seismic data size could be controlled and reduced without losing any of the geoscientists’ valuable information? Paradigm, in collaboration with Hue, has answered this question by implementing a proven, high-performance seismic compression algorithm within Paradigm's interpretation solution. Geoscientists now have a powerful option to compress seismic data while retaining 32-bit precision thereby minimizing the need for more storage disk space. Efficient seismic data access-transfer and decompression is achieved through optimal use of the workstation’s graphical and CPU capabilities in order to parallelize the decompression of seismic volumes within software environment using prominent Software Package.

This paper dwells on the emerging seismic data compression solution to the challenging growth of data volumes facilitating the G & G users with both efficiency and effectiveness maintaining the data integrity and quality of embedded information with economic benefits to the organization and intellectual drive to the IT solution providers for innovative dynamic thinking. This has been carried out in the backdrop of a sample data set of Mumbai offshore basin demonstrating the efficacy of the SDC solution.

Introduction

During oil and gas exploration and production process, seismic data plays a pivotal role in first imaging the subsurface, then enabling identification of drillable prospects, then guiding dynamically the drilling process, then in case of any discovery delineating and developing the field followed by monitoring and maintaining the reservoir pressure with dynamic production from the field. In each stage, seismic data appears in varied forms and is integrated with other data to sustain the process effectiveness. The imaging of the subsurface begins with seismic survey that produces a huge amount of seismic data. This data is processed at computer center and transformed into subsurface images. A typical seismic survey can produce hundreds of terabytes of data. With the advent of wide-azimuth seismic and other high-density acquisition technologies, surveys are commonly 40-50 terabytes or more. Considering the number of ongoing projects and assets across ONGC, there are massive amounts of data in daily use, specially, in WOB, Mumbai where seismic data are in Broadband and Ocean Bottom node (OBN-Seismic) having multi-fold. This huge seismic data volume need massive amount of network bandwidth

consumption during the process. Co-locating compute, data, and visualization environments minimize data movement outside the data center and maximize performance, but the data storage requirements, network bandwidth and application latency still apply inside the typical data center. Data handling is becoming painfully slow due to Seismic file which exceed 40-50 TB. Table below show the approximate data handling through designed network bandwidth. The data size and network issues negatively impact geoscientist productivity which is already proven. Moreover Geoscientist has a deadline to complete the project in stipulated time for releasing the locations of their respective areas which cannot be modified or shifted because of an issue of disk storage.

SEG-Y file size	100 Mbit line	1 Gbit Line	10 bit line
1 TB	>25.5 hours	>2.5 hours	15 mins
50TB	>53 days	>4 days	12 hours

The Seismic big size issue has typically been to buy bigger and better workstation hardware, faster networks and faster, more capacious disk arrays. The purchase, installation and maintenance of this hardware add significantly to the total cost of ownership and operation of software platforms, especially if those platforms impose local data storage requirements on user workstations. It is apparent that this primitive way of working is not ideally suited to today's industry, which is currently operating under restricted budgets for acquisition, software and hardware, workforce reductions, and with less time available for users to parse through huge amounts of data. This trend is driving an industrywide rethink of seismic data management and storage practices.

The Seismic data compression roaming (SDC) is very useful for loading all future data into compressed format and delete/archive originals seismic migrated data gives Reduce size, Increase speed and full dynamic range over the data.

This paper deals with the emerging seismic data compression (SDC) algorithms to make the storage more efficient thereby reducing the broadcast and cost and enabling the users flexibility to use data with efficient storage space management without playing with the integrity and quality of the data driven information in the complexity of exploration and production scenario all over the industry.

Methodology & Results

SDC software provides a dynamic platform to the users by continually updating its solutions with the latest computing capabilities thereby facilitating users load, display, process, and analyze the huge data volumes. Some of the more effective options for providing an optimal return on investment in computer hardware are the use of multiple, background threads, which can parallelize I/O of data across multiple central processing units (CPUs) and clusters of processing "blades," and utilizing the parallelization capabilities offered by modern graphics cards with their hundreds of onboard cores.

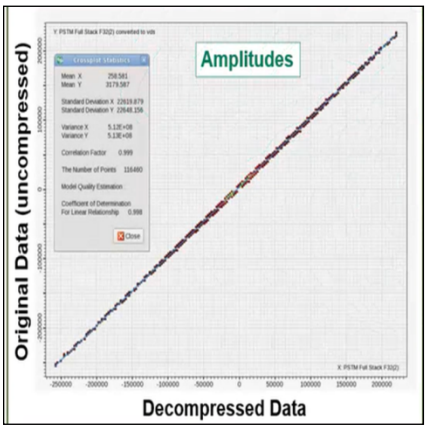
SDC method provides for managing the large data conundrum and Compression roaming enables the compression of seismic data into much smaller volumes on disk. Efficient reading of seismic data into software applications is achieved through optimal use of the graphical and CPU capabilities of a workstation to parallelize the decompression of seismic volumes. Combined with the reduction in data volumes being transmitted across a network, this typically results in significant improvements in data access speeds. Seismic compression helps companies reduce their annual outlay on disk storage and networking, and improve access speed to large 3-D seismic datasets. Typical compression ratios achieved through the use of compression utilities range from 4x to 20x or more. Compression up to about 4x is virtually lossless and adequate for quantitative workflows. Higher levels of compression correspond to higher levels of accuracy loss, but compression ratios as high as 15x or more can be adequate for structural interpretation workflows.

The compression algorithm is based on 3-D wavelet transformation and involves no clipping or truncating of the data. Being insensitive to amplitude variations, it keeps the same quality of compression throughout the data without causing any coherent artifacts. The compressed file is decompressed on the fly prior to

being used by applications to become a brick file, which is non-compressed, multiresolution and optimized for roaming in a 3-D environment.

For seismic data compression (SDC), the most important consideration is how to represent seismic signals efficiently, that is to say, using few coefficients to faithfully represent the signals, and therefore preserve the useful information after maximally possible compression. It is easy to comprehend that compression effectiveness is used for different expansion bases. Many orthogonal transforms have been used for data compression.

The methodology also ensured the quality based compression level which gives same quality of compression throughout the entire volume, not variable on amplitude, keeping continuity in signal.



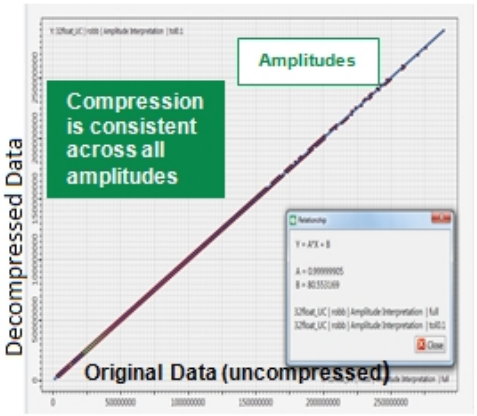


Fig 1 : Compression/decompressed is consistent across amplitude

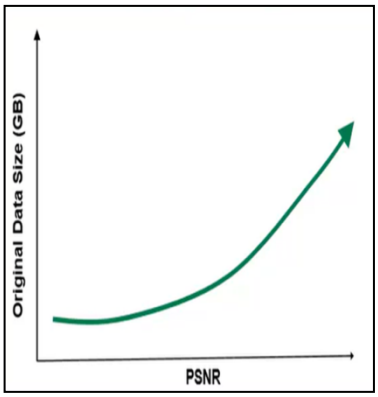


Fig 2 : Compression quality increses with data size

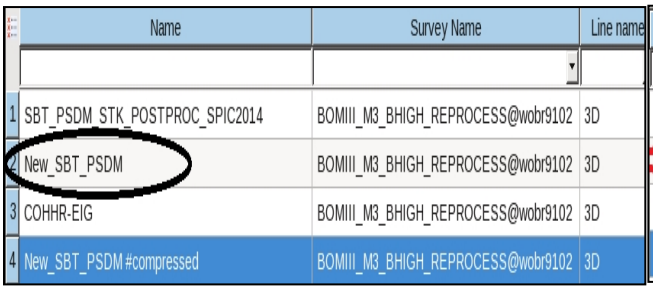
Full precision (Fig-1) of the data is retained resulting in the higher qualtiy than alternate method.

Compression quality is consistent across amplitude (Fig-2) maintaining signal to noise ratio of original data. Infact the compression quality increases with increasing data size.

Highest data quality in the Oil industry, relative to file size and over 10 dB higher quality (measured in PSNR) with same level of compression compared to latest technology.

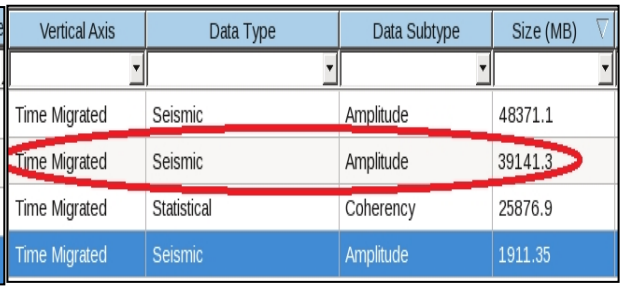
The method in which the Original Post stack data is loaded (full-size) from SEG-Y to brick or Trace format and then compressed depending upon the compression level and saved as a separate compressed file and listed as such in the Seismic Manager. Original full-size file can be deleted**.** Thus saving of huge storage space with no performance difference between small and large files where network bandwidth is a bottleneck, a reduction in data volume transferred over the network will improve performance . System manager can observe drastic improvement using multiple CPUs or GPUs in the system.

This case study dwells, a sample data set of Mumbai Offshore basin broadband PSDM (Fig-3) comprising a volume of around 39 GB (Fig-4) has been compressed in the ratio of 20:1 thereby reducing its size to less than 2 GB demonstrating the efficacy of the SDC solution.



	Name	Survey Name	Line name
1	SBT_PSDM_STK_POSTPROC_SPIC2014	BOMIII_M3_BHIGH_REPROCESS@wobr9102	3D
2	New_SBT_PSDM	BOMIII_M3_BHIGH_REPROCESS@wobr9102	3D
3	COHHR-EIG	BOMIII_M3_BHIGH_REPROCESS@wobr9102	3D
4	New_SBT_PSDM #compressed	BOMIII_M3_BHIGH_REPROCESS@wobr9102	3D

Fig-3 : Seismic data set 'New SBT PSDM'



Vertical Axis	Data Type	Data Subtype	Size (MB)
Time Migrated	Seismic	Amplitude	48371.1
Time Migrated	Seismic	Amplitude	39141.3
Time Migrated	Statistical	Coherency	25876.9
Time Migrated	Seismic	Amplitude	1911.35

Fig-4 :Time migrated data set of 39.14 GB volume

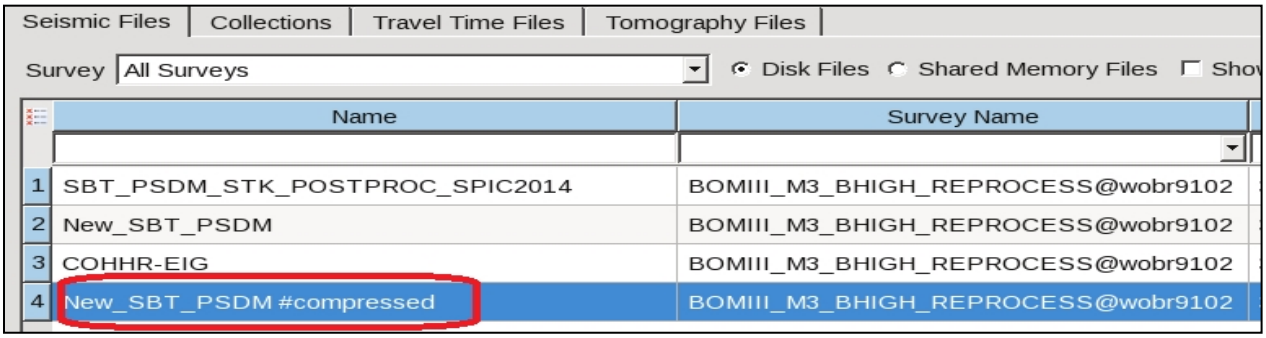


Fig-5 : New seismic data set after compression

After compression the new seismic volume has been preserved (Fig 5) with another name in the same Seismic data Manager with extension 'Compressed". The compressed data has maintained its integrity with undistorted amplitude and quality at much higher levels (4x to ~22x) of compression suitable for common workflows using different applications.

This has been depicted (Fig-6) for comparison of time Slice of original vs. compressed volume demonstrating how the image quality and data sanctity has been maintained.

.

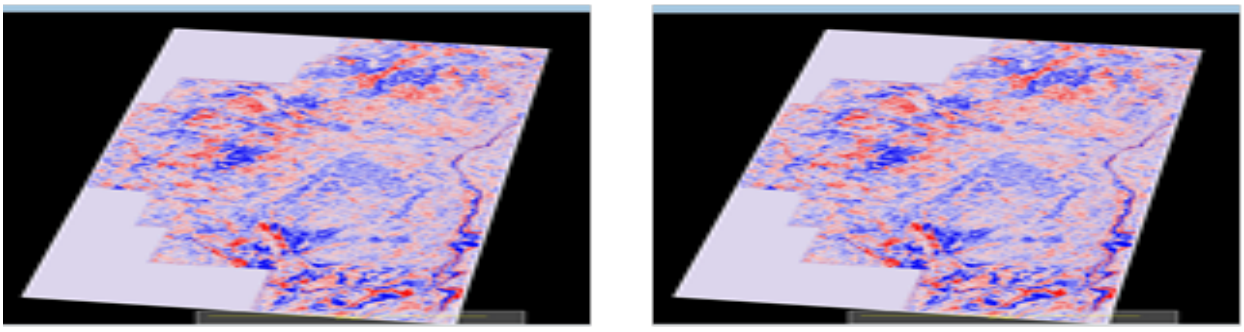


Fig-6 : Comaprison between original(39.15GB) & compressed Time Slice (7 GB) Volume

The original seismic time migrated amplitude volume of around 39 GB is shown against its 22x compressed volume of around 1.9 GB (Fig 7). Given the importance of seismic data quality, the compression algorithm does not negatively impact amplitude, phase and frequency retaining all data at full 32 bit.

In Geophysical terms, compressed volume will load faster in the memory enabling quick seismic interpretation and Visualization. While working on regional perspective, dealing with huge volume,

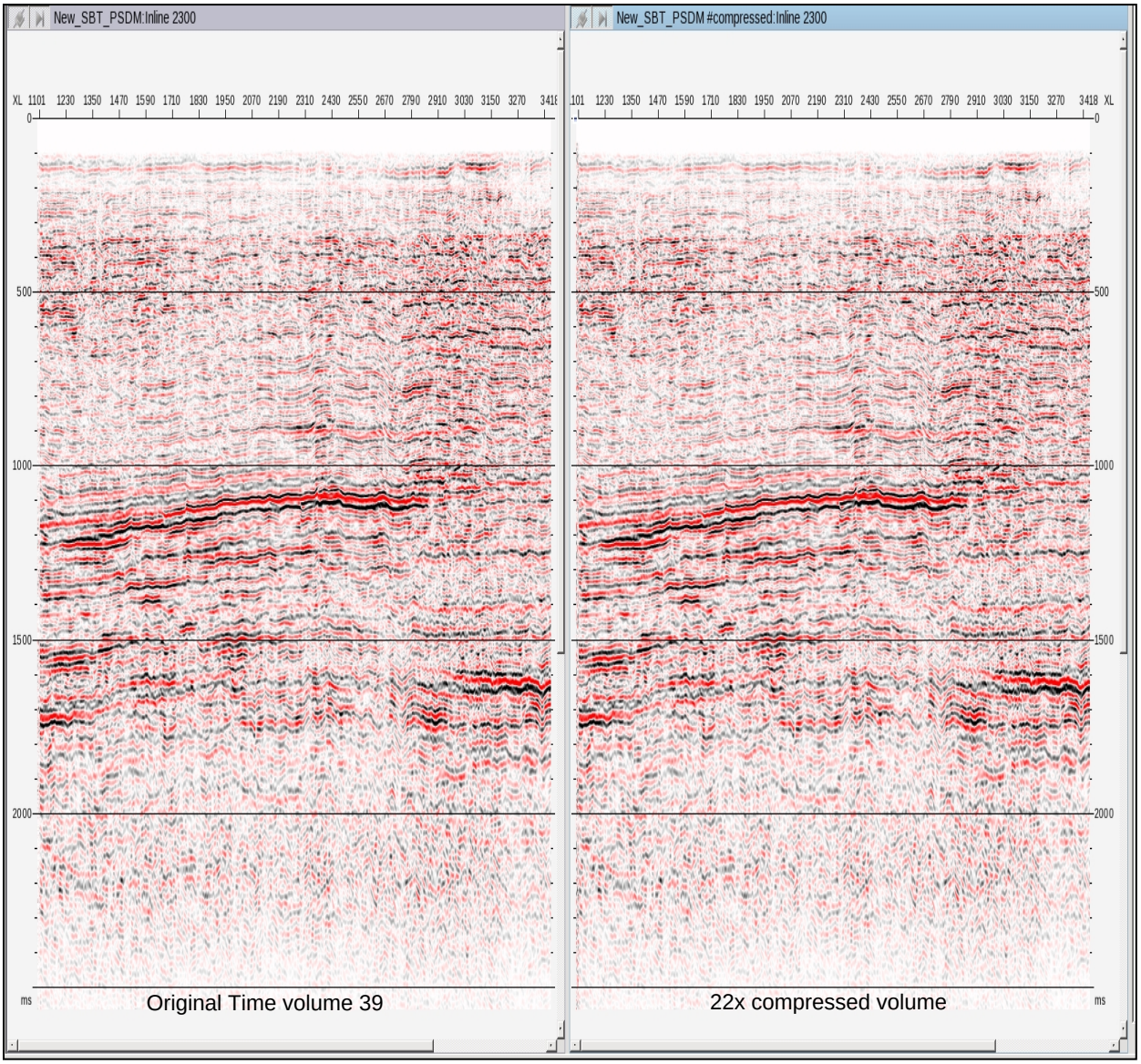


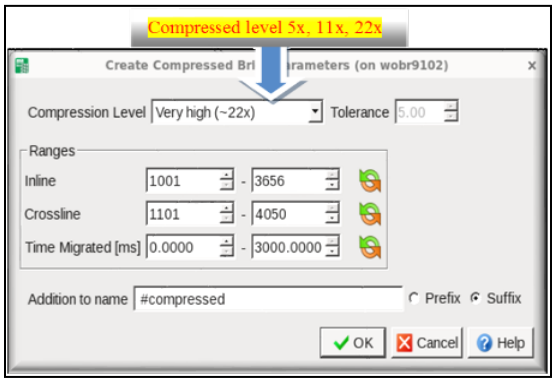
Fig 7 :Comprision between original & compressed volume (~22x Compressed 20:1)

Geoscientists take this compressed volume for quick-look and prepare quality deliverables in limited time without losing seismic information & retaining full precision of seismic dataset.

As regional prospect becoming more complex, for Geoscientists no need to look for any workaround if working on huge dataset. In that sense Geoscientist will work on existing database without wasting time in creating separate Survey and project. The option (Fig-8) of compressing to 5x, 11x, 17x and 22x levels prevail depending upon the storage requirement/adjustment.

In the process of creating the brick compression, while adding the name to the Seismic compressed file, we can add a suffix or a prefix as per the requirement of the user.

In the table below (adjutant to Fig: 8) are comparisons of various options of seismic data where the original migration data is around 39 GB. Various compression results in the compressed data sizes represented below along with the time taken depending upon the Compression ratio. A complete data storage solution and 'virtually lossless' compression of 3D poststack and prestack seismic data. A means by which to save disk space, time and seismic precision.



Original	Compress Level	Bandwidth	Time in sec	Data size
39.29 GB	5x	10 Gbps	320	12 GB
	11x		422	7.7 GB
	15x		680	2.8 GB
	22x		930	1.9 GB

Fig 8 Create compressed volume from Original Seismic Volume

In other words, the methodology provides the G & G users with the optimal return on their investment in computer hardware to the latest computing capabilities. Well known and proven technology for today's huge data set design is developed to maintain data precision and minimized disk space usage beating the challenges faced by IT and Geoscientists team. Compression algorithm is embedded directly in the Petro-technical software solution.

Conclusions

The following could be summarized and concluded based on the above study:

- Compression roaming is a complete data storage solution, save disk space. Relatively in short time, compressed volumes loaded into the memory without any compromise with original dataset and visualization.
- Definable compression levels to address G & G challenges with data storage, according to end user requirements.
- To compensate huge dataset, without changing existing infrastructure, compression roaming will not truncate seismic amplitude and retain full precision of seismic dataset.
- Convert existing data, and delete/archive originals after validation which means Reduce size, Increase transfer speed and data access.
- Excellent quality at much higher levels of compression.

Acknowledgment

The authors sincerely acknowledge gratitude and thanks the Permission given by Oil and Natural Gas Corporation Ltd. (ONGC) for publishing this case study.

References

Hue Streams Seismic Compression Technology, hue.no/Hue-streams-compression.

Compression Roaming as a Way to Address Big Data Challenge in Exploration by Robert Bond, Paradigm – EAGE 2016 Vienna

Efficiencies in Large Volume Interpretation: Compression Roaming and Dependencies by Published on Jul 19, 2016 By Bruno de Ribet.

Low bit-rate efficient compression for seismic data. By Averbuch AZ1, Meyer R, Stromberg JO, Coifman R, Vassiliou A. Sch. of Computer. Sci., Tel Aviv Univ.,Israel.