Addressing the Big Data Challenge through Compression

The oil and gas industry is one of the largest producers and consumers of digital data in the world. In the past few years, Petabytes of 3D seismic data have been acquired and processed, and importantly, stored in huge disk arrays, for the most part paid for and managed by oil and gas companies themselves. In many cases, multiple copies of the same seismic data are also stored, often on local workstation disks, to accommodate limitations in some software vendors' data management capabilities.

The industry's answer to the big data 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 adds 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 old 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 industry-wide rethink of seismic data management and storage practices.

Paradigm at the cutting edge of computing capabilities
In order to enable our customers to load, display, process and analyze huge amounts of data, Paradigm continually adapts its solutions to the latest computing capabilities. Whether through the use of multiple, background threads which can parallelize I/O of data across multiple CPUs and clusters

of processing 'blades', or by utilizing the parallelization capabilities offered by modern graphics cards with their hundreds of onboard cores, Paradigm developers work hard to provide customers with the optimal return on their investment in computer hardware. We squeeze as much productivity out of computer systems as we can.

Paradigm addresses additional challenges associated with large data storage through its rule-based, centralized data storage infrastructure, Epos®. This provides users with a 'load your data once' environment in which only one copy of wells, interpretations or seismic data need be stored on the system, reducing the data management load on an oil company's processes and organization. Other techniques, such as automated selection of decimated volumes when roaming from disk, and on-the-fly seismic attribute calculation, help improve data loading performance and reduce the amount of storage space needed in a workstation or in the database.

From Paradigm 15.5 onwards, Paradigm will supply yet another method for managing the large data conundrum: It will enable customers to compress seismic data into much smaller volumes on disk. Efficient reading of seismic data into our 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.

This table compares the value of the different approaches:

Method	Significantly reduces required disk space	Maintains data fidelity and precision	Significantly reduces data access times	Improves manageability of seismic data
Use of different dynamic depths – e.g. 8, 16 bit	Yes	No	Yes	N/A
Use of centralized data storage	Yes	N/A	No	Yes
Use of 'decimated' seismic volumes	No	N/A	Yes	No
Use of compressed seismic data in a centralized storage environment	Yes	Yes	Yes	Yes

Paradigm®

Enabling seismic compression in Paradigm Epos will help customers reduce their annual outlay on disk storage and networking, and improve access speed to large 3D seismic datasets. Typical compression ratios achieved through the use of compression utilities range from 5x (a 'virtually lossless' compression) to 20x or more.

Naturally, customers may be concerned about a loss of data fidelity due to compression, but the statistics and value provided by this approach speak for themselves:

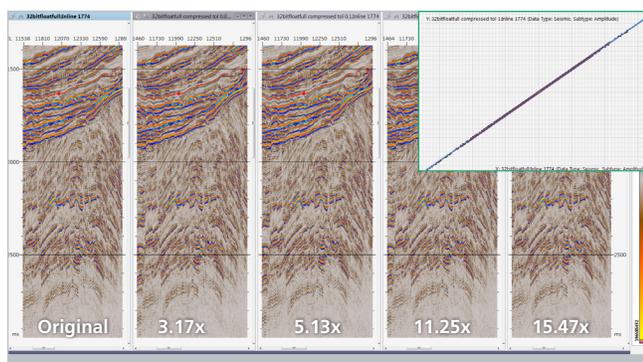*From an industry relevant 24-gigabyte North Sea Quad dataset:*

Resulting file size	1.69 GB
Compression ratio	14.2X
Average difference between compressed/ uncompressed value	0.0245%
Maximum difference	0.211%
PSNR	70.1 dB

Noise-resilient

A very important consideration for any compression technology, in particular for seismic data in which noise is inherent, is how much the compression impacts or is impacted by noise. Paradigm compression is resilient to different levels of noise in the data; the amount of real ambient noise in the dataset does not significantly affect the precision of the decompressed data. Tests indicate that the technology provides safe compression for any quality data set when used at a given compression level.

Characteristics of our approach
- Adjustable, high-speed compression and decompression
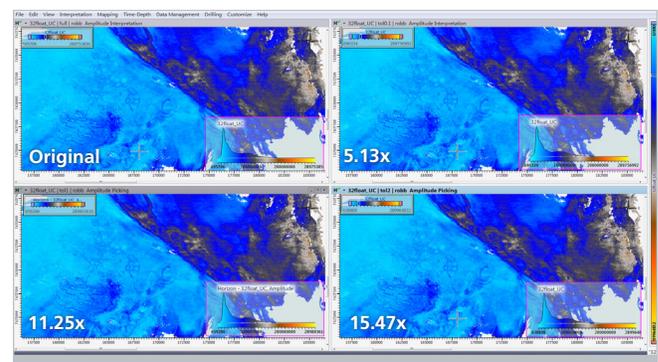- Multi CPU/GPU parallelized wavelet compression/ decompression

- No clipping or truncating, maintains values outside of the value range
- Quality based compression level (gives same quality of compression throughout, insensitive to amplitude variations)

A further benefit of using this compression technology is in the bit-depth of data used by consumers. No longer will data loaders and administrators have to decide in advance on the typical expected use of a seismic volume through its lifecycle, and determine whether to save it as 8, 16 or 32 bit, or in more than one format. Full precision seismic data can now be stored in less space than an 8 bit volume, with minimal loss of fidelity. Attribute extractions and calculations can be performed with the highest precision possible.

Paradigm compression roaming will be made available in the Paradigm 15.5 release, through a licensable plugin to our Epos infrastructure and database. This release will enable customers to compress 3D poststack seismic data and read the compressed seismic in all Paradigm applications that access Epos seismic directly. This includes our flagship interpretation and modeling applications, SeisEarth® and SKUA-GOCAD™, and our Processing & Imaging application suite, GeoDepth®.

In future releases, we plan to extend our capabilities in seismic compression into the prestack domain, facilitating the easy use of gathers in standard interpretation workflows, and of course, improving the throughput of data into processing and imaging workflows.

So how does Paradigm compression roaming respond to the big data challenge? By compressing seismic data into much smaller, more manageable volumes, while saving our customers time, effort and money in the process.

▲ Comparison of sections following compression, with a crossplot of amplitudes before and after compression

▲ Comparison of the same horizon-bound RMS amplitude extraction from 4 volumes with different levels of compression