Data management challenges in the pre-stack era

Philip Neri* argues that the emerging importance of pre-stack seismic data in interpretation and characterization workflows has created a major requirement for innovative data management solutions.

Just when things were settling down, as energy companies, storage vendors, and software companies had come to grasps with the hundreds of terabytes of post-stack migrated seismic and associated data, our industry is moving to the next big thing: access and interactive usage of pre-stack data in routine seismic interpretation and characterization workflows. The storage requirements jump a hundredfold and with that the stress on just about every component of the data chain: loading, storing, referencing, and most importantly feeding to the computation /visualization node of vast, randomly searched data samples in untold amounts. Innovation is the key to a solution that can create explicit pre-stack data objects and organize them in specific formats adapted to random access and streaming.

On the frontier

The seismic industry has a long history of living on the frontier of computer technologies. The abundance of digital data produced most notably by seismic recording, processing, and interpretation makes oil and gas one of the largest consumers of storage, computing power, and graphical display capabilities. Up until the mid-80s, the digital data was confined to the computer rooms, and subsequent work was performed on paper plots, disconnecting the interpretation and all subsequent steps in the workflow from the vast arrays of computational and storage devices used for data processing. 3D seismic made it necessary to go beyond the paper plot, and this spurred the development of interactive, computerized interpretation and modelling technologies. This was the first game-change. It started a separate, but nonetheless aggressive growth in demand for computer and disk resources for the interpretation, characterization, and modelling experts. This impacted storage volume (the 3D surveys were growing in size and density of sampling), bandwidth (to facilitate random access to ever larger datasets), and compute power (to generate attributes, perform analysis and build models). This has defined the past 25 years, with essentially separate domains for data processing and imaging using supercomputers, later replaced with clusters, and the asset teams working on integrated geosciences and engineering application platforms based more recently on networked storage, client/server data access architectures, and very high-spec workstations. Software providers developed optimized file formats for seismic and grid data to facilitate random access and rationalize computer memory usage, keeping pace with demand for multi-survey, multi-resolution, multi-version datasets that more and more companies require to stay abreast of their regional and global activity. While seismic has always been the main driver for improved horsepower, one has to mention well data and the requirement to have on hand increasingly large numbers of wells, each one accompanied by many logs and other borehole data.

As we go into the second decade of the new millennium, it has become possible to work on tens of large 3D seismic datasets, with thousands of regional 2D profiles, and tens of thousands of wells, and all the associated data for such large projects. But just as these benchmarks were reached, and companies were looking at the gradual commoditization of such capabilities through further improvements in hardware and infrastructure, the needs of modern exploration reset the goal posts to a whole new place, namely pre-stack.

The pre-stack era

Pre-stack has always belonged to certain specialized workflows, allowing geoscientists to gain insight into the original information and making use of some of its attributes, such as amplitude versus offset. For such occasional uses, it was acceptable to reach into the data processing domain and retrieve trace gathers as needed. Performance and speed were not a pre-requisite for such tasks.

But interest in pre-stack has grown dramatically both to get a firm grasp on complex imaging situations such as sub-salt, and to extract critical information from azimuth-rich, anisotropy-aware imaging processes that cannot convey all their value in the form of a single stacked trace. In these situations, pre-stack data must be accessed and displayed at speeds equivalent to stack data, and with similar capabilities to look up data in any direction, as a whole set or a sub-set, while positioning the trace data in viewers where correlations can be made to well and other data.

Pre-stack in the interpretation realm

These requirements equate to making pre-stack data not just a static and occasional data object, but one on par with the highly optimized items such as 3D post-stack volumes, 2D seismic sections, large grids, etc.. The logical extension of this

* *Paradigm. E-mail: Philip.neri@pdgm.com*

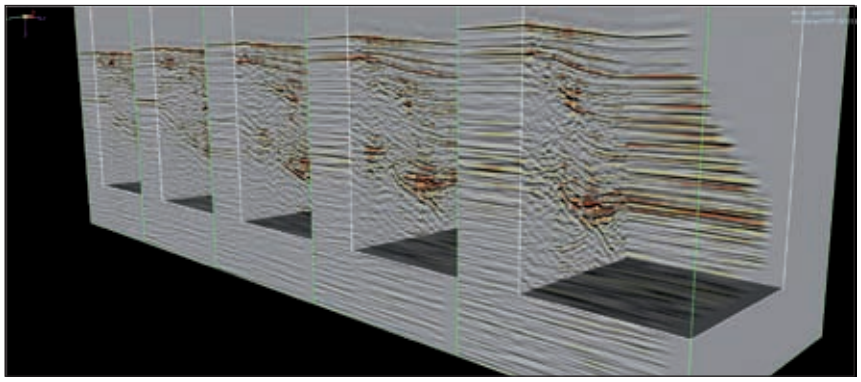

Figure 1 *The interpretation 3D viewer has been modified to show a succession of profiles each with a full set of pre-stack offset traces. As for normal interpretation data (post-stack), the data volume can be cut in any direction and the user can control the movement of the cut-in panels in any direction for data scanning. The pre-stack data has to be brought to the display in real time, sourcing the content from data objects tens if not hundreds of gigabytes in size.*

evolution is to format the pre-stack data in fit-for-purpose file systems that allow for random access at equivalent transfer rates, as well as supporting sub-setting to read, transfer, and store in memory only the portions of data that are actually required. The so-called brick format, implemented in many current interpretation systems, is the basic design component for such file systems.

In addition to addressing the needs of rapid and random access, the pre-stack data, in order to be useable when aligned to the post-stack volumes derived from pre-stack, must be linked to the original velocity model used for processing and imaging. This model must in turn be made accessible in the same project data structures as the other data types, to be used on the fly to correctly position trace gathers.

Adding azimuth

Azimuth has become a very popular attribute for the traces contained in a seismic gather. Land 3D data is very often already azimuth-rich by the very process of modern, large swath acquisition logistics. Offshore data is now being re-acquired in certain areas to gain a full azimuth illumination of complex targets, as well as to assess anisotropy that could point to stress fields and fracture zones. Azimuth and illumination properties can only be accessed if the acquisition pattern is stored and made accessible to the relevant software tools and data categorization tasks. These techniques are not so much an issue of data volumes (although azimuth-rich data tends to have higher trace densities for a given coverage area) as one of managing pointers and indexes to relate the bulk seismic data files to the acquisition geometry.

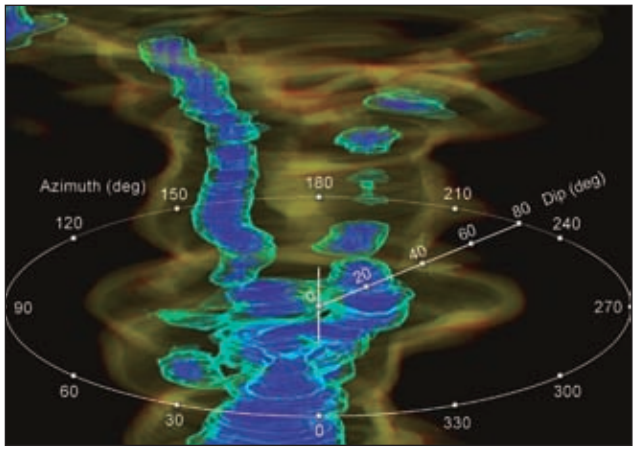

Figure 2 *New visualization tools show reflection strength in a 360^0 azimuth reference system, with distance from the centre representing the local angle of incidence.*

Strategies for storage

As we have just discussed, the data volumes are going to increase dramatically. While storage costs are going down at a steady pace in terms of dollars per gigabyte, the demand of a full pre-stack environment will continue to escalate at a faster pace. This makes it interesting in the coming years to explore a number of strategies to mitigate the impact and the costs. First, compression can be of some help. With the inexorable increase in the compute power of processors, and the ability to harness other compute resources, pursuing compression can yield an order of magnitude or more in terms of lesser storage requirements. The elimination of redundant or oversampled data is another avenue of investigation. In the case of local domain gathers, one often observes that the acquisition pattern is offering appropriate detail in complex areas but excessive numbers of nearly similar traces in areas of lesser intricacy. Analytical tools can be applied to identify such areas and perform intelligent data decimation to reduce data volumes without significant loss of information.

Technology changes

We are already seeing the rapid growth of solid-state storage devices in portable computing devices. While costs are still high, this technology has a good potential to improve storage density and lower its costs, as it is not associated with mechanical technologies, motors, servos, and the other often delicate and difficult to miniaturize components of disk-based technology. It is hard to say whether petabyte-scale data centres will one day transition to solid-state storage, but this would open new vistas in terms of rapidity of access, improved random-access capabilities, and better reliability. At the other end of the spectrum, it is increasingly questionable to use magnetic

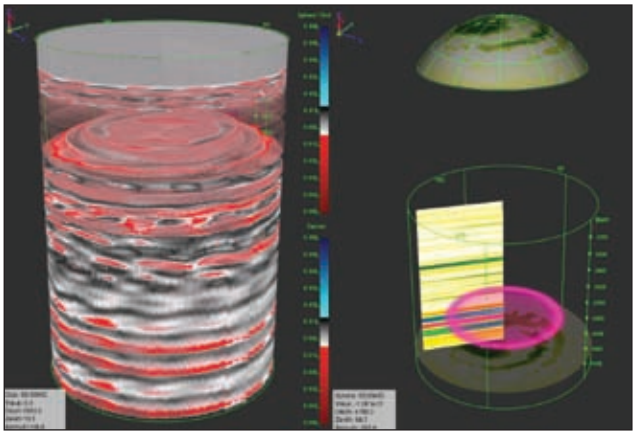

Figure 3 *Where a display of a simple pre-stack gather used to involve only a panel of traces, new methodologies make it necessary to display three-dimensional views, as well as spherical objects that materialize angles of incidence on designated targets. It is desirable to see such views updating in real time as the user navigates the data volume. (Original data courtesy of Devon Energy Corp.)*

tape for archiving purposes. While the tapes are very cheap, the cost of purchase and cost of ownership of tape drives, and the onerous processes needed to manage data that is offline, and therefore not searchable without a well-designed index system, make tape a less desirable solution. Low-cost disk farms, that can be mirrored to remote locations for disaster recovery purposes, can offer similar if not better economics, while making it much easier to access and catalogue the data as an on-going process.

Location

One implication of the broader usage of pre-stack data for companies that perform some in-house imaging or processing of seismic data as well as the interpretation, modelling and characterization, is to seek a common format and localization of this information.

This leads to putting all the large computing resources and all the data in a single centralized environment, including the resources executing the interactive interpretation and modelling applications. In such a context, seismic processing would be reading and writing the same files as the interpretation and modelling applications would later access, thereby avoiding duplication. While many implementations of remote desktop application services are specific to certain vendors or operating systems, the benefits are already of high interest. The main advantage is the proximity of the application execution and graphical rendering to the actual physical storage of the data, allowing for fast and predictable data transfer rates within a controlled environment. Additional benefits relate to the allocation of resources, the system's redundancy, and the ability to perform upgrade and maintenance tasks in a single location. The additional security of a dedicated facility, removing critical data handling from the office environment, is also a factor to consider. The alternative, pushing data out to individual workstations in offices and team rooms, will have a higher dependency on bandwidth along the whole path from storage server to the desk-side workstation, with predictable issues related to bottlenecks and a difficult oversight of the whole network and its performance. Some of the issues of remote access can be improved on, notably by using protocols that are best suited for large volumes of data transfer with minimal overheads.

Global data management

Current assumptions about data localization are predicated on the limitations in bandwidth that can exist in a building, within a city, a country, or across borders and continents. In the technological world, assumptions can often become obsolete if quantum improvements are made possible through new technologies or radical improvements in infrastructure. Data management systems already make it possible to roam, catalogue, and access data around the globe. Current bandwidth limitations do not make it practical to access or synchronize very large data volumes such as pre-stack data across long distances, let alone in some cases over relatively small distances of some tens or hundreds of kilometers. But in the event of a breakthrough in bandwidth capacity, the software and data topology solutions already exist to leverage remote access over very large distances, making it possible for a remote user to access a central data server , thus avoiding data duplication.

Metadata

We have seen that the pre-stack data object, far from being a simple collection of samples, requires a significant amount of metadata to be fully utilized. Map-based search facilities, and geographical indexing of all data, should make it relatively easy to locate and catalogue data so that it can

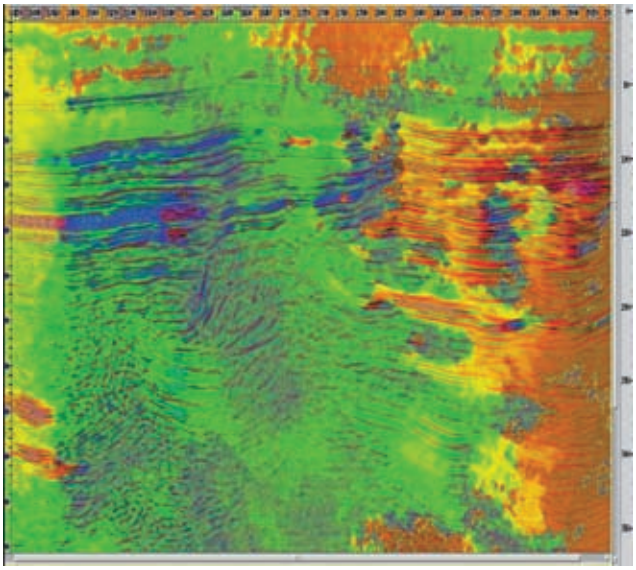

Figure 4 *Post-stack migrated seismic is shown with an color-coded overlay representing the local azimuth angle value derived from the pre-stack data.*

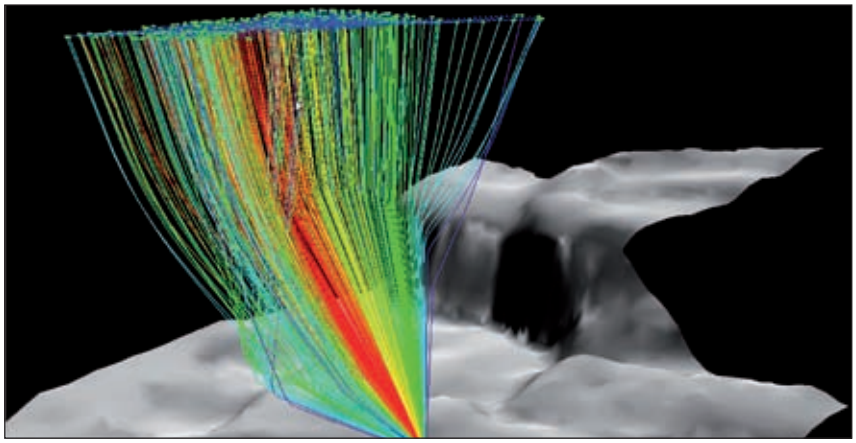

Figure 5 *The real time display of the ray-paths from a subsurface target back into the acquisition pattern is performed with the combined usage of the velocity model, the structural model, and the shot/record geometry.*

be readily included in projects. History information tracing all the transformation steps applied from field data to final product, and localizing where appropriate any other data associated with these steps (e.g., velocity models) will help master what has the potential to be some of the most voluminous amounts of data stored on private computer systems anywhere in the world.

Pre-stack data as an emerging component of interpretation and modelling workflows is going to challenge many of the assumptions currently used to design and manage data systems for exploration and production activities. At stake is the ability to extract further value from seismic data, to reduce uncertainties, improve the accuracy of prospect and reservoir predictions, and mitigate risk associated with drilling and production activities. Rethinking the architecture of data storage and data management will yield great benefits to the user community, and ensure the best possible leverage of new technologies.