SPE-14MMFD-61-MS

Cloud-based Remote Visualization of Big Data to Subsurface Exploration

Alécio Pedro Delazari Binotto, IBM Research
Nicole Sultanum, IBM Research
Renato Cerqueira, IBM Research

This paper was prepared for presentation at the SPE Large Scale Computing and Big Data Challenges in Reservoir Simulation Conference & Exhibition held in Istanbul, Turkey, 15-17 September 2014.

Abstract

Since the first visualization solutions were explored for O&G, major technical improvements enabled gigabyte-sized models to be rendered and manipulated at interactive speeds. Yet, other fundamental aspects such as data access and distribution are often overlooked in the process, even to date. Data movement may often be prohibitive, either due to legal constraints (data is restricted from departing the country) or practical considerations (data is too large to be moved, is a checkpoint tightly connect to imaging processes, and requires costly resources to be manipulated). Collaborative visualization tends to be performed co-locally, following an explicit, manually conducted, data transfer to a dedicated visualization machine.

We propose an alternative based upon **data-centric computing**. The model offers **visualization as-a-service over a multi-tenant cloud-based environment**. Remote visualization enables lightweight access and interaction with generated data readily after the processing, dismissing the need to transfer the whole dataset into analyst's machine. It offloads heavy graphics processing to a cloud server featuring the necessary infrastructure to handle such voluminous data, like GPUs, GPFS, and ultimately sends only a reduced output to lightweight clients (rendered images/geometry). Visualization resources can also be shared among concurrent users in a web-based interface and combined with other data sources, like correspondent well information or velocity models, facilitating effective remote collaboration towards knowledge discovery in subsurface exploration. Ubiquitous on-the-go data access (e.g., in the exploration field itself) is thereby made possible through mobile interfaces.

Concurrently, several challenges emerge with the aforementioned visualization model. The effective resource distribution of different data sources among several clients needs to benefit from the cloud execution platform. The OpenPower Foundation is an example of the future HPC platform that can be customized to O&G characteristics and be offered as a cloud model.

Introduction

With the emergence of the first 3D visualization technologies for oil & gas in the early 90's, like in the work of [1], proponents were striving to provide fundamental understanding on reservoirs, powered by dedicated SGI workstations. Later, PCs had grown powerful enough to enable local graphics processing, thus contributing to a widespread usage of visualization in nearly every discipline in E&P. However, with the exponential growth in computational capacity, there was also an exponential growth in the size, resolution and complexity of the data to be handled. Simulation-driven disciplines such as geophysics and reservoir engineering are now operating with extremely high-resolution models (a scale of billions of cells) which require distributed computing infrastructures and highly optimized computational algorithms to enable interactive exploration, as in [2].

For example, Full Waveform Inversion (FWI) is a numerically challenging technique in exploration based on full wavefield modeling of a geological domain that extracts relevant quantitative information from seismograms. When dealing with realistic physics-based 3D elastic PDE formulation and accurate discretization techniques, such as high order Spectral Element Method, forward modeling becomes particularly challenging from the computational and visualization point of view. This type of modeling often requires grid representation of the order of billions of elements with interpolation polynomials of at least 4th degree in each spatial dimension, which entails hundreds of billions of variables computed for forward solve. Currently, it is not rare to model a survey for depth imaging with more than 2000x2000x1000 cells and correspondent volumetric output for each timestep, representing petabytes of storage that can easily reach the exascale.

Greater insight and understanding is often promoted through the integration of various datasets for improved cross-discipline communication [3], which further increases the burden on computational systems. Just like in the 90's, the sheer size of data still requires the existence of a specialized setup to handle such scale – HPC clusters. While specialized hardware is often unique and bound to limited physical locations, there is a growing need for distributing its benefits enterprise-wide. Particularly, data movement quickly becomes an issue when dealing with such large data (as the FWI example); legal restrictions might also apply, such as when regional laws prevent local reservoir data to leave the country; and smarter visualization techniques that can support a more precise and faster interpretation by the geophysicist is desired. Therefore, personnel relocation is often needed in such cases, which represents additional costs to operation and an inconvenience to workers alike. Another issue impacting relocation is remote collaboration, often required in multinational companies and which can be greatly benefited from visualization (as is serves an effective convergence point for cross-discipline communication [3, 4]).

We propose remote visualization based on a private cloud infrastructure as a potential solution to the aforementioned issues. It can help reduce the impact of data movement as well as personnel relocation, while leveraging dedicated, specialized infrastructure. Having seismic application and visualization together offered as a service in the cloud can bring several benefits to the users. Some benefits focusing remote visualization to Oil & Gas E&P are depicted as follows:

- Allocation of an **exploration cloud infrastructure resources on demand** to empower organizations in meeting their interpretation deadlines;
- Handling of **massive visualization datasets**;
- **Integration of large amounts of data from different sources** (correlation) – visualization as a cross-discipline communication tool;
- Reduce **data movement**;
- Enable **remote collaboration**.

In the following sections, we review existing strategies for large-scale remote visualization and HPC, and follow up with our proposed approach to remote visualization. We conclude with a discussion of emerging issues related to the topic, as well as glimpses into the future challenges to arise and fruitful directions to take.

Remote Visualization of Big Data

Visualization tools are fundamental components to navigate practitioners to insight through the sheer volume of data collected and processed, and help navigate through data among the “boredom” (data with little information), “noise” (interferences in data) and “rubbish” (wrong or missing data) [5]. Big data in the oil & gas industry has volume, velocity, and variety [6], and while it requires specialized infrastructure be handled, remote visualization enables the consuming information to be accessed anywhere – from unsophisticated workstations or mobile devices – and simultaneously analyzed by a variety of collaborating professionals.

The existing remote visualization strategies are variations of a basic client-server architecture, in which the client is sent a rendered visualization output from the server. Client-side applications will typically display the GUI widgets locally, and integrate them with the server-provided data view.

The distinctions rely on the rendering workload distribution between client-server:

- The first option is to have rendering and interaction events fully processed on the server, while image-streaming the rendered output to the client;
- The second option allows for part of the rendering to be performed by the client -- through reduced geometry, or intermediate rendering outputs.

The choice between strategies is heavily context-dependent. The first option relies on the existence of a solid network connection between the endpoints, as the perception of interactivity is heavily dependent on how fast the rendered images are sent across: every change in the view parameters requires new images to be rendered and streamed. To enable interactive rates, images are compressed between client and server, which however affect the final visualized output.

One acceptable tradeoff is to stream compressed renderings during manipulation (e.g. mouse interaction), and send a non-compressed high-resolution image when visualization is static. The server must also be powerful enough to handle all stages of visualization, but can usually do so faster as the interdependent resources (data-GPU interchange) are more easily accessed within the server space. This strategy also enables the usage of very thin clients (a web browser for instance), with minimal local resources to interactively manipulate increasingly complex visualization. Al-Harbi and colleagues discuss an interesting remote visualization infrastructure employed to handle massive billion-cell models generated within their giga-cell simulator [2].

A hybrid client-server rendering, on the other hand, can favor interactivity in environments with less reliable network connection. Given that a pre-processed version of the data is available locally, some manipulations are still possible even if communication with the server is temporarily cut. For instance, in Semotus Visum [7] a simplified version of the geometry can be sent to the client together with a rendered texture that features most of the visual detail; the client side can thus be independent form the server for viewpoint manipulation. This strategy usually requires the maintainance of redundant data representations with different levels of detail (LOD) on the server side. However, this approach is not adequate when client resources are limited, or when reducing data complexity is not possible. Also, poor workload balancing can result in undesirable system overload, with too much data being transferred over and/or too much data being locally handled.

Given that the suitability of each scenario is highly dependent on dynamic requirements (e.g. network bandwidth), both strategies should be dynamically interchangeable depending on local data/network configurations, and remote visualization APIs and frameworks should be able to provide this flexibility. For instance, Paraview[1] – a general-purpose parallel visualization framework – allows for the workload to be automatically balanced between client and server by defining a maximum threshold of data to be stored locally. Semotus Visum [7] is also a flexible framework, and offers three levels of workload distribution between full image streaming and full geometry rendering.

Cloud-based Remote Visualization Model

In O&G exploration, a number of organizations/software packages are involved – each attending specific parts of the imaging workflow (e.g., field acquisition, noise attenuation, velocity model, imaging synthesis, visualization, interpretation). These organizations need to exchange and transfer large amounts of data that in the near future have the potential to reach the magnitude of exabytes. In this scenario, data movement becomes a bottleneck before visualization, thus affecting time-to-production. To overcome this challenge, we rely on a private cloud, the SoftLayer Cloud, in which different users and even organizations can work collaboratively on the same data. Through a user's portal, a secured cloud environment can be provisioned on-demand for the subsurface workflow. Image synthesization and visualization are performed on the same execution environment. This platform is flexible and elastic enough to be provisioned as needed – the HPC cloud environment is flexible in the way one can configure the machine tuned to the problem; and elastic in a way that the number of nodes can adapt to the problem size and deadline of the project. Figure 1 shows the overview of our proposed cloud

[1] http://paraview.org/

infrastructure emphasizing the remote visualization module. The proposed cloud (in the middle of the figure) serves the imaging workflow services (left) and also the distributed visualization processing of final results (right).

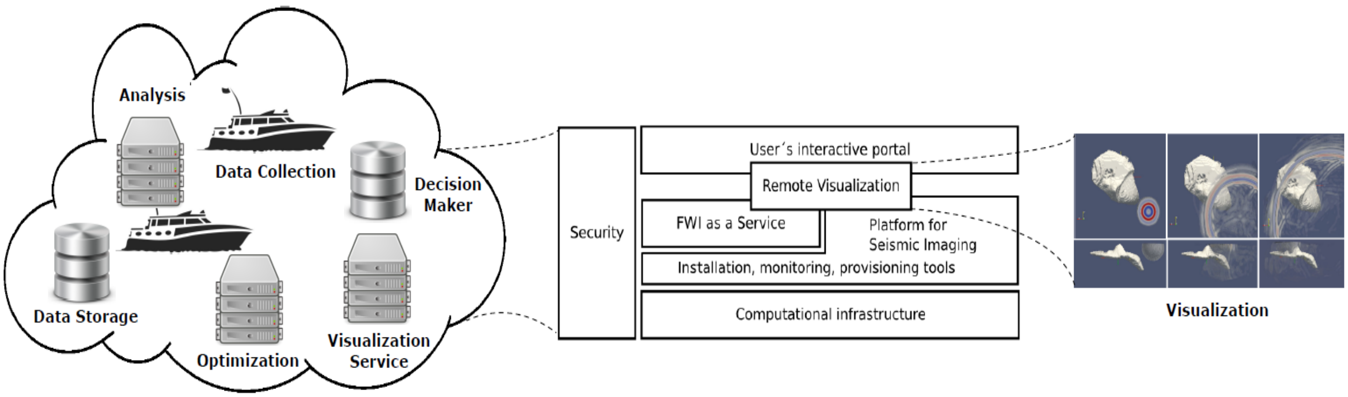

Figure 1 – Cloud Architecture for Subsurface Imaging.

Targeting Remote Visualization, this module is designed to enable lightweight access and interaction with the generated data, without the need to transfer the whole dataset into the analyst's machine. Another advantage is offloading the heavy graphics processing from the client machines to the cloud server, as mentioned, featuring the necessary infrastructure to handle such data. Besides, it sends only a reduced output to a client (e.g., the rendered images). Our initial explorations in remote visualization have employed Paraview[2], an extensible open-source package for parallel and distributed data visualization, to prove the basic concepts. Paraview deals with remote visualization through a hybrid approach, which automatically balances the sending of rendered images or the geometry – when the latter is sufficiently light and simple so it can be transferred over for local processing. Compression techniques, GPU acceleration, and Level of Detail are also taken into account to ensure interactive rates. The infrastructure is compatible with several different user interface platforms, namely Paraview's own graphical interface, customized desktop viewers, and most importantly, web browser applications (built on top of WebGL) and on-the-go apps that can be used over mobile devices. Figure 2 ilustrates the remote visualization of an FWI forward model (using volume rendering and different point-of-view and timesteps). The FWI synthesization and visualization can be performed in the same cloud infrastructure.

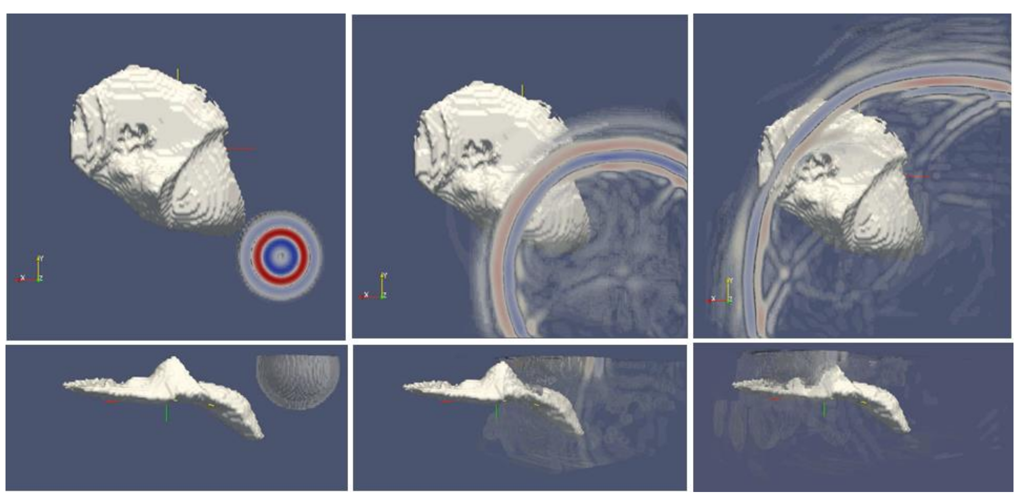

Figure 2 – Remote 4D visualization of a FWI subsurface processing in different timesteps (forward component).

Additionally, remote visualization can also support to deal with a legal issue: O&G companies usually must attend regulatory requirements to handle sensitive data concerning information collected from basins, oil production forecasts, and other

[2] http://paraview.org/

strategic information. To accomplish with this regulation, the private cloud should however be physically installed in the country where the data were acquired, but visualization, when granted, can be manipulated remotely.

In summary, the main advantages of the described architecture model can highlighted:

- **Cloud dynamic allocation on-demand**: given a time deadline for a specific subsurface project and a benchmark for the problem, the proposed model supports the user to project/provision a cloud infrastructure compatible to execute the project in the right timeframe. Eventually, the user can project a hybrid infrastructure to use its own HPC platform and complement with cloud nodes (bursting) targeting to accomplish the deadline;
- **Remote Visualization in the cloud to minimize massive data movement**: integrating the visualization model with the imaging process in the cloud may minimize data transfer. Remote visualization is used to process the image rendering - in parallel and in the cloud – and just the final image/pixels is sent to the analyst machine for interpretation;
- **Enable remote collaboration**: remote visualization also allows for multi-user collaboration. A number of users can interact with the same model collaboratively using distinct devices, like desktops, interpretation machines, mobile devices, or even surface tables;
- **Integration of data from different sources for analytics**: visualization is additionally used as a basis for a cross-discipline communication tool. It can integrate data from seismic processing, optimization, reservoir simulation, among others to better plan the reserve exploration, cost of production, and return of investment.

Potential Directions

This paper described a new architecture for subsurface imaging, targeting remote visualization capabilities, based on a cloud infrastructure. The proposed solution integrates simulation and visualization in the same HPC platform delivered on-demand as cloud, minimizing data movement. Our private cloud architecture provides an environment that enables a set of users and also organizations to cooperate and achieve their goals on the time planned. The allocation of cloud infrastructure resources on demand and the aforementioned solution for the data movement empower organizations in meeting their deadlines. Furthermore, setting up the environment is simplified as the solution provides a platform for seismic imaging, that is, a set of libraries and infrastructure that support seismic applications, like, for example, FWI. Thus, the proposed cloud architecture allows for a collaborative environment with shared resources among a set of users and organizations, which may maximize interactions between stakeholders and increase efficiency.

The aforementioned cloud infrastructure can offer a heterogeneous HPC Cloud environment that can be provisioned with CPUs x86, POWER, accelerators (such as GPUs), Infiniband, and GPFS. It is planned that the OpenPower Foundation[3] will allow a unique infrastructure tuned to subsurface applications and on demand. Here, the POWER processor together with GPU accelerators can be adapted to seismic applications that are reaching the exaslace.

Finally, as stated in [5], big data is not an end to itself and by itself; they should instead complement practitioners' work and assist them in achieving actionable insight to solve real business problems. Current discussions on cognitive computing also emphasize the principle of "assisted insight" while driving the exploration of knowledge-based solutions applied to domain-specific problems. This will continue to push the boundaries of large-scale hardware and software infrastructures [8], and turn visualization into no longer a one-way bridge to domain knowledge, but a powerful mediator between humans and computers.

References

[1] Van, Bul-Tran et al. "3D Reservoir Visualization." *PT Journal of petroleum technology* 43.11 (1991): 1310-1314.

[2] Al-Harbi, Badr, et al. "Advanced Visualization for Reservoir Simulation." SPE Saudi Arabia Section Technical Symposium and Exhibition. Society of Petroleum Engineers, 2013.

[3] Telford, C., M. Burns, and N. Whiteley. "3D visualisation: a common language for the drilling and subsurface communities." SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers, 2004.

[3] http://openpowerfoundation.org/

[4] Love, Frank, and Nicholas Purday. "3D Petroleum Visualization: Workflows and Collaborative Solutions to Identify New Exploration Plays Enhanced Reservoir Characterization and Improved Drilling and Wellbore Placement." Offshore Technology Conference. Offshore Technology Conference, 2009.

[5] Cameron, David B. "Big Data in Exploration and Production: Silicon Snake-Oil Magic Bullet or Useful Tool?." SPE Intelligent Energy Conference & Exhibition. Society of Petroleum Engineers, 2014.

[6] Feblowitz, Jill. "Analytics in Oil and Gas: The Big Deal About Big Data." SPE Digital Energy Conference. Society of Petroleum Engineers, 2013.

[7] Luke, Eric J., and Charles D. Hansen. "Semotus visum: a flexible remote visualization framework." Proceedings of the conference on Visualization'02. IEEE Computer Society, 2002.

[8] Ferrucci, David. "Build Watson: an overview of DeepQA for the Jeopardy! Challenge." *Proceedings of the 19th international conference on Parallel architectures and compilation techniques*. ACM, 2010.