The Big Deal About Big Data in Upstream Oil and Gas

WHITE PAPER
Sponsored by: Hitachi Data Systems

Jill Feblowitz
October 2012

IDC OPINION

The oil and gas industry is a data-driven business. The industry depends on information technology (IT) to increase the speed of finding oil, enhance oil production, and reduce health, safety, and environment risks that come with equipment failure or operator error. Processing large volumes of data for reservoir modeling or simulation is not new. What has changed is the availability of new technology based on commodity hardware — Big Data and analytics. This new technology can process high volumes of a variety of data types with different access patterns at relatively quicker speeds than conventional technology.

The potential for Big Data and analytics lies in accessing previously untapped data, enabling the use of data across disciplines (geology, petroleum engineering, accounting, etc.), providing personnel with access to searchable institutional knowledge, and helping personnel find what they are looking for more quickly. Key findings from IDC Energy Insights’ examination of Big Data and analytics in upstream oil and gas are as follows:

- More sophisticated technologies are producing more seismic drilling and production data than ever before.
- There are some credible potential applications of Big Data and analytics to deliver benefits in exploration, development, drilling, production operations, maintenance, and the enterprise.
- Many of the challenges of accessing data have been overcome with faster networking technologies, but there are still input/output storage challenges such as insufficient capacity, limited upload bandwidth, and multiple access patterns.
- Work on the application of Big Data and analytics in the oil and gas industry is in the experimental stage, with much of the early work focused on data-intensive computing and how I/O data loading can be managed most efficiently.
Oil and gas companies that want to take advantage of data to gain a competitive edge and lower risks should formulate a Big Data strategy that includes the evaluation of decision makers’ requirements, decision processes, existing and new technology, and the availability and quality of data.

Hitachi Data Systems (HDS), with a long history in oil and gas, is advancing two new solutions for the oil and gas industry that show promise in the handling of Big Data. One solution is aimed at performance and I/O storage challenges and is designed to deliver increased efficiency with high availability and reliability along with faster uploads. The other solution worth considering tackles reservoir modeling and interpretation and is designed to allow personnel to work a reservoir in parallel, thereby creating efficiencies.

SITUATION OVERVIEW

A Period of Rapid Change

The oil and gas business environment is in a period of rapid change. With the depletion of older reserves, exploration and production (E&P) has turned to unconventional methods, such as deepwater, tight oil, and shale gas. Extracting resources in these ways is more expensive, and the processes are more complex. Alternatively, the industry is pursuing enhanced oil recovery from existing wells. Enhanced oil recovery takes advantage of new techniques, technology, and awareness of the geology of the field and production history to achieve higher production rates and recovery efficiency. These approaches depend on access to data. In fact, in the words of one oil and gas executive, "Drilling these days depends on data. Data collected becomes the 'eyes' guiding drilling and production."

Oil prices are experiencing greater volatility. The reasons are many, but the reality is that Brent crude prices are no longer in line with WTI crude prices. For example, in less than six months in 2012, front-month prices for Brent crude have gone from over $120 per barrel to less than $90 per barrel and those for WTI have gone from $110 to less than $80. The difference in price can make or break a prospect if the cost per barrel to produce comes too close to the market price. This greater volatility has oil and gas companies maintaining a constant vigil to see whether they should divest assets or continue to develop prospects.

The industry is still feeling the impact of Deepwater Horizon and San Bruno. Considerable attention is being paid to mitigating health, safety, and environment risks that come with equipment failure or operator error. In the United States, the Safety and Environmental Management System (SEMS), based on the American Petroleum Institute's RP-75, calls for owners/operators of offshore rigs to be able to identify hazards, assess environmental impact, optimize operations,
develop safe work practices and training programs, and analyze incidents. The challenging physical environment in deepwater and the Arctic and the need to limit the number of personnel in hazardous and remote locations have led to the development of some fully automated rigs.

The aging of the oil and gas workforce is not new, but it is certainly a reality. Many seasoned petroleum engineers have already retired from the industry, taking with them valuable experience. There is fierce competition between oilfield services companies, owners, and professional service firms for skilled resources. The Society of Petroleum Engineers reports that its membership grew 12% between 2008 and 2012, but year-over-year growth from 2007 to 2008 was 11%. Thus, the SPE's rate of growth has slowed at a time when more engineers are needed. The implication for the oil and gas industry is that there is a need to bridge the talent gap in other ways.

The oil and gas industry is highly competitive. It used to be dominated by a few supermajors, but other competitors have emerged. In the 1970s, national oil companies controlled less than 10% of the world's oil and gas reserves; today, they control more than 90%. Independent exploration and production companies have made inroads into the industry with tight oil, oil sands, and shale gas. This is not to say that there is not cooperation in the industry. Most of the deepwater projects are so expensive that they require joint development. Much of the joint development has also included national oil companies. Still there is a need in the industry to maintain competitive edge.

---

**More Data and New Technology**

The past decade has seen the digitization of the oilfield. Whether the term is smart oil, digital oilfield, i-fields, or smart fields, in the past 10 years, digital oilfields have helped increase oilfield production. With increased digitization has come the availability of more data. According to Jay R. Pryor, vice president, business development, Chevron Corp., in a speech to the World National Oil Companies Congress 2011, "Advanced technology is the spine of 21st century energy development. Chevron's internal information technology traffic exceeds 1.5 terabytes a day…. For example, a large seismic data processing center will gather the power of 20,000 personal computers to crunch a single seismic data set."

Processing large volumes of data is not new to the industry, however. Geologists, geophysicists, and reservoir engineers have been using massively parallel processing capabilities of high-performance computing (HPC) to perform analysis on petabytes (PB) of data to inform exploration since the late 1990s. Exploration uses sophisticated analytics such as seismic processing and reservoir modeling and simulation provided by software divisions of oilfield services companies.
companies and petrotechnical application vendors. 3D visualization is used to identify new resources. 4D visualization is also providing the ability to understand changes in a reservoir over time.

What has changed is that more sophisticated technologies are producing even more data. Starting with exploration and development, a recent innovation called wide azimuth (WAZ) is being used to increase seismic plot accuracy. WAZ involves multiple ships making multiple passes to capture a richer picture. WAZ to support exploration produces three to four times more traces per square kilometer than another acquisition technology, narrow-azimuth towed streamer (NATS). WAZ to support development has six to eight times the data volume of NATS. Onshore, Royal Dutch Shell is deploying another new technology — surface sensors developed by HP. The first test produced by the sensors created 1PB of data.

The industry has also been getting better at improving access to production data. These improvements are motivated in part by the need to more closely monitor and optimize production. Advances in telecommunications technology to access the data have also helped. Communications to remote locations have typically depended on satellite communications. However, some oil and gas companies have laid out fiber to offshore rigs — a recent investment of $80 million for fiber in the Gulf of Mexico is an example — or stood up wireless towers to support onshore wells in remote locations.

What has also changed is the availability of new technology based on commodity hardware — Big Data and analytics — pioneered by firms such as Facebook and Google. This new technology can process high volumes of a variety of data types at relatively quicker speeds than conventional technology. Much of the new technology takes advantage of open source code such as Hadoop and MapReduce to handle large volumes of data so that the data can be processed efficiently and reliably. The ability to handle a variety of access types, including massively parallel direct access to Big Data volumes, is supported by other open source code such as Lustre.

With the discovery of what companies can accomplish with analytics, the oil and gas industry is recognizing that there could be untapped value in data that has been previously unexamined or inaccessible. Then, too, the industry is starting to think about whether there is value in analyzing data across disciplines. For example, could seismic data, typically the province of exploration, be used to enhance oil production? Based on the recognition of the large volumes and value of data, one major oil and gas company is setting up a dedicated group just to focus on management of and access to technical data.
**Defining Big Data and Analytics**

Big Data and analytics represent a new generation of technologies and architectures designed to economically extract value from very large volumes of a wide variety of data (structured and unstructured) by enabling high-velocity capture, discovery, and/or analysis.

IDC expects the Big Data technology and services market to grow from $3.2 billion in 2010 to $16.9 billion in 2015. This represents a compound annual growth rate (CAGR) of 39.4% or about seven times that of the overall information and communication technology (ICT) market. Another point of contrast is the external IT spending for the oil and gas industry. According to IDC Energy Insights’ Worldwide Oil and Gas Industry IT Spending Guide, the CAGR for IT spending on hardware, software, and services for the oil and gas industry is expected to be 5.52% from 2010 to 2015.

Big Data and analytics include infrastructure, data organization and management, analytics and discovery, and decision support and automation interface. Infrastructure comprises a foundation stack that includes the use of industry-standard servers, networks, storage, and clustering software used for scale-out deployment of Big Data technology. Data organization and management refers to software that processes and prepares all types of data for analysis. This layer extracts, cleanses, normalizes, tags, and integrates data. Analytics and discovery includes software for ad hoc discovery and deep analytics and software that supports real-time analysis and automated, rules-based transactional decision making. Decision support and automation interface consists of applications with functionality required to support collaboration, scenario evaluation, risk management, and decision capture and retention.

The oil and gas industry has it all — volume, velocity, variety, and value.

- **Volume.** As discussed previously, volumes of data continue to grow. Besides WAZ, a new technology called Schlumberger’s IsoMetrix using a multisensor towed streamer delivers more sampling data than conventional methods, providing greater granularity in image quality. New technology used in drilling is also creating more data. Nuclear, electromagnetic, and acoustic measurements are transmitted from the bore hole where they are used to monitor drilling in real time. With the traditional mud-pulse technology, data is transmitted at a rate of 2–3bps. With the new technology, it is possible for the data to transmit at a rate of 2Mbps. Optical fiber with embedded sensors is being tested by several companies to monitor strain in well tubulars and for chemical sensing. Optical fibers have been used for temperature sensing and fluid flow around a well.

  - **Value:** Volumes of data can illustrate a clearer view of potential resources, help guide a fracking process, or show the way to enhancing production.
- **Velocity.** Velocity can be associated with discovery/analysis on Big Data captured from SCADA systems, drill heads, flow sensors, or other condition sensors in near real time. The velocity of data processing is stream rather than batch process. Velocity is particularly applicable to the oil and gas industry, which has a vast amount of complexity involving many engineering disciplines and a large amount of data beyond the capability of a single individual or team to process quickly. For example, if Big Data and analytics were applied to the data generated on the Deepwater Horizon, operators on the rig might have been able to understand more quickly what the data was telling them about the condition of the blowout preventer.

  ○ **Value:** Velocity is a requirement when the consequences of failure are great or when a delay in processing data may mean missing a bid for an oilfield.

- **Variety.** Data can be structured, unstructured, or semistructured. Structured data can be time series or transactional data that is conducive to a structured relationship database. Traditional sources of unstructured data at oil and gas companies include CAD drawings, specifications, seismic, well log or daily drilling reports in paper or PDF, Web traffic, and, more recently, social media. Processed data, a staple of the oil and gas industry, is a somewhat gray area that some refer to as "semistructured" data.

  ○ **Value:** Variety means being able to access data that was previously inaccessible due to multiple access patterns or the unstructured nature of the data.

---

**The Challenges of Accessing Big Data**

The industry has recognized some potential value from having access to more data; however, there are technical challenges to gaining that access. Much of the data is locked in applications on the desktop and cannot be shared efficiently. Although there has been greater attention in recent years to the development of standards and data models, such as PPDM, SEG-Y, WITSML, PRODML, and now RESML, there is legacy data that is in proprietary formats. Another challenge that is being addressed is I/O on the network. On one hand, the solution is with the steady progression of faster networking technologies for both Ethernet and InfiniBand. On the other hand, because this is essentially a scalability problem, it can be solved financially by investing in fiber and/or investing in multiple connections that can work in parallel and aggregate bandwidth. However, storage I/O requirements can still be a challenge. There may not be enough capacity or upload bandwidth. In addition, the workloads may be random or sequential or both in an unpredictable pattern. Most storage systems can perform well on either random workloads or sequential workloads, but not both.
FUTURE OUTLOOK

Oil and gas companies look for information technology investments to provide value in improving operations. For oil and gas companies, Big Data and analytics will need to demonstrate value by:

- Increasing speed to first oil
- Enhancing production
- Reducing risks, especially in the areas of health, safety, and environment
- Reducing costs, such as nonproductive time

The potential for Big Data and analytics lies in several areas. In addition to the value mentioned previously — unlocking insights from new data sources, such as microseismic and leveraging untapped unstructured data — there are other potential uses for Big Data and analytics. Big Data and analytics enable the use of data across disciplines (geology and geophysics, reservoir engineering, production engineering, etc.). The idea is to be able to manage technical data so that multiple data sources and types can be accessed, bringing together data that has not been analyzed in conjunction before that could offer new insights. Big Data and analytics, along with collaboration technologies, can provide personnel with access to searchable institutional knowledge that compensates for limited expert staffing. Further, there is a value in achieving more accuracy and helping personnel find what they are looking for more quickly. The time saved in accessing and loading data is important given the shortage of experts.

USE CASES IN OIL AND GAS

Work on the application of Big Data and analytics in the oil and gas industry is in the experimental stage. Investigation is being done by a combination of industry vendors, universities, oil and gas companies, and Big Data vendors and business analytics vendors. Much of the work centers on data-intensive computing and how I/O data loading can be managed most efficiently.

IDC Energy Insights sees the following potential uses for Big Data and analytics.

Exploration

By applying advanced analytics, such as pattern recognition, to a more comprehensive set of data collected during seismic acquisition, geologists may be able to identify potentially productive seismic trace signatures that have been overlooked in newly acquired or archived data.
Using data from other disciplines could enhance exploration efforts. For example, historical drilling and production data from a nearby well could help geologists and geophysicists verify their assumptions in their analysis of a field. This becomes especially important where environmental regulations restrict new surveys.

**Development**

Big Data and analytics could aid oil and gas companies in acreage assessment and prospect generation. Analytics applied to geospatial data, news feeds, oil and gas reports, or other syndicated feeds could provide competitive intelligence on where to submit bids for leases.

**Drilling**

Beyond monitoring and alerting based on limited data, Big Data and analytics could be applied to real-time "big" drilling data to identify anomalies based on multiple conditions or predict the likelihood of drilling success and allow the correlation of information to better predict equipment failure and help prevent operational mistakes.

**Production Operations**

Enhancing oil recovery from existing wells is a key objective for oil and gas companies. Analytics applied to a variety of Big Data at once — seismic, drilling, and production data — could help reservoir engineers map changes in the reservoir over time and provide decision support to production engineers for making changes in lifting methods. This type of approach could also be used to guide fracking in shale gas plays.

**Maintenance**

Predictive maintenance is not a new concept for the oil and gas industry, although if you ask a maintenance executive, it does not get the attention and budget it deserves. In upstream, if pressure, volume, and temperature can be collected and analyzed together and compared with the past history of equipment failure, advanced analytics can be applied to predict potential failures. Additionally, many upstream operations are in remote locations or on ships, so being able to plan maintenance on critical assets is important, especially if work requires purchase of specialized equipment.

**Enterprise**

The shortage of skilled labor continues to be a challenge for the oil and gas industry. Big Data and analytics can make knowledge more easily found and accessible. Another application of Big Data and analytics could be in performing social business scans in the service of recruitment. For example, a company could scan for image and reputation and create a strategy to enhance its reputation as a good place to work. Alternatively, scans could identify likely recruitment targets who fit a desired profile.
IDC Energy Insights predicts that Big Data and analytics will be deployed first to improve data quality and data loading to feed existing high-performance computing models. Processing power has also moved forward with the development of new and faster chips, but I/O is a problem with the growth in data.

Overview of HDS Offerings

HDS has been serving the oil and gas industry for more than 20 years. The company’s unified storage system has provided scalable storage that can handle mixed and unpredictable file storage workloads and provide high capacity with industry-standard file protocols. Storage capacity supports block-oriented applications such as databases as well as file-oriented protocols. The company is advancing two new solutions for the oil and gas industry that show promise.

Storage, Performance, and Capacity

HDS’ acquisition of BlueArc in September 2011 led to the development of a solution aimed at performance and I/O storage challenges and designed to deliver increased efficiency with high availability and reliability along with faster uploads. BlueArc has had a long history of supporting high-performance computing in oil and gas. The two companies have had previous experience working together on some oil and gas projects. This acquisition brings together the reliability and availability of HDS’ unified storage with BlueArc’s high-performance file management system under the HNAS brand. The new solution complements HNAS and takes advantage of open source Lustre to achieve high performance and massive scalability for very large analysis requirements. For example, on the throughput/sequential side, solutions such as Lustre and Hadoop utilize different architectures to take the scalability far beyond what clusters of NAS heads physically can do. The impact of this combined offering is increased efficiency with high availability and reliability along with faster uploads. The solution also has the potential to support a more efficient workflow.

Enhanced Backup and Recovery in the E&P Environment

HDS, in partnership with Schlumberger, developed a solution that tackles reservoir modeling and interpretation and is designed to allow personnel to work a reservoir in parallel, thereby creating efficiencies. The HDS solution works with Petrel, Schlumberger's application for seismic interpretation and reservoir modeling. It enables snapshots of modeling and interpretation work to be retrieved for further analysis and moved off the local workstation to be shared across disciplines and among larger workgroups. Models can be run repeatedly using multiple scenarios, thus facilitating easy-to-manage "what if" analysis and method comparisons.
**Strengths and Challenges**

HDS can address many aspects of the needs of oil and gas companies. HDS offers servers, storage, processing, and networking as well as the ability to build integrated, converged solutions. HDS is a systems company, but it is a systems company that was born in storage and high-availability environments. At the same time, HDS can work in a heterogeneous environment and has had experience virtualizing storage provided by other vendors. HDS has always been a company that has optimized storage and data access. With the acquisition of BlueArc, the company obtained a set of clients and additional expertise in oil and gas. HDS can now provide converged, supportable solutions across the globe.

HDS will face challenges. The market understands high-performance computing, but it does not yet understand Big Data and analytics. In fact, in IDC’s spring 2012 Vertical IT and Communications Survey of oil and gas companies based in the United States, 70% of the 144 respondents were not aware of the terms Big Data and analytics.

With the terms Big Data and analytics defined, oil and gas respondents see "lack of business support and/or business units do not understand the benefits of Big Data" as a barrier to the adoption of Big Data and analytics (22.5%). Other barriers to adoption include not knowing which data is relevant for analysis (15%) and not having enough or the right IT skills (11.9%). Moreover, the market is still uncertain about the costs and requirements of Big Data and analytics. There have not been enough cases for the industry to weigh the business benefits of Big Data and analytics against how much investment is required to achieve greater reliability and speed.

HDS’ success will depend on the company's ability to fully integrate BlueArc solutions and expertise. In this "show me" industry, HDS will need to develop proofs of concept that show credible results that can be used to move the industry forward in its pursuit of oil and gas resources. This may mean that the company will need to partner with analytics vendors that can work on the data served up by HDS solutions.

**ESSENTIAL GUIDANCE**

Oil and gas companies looking to improve their capabilities in Big Data and analytics should consider the following:

- Recognize the value of untapped data assets in supporting fact-based decisions by geomodelers, reservoir simulation engineers, operations reservoir engineers, production engineers, drill operators, human resources recruiters, and executives of oil and gas companies. Build use cases to address the challenges and connect those use cases to business value.
● Understand the competitive implications of operating without all the information at your disposal. While your company may not be investing in Big Data and analytics, your competitors may be able to obtain an edge as they build capabilities in this area.

● Conduct a gap analysis to determine what new technology and staff investments are required. Understand how Big Data and analytics can work in conjunction with HPC to speed analysis.

● Formulate a Big Data strategy that includes the evaluation of decision makers’ requirements, decision processes, existing and new technology, and the availability and quality of data.

● Take into account the requirements for a shared environment that supports the total workflow of a workgroup. All oil and gas workflows require multiple steps to get from raw data to visualization to a drilling decision.

● Consider a shared services model with other oil and gas companies to reduce costs. This strategy will allow you to experiment prior to investment. Companies with limited resources — pipeline, unconventional entrepreneurs, and so forth — may want to consider this option in order to keep costs low.

ABOUT IDC ENERGY INSIGHTS

IDC Energy Insights is based in Framingham, Massachusetts, and operates throughout the world utilizing IDC's network of global facilities. Our focus is to provide strategic business technology and application advice for the energy industry. IDC is the foremost global market intelligence and advisory firm with more than 1,000 analysts in 110+ countries. IDC forecasts worldwide markets and trends to deliver dependable service and client advice. IDC is a subsidiary of IDG, the world's leading IT media, research, and exposition company.

Copyright Notice

Copyright 2012 IDC Energy Insights. Reproduction without written permission is completely forbidden. External Publication of IDC Energy Insights Information and Data: Any IDC Energy Insights information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Energy Insights Vice President. A draft of the proposed document should accompany any such request. IDC Energy Insights reserves the right to deny approval of external usage for any reason.