Tapping the power of big data for the oil and gas industry

The petroleum industry is no stranger to large volumes of data. Operating in arguably the original sensor-based industry, oil and gas companies have for decades used tens of thousands of data-collecting sensors installed in subsurface wells and surface facilities to provide continuous, real-time monitoring of assets and environmental conditions. These companies closely monitor the performance of their operational assets. They also conduct advanced physics-based modeling and simulation to support operational and business analytics and optimization.

Today, organizations are capturing a greater volume and variety of data, at a faster velocity, than ever before. In addition to sensor data, this big data includes large volumes of semi-structured and unstructured data—ranging from high-frequency drilling and production measurements to daily, written operations logs—that quickly add terabytes of new data. It also contains a massive collection of business data, such as internal financial results, and news on energy and petroleum competitors bidding on leases and making major capital investments. Those organizations accumulate petabytes of such information with the goal of using it to improve performance and increase their competitive edge (see Figure 1).
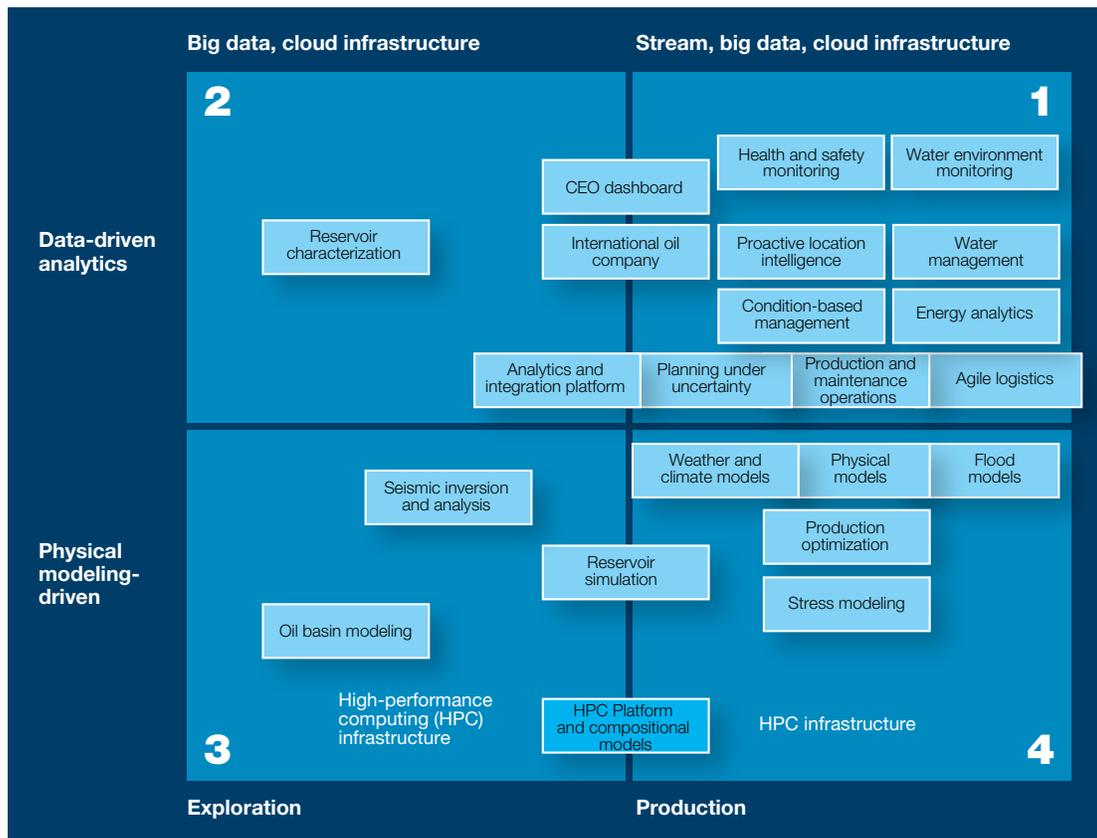

Figure 1: Organizations in the petroleum industry use data for a wide array of exploration and production functions, which can be modeled with established physics-based methods and newer empirical big-data methods, including artificial intelligence and predictive analytics.

With the right technology solutions, these companies can move beyond traditional real-time monitoring to more agile real-time prediction. By rapidly analyzing incoming technical and business data—and applying that information to complex models in real time—they can generate tactical insights that help increase drilling and production performance while preventing problems. By quickly searching and analyzing a large volume and variety of competitive intelligence, such as news about mergers, acquisitions or new investments, they can substantially improve strategic decision making.

Big data can help companies develop the "digital oilfield"—integrated operations that unite operational technology (OT) with information technology (IT) to improve decision making and enhance operational and business performance. Adding empirical analytics to existing physics-based analytics can take the industry to a new level of business improvement.

To capitalize on these opportunities, many oil and gas companies will need to adopt new IT solutions designed to address the specific challenges of big data. They need technology that can collect, manage and analyze large and rapidly growing volumes of data, such as the petabytes of production data generated by oilfield sensors. In addition, they need solutions that can analyze a wide variety of data types—including numerical data streaming in from drilling-rig sensors and unstructured data from logs, microseismic and other sources. New solutions must help integrate business data with technical data (see Figure 2), bringing together multiple OT systems and IT systems. They must enable searches of big data repositories to help companies quickly identify and visualize information among vast quantities of structured and unstructured data, and deliver results to support time-sensitive processes.
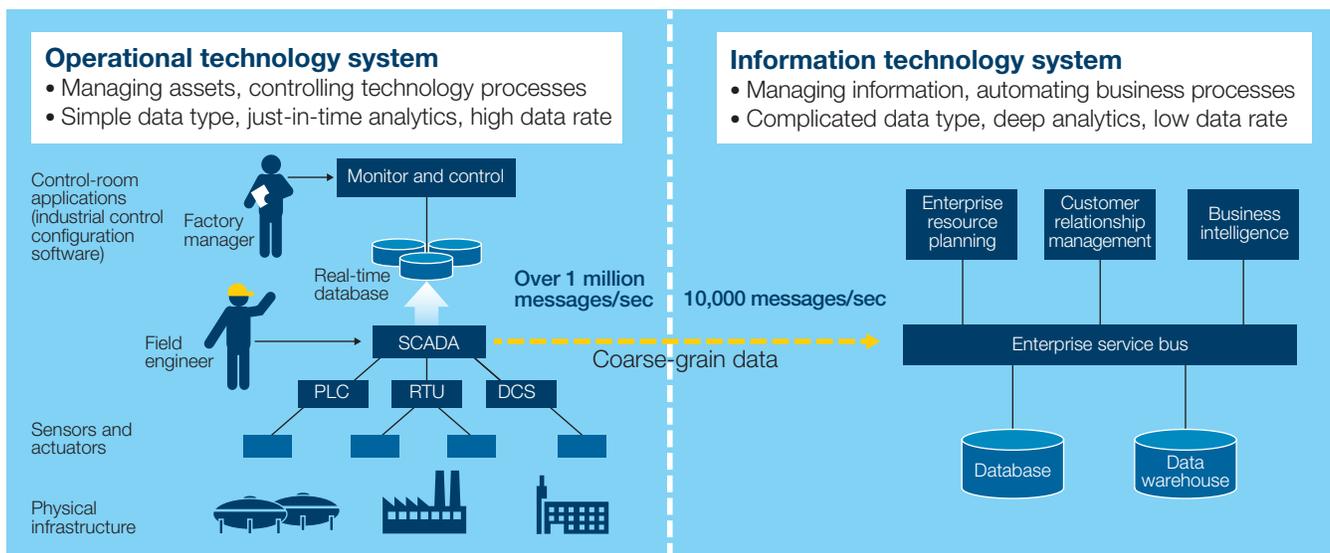
Operational technology system
- Managing assets, controlling technology processes
- Simple data type, just-in-time analytics, high data rate

Information technology system
- Managing information, automating business processes
- Complicated data type, deep analytics, low data rate

Control-room applications (industrial control configuration software)

Factory manager

Monitor and control

Real-time database

Over 1 million messages/sec

10,000 messages/sec

Field engineer

SCADA

Coarse-grain data

PLC   RTU   DCS

Sensors and actuators

Physical infrastructure

Enterprise resource planning

Customer relationship management

Business intelligence

Enterprise service bus

Database

Data warehouse

Figure 2: To achieve a wider range of business-improvement opportunities through analysis and optimization, field OT and corporate IT systems must be integrated.

Building the digital oilfield with the IBM big data platform

The IBM® big data platform provides a broad portfolio of solutions and capabilities designed to help oil and gas organizations capitalize on the potential of big data to optimize operations, improve business performance and facilitate effective, strategic decision making (see Figure 3). The platform-based approach allows organizations to adopt fit-for-purpose solutions that address specific challenges or goals and then easily add capabilities in the future. Pre-integrated capabilities help accelerate the time-to-value.
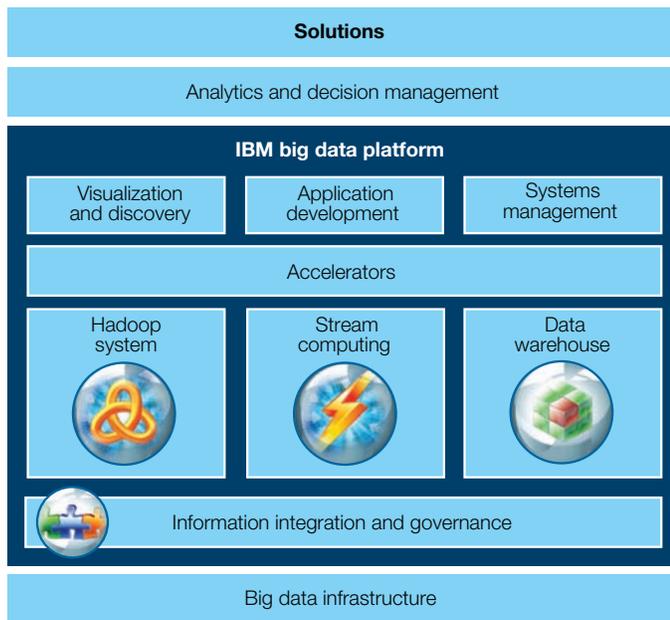
Figure 3: The IBM big data platform offers an array of integrated capabilities to address the tremendous volume, variety, velocity and veracity of big data.

Enterprise-class Hadoop

In the oil and gas industry, only a small fraction of data exists in relational format. The majority of data is in tagged, flat-file format or organized according to XML-based standards. Because there is such detailed characterization of wells, equipment, facilities and other oil and gas subject areas, there is no easy way to convert this data into a relational format.

Building on the open-source Apache Hadoop software framework with unique IBM innovations, IBM InfoSphere® BigInsights™ helps oil and gas companies collect, process, analyze and manage this large volume and variety of data. With InfoSphere BigInsights, organizations can ingest and analyze not only structured data from relational databases but also complex, unstructured data that those relational databases cannot accommodate. InfoSphere BigInsights analyzes all of this data in its native format, without imposing a schema/structure. As a result, companies can bring together and use a broad range of data that might have previously been too expensive or complex to integrate and analyze.

Data consolidation helps streamline collaboration among interdisciplinary engineering teams that might have previously spent excessive amounts of time assembling data for reservoir simulations or other purposes. Data existing in different formats, associated with different disciplines, can now be freely shared and analyzed without an expensive transformation to a relational database. At the same time, InfoSphere BigInsights enables organizations to conduct any transformations necessary to prepare that data for modeling and simulation.

Real-time adaptive analytics and living model management

InfoSphere Streams enables organizations to continuously capture, analyze and cleanse data in motion to rapidly refine models and facilitate real-time, tactical decision making.

Organizations can analyze data streaming in the drilling of wells or the operating of surface-facility equipment without having to first store the data. As a result, engineers can refine models, check the veracity of data and make operational changes at the speed of business.

Workload-optimized data warehousing

IBM PureData™ System for Analytics is an integrated data warehouse appliance that enables organizations to conduct deep, complex analytics on large-scale data volumes, from the production data collected from wells to enterprise financial information.

With an innovative, asymmetric, massively parallel processing (MPP) architecture, PureData System for Analytics supports rapid analysis of hundreds of terabytes of relational information, enabling organizations to explore more variables in production data, find more patterns in operational data and deliver results faster than before. As a single, expert integrated appliance, PureData System can be up and running in hours instead of days—critical for the fast-moving oil and gas industry.

Visualization and discovery

Searching massive volumes of historical reservoir data or unstructured information, such as text-based drilling reports or competitive press releases, can take weeks or months depending on the resources at hand. IBM InfoSphere Data Explorer offers technology to shorten that timeframe; now organizations have the power to search federated and consolidated repositories both internal and external to the enterprise. With secure federated navigation and discovery across a broad range of applications, data sources and file types, organizations are able to quickly find key information and enable rapid decision making.

Improving operations with operational analytics

Big data presents important opportunities for enhancing the efficiency, safety, productivity and cost-effectiveness of oil and gas operations. Yet it comes with an array of operational technology challenges that often impede the use of big data for operational gains. For example, companies need ways to cleanse and verify data generated by sensors on equipment and in wells, as erroneous data can lead to poor conclusions during surveillance and impair decisions based on models.

IBM big data platform at a glance

InfoSphere BigInsights provides an integrated solution for analyzing hundreds of terabytes, petabytes or more of raw data derived from an ever-growing variety of sources.

InfoSphere Streams provides a state-of-the-art computing platform that can help companies turn burgeoning, fast-moving volumes and varieties of data into actionable information and business insights.

InfoSphere Data Explorer provides federated discovery, search and navigation over a broad range of data sources to help organizations get started quickly with big data initiatives and gain more value from their information.

Robust IBM data warehouse software and integrated systems help simplify and accelerate the delivery of insights derived from your data.

IBM PureData System for Analytics is a high-performance, scalable, massively parallel system that enables clients to gain deep insight from their data and perform analytics on enormous data volumes.

IBM PureData System for Operational Analytics—part of the IBM PureSystems™ family—is an expert integrated data system designed and optimized specifically for the demands of an operational analytics workload.

Analysis of streaming data also plays a key role in addressing the challenges of advanced condition-based monitoring and maintenance. Sensor data from equipment and wells is critical for avoiding the downtime, costs and safety issues caused by equipment failures and the subsidence that often follows water-flooding or steam-injection processes. However, many organizations today use only a fraction of available data because they lack tools to effectively handle its volume and velocity.

InfoSphere Streams helps address that issue by enhancing the veracity of data and the analytics based on SCADA and PLC data. Because InfoSphere Streams does not require data storage, organizations can use it to cleanse data before it enters historians and control systems—opening up a larger amount of their big data for analysis while generating faster results. As a result, organizations gain insights to make real-time decisions about equipment and conditions that help improve the availability and utilization of assets, reduce downtime, increase production, reduce nonproductive drilling time and improve safety.

Analysis of streaming data enables organizations to refine the complex models used for drilling and for production. Typically, organizations run detailed physics-based models offline because they lack IT solutions that can apply large data volumes of streaming data to these sophisticated models. Consequently, model refinement is slow because it might take days or weeks to prepare in-and-out data, generate the results and tune the model.

With InfoSphere BigInsights and InfoSphere Streams used together for real-time adaptive analytics (RTAA), organizations can score streaming data (both structured and unstructured) against physics-based models in the field, in real time. By continually refining and updating models while monitoring processes, organizations can optimize drilling and production operations and improve productivity. The RTAA solution pattern can be used for a wide variety of subject areas including drilling optimization, production optimization, advanced condition monitoring and microseismic interpretation for unconventional resources.

Optimizing operations for specific business goals

Solutions designed for big data can help organizations integrate operational analysis with business intelligence so they can optimize processes in order to meet specific business goals. For example, a company might want to determine the best offshore drilling location for maximizing oil recovery while minimizing the tax obligations to nearby countries. Or an organization might want to set future production goals for a particular revenue target or operating expenditure limit. To succeed, organizations need to analyze not only operational parameters but also financial parameters.

The IBM big data platform can integrate data from multiple operational and business sources to facilitate decision making. Using PureData System for Analytics, organizations can analyze a large volume of structured data from production, financial and other subject-area source systems in a single environment. They can then run ad hoc queries to find the best ways to improve production, reduce expenditures, improve safety, and optimize their supply chain or procurement. PureData System for Analytics can also deliver results rapidly—in minutes and seconds rather than days.

Enhancing competitive analysis and facilitating strategic decision making

Harnessing the power of big data is valuable when conducting competitive analyses. Organizations can search and analyze a range of industry information to identify trends, anticipate changes and uncover emerging opportunities. To use big data in this way, organizations need solutions that can quickly search and analyze a wide variety of internal and external data, including the unstructured data of news feeds.

InfoSphere BigInsights enables organizations to pull in large volumes of data from a wide variety of sources, including unstructured sources such as text-based news feeds as well as production and other engineering data from external data providers. IBM content analysis capabilities, which draw on the natural-language processing technology used for IBM Watson™, can help make unstructured content accessible and ready for analysis. InfoSphere Data Explorer enables organizations to search and visualize information from federated and consolidated repositories, including both internal and external sources.

For example, an exploration and production company might use the IBM platform for big data to conduct acreage assessment and appraisal, factoring in scientific data and competitive considerations to enable more competitive bidding. The company could analyze directional survey and other well data, pressure-volume-temperature (PVT) data, well logs, and 2D and 3D seismic data, alongside industry news and even relevant social media to evaluate a particular oilfield. Managers could view all of this data via a competitive intelligence dashboard and develop insights to identify the best drilling locations, optimize the budget and confidently bid for leases with less investment risk.

Analyzing operational data to improve business results

Operating across six continents, one of the world's largest oil and gas producers wanted to obtain a more comprehensive view of key assets to deliver greater value to shareholders. The company wanted to examine specific attributes of wells and evaluate their production levels and well works. To do so, the company needed to integrate information across numerous applications, organizational silos and discrete data repositories.

The company implemented a solution that joined IBM PureData System for Analytics with other core data sources. This IBM appliance loads and stages data to help deliver faster performance than traditional extract, transform and load (ETL) methods. The company ran a record 25,000 business queries in one day and accelerated key processes. For example, geographic information system (GIS) reporting used to take 12 months; with the IBM solution, the company now can generate reports in real time.

By analyzing more data, faster than before, the company is better able to capitalize on available information to achieve business goals. At the same time, the company eliminated 50 percent of the cost of legacy ETL solutions—saving USD2 million per year.

To learn more about this success story, visit:
http://ibm.co/TMGvFW

Gearing up for big data

Big data holds opportunities for increasing oil and gas production, improving efficiency in the field and in corporate offices, reducing costs, enhancing safety, facilitating interdisciplinary collaboration for integrated operations, and boosting strategic decision making and competitiveness. To capitalize on those opportunities, organizations need solutions that can handle the distinct challenges of big data.

The IBM big data platform offers a comprehensive array of capabilities for addressing the tremendous volume, variety, velocity and veracity of big data. With IBM solutions, oil and gas companies can analyze data streaming in from sensors and use insights to refine models, prevent and address problems and boost production. They can optimize operations for specific business goals. And they can mine a wide variety of data sources to enhance competitive decision making. The IBM big data platform helps organizations build a foundation that supports a digital oilfield designed to enhance efficiency, improve safety and maximize business performance.

For more information

To learn more about how IBM solutions can help you capitalize on big data, please contact your IBM representative or IBM Business Partner, or visit: **ibm.com**/software/data/bigdata/industry-oil.html

To read more about the use of big data in real-world scenarios, please visit: **ibm.com**/services/us/gbs/thoughtleadership/ibv-big-data-at-work.html