Big Data in Oil and Gas Industry

Author: Elvirosa Brancaccio - Serintel Srl - Rome (Italy)

1 Introduction

Big Data or Big Data analytics refers to a new technology which can be employed to handle large datasets which include six main characteristics of volume, variety, velocity, veracity, value, and complexity.

With the recent advent of data recording sensors in exploration, drilling and production operations, oil and gas industry has become a massive data intensive industry.

Analyzing seismic and micro-seismic data, improving reservoir characterization and simulation, reducing drilling time and increasing drilling safety, optimization of the performance of production pumps, improved petrochemical asset management, improved shipping and transportation, and improved occupational safety are among some of the applications of Big Data in oil and gas industry.

In fact, there are ample opportunities for oil and gas companies to use Big Data to get more oil and gas out of hydrocarbon reservoirs, reduce capital and operational expenses, increase the speed and accuracy of investment decisions, and improve health and safety while mitigating environmental risks.

Figure 1 Big Data in Oil and Gas Exploration and Production

One of the key enablers of the data-science-driven technologies for the industry is its ability to convert Big Data into “smart” data. New technologies such as deep learning, cognitive computing, and augmented and virtual reality in general provide a set of tools and techniques to integrate various types of data, quantify uncertainties, identify hidden patterns, and extract useful information enormously reducing the data processing time. This information is used to predict future trends, foresee behaviors, and answer questions which are often difficult or even impossible to answer through conventional models.

2 Big Data Definition

Big data includes unstructured (not organized and text-heavy) and multi-structured data (including different data formats resulting from people/machines interactions)[1].

The term Big Data (also called Big Data Analytics or business analytics) defines the first characteristic of this method and that is the size of the available data set. There are other characteristics related to the data which make it viable for Big Data tools. Those *6Vs* characteristics refer to[2]:

- *Volume*: refers to the quantity of data or information. These data can come from any sensor or data recording tool. This vast quantity of data is challenging to be handled due to storage, sustainability, and analysis issues[1]. Many companies are dealing with huge volume of data in their archives; however, they do not have the capability of processing these data. The main application of Big Data is to provide processing and analysis tools for the increasing amounts of data [3].

- *Variety* and *Velocity* as a characteristic of Big Data refers to the speed of data transmission and processing. It also refers to the fast pace of data generation.
  - Variety refers to the various types of data which are generated, stored, and analyzed. The data recording devices and sensors are different in types and as a result the generated data can be in different sizes and formats. The formats of the generated data can be in text, image, audio, or video. The classification can be done in a more technical way as structured, semi-structured, and unstructured data.
    The majority of oil and gas generated data from SCADA systems, surface and subsurface facilities, drilling data, and production data are structured data. These data could be time series data which have been recorded through a certain course of time. Another source of structured data includes the asset, risk, and project management reports. There would be also external structured data sources such as market prices and weather data, which can be used for forecasting.
    The sources of unstructured data in oil and gas industry include well logs, daily written reports of drilling, and CAD drawing.
    The sources of semistructured data include processed data as a result of modeling and simulation.
  - The challenging issue about the velocity component is the limited number of available processing units compared to the volume of data. Recently, the data generation velocity is huge, as a data of 5 exabyte is generated just in two days. This is equivalent to the total amount of data created by humans until 2003[4]. The velocity characteristic is even more prominent for oil and gas industry due to complex nature of various petroleum engineering problems. There are many cases in which real time and fast processing of data is crucial in oil and gas industry. For example, fast processing of well data during drilling can result in identifying kicks and preventing destructive blow-outs efficiently [5].

---

[1] Trifu MR, Ivan ML. Big Data: Present and Future n.d.:32–41.

[2] Mohammadpoor, M., Petroleum, https://doi.org/10.1016/j.petlm.2018.11.001

[3] J. Ishwarappa, J. Anuradha, A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology vol. 48, (2015), pp. 319–324, https://doi.org/10.1016/j.procs.2015.04.188

[4] M.S. Sumbal, E. Tsui, See-to EWK, M.S. Sumbal, E. TsuiSee-to EWK, Interrelationship between Big Data and Knowledge Management : an Exploratory Study in the Oil and Gas Sector, (2017), https://doi.org/10.1108/JKM-07-2016- 0262.

[5] J. FeblowitzInsights IDCE, Analytics in Oil and Gas: The Big Deal about Big Data, (2013), pp. 5–7.

- *Veracity* refers to the quality and usefulness of the available data for the purpose of analysis and decision making. It is about distinguishing between clean and dirty data. This is very important as the dirty data can significantly affect the velocity and accuracy of data analysis. The generated data should be professionally and efficiently processed and filtered to be used for data analysis; otherwise the results will not be reliable. The veracity of data is challenging in oil and gas industry specifically due to nature of data, which mainly comes from subsurface facilities and it might include uncertainty. Another challenge comes from the data collected by conventional manual data recording, which is done by human operators.

- *Value* is a very significant characteristic of the Big Data. The returned value of investments for Big Data infrastructures is of a great importance. Big Data analyzes huge data sets to reveal the underlying trends and help the engineers to forecast the potential issues. Knowing the future performance of equipment used during operation and identifying the failures before happening can make the company to have competitive advantage and bring value to the company.

- *Variability* refers to the fact that data changes during processing and lifecycle.

The Figure 1 summarizes the above-mentioned characteristics of Big Data.

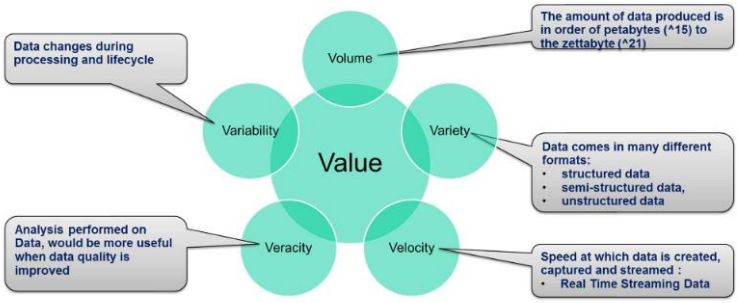

Figure 2 Big Data 6Vs Characteristics [6]

Complexity of the problem for which the data gathering is conducted is another important characteristic [7]. Dealing with large data sets which are coming from a complex computing problem is sophisticated and finding the underlying trend can be challenging. For these problems Big Data tools can be very helpful.

3 Big Data Methodologies

As the Big Data is involving huge data sets and, in some cases, complicated problems, it is very important to have access to innovative and powerful technologies. These robust technologies should be very fast and accurate processors.

3.1 Technologies

Apache Hadoop

This tool is an open-source framework which is created by Doug Cutting and Mike Caferella in 2005 which is named after a toy elephant it has the capability of parallel processing of huge data sets, which results in scalable computing.

---

[6] How to use Big Data technologies to optimize operations in Upstream Petroleum Industry, 21st World Petroleum Congress;

[7] D. Khvostichenko, S. Makarychev-mikhailov, Effect of fracturing chemicals on well Productivity: avoiding pitfalls in Big data analysis, SPE Int. Conf. Exhib. Form. Damage Control, Lafayette: Society of Petroleum Engineers, 2018, https://doi.org/ 10.2118/189551-MS

MongoDB

This is a NoSQL (non-relational) database technology which is document-orientated, based on JSON and written in C++. JSON is data processing format based on a JavaScript and is built on a collection of name/value pairs or an ordered list of values. NoSQL database technology can handle unstructured data such as documents, multimedia, and social media. Moreover, MangoDB provides a dynamic and flexible structure to be customized to fit the requirements of various users[2].

Cassandra

This is another NoSQL database technology which is key and column orientated. Cassandra was first a Facebook project that became open sourced few years later. It is especially efficient where it is possible to spend more time to learn a complex system which will provide a lot of power and flexibility.

3.2 Processing

Big data sets which are collected need to be analyzed to extract the valuable underlying information. There have been different processing tools which translates the large data sets into meaningful results and outcomes. Following is a list of common processing tools for Big Data [8].

R

R is a modern, functional programming language that allows for rapid development of ideas, together with object-oriented features for rigorous software development initially created by Robert Gentleman and Robert Ihaka. The powerful set of inbuilt functions makes it ideal for high-volume analysis or statistical simulations. It also supports the packaging system, which means that the code provided by others can easily be shared. Finally, it generates high-quality graphical outputs, so that all stages of a study, from modelling/analysis to publication, can be undertaken within R.

Datameer

Datameer is an easy to use programming platform which uses Hadoop to improve its data processing. It comes with user-friendly data importing and output visualization tools. It is estimated to gain more interest as it uses a user-friendly interface to conduct various data processing tasks.

BigSheets

IBM has offered a web application called BigSheets, which helps less expert and nontechnical users to gather unstructured data from various online and internal sources and then conduct a data analysis and present the results with simple visualization tools. BigSheets also utilizes Hadoop to process massive datasets. It also employs some additional tools such as OpenCalais to facilitate the extracting of structured data from a pool of unstructured data. This tool should be used for data analysis individually and it is easier to be used by the users familiar with spreadsheet applications.

4 Big Data in the Oil and Gas Industry

As written above, large volume of data is generated from day to day operations in Oli & Gas industry. Following are three big oil industry problems that consume money and produce data:

1. Oil is hard to find. Reservoirs are generally 5,000 to 35,000 feet below the Earth's surface. Low-resolution imaging and expensive well logs (after the wells are drilled) are the only options for finding and describing the reservoirs. Rock is complex for fluids to move through to the wellbore, and the fluids themselves are complex and have many different physical properties.

[8] https://doi.org/10.1016/j.petlm.2018.11.001

2. Oil is expensive to produce. The large amount science, machinery and manpower required to produce a barrel of oil must be done profitably, taking into account cost, quantity and market availability.
3. Drilling for oil presents potential environmental and human safety concerns that must be addressed.

Finding and producing oil involves many specialized scientific domains (i.e., geophysics, geology and engineering), each solving important parts of the equation. When combined, these components describe a localized system containing hydrocarbons. Each localized system (reservoir) has a unique recipe for getting the most out of the ground profitably and safely.

Big data analytics assists in streamlining key oil and gas operations, such as exploration, drilling, production and delivery, in the three sectors – upstream, midstream and downstream.

Figure 3 Global Big Data Services Market for Oil & Gas Industry, by Application Segment, 2016 (US$ Mn)[9].

Among all business segments upstream segment is the most dominant segment owing to increasing use of big data analytics for the discovery of non-conventional shale gas. Moreover, volatile oil prices increasing uncertainty in oil & gas business. Thus, organizations are adopting data driven approach to minimize their risk.

Followed by it, downstream is expected to be the second largest segment due to increasing use of product analytics solution which assist refineries to standard chemical composition of the finished products. Further, increasing use of renewable energy sources such as solar, wind is anticipated to negatively impact the demand for fossil fuels which in turn has influenced organizations to improve their relations with customer. Growing use of social media to drive value to improve customer relationship is projected to fuel the demand for big data services in downstream business segment over the forecast period (2016-2023).

4.1 Big Data in the Oil & Gas Upstream Sector

Oil & Gas companies use thousands of sensors installed in subsurface wells and surface facilities to provide continuous data-collecting, real-time monitoring of assets and environmental conditions.

The data volume is coming from sensors, spatial and GPS coordinates, weather services, seismic data, and various measuring devices. "Structured" data is handled with specific applications used to manage surveying, processing and imaging, exploration planning, reservoir modelling, production, and other upstream activities. But much of this data is "unstructured" or "semi-structured" such as emails, word processing documents, spreadsheets, images, voice recordings, multimedia, and data market feeds, which means it is difficult or

[9] https://www.credenceresearch.com/report/big-data-services-market

costly to either store in traditional data warehouses or routinely query and analyse. In this case, appropriate tools for Big Data should be used.

To support the real-time decision-making, Oil & Gas companies need tools that integrate and synthesize diverse data sources into a unified whole. Being able to process Big Data makes it possible to derive insight from the relationships that will surface when all of these sources are processed as a whole. But to unlock this value, Oil & Gas companies need access to the appropriate technology, tools, and expertise.

Like generic Big Data, the Upstream Data is also characterized by the *6Vs* as shown in the next Figure:

	Exploration	Resevoir Engineering & Development	Drilling and Completion	Production
Volume	Seismic acquisition SEGD	Facilities Reservoir Engineering	Sensors : Flow Pressure ROP	SCADA Sensors : Flow Pressure
Variety	Structured data : • SEGD • Pre-stack • Post-stack Semi-structured : • Implantation Report …	Structured data : • WITSML (XML) • PRODML • RESML Unstructured data : • Log Curves / Drilling & Test / Lithology /Cores …	Structured : • WITSML Semi-structured : • Final Well Report, • Daily Drilling Report Unstructured : • Drilling Log / Gas Log .. etc.	Structured Production data : • PRODML • RESML Semi-structured : • Crude Analysis Report
Velocity	Real Time Data Acquisition : Wide azimuth data acquisition		Real Time Data Acquisition : Mud Logging / LWD / MWD	Real Time Data Acquisition : SCADA Sensors
Veracity	Seismic processing	Reservoir Modeling	Sensor calibration	Sensor calibration
Variability	Seismic Interpretation Geology Interpretation	Reservoir Simulation Combination of seismic, drilling and production data	Data Interpretation & Optimisation	Data Interpretaion
Value	**Navigation, Visualization & Discovery Run integrated asset models**	**Improve Drilling Program Drive innovation with unconventional resources (shale gas, tight oil)**	**Reduce costs Reduce Non Productive Time (NPT) Reduce risks Improve HSE performances**	**Increase speed to first oil Enhancing production**

Figure 4 Upstream Big Data[9]

Work on the application of Big Data and Analytics in the Oil & Gas industry is in the experimental stage. Only some companies have adopted Big Data in the field.

At present, companies' experience in using this tool is at the level of pilot projects, and efforts are made to test this technology in practice and to assess its potential benefits [10]:

- Chevron proof-of-concept using Hadoop (IBM BigInsights) for seismic data processing. The company hopes to reduce the cost of sending deep water drillships out into the ocean by doing a better job of processing the vast amounts of data that can help identify reservoirs of oil, using open source software. Currently, the daily internal traffic generated by Chevron accounts for more than 1.5 terabytes.
- The BP's "Field of future" project began from 2003, program was established with an initial focus on engagement and deployment, the objective being to deploy core technologies in a limited number of assets in order to build a track record, to reaffirm the prize and to build a technical and architectural foundation for subsequent 'bigger moves'. In this project the big data related application contains: big data application framework, CoRE (Collaborative Real Time Environment), ACEs (Advanced Collaboration Environment Centre). Based on these big data applications, BP effectively accomplish the security control of overseas drilling and enhance the efficiency of drilling. Based on the key performance indicators, BP identified the potential big data applications, such as preventing equipment down time, optimizing field scheduling, increase efficiency, etc.

- Shell piloting Hadoop in Amazon Virtual Private Cloud (Amazon VPC) for seismic sensor data.
  Smart Fields is Shell's global program to address the "digital oilfield of the future", it is an asset that Shell can continually optimize 24 hours a day, 7 days a week.

[10] Big data strategy for the oil and gas industry: general directions, 2017 https://www.researchgate.net/publication/319975532

- Cloudera Seismic Hadoop project combining Seismic Unix with Apache Hadoop. The company implemented an architecture of Hortonworks Connected Platforms running on Amazon Web Services (AWS) in the cloud. It is leveraging a mix of AWS services to run these platforms. Another key component has been the Operational Services team for the big data implementation. The company also developed a new product offering, Saismic™. This cloud-based service provides global seismic data on-demand with native support for deep learning and advanced analytics. Figure 3 Cloudera: Complete, Integrated Hadoop Stack

- PointCross Seismic Data Server and Drilling Data Server using Hadoop and NoSQL. PointCross Inc. introduced the addition of its petabyte–scale Seismic Data Server and Repository (SDSR) using NoSQL and Hadoop™ technologies. The SDSR solves the problem of storing and managing very large data sets and of delivering on-demand and reduces the time it takes to locate and load the data needed by geophysicists at their workstation from a centralized store or federated stores in remote countries to meet local laws. All available geo-physical data is geo-tagged, categorized, and faceted making them ready for searching using a GIS front end. The technology is available to the organization either on the cloud or in a private virtualized Hadoop Distributed File System with the cataloging managed in Hbase™.[11]

- ENI HPC5 High Performance Computing - layer 5. HPC5 is a set of parallel computing units with a peak processing power of 51.7 petaFlops. Combined with the supercomputing system in operation since 2018 (HPC4), the peak computational capacity of the infrastructure totals 70 petaFlops: that is, 70 million billion mathematical operations performed in a single second. This is the latest-generation supercomputer and also one of the "greenest" on the planet – or in other words, with the lowest electricity consumption per petaFlop. The use of a supercomputer like HPC5 enables to increase the accuracy of the studies of underground rocks, reducing the margin for error in prospecting operations and decreasing time-to-market, or the time between the identification of the field and the launch of production. This also has a positive impact on sustainability, as it reduces waste, both in terms of energy and resources. HPC5 is a computing cluster, or in other words, **a set of computers** that work together to multiply overall performance[12].

Figure 5 HPC5, ENI Supercomputer

Much of the software innovation that's key to the digitization of big oil is happening at oil service contracting companies, such as Halliburton and Schlumberger, and big IT providers including Microsoft, IBM, Oracle and Open Source Projects.

[11] https://www.hartenergy.com/news/pointcross-unveils-updated-seismic-data-technology-92750

[12] https://www.eni.com/en-IT/operations/green-data-center-hpc5.html

Exploration and Development

By combination of Big Data and advanced analytics in Exploration and Development activities, managers and experts can perform strategic and operational decision-making. The areas where the analytics tools associated with Big Data can benefit Oil & Gas exploration include:

- Enhancing exploration efforts: historical drilling and production data help geologists and geophysicists verify their assumptions in their analysis of a field where environmental regulations restrict new surveys
- Combine enterprise data with real-time production data to deliver new insights to operating teams for enhancing exploration efforts
- Assessing new prospects: create competitive intelligence using Analytics applied to geospatial data, oil and gas reports and other syndicated feeds in order to bid for new prospects. Identifying seismic traces: using advanced analytics based on Hadoop and distributed Database for storage, quick visualization and comprehensive processing and imaging of seismic data to identify potentially productive seismic trace signatures previously
- Build new scientific models: By using high performance computing and based on combination between "historical data" and "Real Time Data Acquisition": Petabyte seismic data sets, Mud Logging, MWD, LWD, Testing, Gamma Ray.

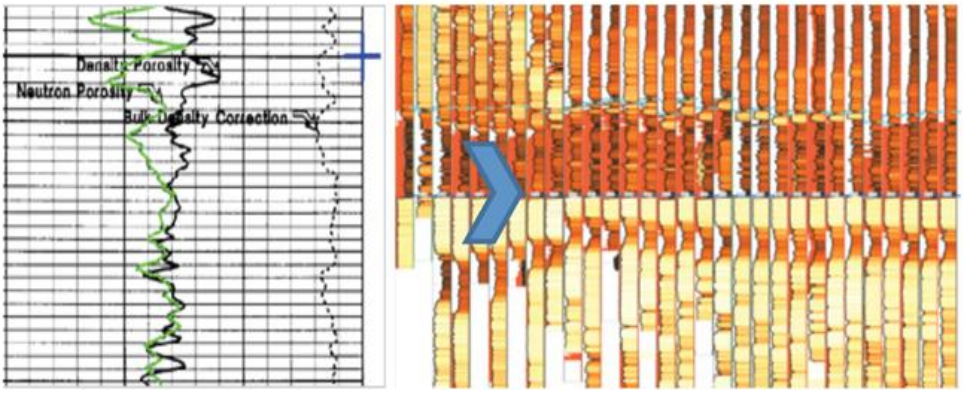

Figure 6 Raw well-log (left) and processed well log images showing the rock type (right)[13]

Drilling and Completion

It can be valuable making the drilling platforms and pipeline infrastructure smart to anticipating issues and acting to prevent failures and increase productivity. In this case, Big Data is used to identify conditions or anomalies that would impact on drilling operations can save human lives and equipment.

Real-time information returned from supervisory control and data acquisition systems on well-heads can be used to grasp opportunities that maximize asset performance and optimize production.

Related areas where analytics can improve drilling and completion operations are:

- Build and assessment of drilling models: based on all existing well data. These models are incorporated with geologic measurement into drilling processes, such as shale development. This will refresh models based on incoming sensor data from drill rig and help to optimize drill parameters.
- Improve Drill Accuracy and Safety: by early identifying anomalies that would impact drilling and prevent undesired events: kicks, blowout, etc.
- Drilling optimization: predictive Analytics help to reduce NPT (Non Productive Time), by early identifying the negative impacting factors of drilling operations.

[13] http://analytics-magazine.org/how-big-data-is-changing-the-oil-a-gas-industry/

- Cost optimization: by using scalable compute technologies to determine optimum cost.
- Real-time decision-making: must act in real time on drilling data and formation evaluation data and use this data for predictive modelling to facilitate real-time decision-making.
- Predictive maintenance: predict drill maintenance/downtime.

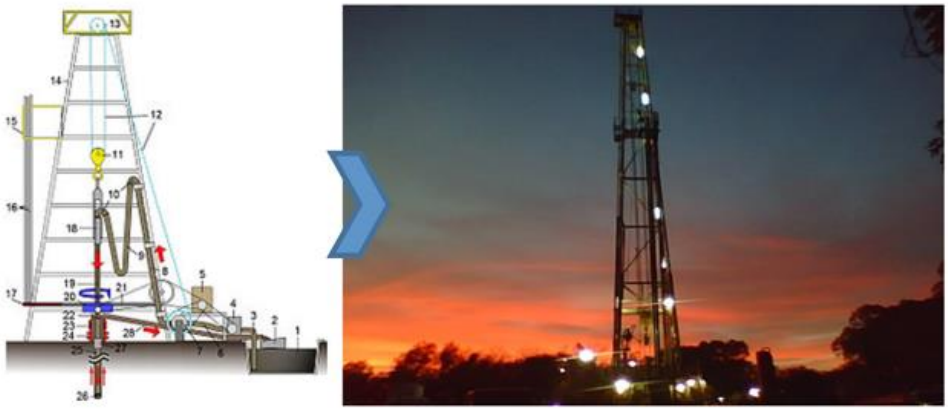

Figure 7 Simple well diagram (left) and an onshore drilling rig in south Texas (right)

Production

Producing oil and gas involves drilling and completing the wells, connecting them to pipelines and then keeping the flow of the hydrocarbons at an optimum rate, all integrally related to the subsurface environment. The path to optimizing production is dependent on the type of rock and structure of the reservoir. These decisions depend heavily on models created in the exploration phase.

Big Data is of great interest to production and operation work. Being able to predict future performance based on historical results, or to identify sub-par production zones, can be used to shift assets to more productive areas. Oil recovery rates can be improved, as well, by integrating and analysing seismic, drilling, and production data to provide self-service business intelligence to reservoir engineers.

- Enhanced oil recovery: 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.
- Performance forecasting: forecast production at thousands of wells. Aging wells where the forecast does not meet a predetermined production threshold are flagged for immediate remediation.
- Real-time production optimization: Real-time SCADA and process control systems combined with analytics tools help Oil & Gas producer to optimize resource allocation and prices by using scalable compute technologies to determine optimum commodity pricing. They also, help to make more real time decisions with fewer engineers.
- Improve Safety and prevent risks: by early detecting well problems before they become serious (slugging, WAG (Water Alternating Gas) gas breakthrough).

Equipment maintenance

Predictive maintenance is not a new concept for the oil and gas industry. In upstream, if pressure, volume, and temperature can be collected and analysed 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. Technicians often use data collected from pumps and wells to adjust repair schedules and prevent or anticipate failure. Better predictive maintenance also becomes possible[14]:

- Preventing down time: understand how maintenance intervals are affected by variables such as pressure, temperature, volume, shock, and vibration to prevent failure and associated downtime.
- Optimizing field scheduling: use this insight to predict equipment failures and enable teams to more effectively schedule equipment maintenance in the field.
- Improving shop floor maintenance planning: integrate well and tool maintenance data with supply chain information to optimize scheduling of shop floor maintenance.

Potential drilling errors or equipment failures can be identified in advance by analysing the sensor data (from equipment such as drill heads, down hole sensors, etc.) as well as geological data and thus telling what equipment works best in which environment.

Reservoir Engineering

Oil & Gas companies improve understanding of future strategy based on available oil for a better identification of reservoirs and reserves by integrate real-time data into the earth model both on the rig and in the office.

Also, they predict the chances of success of turning reservoir into a production well by:

- Improving engineering studies: engage sophisticated subsurface models and conduct detailed engineering studies on wells to identify commercial prospects earlier and with less risk
- Optimizing subsurface understanding: use Big Data tools to understand the earth's subsurface better and to deliver more affordable energy, safely and sustainably.
- Experiences and learned lessons from drilling operations: such determination of drilling coordinates through oil shale to optimize the number of wellheads needed for efficient extraction of oil, optimization of drilling resources by not over drilling well site, reducing waste of drilling exploration wells, etc.

4.2 Big Data in the Oil & Gas Downstream Sector

Big Data in refining

Oil and gas enterprises can employ big data predictive analytics and data management solutions to help downstream energy businesses streamline operations, improve efficiency, and minimize risk and reduce downtimes and maintenance costs of the refining equipment, thus improving asset management.

In a recent project by Repsol SA, Big Data analytics is utilized to conduct management optimization for one of the company's integrated refineries in Spain. For this project, Google Cloud would provide Repsol with data analytics products and consultation as well as Google Cloud machine learning services[15].

- Optimize Pricing, Minimize Financial Risk
  The downstream energy sector maximizes profits by tracking and analyzing a vast amount of data such as resource availability, usage, and predicted changes. Big Data management instrument can help with Big Data analytics, featuring the Hadoop file system.

- Efficient Data Collection and Delivery

[14] Hems A., Soofi A., Perez E., 2013, Drilling for New Business Value: How innovative oil and gas companies are using Big Data to outmanoeuvre the competition, A Microsoft White Paper, May 2013

[15] R. Brelsford, Repsol launches Big data, AI project at tarragona refinery, Oil Gas J. 116 (2018

It is needed uninterrupted access to all data all the time. Big Data analysis gives a stable storage infrastructure for enterprise data.

- Regulatory Compliance:
  Real-time monitoring of pipeline and refinery equipment health helps downstream energy companies comply with environmental and safety regulations. Cost-effective solutions can help store and manage the massive amounts of data these apps produce.

The use of Big Data Analytics in processing, logistics and sales is predicted to cause a breakthrough. Big Data can contribute to predicting the demand for oil products in the retail sales network and to analyse the pricing and the changes in prices by the competitors and the regions. Due to the discovered patterns, the opportunities of increasing the sale of related products and reducing delays in retail network (through more precise logistics of fuel-carrying vehicles) may also be attractive [16].

4.3 Big Data in the Oil & Gas Midstream Sector

The 'midstream' segment of the oil and natural gas industry refers to anything required to transport and store crude oil and natural gas before they are refined and processed into fuels and key elements needed to make a very long list of products, we use every day. Midstream includes pipelines and all the infrastructure needed to move these resources long distances, such as pumping stations, tank trucks, rail tank cars and transcontinental tankers.

Logistics in the petroleum industry is incredibly complex and the major concern is to transport oil and gas with the lowest risk possible. Companies utilize sensor analytics to ensure the safe logistics of their energy product. Predictive maintenance software analyzes sensor data from pipelines and tankers to detect abnormalities (fatigue cracks, stress corrosion, seismic ground movements, etc.), which allows preventing accidents [17]. Oil and gas enterprises can employ big data predictive analytics to reduce downtimes and maintenance costs of the refining equipment, thus improving asset management.

5 Big Data Components

On the basis of component, global big data services market for oil & gas industry is categorized into following types:

- Hardware
- Software
- Services

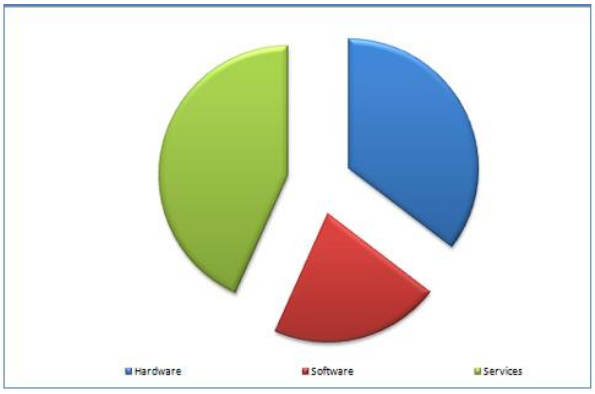

Figure 8 Global Big Data Services Market, by Component Type, 2016 (US$ Mn)

[16] How does Google search affect trader positions and crude oil prices? // Economic Modelling, 2015, vol.49, pp.162–171

[17] How To Benefit from Big Data Analytics in the Oil and Gas Industry? big-data-analytics-oil-gas, 2020

Oil & gas industry has invested for long in seismic software, sensors, visualization tools and other digital technologies for real-time monitoring of their operations. Massive financial operational data is generated from various financial and operational transactions which can be utilized to derive value to improve efficiency and to deliver value to the customer. Leading players in this space offering various products and services to facilitates organizations to utilize this data to derive value.

On the basis of component type, the global big data services market for oil & gas industry can be segmented into hardware, software and services.

Among product component type the services segment is expected to be the most dominant and fastest growing segment due increasing adoption of various services such as consulting, cloud and integrated services. Increasing competition among organizations in oil & gas industry has influenced them to seek assistance from consulting companies to implement various big data solution to improve their operations across value chain. Increasing adoption of cloud services due to low data storage cost is anticipated to fuel the demand of big data services market for oil & gas industry over the forecast period (2016-2023).

On the basis of geography, the global big data services market for oil & gas industry is segmented into following categories:

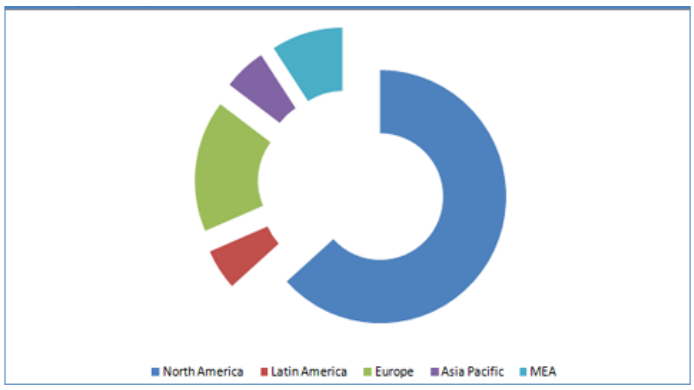

Figure 94 Global Big Data Services Market for Oil & Gas Industry, By Geography, 2016 (US$ Mn)

Big data services market is expected to witness robust growth over the forecast period due to wide adoption of big data technologies such as Hadoop, DataStax, Splunk, and HPCC Systems.

Global big data services market for oil & gas industry is segment on the basis of geography. North America is expected to contribute highest revenue share in global big data services market for oil & gas industry by end of 2016 due to wide adoption of digital technologies with the focus to improve efficiency and explore new as well as existing resources efficiently. Leading organizations of oil & gas industry in the region are adopting big data services to utilize operational data improving recovery rate and to reduce cost. The region has witness highest IT investment among all regions and anticipated to maintain its dominance over the forecast period (2016-2023). Followed by it, Europe is anticipated to be the second largest regions in the global big data services market for oil & gas industry in 2016. Strict government regulations coupled with increase awareness about the associated advantages of big data services is projected to intensify the significant growth of the region over the forecast period (2016-2023)[18].

6 Security

[18] Big Data Services Market By Application (Upstream, Midstream, Downstream, Attractive Investment Proportion By Application), By Component (Hardware, Software, Services Attractive Investment Proportion By Technology) - Growth, Future Prospects & Competitive Analysis, 2016 - 2023

Oil & Gas companies anticipate IT security breaches by using predictive analytics and bolstering security with data from the global protection systems including video monitoring, access control et anti-intrusion.

Also, there is particular interest in deploying complex event processing (CEP) technology to monitor security concerns in the Oil & Gas industry in real time by:

- Combining data from multiple sources to infer events or patterns that indicate a current or imminent threat.
- Making faster decisions, supported by faster delivery of decision support information, to identify possible threats.
- Predict/prevent cyber-terror acts.

Oil companies need to identify events or patterns that could indicate an impending security threat or cyber-terrorist act in order to keep their personnel, property and equipment safe. Predictive analytics is central part of identifying patterns that can help detect these threats beforehand. The MapR Distribution for Hadoop can help in identifying threats in real-time through machine learning and anomaly detection techniques and reducing the likelihood of such incidents.

7 Health, Safety & Environment

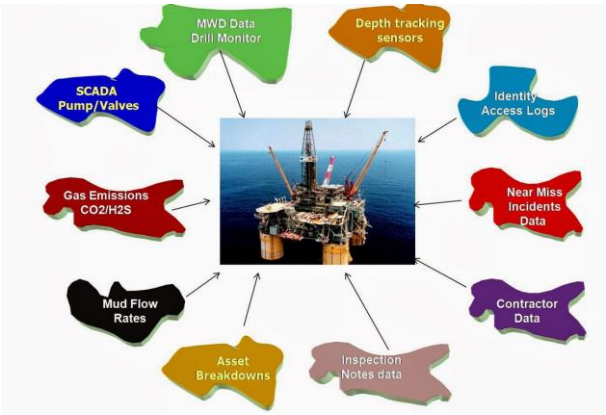

Figure 10 Upstream Sensor Data and Big Data Analytics in Oil and Gas Industry

Big Data contributes significantly to reduce risk and optimize costs related to operations and Health, Safety and Environment (Baaziz & Quoniam, 2013b):

- Prevent undesired events while drilling like kicks,
- Predict drill maintenance/downtime, optimize drill parameters and prevent blowout accidents,
- Using weather or workforce scheduling data to avoid creating dangerous conditions for workers and mitigating environmental risks.

By using Hadoop to examine data from a variety of sources, well problems can be detected before they become serious, anomalies in drilling can be identified in real time , and drills can be proactively shut down in order to prevent environmental risks. The MapR Distribution for Hadoop can be used to enhance the environmental health and safety of oil rigs and drills through identification of patterns and outliers before any catastrophic incidents occur.

8 Big Data Challenges

Figure 11 Data are the Fuel of the Future[19]

One of the major challenges of Big Data's application in any industry including oil and gas industry is the cost associated with managing the data recording, storage, and analysis. With the recent technological improvements, fog computing, cloud computing, and Internet of Things (IoT) have become available to fix the issues regarding data storage and computations[20]. Costly and limited cloud computing facilities are not suitable options for non-fixed location or latency-sensitive applications. On the other hand, fog computing facilities provide storage and computing facilities closer to data generation sources, which resolves the mentioned challenges to some extent. However, IoT is a newer technology, which is more mobile and fixes the latency issues as well.[21]

The challenges of using Big Data for oilfield service companies include the knowledge of personnel in oil companies and the data ownership issues. Big Data can be used for seismic analysis, reservoir modeling, drilling services, and production reporting. Furthermore, were defined some factors for a successful application of Big Data for oil and gas industries including accurately defining the business problem, combining Big Data methods with physics-based data analysis, using interdisciplinary team of computer scientists and petroleum engineers, delivering the results as a user-friendly interface, being need driven, and addressing exactly how the solved problem is related to the whole picture [22,23]

The emergence of Big Data in oil and gas industry has become more prominent by evolution of digital oilfields, where various sensors and recording devices are generating millions of data each day. One of the critical challenges in digital oilfields is the data transfer from the field to data processing facilities based on the type of data, amount of data, and data protocols[24].

Other technical challenges facing the application of Big Data are:

- related to the limitations associated with the data recording sensors.
- the frequency of data recording
- the quality of the recorded data.
- the thorough understanding of the physics of the problem.

---

[19] https://www.economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy

[20] https://doi.org/10.2118/1011-0042-JPT.

[21] S. Konovalov, R. Irons-mclean, Addressing O & G Big Data Challenges at the Remote Edge Fog Computing and Key Use Cases, (2015), pp. 3–5.

[22] Cameron, S. As, Big Data in Exploration and Production : Silicon Snake-oil , Magic Bullet , or Useful Tool ? (2014).

[23] Y. Gidh, N. Deeks, L.O. Grovik, D. Johnson, J. Schey, J. Hollingsworth, Paving the Way for Big Data Analytics through Improved Data Assurance and Data organization, (2016).

[24] P. Neri, Big Data in the Digital Oilfield Requires Data Transfer Standards to PerformIndustry Environment a Trend towards More Collaboration Standardization the Critical Role of Metadata, (2018), pp. 1–6.

9 Conclusion

Figure 12 Connected Oil&Gas Industry

The oil and gas industry are intensively using data and it necessitates Big Data technologies that enable the maximum use of different types of rapidly increasing data. Real time Big Data analytics may ensure more efficient oil production, reduce the costs and risks, improve security and compliance with regulatory requirements and the quality of decisions. The introduction of Big Data technologies can be an important step toward the implementation of corporate strategies for the complex intellectualization of the production of the of oil and gas companies

Leading Oil & Gas companies are already beginning projects to deploying Big Data technologies that can help them track new business opportunities, reduce costs and reorganize operations.

Oil & Gas companies need first to proceed to a gap analysis to determine the major requirements of technology and data-management expert staff. This allows a focused investment in mature and proven technologies as well as those who will face the exponential growth of data volumes. Then Oil & Gas companies can create new strategies that will help them manipulate these data and use them to support experts in their business process and managers in their decision-making process.

By recognizing the value of the unexploited data assets in supporting fact-based decision-making, Oil & Gas companies establish real cases of Big Data uses. Then they can create enhanced business value based on innovation and able to lead towards a sustainable competitive advantage.

With big data analytics, companies transform enormous datasets into sound oil and gas exploration decisions, reduced operational costs, extended equipment lifespan, and lower environmental impact[25]:

To survey and monitor oil exploration areas.

The company employs a seismic analysis to survey the area and indicate whether the given area contains oil and gas deposits. The more sophisticated big data analysis allows understanding the nuances of a particular drilling site before deciding to drill.

To forecast production.

For example, the installation of optical fiber cables with sensors within the wells to measure seismic data. This data is further analyzed using artificial intelligence technologies to create 3D and 4D maps of the oil reservoirs to find out how much oil and gas is still left in the reservoir.

[25] https://www.scnsoft.com/blog/big-data-analytics-oil-gas#midstream; Feb 2020

To extend equipment lifespan.

Generating tons of sensor data, companies run advanced analysis on drilling sites machinery to improve its performance and proactively understand what equipment requires maintenance. This stimulates a longer drilling time with fewer maintenance stops and. Solely in Nigeria, Shell has managed to save over $1 million by leveraging sensor analytics.

To increase logistics efficiency.

Industry utilizes complex algorithms to analyze transportation and production costs, economic factors that drive demand as well as weather patterns to determine how and where to move refined products and how to set the prices.

As oil and gas companies awake to the potential of analytics, many jobs will be created for data scientists, opening a portal for new applications and ideas to enter the industry.