

Ramiz M. Aliguliyev¹, Yadigar N. Imamverdiyev²

DOI: 10.25045/jpit.v08.i2.04

Institute of Information Technology of ANAS, Baku, Azerbaijan

¹a.ramiz@science.az, ²yadigar@lan.ab.az**BIG DATA STRATEGY FOR THE OIL AND GAS INDUSTRY: GENERAL DIRECTIONS**

Big Data technologies provide approaches and tools that are essential for the competitive development of the oil and gas industry. Interests of the oil and gas companies towards the Big Data are growing against the backdrop of the plummeting oil prices in the global energy market. An imperative prerequisite for the effective implementation on this steer is to establish a strategy organically correlated to the general corporate strategy. To this end, in this paper, we consider the development of Big Data strategy for the oil and gas industry. Moreover, the paper analyzes the potential of Big Data technology and Big Data sources in the oil and gas industry, Big Data application experience in oil and gas companies and the existing problems in the field of data management. We also define the general principles and direction of the formation and implementation of Big Data strategy.

Keywords: oil and gas industry, Big Data; Hadoop, Big Data strategy, Big Data Analytics.

Introduction

Big Data is often referred to the oil of the information age, however the connection between oil and Big Data is not limited to this. Big Data technologies and the oil and gas industry are not strangers to each other, and they are closely related. Interests of the oil and gas companies towards the Big Data are growing against the backdrop of the plummeting oil prices in the global energy market [1–3].

The performance of the oil and gas industry depends on the multiple technological and business processes generating large volumes of information flows [4, 5]. The digital data about the oil and gas fields, exploration, drilling and production processes, marketing of oil and oil products, and the status of the oil and gas companies in the market are continuously entering into the management centers of the companies. Decision-making on the hierarchical levels of the oil and gas industry is based on the results of the modeling of specific situations and processes on the basis of this data. Sophisticated modeling methods on the large scale data are used for the exploration and development of the oil and gas beds, investment decisions, and the development of oil and gas production [5, 6]. To reduce the duration of such a large-scale data processing, speed up geological and hydrodynamic modeling and perform more complex modeling, the oil and gas companies have to build High Performance Computing (HPC) systems [7]. Big Data presents approaches and technologies in this regard. In addition, Big Data technology has potentials to create new values [8].

Big Data technologies were previously reviewed through the prism of the three “V”s, which included [9]: Volume, Velocity and Variety. Although, at least two more “V”s are also required for a wide range of business applications of Big Data technologies [9]: Value and Veracity. In other words, Big Data technology is regarded as a real business tool for the storage and processing of poorly structured large volumes of data, which also generates income for the organization.

Various general statements are available on how to transform the data into a value with the help of Big Data technologies. The oil and gas industry is no exception [10–13]. In some cases, due to the lack of innovation and experience in the implementation of such projects, the importance of Big Data technologies might be over evaluated. In 2015, after the exclusion of Big Data out of “breakthrough technologies” from the Gartner analysts’ “HypeCycle” schedule, some authors even began to talk about the “Death of Big Data” [14].

However, Big Data technologies are the reality. Big Data technologies have a number of perspectives for the oil and gas industry. Although these technologies have only recently appeared, real effective experiences on their use already exist in various fields [15]. Big Data provides access to absolutely new quality knowledge and opportunities; it does not only give companies a

competitive advantage in the market, but also develops the industry as a whole by revealing its latent potential. Therefore, Big Data strategy should be developed taking into account the specific features of the specific company and specific business to steer Big Data technologies toward the implementation of more profitable business solutions.

The key feature of Big Data strategy for the oil and gas industry should be noted in advance. Nowadays, the development of hydrocarbon fields is associated with the implementation of large-scale complex projects, which require significant investments, the establishment of the diversified infrastructure and the development of advanced technologies. In order to gain a competitive advantage, oil and gas companies are developing the strategies for 10 years or longer term. According to the management of level, Big Data strategy refers to the functional (supportive) strategy [16, 17], and it has to take into consideration the overall corporate strategy and superior strategies in the hierarchical structure for the strategic directions. Moreover, it has to provide the implementation of the main objectives. In this regard, the use of Big Data technologies should not meet the needs of the scientific project, but focus on the business world. In terms of certain management, the formulation and implementation of Big Data strategy cover the decision-making on the priorities for the application of these technologies, the specification of clear and precise objectives, the creation of an appropriate structure and the implementation of investment programs, effective feedback and evaluation.

The current study analyzes the development issues of Big Data strategy for the oil and gas industry and proposes a number of certain recommendations for the formulation and implementation of this strategy.

A brief review of the oil and gas industry

Oil and gas industry –is the field of industry dealing with oil, gas and oil production, refinery, transportation, storage and sale [18]. Related areas of industry are geophysics, drilling and production of oil and gas.

The production processes of oil and gas industry are divided into the stages, such as exploration and production (*upstream*), preparation and transportation (*midstream*), processing and sale (*downstream*) [19].

The basis of the oil and gas industry is the vertically integrated oil companies. International and national oil companies are the leading players in this field [18]. The business ecosystem also includes oilfield-service companies, which perform a significant part of the work of the oil and gas industry; the innovation of the field dramatically depends on the technology (including information technology) applied by the contractors in many respects.

The development of computing and communication technologies has led to the emergence of new innovations in the oil and gas industry. Since 1980-90s, as a result of the widespread introduction of information technologies in the chain of production cycle, the oil and gas industry has been digitized and intellectualized. Remote monitoring, modeling and management systems are applied in the technological processes for the optimization and high security provision, and risk minimization [4, 19]. Consequently, the main part of the work is carried out a distance, decision accuracy and efficiency rises, effective interaction between the functional units is provided, the maximum oil production and reduction of operation and maintenance costs, becomes possible. Some authors often call this situation in the oil and gas industry as a “digital revolution” [20].

Most software innovations playing a key role in digitalization of oil fields are offered by the oil-service providers (Halliburton, Schlumberger) and large IT service providers (HP, Oracle, Microsoft, IBM) [21-24]. The leading global players in Big Data market in the oil and gas sector are HP, Hitachi Data Systems, IBM, Oracle, Cloudera, EMC, MapR Technologies and SAP.

Thus, the development of intellectual oil and gas complexes is necessary for high profit of oil and gas fields. Such complexes are characterized by high automation, unmanned production facilities, and the intelligent management systems of the technological, energy, transport and industrial processes.

The progress made in recent years in the fields of computing technology, the Internet of

Things, cloud computing, mobile communication technology, robotics, Big Data analytics, and the artificial intelligence brings new innovations to the oil and gas industry [25-27]. In the near future, the main target of the innovative development of oil and gas industry is to shift the oil and gas complex to the real-time management mode. This includes the development of high performance computing systems that realize collection of geological-mining and other data from the sensors at all the technological chain of oil and gas production and processing, and the management and analysis of large volumes of data [4].

Large-scale data of the oil and gas industry

Typically, large-scale data are received from three sources:

- sensors;
- corporate document archive;
- Internet (social networks, media, forums, blogs and other web-sites).

The oil and gas industry also has the same and rather diverse data sources. We will take a brief look at them.

The oil and gas industry is the largest worldwide consumer of the different types of sensor devices (including intelligent meters of gas supply network) [28]. The 4C (4-component) sensors in the oil and gas seismic exploration and 4D seismic sensors (4 dimension time, the seismic measurements sequence the mine specified by a certain time interval) in the field of geophysical surveys, as well as the optical fiber sensors in the wells, oil and gas extraction, processing and transportation systems are widely used; and they receive large scaled data. For example, the volume of the data obtained during the seismic exploration of a field can reach tens of terabytes [29].

A fiber optic sensor in a well measures temperature, pressure and other parameters at each 100 m, 10 m, 1 m, and even 10 cm [30]. The real time control the pipe's performance and its condition can be ensured through the use of sensors. In practice, the most commonly requested tool can be a pipe visualizer, the application of which can make the field workers' dream come true, i.e., they can view the status of the mechanical and physical characteristics of the pipe at each point throughout its entire life-cycle. A daily volume of data received from such sensors from a single well can reach several terabytes.

To control the oil and gas pipelines and determine the location and size of different defects, a large number of *Magnetic Flux leakage sensors (MFL)* are used in these pipelines. The sensors are placed at an equal distance along the bounds of the pipe and measure MFL signals at each 3 mm. As a result, the volume of collected data becomes extremely large [31, 32].

3D geological and hydrodynamic models of oil and gas fields also generate large amounts of synthetic data. Geological modeling of oil and gas fields shapes 3D model of a layer, on the basis of which the hydrocarbon reserves in the layer are estimated. The 3D hydrodynamic model developed according to the geological model shows the changes in the properties and volume of reserves in the layer of the field and the pace of oil (gas) extraction from the wells. Grounding on the information received from sensors installed in the well, the geological and hydrodynamic models are adapted. Various texts, tables and graphics reports are set out based on the geological and hydrodynamic models of oil and gas fields. As a result of the geological and hydrodynamic modeling of oil and gas fields, terabytes, even petabytes of seismic, geophysical and mining data are processed [26].

Another source of data is the digitization of the historical data on the geological and geophysical surveys conducted by the company. The data collected on all mines and wells (including drilling and production) of the company and the technical passports of the objects are digitalized according to the area standards and included to the centralized archive.

In addition to automate complex production processes of oil and gas companies, the management processes and business processes are also informatized on a large scale (SAP system projects for several companies). However, the data on the management processes are processed in the systems for budgeting, management reporting, efficiency coefficient system, consolidation of

financial reports, and the management of contracts, treasury, risks and human resources. The data on the operational processes are collected in the systems for projects, repairs, refining operations, procurement and order management, production planning, environmental protection, transportation management and logistics planning.

For obvious reasons, the oil and gas companies are in the focus of the media, commercial structures and authorities. Therefore, collection and analysis of the large volume of unstructured data received from outside sources (social networks, media, web-sites, e-mail, various reports, photos and multimedia) become very relevant. It should be noted that these data are unstructured or semi-structured. Consequently, their storage in the traditional data warehouse, regular access and analysis become very difficult.

The complexity of the organizational structure, the complex production processes scattered throughout wide areas of land and a diversity of the activities require the oil and gas companies to provide solutions that connect all the data necessary for the effective management of the enterprise in a unified information space. The traditional IT infrastructures, particularly the data storage systems infrastructures operating in different directions or sections in unlinked way are increasingly diversifying, and their service is getting complicated and the costs are increasing, their management is getting more difficult, and the limited capabilities of these systems do not provide the required economic efficiency of the modern oil and gas industry enterprises. This is one of the main reasons urging the use of Big Data technologies in the oil and gas industry.

A brief description of the Apache Hadoop ecosystem

Hadoop ecosystem is considered to be synonymous to Big Data technologies. Initially, Hadoop was a tool for the data storage in clusters and parallel data processing with MapReduce method, however, now it is a constantly growing large stack of related technologies in one or another way with large-scale data processing. This happens first and foremost due to the software programmers contributing to the open-source Apache Foundation Hadoop repository, which is the basis of numerous commercial products. Leading manufacturers offer different tools for large-scale data processing and analysis: EMC GreenPlum, IBM Netezza, Oracle Exadata, Teradata. Although, such tools are extremely expensive. Hence, at present, the demand for the components for the open source Apache Hadoop-based distributed computing is particularly high. Hadoop includes a number of core components [33]:

- **Hadoop Distributed File System (HDFS)** - distributed file system that enables unlimited data storage.
- **Hadoop YARN** (Yet Another Resource Negotiator) - platform for managing cluster resources and issues.
- **HadoopMapReduce** - program framework for distributed computing overlarge-scale data.
- **HadoopCommon** - collection of utilities and libraries used by other modules in Hadoop ecosystem. For example, HBase and Hive use Java archives (JAR files) stored in Hadoop Common modules in order to access HDFS.

Hadoop – a cluster based on the principle MPP (*Massive Parallel Processing*). NameNode stores and distributes the metadata; DataNode includes the data blocks, where the calculations are carried out. Besides the main components, Hadoop cluster may use many other components as follows:

- **Hive** - data storage infrastructure applied to access the large-scale data located at HDFS through SQL (HiveQL);
- **Pig** - top-level programming language and parallelizing environment for large-scale data analysis;
- **HBase** (HadoopDataBase) - distributed non-relational column data base (Google's BigTable);
- **ZooKeeper** - distributed configuration storage and configuration changes synchronization service;

- **Sqoop** - exchange system between Hadoop and structured data warehouse, such as relational database (DB);
- **Oozie** – performs the task planning for Hadoop cluster;
- **HCatalog** - set of interfaces that provides access to Hive warehouse metadata from external systems;
- **Ambari** – Hadoop cluster management and monitoring system that supports HDFS, MapReduce, Hive, HBase, HCatalog, ZooKeeper, Pig, Sqoop, Oozie;
- Mahout – machine learning application library on large-scale data that enables developing intelligent applications.

Apache Hadoop has the following advantages [34]:

Low price - Hadoop ecosystem is the product of open-source Apache project, it can be downloaded for free. At the same time, it works on cheap simple servers.

Linear scaling – power of the cluster is increased when the volume of data increases and processing speed needs to be increased, the program does not require any additional code.

High Denial resiliency – accidents and denials occurred in the nodes' performance do not pose any problems of the service delivery, since the data is stored and backed up in at least three nodes, if necessary, they can be accessed without notifying the user.

Tolerance to different types of data -Big Data is applied to the structured, quasi-structured and unstructured data.

In addition, *Plain Hadoop* has several disadvantages [34]:

- *configuration complexity* - the configuration of the system without sufficient qualification is very difficult, and finding Hadoop expert is another problem;
- *management shortage*—due to the lack of comfortable administration tools with the interface;
- *low security* – Hadoop was not initially designed to be installed in the complex multi-user corporate environment, therefore, the files are not protected. However, the traditional relational databases have sufficiently powerful and flexible security mechanisms;
- *lack of equipment optimization* – Hadoop does not take full advantage of the equipment resources.

To eliminate the listed shortcomings of Plain Hadoop the open-source projects and their commercial products are also available. In practice, one of the distributives of three young companies is applied most: Cloudera (founded in 2008), HortonWorks (founded in 2011) and MapR. Unlike the free versions, the commercial distributives include the cluster installation and management, as well as comfortable graphic tools for the monitoring of availability, loading and implementations. The cluster installation lasts only a few minutes. The distributives of the first two companies are shared free of charge: the exploitation and industrial deployment of the cluster is possible without any costs. The licensing fees are paid during the technical support.

It is difficult to make a right choice as there are enormous variety BD tools. It is particularly important when speeding up and optimizing the data access, large-scale data processing and movement. Open source tools are not always performance effective. The optimized solutions developed by the vendors accelerate the data transfer between different storage facilities and technologies to several times.

The potential of Big Data technologies

The promised potential of Big Data technologies can be grouped as follows [35, 36]:

1) Significant reduction in the cost of data storage and processing. The main business goal is to store all the data. The cost of data storage of 1 TB (terabytes) in the traditional systems is very high. Due to the use of cheaper equipment, the application of Big Data solutions may significantly reduce the storage costs of hundreds of terabytes or petabytes of data. The fall in the price of the data storage and processing opens up new opportunities for their analysis.

The main trend in the data storage systems is the transition from special equipment to SDS-systems (*Software Defined Storage*) [37]. At the same time, the hybrid solutions combining the data storage and processing functions are also becoming popular. Companies that endeavor to increase the speed of data processing without exceeding the constraints of the operational costs use rather the hybrid storage systems.

In general, Big Data predicts the transition from data storage systems with multiple controllers to distributed computing, Hadoop-clusters and other similar solutions.

2) Unstructured data integration enables to work with Hadoop data with the help of structured data processing tools (SQL etc.). As a result, the system architects are able to choose the most efficient location for the storage of these and other data (relational DBMS or Hadoop). Most Big Data are unstructured data such as text, video, audio, images, multimedia and so on. The organization of unstructured data processing and analysis is one of the complex research issues. Unstructured data mining is relatively young field of scientific research while Text Mining has been studied more [38].

3) High-performance scalable data processing. The software developed within Apache Hadoop technology enables to provide high performance distributed processing at all stages of data analysis using horizontal scaling, which is one of the main advantages of these systems. Available algorithms can effectively work even at the cluster with one server and be scaled when the volume of processed data is increased hundred times through the servers running in an emergency mode.

4) Data Mining. As a result of the increase in the performance of computing systems and the volume storage systems, as well as decline in the data storage and processing costs, the data storage and analysis make an economic sense. New approaches, new architectural solutions, software and hardware products are emerging.

Statistical analysis, data mining, machine learning, simulation models, optimization methods, data visualization, data aggregation and integration and other methods are used for Big Data analysis. Predictive analytics is distinguished as a separate field [36].

Evidently, the obtained data cannot be properly understood without the appropriate visualization. A widespread use of visualization tools, which enable easy comprehension of all the data necessary for operative management decisions, is possible due to the Big Data technologies [37].

5) Easy implementation with cloud services. Big Data and cloud are developing in parallel and mutually complement each other. Large clusters of servers are required for large-scale data processing, and the clouds provide such clusters. Cloud technology provides the flexibility of Big Data and reduces the costs. Therefore, interests in Big Data cloud services are growing [39, 40].

Software and hardware sets can be rented in the cloud for Big Data. *IaaS (Infrastructure as a Service)* or *PaaS (Platform as a Service)* can be selected. *IaaS* option opens up great opportunities for the choice of service providers, global or private clouds, however, additional responsibilities related to the software installation and cluster management arise. *PaaS* option avoids a number of service burdens, although its possibilities are limited to the functions offered by a provider. The transmission of large-scale data in cloud can cause problems. The easiest approach is to collect and analyze data directly in the cloud. Solutions for data loading to clouds are also available – they periodically collect the data from various sources in accordance with the specified rules, and place them in the given locations, including the clouds.

In addition, there are solutions that read the log-files of databases, which facilitates obtaining information about all the changes taking place in the databases. In the case of any changes in the database, they can make the same changes in the database in the cloud, as a result, local and cloud databases will be synchronized in real time.

The status of implementation of Big Data in the oil and gas industry

Some companies in the oil and gas industry attempt to use Big Data technologies to solve their problems [12].

The concept of Big Data can be applied in various functional areas of corporate management such as production, logistics, and marketing. In this vein, Big Data can be used in management – from real-time analysis of business parameters of workshop equipment to the development of information foundations for the strategic decisions about business development.

The use of Big Data Analytics in processing, logistics and sales (*Downstream*) is predicted to cause a breakthrough. Big Data can contribute to predicting the demand for oil products in the retail sales network and to analyze the pricing and the changes in prices by the competitors and the regions [41]. 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.

Big Data enables the analysis of hydrocarbon deposits, the detection of its non-optimal areas, the selection of production program and the prediction of the outcome [42].

Engineers have the experience in working with surrogate models processing and therefore, they are well aware of the advantage of the use of analytic functions in addition to the physical models that use machine learning approaches. In terms of approaches to data, the drilling and production remind the manufacturing, i.e., the engineers understand the importance of reliability forecasting, surface and underground work modeling based on the collected data.

In particular, the automation of the management of the drilling process requires an in-depth analysis of data in a close mode to the real time. There are very large amounts of data here and the predictive analytics based on them is important. Therefore, the study of the use of Big Data technologies in this issue is exceptional.

Big Data technology can also be useful in the process of production. The system monitors the drilling process and records gas leaks, water passage, pressure change, and any other changes resulting from the seismic activity. With the help of this information, engineers are able to envisage and control the incidents, which can damage equipment [43].

Analysis shows that the use of Big Data analytics in the oil and gas industry is still at an experimental level [12]. Only a few companies have been trying to use Big Data technology [44]:

- The Chevron uses Hadoop for seismic data processing (IBM BigInsights); currently, the daily internal traffic generated by Chevron accounts for more than 1.5 terabytes.
- The Shell implemented a pilot project on the use of Hadoop for seismic sensor data; Fiber-optic cable sensors are placed inside the well to measure the data. The collected data analyzes how the wells work and how much oil/gas remains. Large volumes of data from sensors are located in Amazon Virtual Private Cloud (Amazon VPC). Shell is planning to deploy fiber-optic cables inside the 10 thousand oil wells.
- Seismic Cloudera Hadoop project combines Apache Hadoop and Seismic Unix [45];
- Point Cross Seismic Data Server and Drilling Data Server are using Hadoop and NoSQL;
- The University of Stavanger has studied the effectiveness of data collection by using Hadoop [46].
- “Газпромнефть” analyzed the operation of the well fund using Teradata Aster system within the framework of Big Data pilot project completed in 2015 [47]. The analysis used about 200 million logs recorded in 1649 wells in 2014 and the reconnection logs of the electrical voltage taken from accident journals. These are the huge amounts of data, which is not possible to be analyzed manually or by traditional tools. The results of the analysis have very clearly showed a new quality obtained with the help of Big Data. As a result of the application of Big Data tools, the previously unknown mutual dependencies occurred in the pumping equipment have been identified, for example, in some cases, the turbine rotation effect - oil flew back when the power supply of the pump has been identified.

Despite these examples, the vendor solutions in the area of Big Data for the oil and gas industry are still at the early stage. 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.

It should be noted that this situation is not specific only to the oil and gas industry. The integration of Big Data with traditional IT systems is one of the questions in other areas. The author of the work [48] points out the fascinating outcome of the survey about the practice of Big Data usage at the companies listed at Fortune 500. Contrary to the expectations of the author about the infrastructures and complex analytics applications, most companies have small clusters of Big Data type comprising only 8-10 nodes. IT company managers call the emergence process of such clusters as "analytics extension". Big Data infrastructure was built in isolation from the traditional IT systems and the data were moved for analysis from the main warehouse to HDFS, and then, moved back to the main warehouse from HDFS. Such isolation leads to an increase in operating costs and an inefficient waste of resources.

According to the interpretation, the massive advent of Big Data in the oil and gas industry may not occur. Even Big Data projects in the oil and gas industry may not be developed, however, IT tools using Big Data approaches may arise, and geologists, oilmen and IT specialists should be aware of them.

The challenges in the data use in the oil and gas industry

Data is the most valuable asset and its proper management directly impacts the business efficiency and the success and risk of new projects. To establishing Big Data strategy in the oil and gas industry, it is necessary to take into account the existence of the following problems in the field of data usage:

- Data acquisition in the field of investigation and production costs great amount of money (\$ 10 billion per year), however, the resulting data is not handled to be properly maintained [49].
- The oil and gas industry does not treat the data as an asset, but as the data that describes the assets. Big Data leaders in the field - social networking providers, handle the data as the underlying assets [50]. The data should be included as an asset in the business concept of the organization in the oil and gas industry; which data is of value should be clarified, and the holistic approach to the data assessment should be applied. As a result, the value of the data will be clarified for the company, and the data management will be paid greater attention.
- Most companies do not have a structured approach to the data management. It is proven by many different approaches applied within a company.
- Currently, a huge archive of old unprocessed data is available and they can be converted into more valuable knowledge with the help of Big Data technologies. It is required to find new ways of conducting business transactions based on the collection and analysis of large-scale data [51].
- Functional approach prevails the systematic usage of data now. There are huge amounts of data in the oil and gas industry, although they are at disposal of different legal entities. The importance and difficulty of the data subject areas (marketing, finance, manufacturing and etc.) extracted from "Silage pits" into the common corporate space have rigorously been discussed, however, no practical results have been achieved yet.
- Most digital tools and devices have not been integrated into a single system.
- Capital-intensive new projects involve the collection of huge amounts of mining data, nevertheless, integrative approaches to the use of collected data are not available.
- Smart tools are on demand, but the opportunities to improve the production processes have not been thoroughly studied.

The main trends of Big Data strategy

Before describing the main trends, several useful principles of an effective strategy for Big Data should be noted:

- *Completeness*—for long-term success, the companies need to view a complete picture covering different components of an effective Big Data ecosystem (including a well thought-out architecture and secure Big Data tools), and various aspects that could benefit Big Data;
- *Business focus* - Big Data strategy needs to be related to the specific business problems;
- *Flexibility* - the state-of-the-art and future use of Big Data needs also be considered; the strategy shall avoid common restrictions such as the excessive dependence on several technologies and partners;
- *Evolutionary approach*—big changes promised by Big Data will not happen immediately; the strategy needs to be accounted for the creation of value and the gradual evolutionary process.

In general, as a result of strategic planning processes, the general trends providing the achievement shall be set. Strategic management processes involve: 1) analysis; 2) strategy formulation; 3) imposing the goals; 4) realization and 5) feedback processes. Stating from this methodology, the key trends in Big Data strategy for the oil and gas industry are defined as follows:

1.1. Raising awareness in the field of Big Data. Oil and gas companies have to understand the notion of “Big Data”, its potential applications and clarify which technological updates are required. The main issue here is to clear out which data is of value and to analyze them in order to gain an advantage in the market and to maintain the competitiveness.

1.2. The evaluation of readiness to Big Data. Before the start of Big Data initiative, oil and gas companies shall comprehend their own IT environments. The organization should assess the level of Big Data maturity. They need to identify business goals and objectives, evaluate the potential of Big Data and determine the further steps to be taken by the company.

2.1. Creating Big Data business environment. To discover the full potential of Big Data technologies, the business processes have to be improved and a certain business environment to be created. For many companies this means forming data management and applying more advanced processes for their management, and obtaining new institutional opportunities supporting the management of ever-growing data stream.

To get feasible benefits, the implementation of Big Data projects should be invested, besides, the management approaches need to be changed. Top-managers are not able to deal with a large number of details, therefore, they often oblige low levels to make decisions being forced to give the employees more freedom.

2.2. Big Data application risks. The expectations and risks of the profitability of the investments on Big Data technologies shall be analyzed (SWOT-analysis (*Strengths, Weaknesses, Opportunities, Threats*)).

3.1. Identifying the business objectives of the use of Big Data. The success of Big Data project depends on the positive balance between the projects costs and the business benefits of the company obtained as a result of the project. It is important to clear out at what expense the organization gets the revenue. Unfortunately, this task is often difficult in the field of big data. Objectives should appropriately be set, and benefits obtained from the use of Big Data should be calculated.

3.2. Unified architecture. The corporate Big Data strategy should be based on the unified architecture of the data management platforms. This architecture should cover data management, analysis and search technology. This will solve one of the main problems of Big Data projects - ensuring data collection and quality and using more complex architectures of data sources.

3.3. Consolidation of the unstructured information. One of the most effective Big Data methods is to find new patterns and correlations, through combining and coordinating different data - internal and external data, data from various corporate functions.

4.1. Pilot project on Big Data. The pilot project on Big Data has to be a part of the company's corporate strategy. Accompany follows the trends of Big Data when it intends to use the information for analysis and forecasting. It is required to start from the area with the most economic significance. Upon the completion of the project in this application level, the further implementation of the project is explored and reviewed.

4.2. Scaling—for the full implementation of Big Data strategy and not obtaining one more “data silo”, the pilot project needs to be developed on a larger scale. Successful step forward is possible due to the strong and adapted Big Data ecosystem.

4.3. Human resources in Big Data. Another most important factor is a human. The people “infected” with the idea of the organizing similar projects should be revealed [52].

4.4. Support for the development of the ICT industry. The oil and gas industry is the leading economic field of Azerbaijan, therefore, reducing the dependence on oil and developing other fields of economy, including ICT sector are of the priorities of the state policy. Major oil and service companies try to provide their monopolistic status in the field of Big Data technologies for the oil and gas industry. On the other hand, favorable conditions should be created for small local software producers in Big Data. They would prevail their monopolist partners in terms of price and could provide a sufficient level of quality of the field in the current conditions.

4.5. Ensuring information security. Transition of the oil and gas companies to Big Data technologies raises serious information security threats, which require these systems to be protection from cyber-attacks. The cyber-attacks to the oil and gas industry facilities are quite diverse; they may be targeted at data, production operations or management systems, the consequences of which may be quite devastating. For example, in August 2012, Saudi Aramco was attacked with computer virus that infected 30 thousand working stations. IDC Energy Insights' research shows that the oil and gas companies lag behind other sectors in the field for the development, approval and implementation of information security policy.

5.1. Strategy implementation control and feedback. Due to some uncertainties, the implementation of Big Data strategy can also yield negative effects. The control points in the interim period of the strategy term should be defined. The question here is not to control the budget in the practice of traditional production. Self-affirmation of the costs in these control points should be evaluated and the necessary corrections have to be made.

Conclusion

The modernization of the economy requires the introduction of new innovative technologies in the oil and gas industry too. 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.

In this case study, the potential and application experience of Big Data technologies in the oil and gas industry were analyzed, the current problems in the data management in the oil and gas industry were examined, and the general trends in Big Data strategy for the oil and gas industry were proposed.

References

1. Baaziz A., Quoniam L. How to use Big Data technologies to optimize operations in upstream petroleum industry // International Journal of Innovation, 2013, vol.1, no.1, pp.19–25.
2. Hajirahimova M.S. Opportunities and challenges big data in oil and gas industry /National Supercomputer Forum (NSCF-2015), Russia, Pereslavl-Zalessky, 24-27 November, 2015.

3. Alguliyev R.M., Imamverdiyev Y.N., Abdullayeva F.J. Study of the opportunities of implementing Big Data analytics for the oil and gas industry on cloud computing platform as analytics-as-a-service // *Problems of Information Technologies*, 2016, No 1, pp.11-26.
4. Speight J.G. *Handbook of offshore oil and gas operations*, 2015, 444 p.
5. Hilyard J. *The oil & gas industry: A nontechnical guide*. 2012, 322 p.
6. Sasson A., Blomgren A. *Knowledge Based Oil and Gas Industry*. Research Report - 03/2011. BI Norwegian Business School. March 2011, 5 p.
7. Taneja P., Wate P. Big data enabled digital oil field // *CSI Communications*, 2013, vol.37, no.1, pp.18–20.
8. Gasimova R.T. Big Data analytics: current approaches, problems and solutions // *Problems of Information Technologies*, 2016, No1, pp.75-93.
9. Alguliyev R.M., Hajirahimova M.S. “Big Data” phenomenon: Challenges and Opportunities // *Problems of Information Technologies*, 2014, No2, pp.3-16.
10. *Open Oil: Exploring oil data: a reporter’s handbook*, 2012, 100 p.
11. Brulé M. Tapping the power of Big Data for the oil and gas industry. IBM Software White Paper for Petroleum Industry, May 2013, 8 p.
12. Febowitz J. The Big Deal about Big Data in upstream oil and gas. Paper & Presentation, IDC Energy Insights, October 2012, 11 p.
13. Sangvai P. Impact of Big Data in oil and gas industry /Proc. of the 10th Biennial International Conference & Exposition, 2013, pp.439–440.
14. Heudecker N., Beyer M. A., Edjlali R. The demise of Big Data, its lessons and the state of things to come. Gartner Technical Report, 2015.
www.gartner.com/doc/3115022/demise-bigdata-lessons-state.
15. Thomson D. Big Data is dead, long live Big Data / *The Cyber Defense Review*, 2016.
www.cyberdefensereview.org/2016/03/10/big-data-is-dead/
16. Nag R., Hambrick D. C., Chen M.-J. What is strategic management, really? Inductive derivation of a consensus definition of the field // *Strategic Management Journal*, 2007, vol.28, no.9, pp.935–955.
17. Hitt M. A., Ireland R.D., Hoskisson R. E. *Strategic management: Concepts: Competitiveness and Globalization* (11th Edition). Cengage Learning. 2014, 480 p.
18. Inkpen A.C., Moffett M. H. *The global oil & gas industry: Management, strategy & finance*. 2011, 455 p.
19. Devold H. *Oil and gas production handbook* (Edition 3.0). Oslo: ABB, 2013, 162 p.
20. Barber A., Shippen M. E., Barua S., Velázquez J.C., et al. Optimizing production from reservoir to process plant // *Oilfield Review*, Winter 2007/2008, pp.12–29.
21. Hollingsworth J. Big Data for oil & gas. Oracle Oil & Gas Industry Business Unit. March 2013, 24 p.
22. Febowitz J., Rice L., Beals B., Andersson B. Big Data in oil and gas: How to tap its full potential. Hitachi Data Systems Corporation, 2013, 45 p.
23. Hems A., Soofi A., Perez E. Drilling for new business value - How innovative oil and gas companies are using big data to outmaneuver the competition. Microsoft White Paper, May 2013, 13 p.
24. Seshadri M. Big Data science challenging the oil industry, 2012, 32 p.
25. *The Internet of Things: Mapping The Value Beyond The Hype*. McKinsey Global Institute, McKinsey & Company, June 2015, www.mckinsey.com/mgi
26. Kamal S. Z., Williams J., Liddle J. *Continuous Improvement of Assets Through Existing and New Digital Oilfield Technology*. Society of Petroleum Engineers. 2014.
<http://dx.doi.org/10.2118/167908-MS>.
27. Bravo C. E., Saputelli L., Rivas F., Perez A. G., et al. State of the Art of Artificial Intelligence and Predictive Analytics in the E&P Industry: A Technology Survey. Society of Petroleum Engineers, 2013. <http://dx.doi.org/10.2118/150314-PA>.
28. Onajite E. *Seismic Data Analysis Techniques in Hydrocarbon Exploration*. Elsevier Inc., 2014, 256 p.
29. Hyne N. *Dictionary of Petroleum Exploration, Drilling & Production*. 2nd Edition. 2014, 625 p.

30. Zhang M., Ma X., Wang L., Lai Sh., Zhou H., Zhao H., Liao Y. Progress of optical fiber sensors and its application in harsh environment // *Photonic Sensors*, 2011, vol.1, no.1, pp.84–89.
31. Shi Y., Zhang C., Li R., Cai M., Jia G. Theory and Application of Magnetic Flux Leakage Pipeline Detection // *Sensors*, 2015, 15, 31036–31055; doi:10.3390/s151229845.
32. Mohamed A., Hamdi M. S., Tahar S. A machine learning approach for Big Data in oil and gas pipelines /Proc. of the 3rd International Conference on Future Internet of Things and Cloud (FiCloud), 2015, pp.585–590.
33. Imamverdiyev Y. N. Great potential and challenges of Big Data technologies // *Problems of Information Society*, 2016, No1, pp.23–34.
34. White T. Hadoop: The definitive guide. O'Reilly Media, Inc., 2012, 688 p.
35. Holdaway K. Harness oil and gas Big Data with analytics: Optimize exploration and production with data driven models. Wiley, 2014, 384 p.
36. Gandomi A., Haider M. Beyond the hype: Big data concepts, methods, and analytics // *International Journal of Information Management*, 2015, vol.35, no.2, pp.137–144.
37. Xia W., Wen Y., Foh C. H., Niyato D., & Xie H. A survey on software-defined networking // *IEEE Communications Surveys & Tutorials*, 2015, vol.17, no.1, pp.27–51.
38. Steed C.A., Ricciuto D.M., Shipman G., Smith B., Thornton P.E., Wang D., Williams D.N. Big data visual analytics for exploratory earth system simulation analysis // *Computers and Geosciences*, 2013, vol.61, pp.71–82.
39. Perrons R. K., Hems A. Cloud computing in the upstream oil & gas industry: A proposed way forward // *Energy Policy*, 2013, vol.56, pp.732–737.
40. Febowitz J. Oil and gas: Into the cloud? // *Journal of Petroleum Technology*, 2011, vol.63, no.5, pp.32–33.
41. Li X., Ma J., Wang S., Zhang X. How does Google search affect trader positions and crude oil prices? // *Economic Modelling*, 2015, vol.49, pp.162–171.
42. Wisniewski W., Danner G.E. Optimizing big data in the oil field, 2014. www.epmag.com/optimizing-big-data-oil-field-709906
43. Tan K.H., Ortiz-Gallardo V.G., Perron R.K. Using Big Data to manage safety-related risk in the upstream oil & gas industry: A research agenda // *Energy Exploration & Exploitation*, 2016, vol.34, no.2, pp.282–289.
44. Nicholson R. Big Data in the oil & gas industry. IDC Energy Insights, September 2012, .
45. Wills J. Seismic data science: Reflection seismology and hadoop.2012. <http://blog.cloudera.com/blog/2012/01/seismic-data-science-hadoop-use-case/>
46. Lee J.-G., Kang M. Geospatial Big Data: Challenges and opportunities // *Big Data Research*, 2015, vol.2, no.2, pp.74–81.
47. Korolkov M. In the stream // *Siberian oil*, 2015, No9 (126), pp.36–39.
48. Warfield A., Hadoop and Big Data storage: The challenge of overcoming the science project. Data Center Knowledge, May 2015. www.datacenterknowledge.com/archives/2015/05/26/hadoop-big-data-storage-challenge-overcoming-science-project/
49. Perrons R.K., Jensen J. The unfinished revolution: what is missing from the E and P industry's move to “Big Data” // *Journal of Petroleum Technology*, 2014, vol.66, no.5, pp.20–22.
50. Irving D. 4 reasons oil & gas companies are going to fail in a Big Data world. November 11th, 2015. <http://blogs.teradata.com/international/4-reasons-oil-gas-companies-are-going-to-fail-in-a-big-data-world/>
51. Perrons R., Jensen J. Data as an asset: What the upstream oil & gas industry can learn about “Big Data” from companies like Facebook // *Energy Policy*, 2015, vol.81, pp.117–121.
52. Sicular S. Data Scientist – Mystified. June 29, 2012. <http://blogs.gartner.com/svetlana-sicular/data-scientist-mystified/>