IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 13, 2020, accepted March 3, 2020, date of publication March 9, 2020, date of current version April 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2979678

A Systematic Review of Big Data Analytics for Oil and Gas Industry 4.0

TRUNG NGUYEN[1], RAYMOND G. GOSINE[1,2], AND PETER WARRIAN[2]
[1]Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada
[2]Innovation Policy Lab, Munk School of Global Affairs and Public Policy, University of Toronto, Toronto, ON M4P 1A6, Canada

Corresponding author: Trung Nguyen (tn0432@mun.ca)

This work was supported in part by Petroleum Research Newfoundland and Labrador (PRNL), Atlantic Canada Opportunities Agency (ACOA), Mitacs, University of Toronto's Munk School of Global Affairs and Public Policy, and Memorial University of Newfoundland.

ABSTRACT Big data (BD) analytics is one of the critical components in the digitalization of the oil and gas (O&G) industry. Its focus is managing and processing a high volume of data to improve operational efficiency, enhance decision making and mitigate risks in the workplace. Enhanced processing of seismic data also provides the industry with a better understanding of BD applications. However, the industry still exercises caution in adopting new technologies. The slow pace of technology adoption can be attributed to various causes, from the obstacles to the integration with existing systems, to cybersecurity for defending the BD system against cyber attacks. In some applications using wearable devices, physiological and location-tracking data also causes concerns related to workplace privacy implications. These shortcomings give rise to uncertainties about the practical benefits and effectiveness of applying BD in O&G activities. The objective of this paper is to perform a systematic review of BD analytics within the context of the O&G industry. This paper attempts to evaluate technical and nontechnical factors affecting the adoption of BD technologies. The study includes BD development platforms, network architecture, data privacy implications, cybersecurity, and the opportunities and challenges of adopting BD technologies in the O&G industry.

INDEX TERMS Big data analytics, O&G digitalization, Industry 4.0, data privacy and security.

I. INTRODUCTION

In 2019, BP's Energy Outlook predicted oil demand will continue to rise through the next 20 years, which may result in a big gap in the petroleum supply. The outlook emphasizes the role of improving the efficiency of the production chain through innovation to meet the increasing demand [1]. Many reports have mentioned the role of digital technology to achieve higher value in the industry [2], [3]. Notably, McKinsey & Company indicated that the typical offshore platform only operates at 77% of its maximum production potential. This shortfall amounts to $200 billion US dollars in annual industry-wide revenue, which necessitates embracing advanced digital technologies to improve operational efficiency [3]. Digitalization also includes the rigorous use of big data (BD) analytics to bridge the performance gap in oil and gas (O&G) operations, which is attributed to a prolonged period of low oil prices and the inefficiency of existing systems.

The associate editor coordinating the review of this manuscript and approving it for publication was Vlad Diaconita.

The continuous advancements of data-gathering techniques, the advent of smart wearables and the increasing popularity of the Internet of Things (IoT) devices in the field have significantly increased the amount of relevant data [4]. The improvements in pervasive computing devices, advanced storage capabilities and the new generation of wireless networks have been opening more possibilities for real-time applications, such as remote monitoring of oilfields, illustrated in Fig.1 [2], [3], [5]–[7]. Embracing advanced BD analytic tools can maximize the production potential of a company's assets as well as fill performance gaps [3], [8], [9]. In terms of risk management, it also contributes to the mitigation of the infectious disease outbreak, such as COVID-19 [10]–[12]. Transparency Market Research predicted that the value of the global O&G market with BD could reach $10,935 billion US dollars by 2026 [13].

Unlike other industries, dealing with large quantities of data is not a new issue for the O&G sector. For many years, petroleum companies have invested in software for seismic data processing and visualization, which helps them

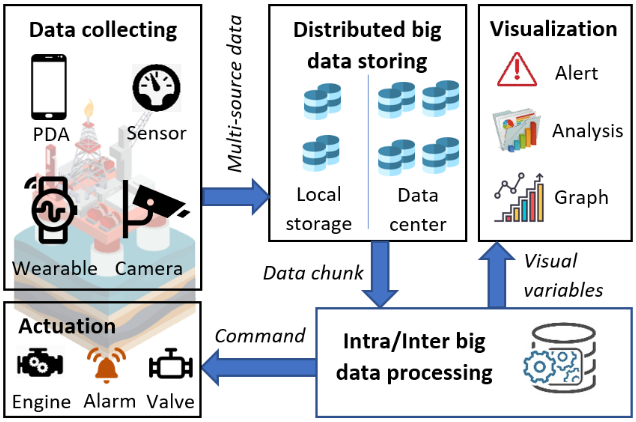

FIGURE 1. Concept of BD application for smart oilfield.

understand what lies below the earth's surface and plan how to extract it. This paves the way for the adoption of BD technologies in the O&G sector at a faster pace with more confidence. However, the digital transformation and promotion of BD technologies in the O&G industry lags behind other sectors. The connected services and networks associated with BD deployment also raises questions related to cybersecurity and data privacy [14], [15]. The legacy asset base of an O&G operation is not built for cybersecurity, while a lack of monitoring tools in the existing networks makes the system vulnerable to cyber attacks [16]. Analyzing these issues helps to develop a better understanding of BD benefits and foster technology adoption.

In this paper, we conduct a systematic literature review of BD analytics in the context of the O&G sector. This paper aims to develop an understanding of technical and nontechnical factors affecting the BD technology adoption. The focus of this paper is to evaluate the readiness of BD technologies to support O&G operations. The following points summarize the key contributions of this paper:

- This paper analyzes the opportunities and challenges of BD analytics for the O&G industry. The analysis utilizes evidence associated with recent BD deployments in O&G to demonstrate the discussion points.
- Some standard tools and platforms for BD deployments, including software database management and network architecture, are reviewed. The review aims to evaluate the readiness of BD technologies to apply in O&G applications.
- Some critical concerns about data privacy and cybersecurity, which limit the BD deployment, are examined. The discussion focuses on the weaknesses of the current regulatory framework and organizational management in technology adoption.

The remainder of the paper is organized as follows. The next section introduces the methodology to conduct this systematic review. Section III describes the specific properties of BD and opportunities to apply it in the O&G sector. Section IV presents review findings from the filtered articles. Section V summarizes standard tools and platforms to develop BD applications. Regarding the importance of data protection, section VI and VII explore data privacy and cybersecurity issues. Finally, lessons and conclusions are presented in section VIII.

II. METHODOLOGY

The earliest records of applying data analytics were in Mesopotamia about 7000 years ago where it was used to monitor and control the cultivation of crops and herds. The principle continued to grow over centuries with more applications and contributions to human history. In 2005, the term "big data" was coined by Roger Magoulas from O'Reilly media [17]. The literature on big data is vast, with an increasing number of research and development projects across multiple industries. The scope of the literature review described in this paper is the O&G sector. We apply the method of an academic and general search outlined in Table 1 to investigate the existing literature. The initial keywords-based search results in 2150+ articles. Some of these articles may not discuss BD research or implementation, but may only make a passing reference to BD. We identify and remove these articles from our article list using Voyant Tools [18]. Then, the resulting list of articles is manually reviewed, using the paper's abstract, introduction and conclusion. This manual filtering resulted in 115 articles, which were reviewed comprehensively to answer the following questions:

- RQ1: What is big data in Industry 4.0?
- RQ2: What are the publication patterns of BD related to the O&G industry?
- RQ3: What are the key applications of BD in the O&G sector?
- RQ4: What are the key challenges for adopting BD technologies by the O&G industry?

TABLE 1. Article searching procedure description.

Searching Index	Specific Content
Article type	Publications in books, journals, and conferences
Database	OnePetro, IEEE Xplore, Springer, Elsevier, Springer
Keywords	Big data analytics, cloud computing, edge computing, fog computing, digitalization, oil and gas 4.0
Classification	By the type of publication (i.e. concept, case-study and review), nationalities, application segments, enabling technologies, and affiliations (i.e. universities and industries)
Focus	Determine opportunities and challenges related to BD analytics in the context of O&G

For RQ2, we extracted metadata, such as the title, year, affiliation, type of publication, and nationality for further analysis, which is presented in section IV. Also, we classify the 115 papers into three groups:

- Concept paper: is an article presenting a theoretical and simulation-based study to develop new solutions for any specific problem. In some cases, the proposed concept is further validated by real-time experiments or field deployment.

- Case-study paper: is an article studying the industrial implementation of technologies, product development and field tests.
- Review paper: is an article describing the technical/non-technical review of the existing literature and surveys.

Some articles describe both theoretical concepts and field deployment. We classify these articles in both the concept and case-study paper group with similar weights.

III. BIG DATA OVERVIEW

A. DEFINITIONS

Big data refers to voluminous sets of data. The data size usually reaches Petabyte (=1024 Terabyte) or Exabyte (=1024 Petabyte) [4], [19]. Data analysts attempt to extract meaning and insights from raw data that will be useful for decision making in different applications in industry. In June 2018, Equinor initially disclosed all subsurface and operating data from a field on the Norwegian continental shelf to support learning, innovation and new solutions for future energy extraction [15]. The complete set of data consists of approximately 40000 files for the production from the Volve field from 2008 to 2016. The size of the entire dataset is about 4,206 Gigabytes, including data of geophysical interpretations, the GeoScience OW Archive, seismic, well log, production, reservoir models, and real-time drilling data.

Many studies attempt to define characteristics of BD, including the procedure for collecting data and analyzing it to extract valuable information. IBM focuses on three Vs: volume, velocity and variety for their applications. Some studies add veracity to highlight the trustworthiness of the data, termed the four Vs. Value is also considered because of the cost-and-benefit relation in data collection and analysis, titled the five Vs. Other characteristics of big data are also mentioned in Fig.2 [4].

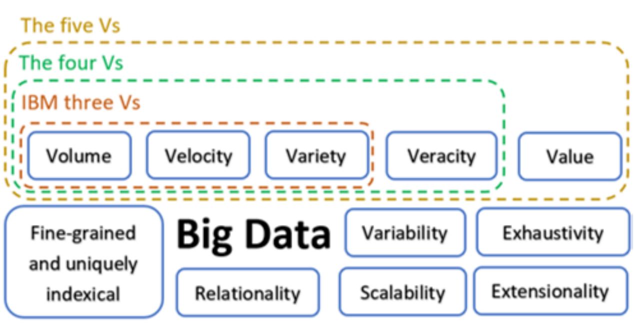

FIGURE 2. Characteristics of big data.

B. OPPORTUNITIES IN THE O&G SECTOR

Many studies have indicated the benefits of BD deployment in the O&G industry. Table 2 summarizes some notable applications of BD in the O&G supply chain. In upstream O&G, the recent improvements of seismic devices, channel counting, fluid front monitoring geophones, logging-while-drilling and measurement-while-drilling tools have boosted the amount of data significantly. BD analytics become an effective solution to manage and analyze these data. For example, Joshi *et al.* replaced conventional tools with the BD technique and Hadoop platform to analyze the massive seismic datasets, identify critical geological features, characterize the reservoir, and define geological issues [20].

In midstream O&G, BD reduces the cost and carbon dioxide emissions of ship transportation. It also supports the monitoring and maintenance of pipeline systems. By comparing the current data with other data sources (e.g. the historical data, maintenance reports, operator data, etc.), the analytic system can detect and localize a leak in a pipeline [44], [45], [59]. This strategy is also applied for equipment maintenance in other segments. Data analysis can provide the performance trends of equipment and forecast hazardous events. These advanced functions are beneficial for the maintenance scheduling, risk mitigation and safety management during O&G operations. In some cases, BD can optimize the large-scale management of O&G assets. For example, Repsol collaborated with Google to launch a BD AI project for the Tarragona integrated refining complex in Eastern Spain [49]. The target of this project is to maximize operational efficiency, including energy consumption and consumption of

TABLE 2. Some potential BD applications in O&G.

Upstream	Exploration & Scouting	Analyzing the seismic data [21], [22] Micro-seismic data [20] 1D, 2D, and 3D geological maps [23]
	Drilling	Efficiency in drilling rig [24] Drilling performance [25], [26]
	Reservoir Engineering	Reservoir management [27] Closed-loop reservoir management and integrated asset modelling [28] CO2 sequestration [29] Reservoir modelling [30]–[32] Enhanced oil recovery [33]
	Production Engineering	Automated decline analysis [34] Production allocation technique [35] Electric submersible pumping [36], [37] Rod pump wells [38] Hydraulic fracturing projects [38] Field development [39] Hazard event forecast [40] Well casing damage prediction [41] Remote smart oil field management [5]–[7] Flaring event management [42]
Midstream	Pipeline Transport	Pipeline monitoring and maintenance [43]–[45]
	Ship Transport	Shipping performance [46] Energy efficiency [47]
Downstream	Refining	Petroleum asset management [48] Comprehensive refinery [49] Well productivity [50], [51]
	Health & Safety Executive	Occupational safety [52], [53] Safety predictive analytics [54] Infectious disease mitigation [10]–[12], [55]
	Trade & Sales	Analyzing market volatility [56] Forecasting crude oil prices [10], [57], [58]

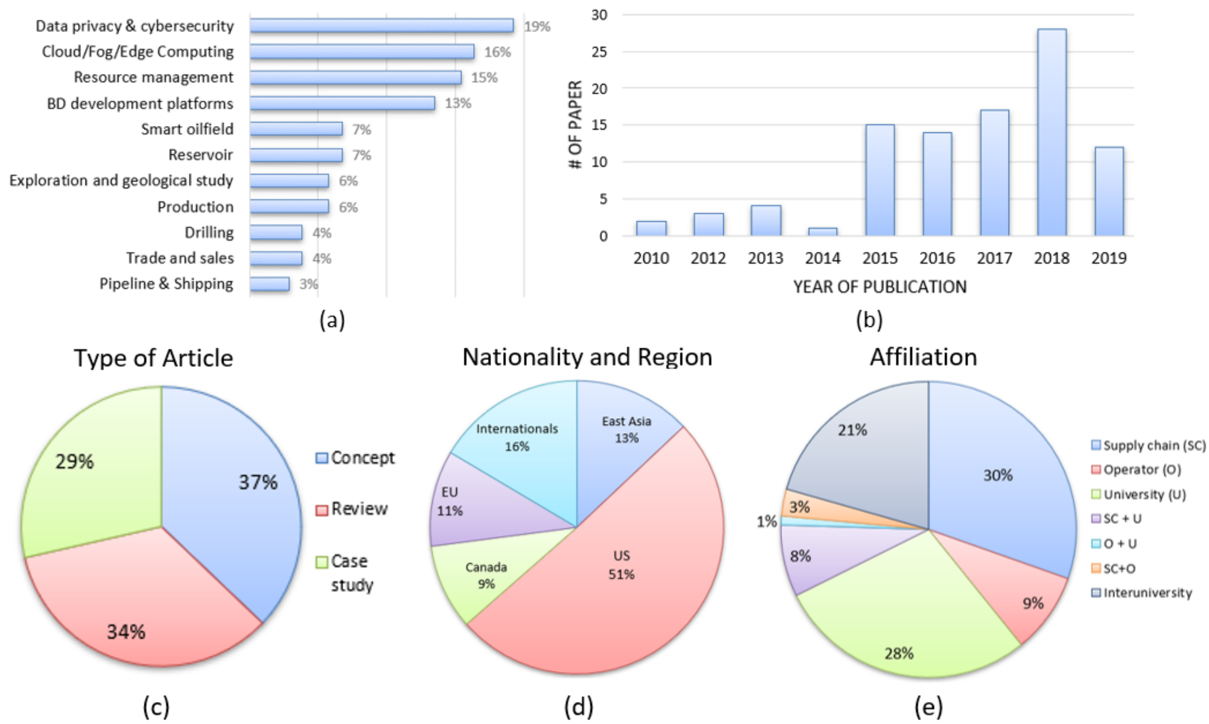

FIGURE 3. The distribution of scientific articles to conduct the literature review in this paper.

other resources. It also attempts to increase unit reliability and improve the economic performance of the refinery.

Many researchers have utilized the analytical capability of BD in O&G trade and sales [57], [58]. BD analytics has been used to understand the volatility of the financial market and even forecast the price trends of various energy sources. The crude oil price significantly affects the economic climate, politics and financial markets at the macro level. The level of expertise, the identification of critical parameters and the complexity of problem formulation are other concerns in this implementation [58].

C. CHALLENGES

The technology's adoption demands considerable capital investments and intensive efforts at every level of the organization, legal system and government. Although the O&G sector has experience in processing massive amounts of data, integrating BD analytics with existing systems raises many technical and nontechnical challenges.

Technical challenges include how to deploy BD technologies effectively using available software tools and hardware computing platforms. At present, there is no perfect model of BD employment, which guarantees a high profit improvement with the constraints of time and budget. More technical discussions about BD deployments can be found in section V. The BD system operation also raises many issues of functionality, cybersecurity and maintenance. Further study of cybersecurity is presented in section VII.

Within an organization, the deployment should also consider the ethical dimension to satisfy the competency and ethical standards of individuals and organizational management [60], [61]. This helps to prevent incidents that affect human lives, the environment, and the credibility of the O&G industry [61]. Nontechnical challenges also include collaboration between departments to deploy and operate the BD system effectively. Any issue, such as password resetting to access personal accounts of field workers, can be time-consuming because of the lack of communication between the IT department and others. At a higher level of innovation management, the implementation also gives rise to concerns, such as standardization, data privacy, data ownership and intellectual property rights [62]. The development of legislation cannot keep up with the rapid development of technology. New kinds of information have caused many uncertainties in privacy protection. Surveillance data and cross-border data are excellent examples of legal gaps in the existing systems. More discussion of data privacy is presented in section VI.

IV. REVIEW FINDINGS AND DISCUSSION

Our systematic review is conducted using 115 filtered articles covering multiple BD topics (Fig. 3(a)). In this section, we attempt to analyze the publication pattern of these articles. This work helps to reveal the current state of BD employments in the O&G sector.

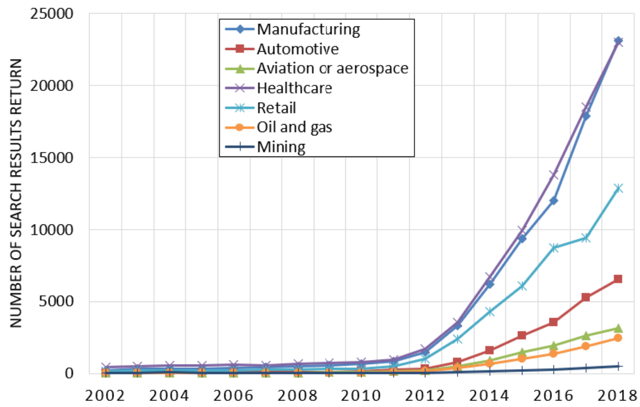

FIGURE 4. Google search results for "big data" in each industry.

A. SORTING RESULTS

1) PUBLICATION TYPES

Figure 3(c) presents different types of publications, including concept, case study and review. Fig. 3(b) shows that the number of publications has increased significantly since 2015. This feature partly shows the increasing interest of the O&G industry in the adoption of BD technologies. The number in 2019 represents only a partial year.

2) LEADING COUNTRIES

Figure 3(d) presents the paper distribution in terms of nationality and region. Notably, 16% of papers are from international projects involving many countries, such as [15], [63], [64]. This feature indicates the critical role of international research collaborations in developing BD technologies. Also, a large percentage of papers are from the US and Canada, indicating significant interests in BD technologies.

3) AFFILIATIONS

Figure 3(e) presents the distribution of article affiliation (i.e. industry, academia and their collaboration). Most articles from academia explore new BD technologies and concepts, which are researched and developed before being adopted by industry. Notably, the interuniversity articles contribute 28%, which points to the critical role of academic collaborations for the effective utilization of shared datasets and resources. Although some companies in the supply chain exercise caution due to the disruptive changes of digitalization, a large number of selected articles in our paper are derived from the cooperation between academia and industry. This indicates the industry's efforts to capture the benefits of BD technologies. This pattern of collaboration also shows that the supply chain drives BD adoption in the O&G industry.

B. LEADING OR LAGGING?

All the industrial sectors, including manufacturing, automotive, aviation, healthcare, retail, O&G and mining industries, have developed different projects to deploy BD technologies in their operations. However, the extractive industries (i.e. O&G and mining) have taken a longer time to adopt BD technologies. This can be verified using the Google search engine, which is an attempt to define the number of articles mentioning ''big data'' together with a given industry, as described in Table 3. Notably, for the mining industry, we use the string ''mining industry'' to distinguish it from ''data mining''. The number of Google search results returned indicates the attention level and popularity of BD technologies for the corresponding industry.

TABLE 3. String for Google search by industry.

Industry	Search String
Manufacturing	"big data" AND "manufacturing"
Automotive	"big data" AND "automotive"
Aviation	"big data" AND "aviation"
Healthcare	"big data" AND "healthcare"
Retail	"big data" AND "retail"
Oil & Gas	"big data" AND "oil and gas"
Mining	"big data" AND "mining industry"

The results of this Google search test are illustrated in Fig. 4. The healthcare and manufacturing industries show the highest interest in BD technologies. Recent trends in autonomous vehicles and personalized healthcare indicate interest in BD technologies. Aerospace, O&G and mining industries show less interest in BD adoption.

V. BIG DATA PLATFORM DEVELOPMENT TOOLS

This section discusses the readiness of technologies for BD deployment. Many petroleum installations are in remote hard-to-reach areas (i.e. deserts and open sea) and connected by the Internet. A data management system is necessary to organize these massive data for sharing and interoperation. Exponentially growing data volumes also necessitate more advanced resources for storing and accessibility. The multi-spatial and multi-temporal data from multiple sensors also challenges data management. Practitioners need to define how to organize and map the multi-dimensional data to a 1-D data array for wireless transfer. Handling metadata and indexing efficiently are other challenges [63]. Selecting an appropriate computing platform affects the utility of big data technologies. Especially in the beginning phase, organizations mostly avoid the potential complexities and considerable investments of on-premises systems. In recent years, a range of technologies has been developed for BD analytics. These tools, frameworks and algorithms can be classified into three main areas, including data storage, data processing and machine learning. The first two areas focus on managing, accessing and processing massive amounts of data. Machine learning is used to extract useful information and knowledge from the processed data for future exploitation [65], [66]. A further systematic review of O&G machine learning applications can be found in [67], [68]. Many options for storing and computing platforms are discussed below.

A. BIG DATA APPLICATION ON A SINGLE SERVER

BD implementation involves a single instance of a server and an operating system. An excellent example of this implementation is portable condition-based maintenance using mobile and wearable devices [69], [70]. Instead of installing a plethora of sensors throughout the plant, this strategy utilizes the mobility of wearable devices for data gathering. These data are transmitted wirelessly to a single server, to estimate equipment performance trends and detect any defects in the system [71]. As most of the computations are executed on a single server, scalability can be achieved by installing more computing units, memory, and faster processors within a single server. This kind of vertical scaling platform (VSP) is likely applicable to most of the software and easy for hardware management. However, additional hardware installations requires considerable capital investment. It is also challenging to predict expected future workloads to avoid unnecessary installation [72]. We discuss some standard hardware accelerator designs for BD on a single server.

1) HIGH-PERFORMANCE COMPUTING (HPC) CLUSTERS

This solution refers to the use of supercomputers having thousands of cores to deliver massive computations. The systems are optimized for speed, throughput and hardware-failure avoidance. HPC is deployed in different disciplines, such as seismic research for exploration, reservoir simulation, high-resolution solid and fluid mechanics and high-fidelity seismic data visualization. The computation executed on HPC not only produces more accurate results for finding new wells, but also boosts production from existing wells. As it requires high-end hardware, the large initial capital investment is one of the solution's drawbacks. It is not preferable for some pilot BD projects [73]–[75].

2) MULTICORE CPU AND GRAPHICS PROCESSING UNIT

The more economical solution is a multicore CPU with dozens of processing cores [72], [76]. CPUs can accelerate the execution of BD analytics using internal parallelism. The computational tasks are broken down into many threads. Each thread is executed separately in parallel on each core. The processing power of the CPU can achieve close to 10 Gflops. For example, the Intel Xeon Processor E7 v4 Family has 24 cores, which allow 48 threads of execution simultaneously [77]. An alternative solution to massive parallelism is a graphics processing unit (GPU) with more than 2500 processing cores, and achieving 1000 Tflops of processing power. GPU also offers a higher degree of parallelism, including the ability to communicate and synchronize many threads. For the software implementation, most of the programming languages provide libraries to create computational threads and achieve parallelism, such as the Parallel Computing Toolbox for MATLAB, Multiprocessing for Python, and PASL for C++. Notably, NVIDIA has released many GPU-accelerated software libraries for machine learning, deep learning and analytics with a CUDA framework [78], [79]. These libraries will assist both the practitioner and the programming language compiler in building and executing BD software.

Although CPUs and GPUs offer parallel computing, these two hardware solutions have limited memory capacity and access. In some cases, the low speed of memory access can become a huge bottleneck, which makes the parallelism inefficient [80]. New generations of GPU have attempted to mitigate these issues. Upgrades use DDR5 memory, which is many times faster than the traditional DDR3 memory. A faster cache for each multiprocessor is also developed to accelerate the data access. Currently, the maximum memory per GPU is 12 GB. If the data size is greater than the GPU memory or if the data size is on a terabyte scale, the GPU performance will decrease significantly [72].

3) SPECIALIZED HARDWARE UNITS

Field programmable gate arrays (FPGA) and application-specific integrated circuits (ASIC) also show the potential to enhance parallelism for BD applications [81]–[84]. Online recognition of mixed natural gas is an excellent example of using FPGA to accelerate machine-learning computation [85]. The FPGA based system can classify the methane, propane and ethane components of mixed natural gases with the test error of less than 0.5% and provides responses in several seconds. These solutions focus on integrated circuits, hardware design concepts and hardware-description languages (e.g. Verilog and VHDL) to develop applications, which in turn achieves better energy efficiency compared to GPU [82], [84]. However, in some cases, the customized hardware and the requirement of specific understanding of low level hardware development affect the cost effectiveness of the project. [86].

In 2016, Google released a series of ASICs for ML applications named the Tensor Processing Unit (TPU). TPUs were developed to run open-source TensorFlow ML software [87], [88]. Compared to GPUs, TPUs have shown better performance in computational speed and power consumption, which allows the rapid execution of the inference process of machine learning [84], [86]. Similarly, the neuromorphic chip is the other option to accelerate the computation of BD training and the inference process. This technology attempts to mimic the human brain and uses both analog and digital signals for operation. An analog subsystem offers the energy savings with a higher processing capability, while the digital subsystem allows programming flexibility and better precision and interfacing capability than do other digital systems. At the moment, these technologies are still far from mature, and require significant research and development efforts for future industrial applications [86], [89].

B. BIG DATA APPLICATION ON MULTIPLE SERVERS

This strategy distributes massive workloads across many separate computing machines. The remarkable benefit of this horizontal scaling platform (HSP) is that the system can be scaled as required. It relatively requires less capital investment than VSP for hardware upgrading. However, software implementation needs to handle all the data distribution and complex parallel processing. Currently, there is a limited number of available software supporting this kind of use [72]. In this review paper, we discuss peer-to-peer networks, Apache Hadoop, MapReduce and Apache Spark. There are other BD tools which are not discussed, such as Apache Kafka for data ingestion, and Apache Cassandra for storing raw data and processed results [90]–[92].

1) PEER-TO-PEER NETWORKS

This strategy utilizes a decentralized and distributed network architecture connecting millions of machines. Each node in the system (known as a peer) can store data and the size of data storage is almost unlimited. Also, each node can communicate and exchange data using a message passing interface. This interface implements the hierarchical master/slave paradigm, which allows the dynamic resource allocation to use the slave's computational resources effectively [72], [93], [94]. The main drawback of this interface is fault intolerance because there is no mechanism for fault handling. Consequently, a single node failure can cause the entire network to shut down. The system also has issues related to communication bottlenecking, limiting data synthesis and realtime processing.

2) APACHE HADOOP

This is an open-source framework, using clusters of commodity hardware and computers, to store and distribute processing tasks. This framework offers scalable computing, performing parallel processing of massive datasets [4]. In other words, Apache Hadoop is a development framework to implement the MapReduce programming model presented in the next subsection [95]. Apache Hadoop consists of two main components, including the Hadoop distributed file system and MapReduce. The first is the database to store a range of gigabyte, terabyte and petabyte files across multiple computing machines. The general architecture of Hadoop includes NameNode, DataNode, TaskTracker, and JobTracker. MapReduce is typically a primary data processing scheme with a JobTracker, to receive job submissions from client applications. JobTracker will push work out to all available TaskTrackers, as shown in Fig. 5. The advantage of this Hadoop framework is that it maintains the same performance of the computing and storage nodes. This advantage allows us to schedule tasks on already existing nodes and avoids moving data between nodes. However, the centralization in JobTracker also creates performance bottlenecks and scalability problems when the cluster size and the number of applications related to TaskTrackers increase [96], [97]. These limitations resulted in the advent of Apache Hadoop YARN to decentralize execution and monitoring. This provides more support for realtime processing and other applications that cannot wait for batch jobs to finish. Notably, YARN also allows running multiple MapReduce processing engines,

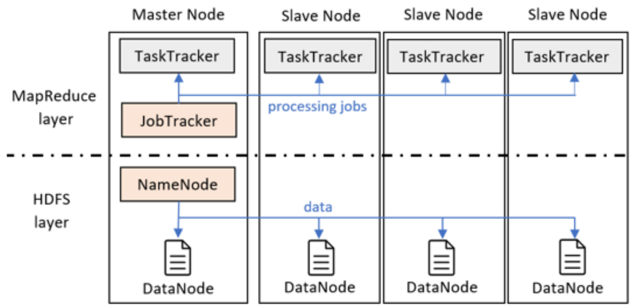

FIGURE 5. High-level architecture of Apache Hadoop.

which in turn increases the data processing capability significantly [96].

Apache Hadoop has been employed in multiple O&G applications to handle a large amount of data, improve the quality of analysis and provide strong supports for decision making [98], [99]. For example, in the study of drilling information, the conventional methods [98]–[100] aim to adopt centralized and partial data, which is preferable to no realtime and small-scale data processing. The development of drilling technologies has increased the volume and variety of drilling data information. These demands motivate the use of Apache Hadoop for multidimensional analysis.

3) MAPREDUCE

MapReduce is designed as a simplified programming model. Generally, MapReduce splits the data into chunks, and processes these chunks by laying them out across all computing nodes with a mapper. These resulting chunks of data are then sorted and go through reduced operations before producing outputs for the entire system [97]. MapReduce can process large numbers of datasets in parallel using multiple scalable clusters. These clusters are flexible and highly fault-tolerant. MapReduce also provides the dataset services of mapping (i.e. sorting and filtering) and reducing. Processing pre-stack seismic data for oil exploration can amplify the benefits of MapReduce. The MapReduce model has been utilized to solve the parametric inversion problem of seismic BD, which further reduces the uncertainty and extracts more reliable seismic attributes in the analysis [97], [101], [102]. Additionally, the use of parallel technology in MapReduce not only reduces the processing time effectively but also guarantees similar fitness value computing times to those of a single-machine deployment. One of the significant drawbacks of MapReduce is its inefficiency when executing iterative algorithms. After each iteration, the data need to be written to the disk to pass them to the next iteration. This structure creates a data bottleneck that degrades the performance significantly [72]. MapReduce is recommended for large-scale fault-tolerant data analysis.

4) APACHE SPARK

This is an open-source BD processing framework using multiple programming languages (i.e. Java, Scala, Python, and

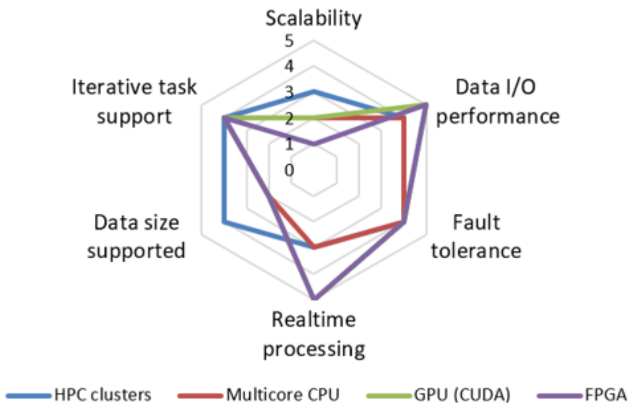

FIGURE 6. Characteristic comparison of different platforms to enhance vertical scaling.

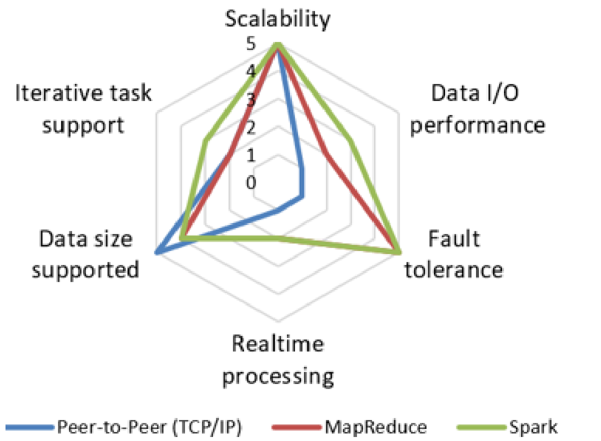

FIGURE 7. Characteristic comparison of different platforms to enhance horizontal scaling.

R algorithm) and is designed for fast computation. In some classification applications, Spark is reported to work 100x faster than Hadoop [103] because of its ability to perform in-memory computations. This helps to eliminate the aforementioned limitations of disk I/O for the Hadoop system and to handle iterative tasks efficiently. This unique feature makes Spark a next-generation data analysis paradigm.

In O&G applications, many projects attempt to replace Hadoop with Spark for faster analysis of tasks such as industrial alarm management [104], seismic data analysis [105], production forecasting of unconventional wells [64], and well casing damage prevention [41]. Remarkably, for gas lift well surveillance, Spark has been employed to develop new online realtime visual analytics of distributed temperature sensor measurements [106]. The system can statistically learn features from the measurements, predict the performance trends of the gas lift valve and detect any anomaly. The fast computation has allowed the generation of realtime results to enable immediate responses and prevent harmful consequences of the valve failures on oil well performance.

C. COMPARISONS BETWEEN PLATFORMS

To better understand the general advantages and disadvantages of various platforms, the work described in [72] included some evaluations for the VSP (hardware based) and HSP (software based). In this paper, we present their evaluations using radar charts (Fig. 6 and 7). The rating is on a 1 (lowest possible) to 5 (best possible) scale in terms of performance, including system/platform-dependence (i.e. scalability, data I/O performance and fault tolerance) and application/ algorithm dependence (i.e. real-time processing, data size supported and iterative task support).

The scalability refers to the system's capability to accommodate an increasing workload. HSPs receive better ratings than VSPs because they require lower investment and effort. Data I/O performance considers the data rate to transfer to/from a peripheral device. In this case, VSP solutions, such as GPU and FPGA, receive the highest rating because of their high throughput memory and high-speed data I/O operations. The virtual clusters (i.e. Hadoop and Spark) receive a lower score because of the bottlenecks and the use of limited network access. Fault tolerance is also one of meaningful metrics to evaluate the system's operation in the case of component failures [90]. Software developers of virtual clusters have paid more attention to this element. Mechanisms for efficient in-built fault-tolerance have been developed for the platforms which receive the best rating among BD platforms. In terms of realtime processing, although Spark and MapReduce can provide fast computing, its acceleration cannot be as fast as that of vertical-scaling solutions, which conduct executions with a single machine. Data size supported reveals the volume of a dataset that a platform can process and manage efficiently. In this case, VSPs cannot achieve better evaluation than HSPs because of the limitations of onboard memories. Lastly, the comparisons consider the ability of iterative task support. The result of one interaction on VSPs can be transferred easily to the next interaction. However, the rating score of VSPs is only four instead of five, because most of the iterative algorithms implemented in VSPs are not easy to modify and update. The iterative tasks can be run on HSPs, but this is not efficient compared to VSPs [72].

D. CLOUD, EDGE AND FOG COMPUTING

The network architecture design is also an important factor in BD projects. This affects the speed of data transmission and processing, especially in the case of realtime applications. Three popular designs, including cloud, edge and fog computing, are discussed below.

1) CLOUD COMPUTING

Cloud computing delivers computing services over the Internet, including for servers, networking, storage, analytics databases and intelligence. It is considered a robust architecture to perform complicated large-scale computing tasks [92]. HPC clusters are utilized to develop cloud applications

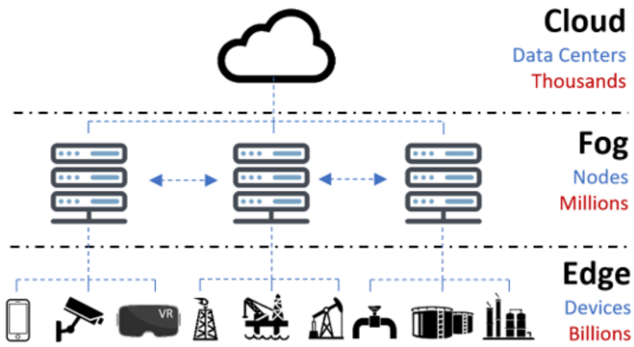

FIGURE 8. Networks of cloud, fog and edge computing with quantity scale (red), implementation location (blue).

because of massive computing capabilities. Compared to traditional IT resources, the remarkable benefits of cloud computing technologies include parallel processing, virtualized resources, security and data service integration with scalable storage. Cloud computing development aims to minimize the cost of capital investments in infrastructure set-up, reduces maintenance efforts and achieves efficient management [92], [107]–[109].

Limitations of available mobile networks cause traffic congestion, large end-to-end delay of data transfer and the considerable cost of processing massive data and communication. These issues necessitate the development of fog and edge computing to efficiently manage and transfer a large amount of data [110], [111]. Figure 8 demonstrates the differences among cloud, fog and edge computing in terms of where the data is processed and stored, as well as the locations of computing capability. In this paper, we distinguish between edge and fog computing, although they have similar features for using local networks to leverage the computing capabilities. The main difference between them is the location where the data processing is performed. Edge computing is deployed directly on the devices and on a gateway that is physically close to sensors. Fog computing moves the activities of edge computing to advanced processors.

2) EDGE COMPUTING

At the bottom level or downstream of cloud services, as illustrated in Fig. 8, edge computing focuses on centralized architecture to arrange computation to be performed at the edge of the network. For example, a smartphone, gateway, micro-data center and a cloudlet constitute the edge between the cloud and mobile devices (i.e. sensor nodes, wearables, etc.) [112], [113]. The rationale of edge computing is carried out at the data sources rather than on the data centers. In this structure, the edge components can process data locally, instead of outsourcing them to the cloud [14], [113].

Consequently, system development can mitigate the limited bandwidth which causes a bottleneck of data transportation to the cloud. In some applications of monitoring and control O&G equipment, the large amount of data needs to be processed in realtime to make accurate decisions [113], [114]. With the current network facility, it is challenging to upload data to the cloud for an instant response for just one vehicle. It is better to process these data at the edge for a faster response time, lower network pressure and more efficient processing. Medical wearable devices to track the health status of a field worker are excellent examples of edge computing [115], [116]. The health monitoring system can utilize physiological data to achieve real-time monitoring. The database is also useful for early diagnosis of chronic diseases, which in turn prevents future incidents. As computing is performed at the data sources, the response time of the system can be reduced significantly. Processing data at the edge also helps to protect user privacy better than transporting raw data to the cloud [117].

3) FOG COMPUTING

The fog is generally a cloud close to end users [118]. In the middle level between the cloud and the edge in Fig.8, fog computing is developed using decentralized computing architecture between the source and the cloud infrastructure to process and store data [119], [120]. In this architecture, a very high number of geo-distributed devices generate a "mini-cloud" at the edge of the network. The edge devices in the fog network will release some computing and storing resources to support the demands of their neighbours [111]. In each fog node, a decision is made to process data from various sources or send them to the cloud for further analysis. Fog computing attempts to organize and manage at the local level, while cloud computing aims to optimize the use of global resources.

Goscinski *et al.* [121] noted the critical role of fog computing between a central cloud and IoT in BD applications to enable the quick discovery of new hydrocarbon reservoirs. A three-month hydrocarbon exploration program can generate one Petabyte of seismic data. Conventionally, these data are transported manually in tape drives, using helicopters and runner boats to centralized processing centers for seismic data processing. The processing time takes a few months, and the processing cost can reach $10 million USD for 1200 km^2 of survey area [121]. Alternatively, the data can be pre-processed at the fog level before sending them to the cloud. This strategy makes data discovery faster and less expensive. Additionally, the implementation of a smart offshore platform can benefit fog computing [5]–[7], [117]. BD analytics for smart oilfields needs to process a diverse set of sensors, such as gas density, fire/gas/H2S alarms, pipeline pressure, temperature sensors flow monitoring and tank levels. BD analytics needs to enhance the real-time processing of resource-intensive emergencies and real-time monitoring of sites. However, these benefits also result in higher investment in infrastructure and concerns related to data consistency across a large network and the traffic going to the cloud [110], [111].

E. FACTORS TO CONSIDER WHEN SELECTING PLATFORMS AND TOOLS FOR DEVELOPING BD APPLICATIONS

The comparisons between platforms and discussions of specific features of cloud, edge and fog computing have provided useful information for BD analytics deployment. Three main technical factors have been identified for better BD platform selection [72]:

- *Data size:* is the most critical factor to predict the computational demand and to decide whether to use single or multiple servers. If a single machine can manipulate the data, multicore CPU and GPUs are the best options to accelerate the computation. If the data require the use of multiple computing machines, Hadoop and Spark are better solutions but have slower processing speeds, particularly for iterative tasks.
- *Processing speed or throughput optimization:* refers to the realtime processing capability of the platform. This feature is essential in applications that require the results to be generated in realtime for online monitoring and maintenance [106].
- *Model development:* includes how the system performs training and predicting phases. The training phase is usually conducted offline and takes some time to optimize the parameters of the model. This phase also requires increased effort, because the user needs to deal with a large amount of data. In contrast, the prediction or inference phases handle fewer data and use less computational time. Appropriate selection of the platform will improve the success of model development.

On the other hand, BD service providers also play a critical role in BD deployment. Enverus (formerly Drillinginfo) [8] and Omnisci [122] are excellent examples of companies providing BD tools for data analytics. Enverus provides the service of rig analytics, which tracks more than 95% of the US rig fleet daily. This real-time tracking allows users to have access to information first, which in turn makes decisions faster and outperforms the competition. The analytics also tracks drilled-but-uncompleted (DUC) wells and rig activity over time to predict future production performance.

VI. DATA PRIVACY

Workplace surveillance and monitoring using cameras can reduce safety risks and undesirable incidents, especially in hazardous areas in O&G facilities (e.g. drilling rig [123]). Apart from safety benefits, these collected camera images also introduce significant challenges for privacy and security protection [124]–[127]. Some non-personal data can become personal data if these data are associated with a specific individual (e.g. the tracking of the field vehicle location). In that case, how does the data management system classify the data sources and assign appropriate safeguards? This section attempts to discuss some concerns about data privacy when deploying BD technologies.

A. NOTABLE LAWSUITS ASSOCIATED WITH DATA PRIVACY

The Arias v. Intermex Wire Transfer lawsuit filed in 2015 was related to the Xora app, which allowed the company to track employee GPS location [128]. The use of the app by employees was one of the requirements of the job. The plaintiff expressed her discomfort to her supervisor and uninstalled the app. Consequently, she was fired. The plaintiff sued Intermex for privacy invasion and unfair business practices. Although the case was ultimately settled out of court, it raised the awareness of employee privacy issues. Similar cases associated with location tracking devices are the Elgin v. Coca-Cola Co, and Cunningham v. New York Department of Labor lawsuits [128]. These cases have established some possible parameters for the privacy claims of electronic device monitoring. Employers have a legitimate interest to track employees during on-duty hours and no legitimate interest during off-duty hours. It is challenging to define the differences and distinguish between on- and off-duty status. However, electronic devices are likely to be found with the employee. How should privacy protection be applied when the company would like to track the company-owned devices of the off-duty employee? Moreover, the study also indicated that the courts might be more amenable to privacy claims arising from electronic device tracking [128].

B. GOVERNMENT REGULATIONS

These lawsuits raised the awareness of privacy protection when adopting technology and offer good opportunities to improve the quality of existing regulations. Even though many countries have developed regulatory frameworks for data protection and privacy implications (presented in Table 4), these regulations still show many limitations and need more clarification, especially with the involvement of multiple companies in the processing of shared data. For example, regarding the operational data from upstream O&G, Vega-Gorgojo *et al.* reported the complication of data management, due to the contract determining the data ownership [129]. Also, the legislation of data is unclear and must be clarified to address issues of liability. Lu *et al.* have indicated the demand for privacy requirements during BD collection, storage and processing [127]. These requirements aim to balance BD benefits and individual privacy preservation during data collection, storage and processing. In the collection procedure, the system attempts to collect data from different sources, which increases the possibility of eavesdropping and accidental leaks. The collection procedure should apply physical protection methods and information security to protect personal and sensitive data. In storage, privacy protection is more critical. If the protection fails in one user account, more individual information is likely disclosed. The processing procedure is more complicated because of the engagement of other organizations and data sharing.

Cross-border data flows are another concern in privacy protection because machines for storing and processing

TABLE 4. Data privacy regulation in some countries.

Country	Regulation
U.S.	Health Insurance Portability and Accountability Act (HIPAA); Patient Safety and Quality Improvement Act (PSQIA); Health Information Technology for Economic and Clinical Health (HITECH)
Canada	Personal Information Protection and Electronic Documents Act
Europe	General Data Protection Regulation
U.K.	Data Protection Act
Russia	Russian Federal Law on Personal Data
China	National Information Security Standardization Technical Committee (TC260)

can be located in different countries. Regulatory gaps still exist because legislation is different for each country [129]. Fietkiewicz *et al.* highlighted the unequal protection of data by US and EU legislation systems [130]. Instead of developing a comprehensive and single regulation system, the US legislative system considers data protection in specific domains (i.e. healthcare, communications or financial services). This raises more uncertainties in the accurate classification of a domain for the work of data processing, and for the company providing analytic services.

C. STANDARDIZATION

Apart from government regulations, many studies have pointed to the pivotal role of developing new standards to support BD management [131], [132]. To date, standardization has not been practiced for field data. This results in structured (e.g. portable document format), semi-structured (e.g. log files) or unstructured data (e.g. Excel spreadsheets). It also leads to the development of data integration platforms using different standards and visualizations. In some cases, the data are typically stored in separate locations without a connection to a shared database. As a result, data integration requires additional effort for interpretation, and becomes impractical with limited budgets and time [131], [133].

The standardization process involves engineers, technical experts, and regulators to generate a technical solution that balances the competing interests and benefits of technological compatibility. The US issued the first standardization documents for BD in 2015 (i.e. NIST Special Publication 1500). Canada and other countries have taken a longer time to prepare standardization roadmaps because of limited experience and support [131], [133]. The lack of standards is a barrier to the growth of BD applications in the O&G sector and other industries. Many worldwide organizations involved in information and communication technologies may be sitting on the sidelines but are compelled to adapt to the rules set by US tech giants. This limitation raises many regulatory uncertainties regarding the deployment of advanced analytic tools, contributing to the low pace of technology adaption [131]. The development of foundational standards is necessary to support BD deployment and collaboration in different organizations. The standardization roadmaps also identify gaps and make recommendations for action nationally, regionally, and internationally.

D. DATA PRIVACY PRESERVATION IMPLICATION IN THE WORKPLACE

Although outsourcing data to the cloud helps to offload massive computation, field operators and end-users lose physical control of the data and allow access to other parties to perform analysis [134], [135]. In the study of wearable devices (e.g. Fitbit, Jawbone) for fitness applications, studies have reported that the company failed to disclose enough information to the end-users about the amount of collected data and the engagement of third-party companies in data processing [130], [136], [137]. This lack of information may give rise to conflicts between employees and employers, which affects consumer trust and company reputation [124]. For example, the employer can utilize surveillance data to evaluate employee performance even when off-duty. Health-tracking data can be the basis for an employee to sue an employer for unhealthy working conditions [136]. Richardson *et al.* recommended consideration of these concerns as new opportunities to upgrade existing systems rather than stifling innovation to avoid lawsuits or other unwanted consequences [136]. An effective management system can navigate the privacy implications of new technology in the workplace. The ownership of data and the information on data usage should be clarified clearly in the employment contract. In some cases, within a company structure, a new department is necessary to deal with data privacy. To prevent a breach of privacy, Chibba *et al.* proposed embedding protective privacy measures into the design of BD systems, networked facilities and business practices [124]. Similar cases of privacy harm can occur in O&G applications. Experiences from other industries are useful lessons for the deployment of BD analytics in the O&G sector. In terms of technical solutions, developing mechanisms for access control and tools for data protection is necessary. This allows sharing the data only with authorized users and accomplishes the policy of preserving privacy. The next section will evaluate the cyber-security implications in the O&G sector as well as introduce some standard tools to protect data against cyber attacks.

VII. CYBERSECURITY ISSUES IN BIG DATA SYSTEMS

The discussion of information confidentiality and data privacy necessitates the review of cyber-security implications. In recent years, cyber attackers have targeted many organizations and companies with increasing frequency and sophistication. In different countries, various strategies have been applied to govern cybercrime as well as support the economy and public safety [138]. However, cybersecurity is relatively immature, and strategic plans are limited in the handling of these cyber issues such as cyberterrorism, industrial espionage, disruption operations, and the theft of company data [16], [139], [140]. The exponential growth of cyber-physical systems also magnifies security

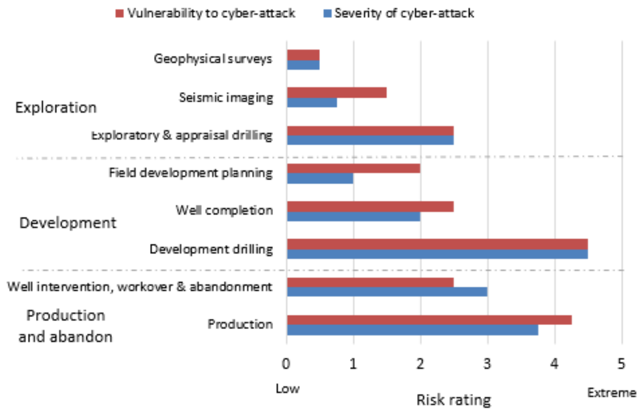

FIGURE 9. Severity of cyber attack evaluation for upstream O&G [16] presented in bar chart.

challenges [141]. Technological adoption requires O&G regulators to pay more attention to security and cyber attack issues. Srinivas *et al.* have reviewed the roles of government regulations in cybersecurity [142]. In this paper, we discuss the role of O&G companies in the deployment of BD technologies and cybersecurity solutions to defend a system against cyber attacks.

A. SEVERITY EVALUATION OF EACH O&G SEGMENT

1) UPSTREAM O&G EVALUATION

The authors of [16] conducted a cyber vulnerability and severity evaluation for upstream O&G operations. In this paper, we re-present their evaluation in the form of a bar chart as Fig.9 before conducting further analysis. This helps to prioritize security for the most critical and risk-prone activities.

The exploration stage has the lowest profile of cyber vulnerability and severity. The first two operations, geophysical surveys and seismic imaging, deploy a closed-data collection system and a single vendor ecosystem. A cyber attack on these two operations would have a low financial impact and low probability of causing business disruption. Exploratory and appraisal drilling receive higher risk profiles because this operation includes many elements of the development stage. An excellent example of cyber attack in this stage is the 2011 Night Dragon [143]. Utilizing remote administrative tools, the hacker disabled proxy settings and stole the data associated with field exploration and bids from many O&G companies for many years. Generally, the existing exploration workflow operates with a relatively low cyber risk profile. However, technology adoption may increase the risk because of the significant boost of data processed in HPC clusters. In future realtime applications, these exploration data will be fed into the next stages of upstream O&G (e.g. drilling plans, completion designs, and reserve estimates), which in turn multiplies the cyber attack effects [16].

In the development stage, O&G well development is also exposed to cyber incidents. The development drilling operation receives the highest severity evaluation of cyber attack because of the high number of drilling activities, the complicated ecosystem and the involvement of multiple engineering firms. The diversity of this business makes it challenging to build a single cybersecurity protocol. Currently, drilling and computer systems are designed following the concept of an isolated network (e.g. located in remote hard-to-reach places), providing a natural defence against cyber attacks. However, the future scenario of real-time monitoring offshore rigs [5]–[7], when all subsystems are interconnected for data visualization and transportation from anywhere in the world, allows hackers to access more information [144]. Notably, the delayed responses are partly attributed to the lack of the workers' knowledge about the inputs and outputs of the computer systems controlling rig operations.

The final stage refers to the operations of production and abandonment. Similarly, the legacy asset base, which was not built with consideration of cybersecurity, and the lack of cyber defence tools can explain and magnify the cyber vulnerability of a production operation. The cyber attack on Saudi Aramco in 2012 is an excellent example of the weakness of legacy IT system. The attack, which resulted in the damage of 30,000 computers, attempted to stop O&G production in Saudi Arabia and prevent resource flow to international markets [145]. The final stage of the upstream chain, including well intervention, workover and abandonment, receives a lower profile of cyber vulnerability [16], [145], [146]. The work described in [147] indicated that 75% of global petroleum production is controlled by resource planning systems, and these systems encounter many cyber risks both from the top segment (i.e. IT systems) and the bottom segment (i.e. the field hardcore legacy operating technology systems). As a result, the cyber attack consequence could be severe in both the top and bottom segments. Well intervention, workover and abandonment operations, which are ranked with lower cybersecurity risks, focus on well diagnostics, mechanical replacement and maintenance work. However, the increasing use of interoperable equipment, developed by open-source software platforms with human-machine interfaces, can raise the cyber risks.

2) MIDSTREAM O&G EVALUATION

Data related to the flow from production reaching a refinery are also attractive to the hackers. Protecting pipeline networks against remote cyber attacks is one of the biggest concerns for midstream O&G [148]. Any manipulation of these flow values can create reporting discrepancies between producer and refinery companies, which negatively affects the company relations and stock prices [149]. The costs of cybersecurity pipeline incidents also include the cost of business interruptions and damage to third parties, the physical plant, equipment and control systems. The pipeline software management system is commonly built on supervisory control and data acquisition (SCADA) systems that allow operators to monitor and control many aspects of operations. The benefits of the SCADA system are to reduce operating costs and improve system efficiency. However, hackers can

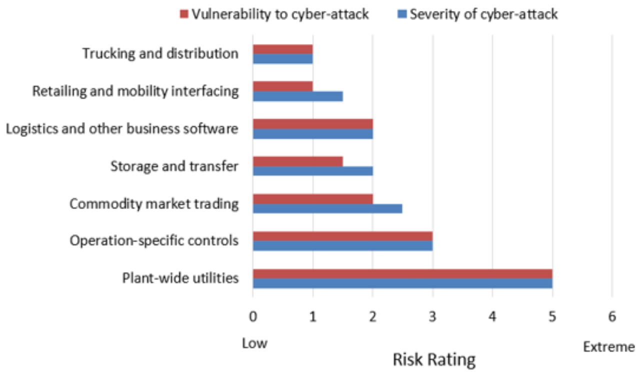

FIGURE 10. **Severity of cyber-attack evaluation for downstream O&G [16] presented in bar chart.**

take advantage of the SCADA system to disrupt pipeline services and cause spills, explosions or fires [150]. For example, in 2012, the UglyGorilla group conducted a significant attack on two dozen US natural gas utilities to steal sensitive data from gas pipeline companies. Similarly in Canada, Telven, which manages about 60% of the O&G pipelines in the Western hemisphere, also experienced mysterious cyber attacks to the IT systems and internal firewall [151]. In 2018, Energy Transfer reported a cyber attack that shut down a main pipeline data network in the US [152], [153]. The attack raised the awareness of the vulnerability of gas pipeline networks and the demand for sufficient cybersecurity defences to ensure resilience.

3) DOWNSTREAM O&G EVALUATION

Similarly, cyber attackers have targeted downstream O&G companies for years with increasing frequency and sophistication. The work described in [16] evaluated cyber vulnerability and severity for downstream O&G operations. Their evaluation is re-presented in Fig. 10 before conducting further analysis.

The highest risk is for the operation of plant-wide utilities and operation-specific controls, which are associated with safety equipment, and high pressure and high temperature processes (e.g. compressors, pumps, reboilers, electric power generator). Any damage to the plant utilities and the loss of control of equipment can lead to process failure or a shutdown of the entire refinery system. Generally, the equipment in a refinery is monitored and controlled by human operators. Sensor feedback, algorithms and setpoints are utilized to input parameters and supervise the process. In the deployment of BD, this equipment is interconnected, which in turn increases the complexity of the process significantly. The interconnectedness is beneficial from the operational viewpoint but provides additional attack surfaces and exposes large parts of the facility to a cyber attack [154]. The worldwide hijacking of the Stuxnet computer worm on industrial control systems (ICS) in 2012 exemplifies the downstream cyber attack. The group targeted computers that are used to operate refineries, pipelines and power plants [145]. Compared to the virus, the computer worm can be self-executable and manipulate an operating system until it achieves the intended target. Additionally, the lack of public information regarding the damage caused by Stuxnet (particularly in Iran, China, and Pakistan) also raised more challenges for the evaluation of the malware's potency [155].

Logistics software and activities may show relatively less risk. Their incidents are limited to delays and communication challenges which affect third-party personnel. For example, in 2015, hackers targeted at retail gas stations across the US. The automated tank gauges, which are used to measure gasoline levels, were manipulated to cause alerts and shut down the fuel flow [145].

B. CYBERSECURITY SOLUTIONS FOR BD NETWORKS

Opportunities for cyber attacks are enhanced in various forms with the increasing levels of complications and include the virus, phishing attack, trojan horse, worm, ransomware, spyware, unauthorized access and control system attacks [142]. The cyber attacks attempt to invade the operational, power and cyber-physical systems. They allow intruders to steal sensitive information and damage the system's operation [156]. Many cybersecurity solutions have been implemented to protect the system against cyber attacks.

1) IDENTITY AND ACCESS MANAGEMENT SOLUTION

This protection solution is necessary, especially at the endpoints or the edge of the network (i.e. mobile devices, laptops and servers). It allows the individuals to access the resources for appropriate reasons at the right time [139], [157]. This solution encounters challenges in increasingly heterogeneous technology environments.

Password, passcode and personal identification number (PIN) protection are a conventional approach to digital identity and widely used in different areas. A password is generally a string of characters allowing a person to be identified and log into the system. Although this technique is simple, password management is also challenging, particularly with decentralized operations extending to remote oil fields and offshore platforms. In these cases, the field personnel tend to share passwords, which can result in unauthorized access to the system. Further, the high reliance on the IT department likely affects operational productivity, especially in the case when remote employees forget their passwords [158]. For hackers, a brute-force attack will be useful to obtain the password information. This is a trial-and-error method submitting all possible passphrases to the system until the correct one is found [159]. For example, the Lyceum threat group conducted many cyber attacks targeting O&G firms in the Middle East in May 2019. This method allowed the group to access compromised email accounts from within victim companies before sending the emails with a malicious attachment to other company departments. If a person clicks on the malicious attachment, DanBot malware is downloaded directly to the network, which provides remote higher-access capability to the attacking group [160]. Limiting the number of failed login attempts can prevent a brute-force attack, but the resulting delays of a system reset can be costly because of the associated downtime.

Biometric matching is an advanced approach to identification and authentication [161], [162]. The method utilizes a set of recognizable and verifiable data that are specific and unique to the person, including physiological (e.g. fingerprint, hand, face, iris, DNA) and behavioral (e.g. signature, voice, keystroke) features. Matching the template data with the person's characteristics will determine resemblance and identify the person. Compared to the password solution, biometric keys are more difficult to copy, share, forget, lose and guess [142]. These notable features make the biometric technique more reliable and secure, satisfying authentication requirements. In the last few years, fingerprint and face recognition techniques have been employed to increase the security level of some commercial smartphones. However, O&G field workers need to wear protective gloves and goggles, which challenges the use of biometric matching for industrial mobile devices.

2) DATA ENCRYPTION SOLUTION

If the attackers remain in the network for a longer time without detection, the breach will cause more damage. The purpose of the encryption solution is to prevent privacy leakage via access and mitigate cyber attack damage [134]. Encryption transformation makes data unreadable without authorized access and the decryption key [163]. Ciphertex [164] and Thales [165] are examples of software companies providing encryption services for the O&G sector. Solutions should be scalable and easy to set up without disruption, replacement or relocation of the network facilities [166]. Encrypting seismic data is an excellent example of BD applications. Although the significant volume of seismic data may overwhelm the hackers, securing the sub-surface cloud data to prevent industrial espionage is necessary [16]. Encryption is enhanced using an isolated token generation system, which replaces the sensitive data with a non-sensitive equivalent [167]. The applications are then executed using tokens, which offers no value for hackers, instead of using actual data. Similarly, in Schlumberger's "cyber hygiene" security strategy, the company employed network segmentation and encrypted every machine, server and database. These efforts have rejected the majority of cybersecurity threats [168]. Similar to the password solution, the management of the decryption key is also challenging. In some cases, the exchange of the key might be complicated and impractical because of a key's remote location and the face-to-face meeting requirement.

3) THE INDUSTRIAL CYBERSECURITY PLATFORM SOLUTION

The industrial cybersecurity platform provides many comprehensive tools for protecting and defending the ICS network against malware and hacking-based attacks [139]. It uses software which can track asset changes over time to determine malicious activities and potential threats [169], [170]. Some software companies provide this industrial cybersecurity service for the O&G sector, such as Dragos [171], Sentry [172] and CyberX [173]. Moreover, for further defence, the platform also tracks hacker activities targeting ICS of O&G companies. For example, Dragos is an industrial cybersecurity platform tracking the following five hacker groups [174]:

- Xenotime group caused the disruption at O&G facilities in Saudi Arabia (2017) and in European, Middle Eastern and North American regions (2018).
- Magnallium group has targeted petrochemical and aerospace manufacturers of Saudi Arabia, Europe and North America since 2013.
- Chrysene group initially gained attention in 2012 with cyber attacks on Saudi Aramco. The group's initial targets include petrochemical, O&G and electric generation industries in the Gulf region. Recently, the group evolved to target other areas.
- Hexane group focuses on O&G and telecommunications located in Africa, the Middle East and Southwest Asia. The group has been identified since May 2019, but there is limited publicly available information.
- Dymalloy group performs highly aggressive and capable activities for long-term and persistent attacks on IT and operational systems. This group targets Turkey, Europe and North America.

Other techniques and tools such as firewalls, antivirus softwares, and secure smartphone apps also contribute to cybersecurity enhancement with different protection levels [139]. The use of these security solutions depends on the particular systems and conditions of application.

VIII. CONCLUSION

This paper has presented a systematic review of BD analytics in the context of the O&G industry. Recent deployments of BD technologies have improved operational efficiency and maximized asset value. Also, this paper reviewed the readiness of tools, platforms and network architecture for BD development. Recent improvements in data computing, storing and cloud technologies have improved computation capability, which in turn opens new opportunities for real-time monitoring applications such as smart-oilfield and smart-refinery. BD deployment also raises many concerns about data management and protection. We analyzed the issue of data privacy in terms of regulations, standards and cross-border situations. We analyzed the severity of cyber attacks for upstream, midstream and downstream O&G industry, using recent O&G incidents as practical examples. From our understanding of the reviewed topic, we can summarize the major lessons we learnt as follows:

- The publication pattern from the filtered articles in this paper reveals the demand for international collaborations and the primary role of supply chain companies to develop BD technologies for the O&G industry.
- Governments have been aware of data privacy issues when deploying BD analytics. Despite many efforts to

develop and upgrade legal systems for better privacy protection for Industry 4.0, regulatory gaps among countries still exist and cause many difficulties in managing cross-border data.

- BD deployment needs to address IT problems associated with cybersecurity. The use of advanced software tools and platforms is necessary to detect and defend systems against cyber attacks. Some improvements in the level of management and communication between IT and other departments are essential to reduce the response and reset time of the system under cyber attack.

These issues have affected the pace of technology adoption and the efficient application of BD in the O&G sector. Addressing these issues provides an excellent opportunity to upgrade existing systems and capture the benefits of BD applications. Digitalization also includes the adoption of digital twins, wearable computing, artificial intelligence, blockchain and robotics. Future work will evaluate the possibility of integrating BD with other advanced technologies within a unified environment.

REFERENCES

[1] BP. BP Energy Outlook 2019 Edition. BP, U.K. Accessed: 2019. [Online]. Available: https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/energy-outlook/bp-energy-outlook-2019.pdf

[2] H. Lu, L. Guo, M. Azimi, and K. Huang, "Oil and Gas 4.0 era: A systematic review and outlook," *Comput. Ind.*, vol. 111, pp. 68–90, Oct. 2019, doi: 10.1016/j.compind.2019.06.007.

[3] A. Brun, M. Trench, and T. Vermaat. (Oct. 2017). *Why Oil and Gas Companies Must Act on Analytics*. [Online]. Available: https://www.mckinsey.com/industries/oil-and-gas/our-insights/why-oil-and-gas-companies-must-act-on-analytics

[4] M. Mohammadpoor and F. Torabi, "Big Data analytics in oil and gas industry: An emerging trend," *Petroleum*, to be published. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2405656118301421

[5] R. F. Hussain, M. A. Salehi, A. Kovalenko, Y. Feng, and O. Semiari, "Federated edge computing for disaster management in remote smart oil fields," in *Proc. IEEE 21st Int. Conf. High Perform. Comput. Commun., IEEE 17th Int. Conf. Smart City; IEEE 5th Int. Conf. Data Sci. Syst. (HPCC/SmartCity/DSS)*, Aug. 2019, pp. 929–936.

[6] R. F. Hussain, M. A. Salehi, A. Kovalenko, S. Salehi, and O. Semiari, "Robust resource allocation using edge computing for smart oil fields," in *Proc. Int. Conf. Parallel Distrib. Process. Techn. Appl.*, 2018, pp. 1–7.

[7] R. F. Hussain, M. A. Salehi, and O. Semiari, "Serverless edge computing for green oil and gas industry," in *Proc. IEEE Green Technol. Conf. (GreenTech)*, Apr. 2019, pp. 1–4.

[8] E. Media. (2019). *Drillinginfo Analytics Help Canadian Oil and Gas Hit the Reset Button*. EnergyNow.ca. [Online]. Available: https://energynow.ca/2019/03/drillinginfo-analytics-help-canadian-oil-gas-hit-the-reset-button

[9] V. Padmanabhan, "Big Data analytics in oil and gas: Converting the promise into value," Bain & Company, Boston, MA, USA, Tech. Rep., Mar. 2014.

[10] B. M. Tran, *Oil Strategy: The Artificial Intelligence Approach to Quantifying the Coronavirus*, RBC Capital Markets, Toronto, ON, Canada, 2020. [Online]. Available: https://www.rbccm.com/en/insights/story.page?dcr=templatedata/article/insights/data/2020/01/oil_strategy_the_artificial_intelligence_approach_to_quantifying_the_coronavirus

[11] J. B. Long and J. M. Ehrenfeld, "The role of augmented intelligence (AI) in detecting and preventing the spread of novel coronavirus," *J. Med. Syst.*, vol. 44, no. 3, p. 59, 2020. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/32020374

[12] X.-G. Yue, X.-F. Shao, R. Li, M. Crabbe, L. Mi, S. Hu, J. Baker, and G. Liang, "Risk management analysis for novel coronavirus in wuhan, China," *J. Risk Financial Manage.*, vol. 13, no. 2, p. 22, 2020.

[13] (2018). *Big Data in Oil and Gas Market Expected to Reach US$ 10, 935.2 Mn by 2026*. [Online]. Available: https://www.transparencymarketresearch.com/pressrelease/big-data-oil-and-gas-market.htm

[14] S. O. Settemsdal and B. Bishop, "When to go with cloud or edge computing in offshore oil and gas," in *Proc. SPE Offshore Eur. Conf. Exhib.*, 2019, pp. 1–5.

[15] Y. He, F. R. Yu, N. Zhao, H. Yin, H. Yao, and R. C. Qiu, "Big data analytics in mobile cellular networks," *IEEE Access*, vol. 4, pp. 1985–1996, 2016.

[16] A. Mittal, A. Slaughter, and P. Zonneveld, *Protecting the Connected Barrels-Cybersecurity for Upstream Oil and Gas*. USA: Deloitte Univ. Press, 2017.

[17] Anastasia. (2016). *Brief History of Big Data*. [Online]. Available: https://www.cleverism.com/brief-history-big-data/

[18] S. Sinclair and G. Rockwell, *Voyant Tools*. Canada: Voyant Tools, 2016. [Online]. Available: http://voyant-tools.org/

[19] F. Xia, W. Wang, T. M. Bekele, and H. Liu, "Big scholarly data: A survey," *IEEE Trans. Big Data*, vol. 3, no. 1, pp. 18–35, Mar. 2017.

[20] P. Joshi, R. Thapliyal, A. A. Chittambakkam, R. Ghosh, S. Bhowmick, and S. N. Khan, "Big data analytics for micro-seismic monitoring," in *Proc. Offshore Technol. Conf. Asia*, 2018, pp. 20–23.

[21] A. Alfaleh, Y. Wang, B. Yan, J. Killough, H. Song, and C. Wei, "Topological data analysis to solve big data problem in reservoir engineering: Application to inverted 4D seismic data," in *Proc. SPE Annu. Tech. Conf. Exhib.*, 2015, pp. 3547–3581.

[22] R. Roden, "Seismic interpretation in the age of big data," in *Proc. SEG Tech. Program Expanded Abstr.*, Sep. 2016, pp. 4911–4915.

[23] T. Olneva, D. Kuzmin, S. Rasskazova, and A. Timirgalin, "Big data approach for geological study of the big region west siberia," in *Proc. SPE Annu. Tech. Conf. Exhib.*, 2018, pp. 24–26.

[24] W. Duffy, J. Rigg, and E. Maidla, "Efficiency improvement in the bakken realized through drilling data processing automation and the recognition and standardization of best safe practices," in *Proc. SPE/IADC Drilling Conf. Exhib.*, 2017, pp. 477–486.

[25] E. Maidla, W. Maidla, J. Rigg, M. Crumrine, and P. Wolf-Zoellner, "Drilling analysis using big data has been misused and abused," in *Proc. IADC/SPE Drilling Conf. Exhib.*, 2018, pp. 1–25.

[26] Q. Yin, J. Yang, B. Zhou, M. Jiang, X. Chen, C. Fu, L. Yan, L. Li, Y. Li, and Z. Liu, "Improve the drilling operations efficiency by the big data mining of real-time logging," in *Proc. SPE/IADC Middle East Drilling Technol. Conf. Exhib.*, 2018, pp. 1–12.

[27] O. Bello, D. Yang, S. Lazarus, X. S. Wang, and T. Denney, "Next generation downhole big data platform for dynamic data-driven well and reservoir management," in *Proc. SPE Reservoir Characterisation Simul. Conf. Exhib.*, 2017, pp. 884–921.

[28] M. R. Brule, "The data reservoir: How big data technologies advance data management and analytics in E&P," in *Proc. SPE Digit. Energy Conf.*, 2015, pp. 1–7.

[29] S. A. Haghighat, S. D. Mohaghegh, V. Gholami, A. Shahkarami, and D. A. Moreno, "Using big data and smart field technology for detecting leakage in a CO2 storage project," in *Proc. SPE Annu. Tech. Conf. Exhib.*, vol. 1, 2013, pp. 815–821.

[30] C. Chelmis, J. Zhao, V. S. Sorathia, S. Agarwal, and V. Prasanna, "Semiautomatic, semantic assistance to manual curation of data in smart oil fields," in *Proc. SPE Western Regional Meeting*, 2012, pp. 196–213.

[31] A. Lin, "Principles of big data algorithms and application for unconventional oil and gas resources," in *Proc. SPE Large Scale Comput. Big Data Challenges Reservoir Simul. Conf. Exhib.*, 2014, pp. 87–95.

[32] E. Udegbe, E. Morgan, and S. Srinivasan, "From face detection to fractured reservoir characterization: Big data analytics for restimulation candidate selection," in *Proc. SPE Annu. Tech. Conf. Exhib.*, 201, pp. 1–20.

[33] J. Xiao and X. Sun, "Big data analytics drive EOR projects," in *Proc. SPE Offshore Eur. Conf. Exhibit.*, 2017, pp. 5–8.

[34] D. Seemann, M. Williamson, and S. Hasan, "Improving resevoir management through big data technologies," in *Proc. of SPE Intelligent Energy International*, 2013, pp. 837–847.

[35] B. T. Rollins, A. Broussard, B. Cummins, A. Smiley, and T. Eason, "Continental production allocation and analysis through big data," in *Proc. 5th Unconventional Resour. Technol. Conf.*, 2017, pp. 1–8.

[36] N. P. Sarapulov and R. A. Khabibullin, "Application of big data tools for unstructured data analysis to improve ESP operation efficiency," in *Proc. SPE Russian Petroleum Technol. Conf.*, 2017, pp. 1–11.

[37] S. Gupta, L. Saputelli, and M. Nikolaou, "Big data analytics workflow to safeguard ESP operations in real-time," in *Proc. SPE North Amer. Artif. Lift Conf. Exhib.*, Oct. 2016, pp. 1–14.

[38] T. Palmer and M. Turland, "Proactive rod pump optimization: Leveraging big data to accelerate and improve operations," in *Proc. SPE North Amer. Artif. Lift Conf. Exhibit.*, 2016, pp. 1–15.
[39] M. Ockree, K. G. Brown, J. Frantz, M. Deasy, and R. John, "Integrating big data analytics into development planning optimization," in *Proc. SPE/AAPG Eastern Regional Meeting*, 2018, pp. 1–16.
[40] L. Cadei, M. Montini, F. Landi, F. Porcelli, V. Michetti, M. Origgi, M. Tonegutti, and S. Duranton, "Big data advanced anlytics to forecast operational upsets in upstream production system," in *Proc. Abu Dhabi Int. Petroleum Exhib. Conf.*, 2018, pp. 1–14.
[41] M. Song and X. Zhou, "A casing damage prediction method based on principal component analysis and gradient boosting decision tree algorithm," in *Proc. SPE Middle East Oil Gas Show Conf.*, 2019, pp. 1–11.
[42] M. Giuliani, G. Camarda, M. Montini, L. Cadei, A. Bianco, A. Shokry, P. Baraldi, and E. Zio, "Flaring events prediction and prevention through advanced big data analytics and machine learning algorithms," in *Proc. Offshore Medit. Conf. Exhib.*, no. 1, 2019, pp. 1–17.
[43] A. Mohamed, M. S. Hamdi, and S. Tahar, "A machine learning approach for big data in oil and gas pipelines," in *Proc. 3rd Int. Conf. Future Internet Things Cloud*, Aug. 2015, pp. 585–590.
[44] M. Layouni, M. S. Hamdi, and S. Tahar, "Detection and sizing of metal-loss defects in oil and gas pipelines using pattern-adapted wavelets and machine learning," *Appl. Soft Comput.*, vol. 52, pp. 247–261, Mar. 2017.
[45] Z. Liu and Y. Kleiner, "State of the art review of inspection technologies for condition assessment of water pipes," *Measurement*, vol. 46, no. 1, pp. 1–15, Jan. 2013, doi: 10.1016/j.measurement.2012.05.032.
[46] A. Anagnostopoulos, "Big data techniques for ship performance study," in *Proc. 28th Int. Ocean Polar Eng. Conf.*, Jun. 2018, pp. 887–893.
[47] S. W. Park, M. I. Roh, M. J. Oh, S. H. Kim, W. J. Lee, I. I. Kim, and C. Y. Kim, "Estimation model of energy efficiency operational indicator using public data based on big data technology," in *Proc. 28th Int. Ocean Polar Eng. Conf.*, Jun. 2018, pp. 894–897.
[48] M. von Plate, "Big data analytics for prognostic foresight," in *Proc. SPE Intell. Energy Int. Conf. Exhib.*, 2016, pp. 6–8.
[49] R. Brelsford, "Repsol launches big data, AI project at tarragona refinery," *Oil Gas J.*, p. 116, Jun. 2018. [Online]. Available: https://www.ogj.com/refining-processing/refining/operations/article/17296578/repsol-launches-big-data-ai-project-at-tarragona-refinery
[50] D. Khvostichenko and S. Makarychev-Mikhailov, "Effect of fracturing chemicals on well productivity: Avoiding pitfalls in big data analysis," in *Proc. SPE Int. Conf. Exhib. Formation Damage Control*, 2018, pp. 1–13.
[51] A. Bin Mahfoodh, M. Ibrahim, M. Hawi, and K. Hakami, "Introducing a big data system for maintaining well data quality and integrity in a world of heterogeneous environment," in *Proc. SPE Kingdom Saudi Arabia Annu. Tech. Symp. Exhib.*, 2017, pp. 2181–2195.
[52] M. Tarrahi and A. Shadravan, "Advanced big data analytics improves HSE management," in *Proc. SPE Bergen One Day Seminar*, 2016, pp. 1–7.
[53] M. Tarrahi and A. Shadravan, "Intelligent HSE big data analytics platform promotes occupational safety," in *Proc. SPE Annu. Tech. Conf. Exhib.*, 2016, pp. 26–28.
[54] C. B. Pettinger, "Leading indicators, culture and big data: Using your data to eliminate death," in *Proc. ASSE Prof. Develop. Conf. Expo.*, 2014, p. 689.
[55] M. Diara, S. Ngunjiri, A. Brown Maruziak, A. Ben Edet, R. Plenderleith, M. Modrick, and D. Buford, "Prevention and control of emerging infectious disease outbreaks in global oil and gas workplaces," in *Proc. SPE Int. Conf. Health, Saf., Environ.*, vol. 3, 2014, pp. 1–11.
[56] K. Wu, E. W. Bethel, M. Gu, D. Leinweber, and O. Rübela, "A big data approach to analyzing market volatility," *Algorithmic Finance*, vol. 2, nos. 3–4, pp. 241–267, 2013.
[57] Y. Chen, K. He, and G. K. F. Tso, "Forecasting crude oil prices: A deep learning based model," *Procedia Comput. Sci.*, vol. 122, pp. 300–307, Dec. 2017.
[58] K. Bekiroglu, O. Duru, E. Gulay, R. Su, and C. Lagoa, "Predictive analytics of crude oil prices by utilizing the intelligent model search engine," *Appl. Energy*, vol. 228, pp. 2387–2397, Oct. 2018, doi: 10.1016/j.apenergy.2018.07.071.
[59] W. M. Alobaidi, E. A. Alkuam, H. M. Al-Rizzo, and E. Sandgren, "Applications of ultrasonic techniques in oil and gas pipeline industries: A review," *Amer. J. Oper. Res.*, vol. 5, no. 4, pp. 274–287, 2015.
[60] GSMA. (2017). *Mobile Privacy and Big Data Analytics*. [Online]. Available: http://www.gsma.com/publicpolicy/wp-content/uploads/2017/02/GSMA_Saf ety-privacy-and-security-across-the-mobile-ecosystem.pdf
[61] I. Ershaghi, M. A. Ershaghi, and A. Popa, "Data ethics in oil and gas operations," in *Proc. SPE Western Regional Meeting*, 2018, pp. 22–27.
[62] P. Malik, "Governing big data: Principles and practices," *IBM J. Res. Develop.*, vol. 57, nos. 3–4, pp. 1–13, 2013, doi: 10.1147/JRD.2013.2241359.
[63] Y. Ma, H. Wu, L. Wang, B. Huang, R. Ranjan, A. Zomaya, and W. Jie, "Remote sensing big data computing: Challenges and opportunities," *Future Gener. Comput. Syst.*, vol. 51, pp. 47–60, Oct. 2015.
[64] P. Mohammadmoradi, H. M. Moradi, and A. Kantzas, "Data-driven production forecasting of unconventional wells with apache spark," in *Proc. SPE Western Regional Meeting*, Apr. 2018, pp. 1–12.
[65] K. Nagorny, P. Lima-Monteiro, J. Barata, and A. W. Colombo, "Big data analysis in smart manufacturing: A review," *Int. J. Commun., Netw. Syst. Sci.*, vol. 10, no. 03, pp. 31–58, 2017.
[66] F. Amalina, I. A. Targio Hashem, Z. H. Azizul, A. T. Fong, A. Firdaus, M. Imran, and N. B. Anuar, "Blending big data analytics: Review on challenges and a recent study," *IEEE Access*, vol. 8, pp. 3629–3645, 2020.
[67] H. H. Alkinani, A. T. T. Al-Hameedi, S. Dunn-Norman, R. E. Flori, M. T. Alsaba, and A. S. Amer, "Applications of artificial neural networks in the petroleum industry: A review," in *Proc. SPE Middle East Oil Gas Show Conf. (MEOS)*, 2019, pp. 1–12.
[68] O. Bello, C. Teodoriu, T. Yaqoob, J. Oppelt, J. Holzmann, and A. Obiwanne, "Application of artificial intelligence techniques in drilling system design and operations: A state of the art review and future research pathways," in *Proc. SPE Nigeria Annu. Int. Conf. Exhib.*, 2016, pp. 1–22.
[69] Y. Liu and H. Song, "Levering mobile cloud computing for mobile big data analytics," in *Mobile Big Data* (Lecture Notes on Data Engineering and Communications Technologies), vol. 10. Cham, Switzerland: Springer, 2017, pp. 21–39, doi: 10.1007/978-3-319-67925-9_2.
[70] Fiix Inc. (2019). *Condition Based Maintenance and Monitoring Software*. [Online]. Available: https://www.fiixsoftware.com/condition-based-maintenance/
[71] M. Bennett. (2016). *The Role of Independents in the Oil and Gas Industry*. [Online]. Available: https://www.i-cio.com/strategy/digitalization/item/the-role-of-wearable -tech-in-the-oil-and-gas-industry
[72] D. Singh and C. K. Reddy, "A survey on platforms for big data analytics," *J. Big Data*, vol. 2, no. 1, pp. 1–20, Dec. 2015.
[73] H. G. Russell, "What is high performance computing?" *Concrete Products*, vol. 102, no. 1, p. 94, 1999. [Online]. Available: http://insidehpc.com/hpc-basic-training/what-is-hpc/
[74] M. A. S. Netto, R. L. F. Cunha, and N. Sultanum, "Deciding when and how to move HPC jobs to the cloud," *Computer*, vol. 48, no. 11, pp. 86–89, Nov. 2015.
[75] *High Performance Computing Enhances Oil and Gas Discovery*. Accessed: Feb. 13, 2020. [Online]. Available: https://www.intel.ca/content/www/ca/en/energy/solutions/high-performance-computing.html
[76] R. Bekkerman, M. Bilenko, and J. Langford, *Scaling Up Machine Learning*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[77] *Intel Xeon Processor E7-8890 v4*. Accessed: Feb. 13, 2020. [Online]. Available: https://ark.intel.com/content/www/us/en/ark/products/93790/intel-xeon-processor-e7-8890-v4-60m-cache-2-20-ghz.html
[78] N. Lopes and B. Ribeiro, "GPU machine learning library (GPUMLib)," *Int. J. Comput. Inf. Syst. Ind. Manage. Appl.*, vol. 3, pp. 15–36, Mar. 2015. [Online]. Available: https://www.mathworks.com/help/parallel-computing/decide-when-to-use-parfor.html
[79] *Machine Learning and Analytics: CUDA Primitives Power Data Science on GPUs*. Accessed: Feb. 13, 2020. [Online]. Available: https://developer.nvidia.com/machine-learning
[80] *Decide When to Use Parfor*. Accessed: Feb. 13, 2020. [Online]. Available: https://www.mathworks.com/help/parallel-computing/decide-when-to-use-parfor.html
[81] A. Shawahna, S. M. Sait, and A. El-Maleh, "FPGA-based accelerators of deep learning networks for learning and classification: A review," *IEEE Access*, vol. 7, pp. 7823–7859, 2019.
[82] G. Batra, Z. Jacobson, S. Madhav, A. Queirolo, and N. Santhanam, "Artificial-intelligence hardware: New opportunities for semiconductor companies," McKinsey & Company, New York, NY, USA, Tech. Rep., Jan. 2018.
[83] *Hitting the Accelerator: The Next Generation of Machine-Learning Chips*, Deloitte, New York, NY, USA, 2017.
[84] L. Du and Y. Du, *Hardware Accelerator Design for Machine Learning*. London, U.K.: IntechOpen, 2018.
[85] T. Jia, T. Guo, X. Wang, D. Zhao, C. Wang, Z. Zhang, S. Lei, W. Liu, H. Liu, and X. Li, "Mixed natural gas online recognition device based on a neural network algorithm implemented by an FPGA," *Sensors*, vol. 19, no. 9, p. 2090, 2019.

[86] P. Jawandhiya, "Hardware design for machine learning," *Int. J. Artif. Intell. Appl.*, vol. 9, no. 1, pp. 63–84, Jan. 2018.
[87] *Cloud TPU: Empowering Businesses With Google Cloud AI*. Accessed: Feb. 13, 2020. [Online]. Available: https://cloud.google.com/tpu/
[88] (2019). *TPU Based Dev Board*. [Online]. Available: https://coral.withgoogle.com/products/dev-board/
[89] *Beyond Today's AI: New Algorithmic Approaches Emulate the Human Brain's Interactions With the World*. Accessed: Feb. 13, 2020. [Online]. Available: https://www.intel.ca/content/www/ca/en/research/neuromorphic-computing.html
[90] A. Oussous, F.-Z. Benjelloun, A. Ait Lahcen, and S. Belfkih, "Big data technologies: A survey," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 30, no. 4, pp. 431–448, Oct. 2018, doi: 10.1016/j.jksuci.2017.06.001.
[91] J. Feblowitz, "The big deal about big data in upstream oil and gas," IDC Energy Insights, Framingham, MA, USA, Tech. Rep., Oct. 2012, p. 11.
[92] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of 'big data' on cloud computing: Review and open research issues," *Inf. Syst.*, vol. 47, pp. 98–115, Jan. 2015, doi: 10.1016/j.is.2014.07.006.
[93] S. Datta, K. Bhaduri, C. Giannella, R. Wolff, and H. Kargupta, "Distributed data mining in Peer-to-Peer networks," *IEEE Internet Comput.*, vol. 10, no. 4, pp. 18–26, Jul. 2006.
[94] B. O. Babaoglu and M. Marzolla, "Escape from the data center: The promise of Peer-to-Peer cloud computing scattering the clouds," *IEEE Spectr.*, to be published. [Online]. Available: https://spectrum.ieee.org/computing/networks/escape-from-the-data-cente r-the-promise-of-peertopeer-cloud-computing
[95] M. R. Ghazi and D. Gangodkar, "Hadoop, MapReduce and HDFS: A developers perspective," in *Proc. Int. Conf. Intell. Comput., Commun. Converg.*, vol. 48. Amsterdam, The Netherlands: Elsevier, 2015, pp. 45–50. [Online]. Available: http://dx.doi.org/10.1016/j.procs.2015.04.108
[96] (2016). *Managing Hadoop Projects: What You Need to Know to Succeed*. [Online]. Available: https://searchdatamanagement.techtarget.com/definition/Apache-Hadoop-YA RN-Yet-Another-Resource-Negotiator
[97] L. A. Rahim, K. M. Kudiri, and S. Bahattacharjee, "Framework for parallelisation on big data," *PLoS ONE*, vol. 14, no. 5, 2019, Art. no. e0214044.
[98] J. Yu, H. Pang, Z. Mei, D. Song, G. Zhu, H. Chen, and L. Wang, "Study on geography information OLAP and data mining system based on Hadoop," in *Proc. IOP Conf. Earth Environ. Sci.*, vol. 252, Jul. 2019, Art. no. 052093.
[99] F. Ming, S. Guannan, and L. Shuaishuai, "Research on multidimensional analysis method of drilling information based on Hadoop," in *Proc. 3rd IEEE Int. Conf. Comput. Commun. (ICCC)*, Dec. 2017, pp. 2319–2322.
[100] S. Chaudhuri and U. Dayal, "An overview of data warehousing and OLAP technology," *ACM SIGMOD Rec.*, vol. 26, no. 1, pp. 65–74, Mar. 1997.
[101] X. Yan, Z. Zhu, and Q. Wu, "Intelligent inversion method for prestack seismic big data based on MapReduce," *Comput. Geosci.*, vol. 110, pp. 81–89, Jan. 2018, doi: 10.1016/j.cageo.2017.10.002.
[102] M. Zhonghua, "Seismic data attribute extraction based on Hadoop platform," in *Proc. IEEE 2nd Int. Conf. Cloud Comput. Big Data Anal. (ICCCBDA)*, Apr. 2017, pp. 180–184.
[103] P. Lathar, "Big data analysis: Apache spark perspective Pankaj," *Int. J. Tech. Innov. Mod. Eng. Sci.*, vol. 4, p. 2585, Sep. 2018.
[104] S. A. Jacobs and A. Dagnino, "Large-scale industrial alarm reduction and critical events mining using graph analytics on spark," in *Proc. IEEE 2nd Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Mar. 2016, pp. 66–71.
[105] Y. Yan, L. Huang, and L. Yi, "Is apache spark scalable to seismic data analytics and computations?" in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Oct. 2015, pp. 2036–2045.
[106] O. Bello, D. S. Bale, L. Yang, D. Yang, A. Kb, M. Lajith, and S. Lazarus, "Integrating downhole temperature sensing datasets and visual analytics for improved gas lift well surveillance," in *Proc. SPE Annu. Tech. Conf. Exhib.*, 2018, pp. 1–14.
[107] M. H. ur Rehman, I. Yaqoob, K. Salah, M. Imran, P. P. Jayaraman, and C. Perera, "The role of big data analytics in industrial Internet of Things," *Future Gener. Comput. Syst.*, vol. 99, pp. 247–259, Oct. 2019, doi: 10.1016/j.future.2019.04.020.
[108] C.-W. Lu, C.-M. Hsieh, C.-H. Chang, and C.-T. Yang, "An improvement to data service in cloud computing with content sensitive transaction analysis and adaptation," in *Proc. IEEE 37th Annu. Comput. Softw. Appl. Conf. Workshops*, Jul. 2013, pp. 463–468.
[109] F. Banaie, J. Misic, V. B. Misic, M. H. Yaghmaee Moghaddam, and S. A. Hosseini Seno, "Performance analysis of multithreaded IoT gateway," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3143–3155, Apr. 2019.
[110] K. H. Abdulkareem, M. A. Mohammed, S. Shamini, M. N. Al-Mhiqani, A. A. Mutlag, and S. A. Mostafa, "A review of fog computing and machine learning: Concepts, applications, challenges, and open issues," *IEEE Access*, vol. 7, pp. 153123–153140, 2019.
[111] M. Mukherjee, L. Shu, and D. Wang, "Survey of fog computing: Fundamental, network applications, and research challenges," *IEEE Commun. Surveys Tutr.*, vol. 20, no. 3, pp. 1826–1857, 3rd Quart., 2018.
[112] *Smart Factory Solution*. Accessed: Feb. 13, 2020. [Online]. Available: https://smartplants.io/smart-factory
[113] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016.
[114] M. Finnegan. (2013). *Boeing 787s to Create Half a Terabyte of Data Per Flight*. [Online]. Available: https://www.computerworld.com/article/3417915/boeing-787s-to-create-hal f-a-terabyte-of-data-per-flight–says-virgin-atlantic.html
[115] F. Sanfilippo and K. Y. Pettersen, "A sensor fusion wearable health-monitoring system with haptic feedback," in *Proc. 11th Int. Conf. Innov. Inf. Technol. (IIT)*, Nov. 2015, pp. 262–266.
[116] H. Wang, D. Peng, W. Wang, H. Sharif, H.-H. Chen, and A. Khoynezhad, "Resource-aware secure ECG healthcare monitoring through body sensor networks," *IEEE Wireless Commun.*, vol. 17, no. 1, pp. 12–19, Feb. 2010.
[117] P. Tedeschi and S. Sciancalepore, "Edge and fog computing in critical infrastructures: Analysis, security threats, and research challenges," in *Proc. IEEE Eur. Symp. Secur. Privacy Workshops (EuroS&PW)*, Jun. 2019, pp. 1–10.
[118] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. MCC workshop Mobile cloud Comput.*, 2012, pp. 13–16.
[119] S. Khan, S. Parkinson, and Y. Qin, "Fog computing security: A review of current applications and security solutions," *J. Cloud Comput.*, vol. 6, no. 1, p. 19, Dec. 2017.
[120] Y. Yao, X. Chang, J. Misic, and V. Misic, "Reliable and secure vehicular fog service provision," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 734–743, Feb. 2019.
[121] A. M. Goscinski, Z. Tari, and I. A. Aziz, "Fog computing as a critical link between a central cloud and IoT in support of fast discovery of new hydrocarbon reservoirs," in *Proc. Int. Conf. Mobile Netw. Manage.* Springer, 2018. [Online]. Available: http://link.springer.com/10.1007/978-3-319-90775-8
[122] *Massively Accelerated Analytics and Data Science*. Accessed: Feb. 13, 2020. [Online]. Available: https://www.omnisci.com/
[123] *Red Zone Monitoring for Drilling Rigs*. Accessed: Feb. 13, 2020. [Online]. Available: https://www.rolloos.com/en/solutions/cctv/red-zone-monitoring/
[124] M. Chibba and A. Cavoukian, "Privacy, consumer trust and big data: Privacy by design and the 3 C'S," in *Proc. ITU Kaleidoscope: Trust Inf. Soc.*, Dec. 2015, pp. 1–5.
[125] A. Dev Mishra and Y. Beer Singh, "Big data analytics for security and privacy challenges," in *Proc. Int. Conf. Comput., Commun. Autom. (ICCCA)*, Apr. 2016, pp. 50–53.
[126] K. Abouelmehdi, A. Beni-Hessane, and H. Khaloufi, "Big healthcare data: Preserving security and privacy," *J. Big Data*, vol. 5, no. 1, pp. 1–18, Dec. 2018, doi: 10.1186/s40537-017-0110-7.
[127] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era," *IEEE Netw.*, vol. 28, no. 4, pp. 46–50, Jul. 2014.
[128] K. S. Shaw, *GPS Tracking of Employee Devices: How Much is Too Much?*. OnLabor, 2017, pp. 2017–2019. [Online]. Available: https://onlabor.org/gps-tracking-of-employee-devices-how-much-is-too-much/
[129] G. Vega-Gorgojo, R. Fjellheim, D. Roman, R. Akerkar, and A. Waaler, "Big data in the oil & gas upstream industry-a case study on the Norwegian continental shelf," *Oil Gas Eur. Mag.*, vol. 42, no. 2, pp. 67–77, 2016.
[130] K. J. Fietkiewicz and M. Henkel, "Privacy protecting fitness trackers: An oxymoron or soon to be reality?" in *Proc. Int. Conf. Social Comput. Social Media, User Exper. Behav. (SCSM)*, in Lecture Notes in Computer Science, vol. 10913. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-319-91521-0_31.

[131] M. Girard, "Canada needs standards to support big data analytics," Centre Int. Governance Innov., Waterloo, ON, Canada, Tech. Rep. 145, Dec. 2018.
[132] R. Jain. (2018). *It's Time to Establish Big Data Standards*. [Online]. Available: https://www.oreilly.com/ideas/its-time-to-establish-big-data-standards
[133] *Digital Single Market-Policy-Big Data*. Accessed: Feb. 13, 2020. [Online]. Available: https://ec.europa.eu/digital-single-market/en/big-data
[134] K. Yang, Q. Han, H. Li, K. Zheng, Z. Su, and X. Shen, "An efficient and fine-grained big data access control scheme with privacy-preserving policy," *IEEE Internet Things J.*, vol. 4, no. 2, pp. 563–571, Apr. 2017.
[135] S. Yu, M. Liu, W. Dou, X. Liu, and S. Zhou, "Networking for big data: A survey," *IEEE Commun. Surveys Tutr.*, vol. 19, no. 1, pp. 531–549, 1st Quart., 2017.
[136] S. Richardson and D. Mackinnon, "Left to their own devices? Privacy implications of wearable technology in Canadian workplaces," Surveill. Stud. Centre, Queen's Univ., Kingston, ON, Canada, Tech. Rep., 2017.
[137] O. Arias, J. Wurm, K. Hoang, and Y. Jin, "Privacy and security in Internet of Things and wearable devices," *IEEE Trans. Multi-Scale Comput. Syst.*, vol. 1, no. 2, pp. 99–109, Apr. 2015.
[138] D. Mussington, B. Dupont, and C. Leuprecht, "Governing cyber security in Canada, Australia and the United States," Centre Int. Governance Innov., Waterloo, ON, Canada, Special Rep., Apr. 2018. [Online]. Available: https://www.cigionline.org/publications/governing-cyber-security-canada-australia-and-united-states
[139] H. Peters, "Defense-in-depth: Cybersecurity in the natural gas & oil industry," Natural Gas Council, USA, Tech. Rep., Oct. 2018.
[140] A. Humayed, J. Lin, F. Li, and B. Luo, "Cyber-physical systems security—A survey," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1802–1831, Dec. 2017.
[141] M. Sharma, F. Gebali, and H. Elmiligi, "3-dimensional analysis of cyber-physical systems attacks," in *Proc. 4th Int. Conf. Comput. Commun. Autom. (ICCCA)*, Dec. 2018, pp. 1–5.
[142] J. Srinivas, A. K. Das, and N. Kumar, "Government regulations in cyber security: Framework, standards and recommendations," *Future Gener. Comput. Syst.*, vol. 92, pp. 178–188, Mar. 2019, doi: 10.1016/j.future.2018.09.063.
[143] (2011). *Global Energy Cyberattacks: Night Dragon*. [Online]. Available: https://www.mcafee.com/wp-content/uploads/2011/02/McAfee_NightDragon_wp _draft_to_customersv1-1.pdf
[144] S. Swanbeck. (2015). *Coast Guard Commandant Addresses Cybersecurity Vulnerabilities on Offshore Oil Rigs*. [Online]. Available: https://www.csis.org/blogs/strategic-technologies-blog/coast-guard-comm andant-addresses-cybersecurity-vulnerabilitie
[145] (2017). *Cybersecurity in Energy: The Implications of a Security Breach on an Oil Gas Company Classifying Oil Gas Organizations Rise of Cyber Threats Targeting Oil Gas Organizations*. [Online]. Available: https://energy.sia-partners.com/20170727/cybersecurity_in_energy_implic ations_of_a_security_breach_on_an_og_company
[146] T. Haidar. (2016). *Digital Trenches: On the Front Lines of the Cyber War*. [Online]. Available: https://www.oilandgasiq.com/oil-gas/whitepapers/special-report-digital-trenches-on-the-front-lines
[147] M. Geli. (2015). *Black Hat Amsterdam: Oil & Gas Cyber-Vulnerabilities*. [Online]. Available: https://www.scmagazineuk.com/black-hat-amsterdam-oil-gas-cyber-vulnerabilities/article/1479359
[148] C. Krauss. *Cyberattack Shows Vulnerability of Gas Pipeline Network*. [Online]. Available: https://www.nytimes.com/2018/04/04/business/energy-environment/pipeline-cyberattack.html
[149] P. Smith. (2019). *Operational Visibility Boosts Cyber Security for Midstream Operators*. [Online]. Available: https://securityboulevard.com/2019/05/operational-visibility-boosts-cyb er-security-for-midstream-operators/
[150] J. R. Dancy and V. A. Dancy, "Terrorism and oil & gas pipeline infrastructure: Vulnerability and potential liability for cybersecurity attacks," *ONE J.*, vol. 2, no. 6, p. 38, 2017. [Online]. Available: http://digitalcommons.law.ou.edu/onej%0Ahttp://digitalcommons.law.ou.edu/onej
[151] (2013). *Cyberattack Threat in Canada's Oil Patch Raises Risk of Disruptions, Stolen Data*. [Online]. Available: https://business.financialpost.com/commodities/energy/cyberattack-threat-in-canadas-oil-patch-raises-risk-of-disruptions-stolen-data
[152] M. Vamburkar, N. S. Malik, and R. Collins. (2018). *Energy Transfer Says 'Cyber Attack' Shut Pipeline Data System*. [Online]. Available: https://www.bloomberg.com/news/articles/2018-04-02/energy-transfer-says -cyber-attack-shut-pipeline-data-system
[153] P. Muncaster. (2018). *US Gas Pipelines Hit by Cyber-Attack*. [Online]. Available: https://www.infosecurity-magazine.com/news/us-gas-pipelines-hit-by-cyberattack/
[154] A. Slaughter, P. Zonneveld, and T. Shattuck, "Refining at risk: Securing downstream assets from cybersecurity threats," Deloitte Insights, New York, NY, USA, Tech. Rep., Nov. 2017. [Online]. Available: https://www2.deloitte.com/insights/us/en/industry/oil-and-gas/securing-downstream-oil-and-gas-assets-cybersecurity-threats.html
[155] P. K. Kerr, J. Rollins, and C. A. Theohary, *The Stuxnet Computer Worm: Harbinger of an Emerging Warfare Capability*. Washington, DC, USA: Congressional Research Service, 2010.
[156] Y. Tang, Q. Chen, M. Li, Q. Wang, M. Ni, and X. Fu, "Challenge and evolution of cyber attacks in cyber physical power system," in *Proc. IEEE PES Asia–Pacific Power Energy Eng. Conf. (APPEEC)*, Oct. 2016, pp. 857–862.
[157] O. Alsaadoun, "A cybersecurity prospective on industry 4.0: Enabler role of identity and access management," in *Proc. Int. Petroleum Technol. Conf.*, 2019, pp. 1–13.
[158] Oracle. (2010). *Strengthening Information Technology Security for the Nation's Energy Infrastructure*. [Online]. Available: http://www.oracle.com/us/solutions/wp-esso-oil-gas-207224.pdf
[159] J. Petters. (2018). *What is a Brute Force Attack?*. [Online]. Available: https://www.cloudways.com/blog/what-is-brute-force-attack/
[160] L. O. Donnell. (2019). *Oil and Gas Firms Targeted by New LYCEUM Threat Group*. [Online]. Available: https://threatpost.com/oil-and-gas-firms-targeted-by-new-lyceum-threat-group/147705/
[161] M. Lee, K. Lee, J. Shim, S.-J. Cho, and J. Choi, "Security threat on wearable services: Empirical study using a commercial smartband," in *Proc. IEEE Int. Conf. Consum. Electron.-Asia (ICCE-Asia)*, Oct. 2016, pp. 1–5.
[162] J. Joshi, S. Mittal, B. Birdi, R. Kumar, D. S. Kurian, S. Mukherjee, and P. Awasthi, "Secure and wearable computing in WBANs," in *Proc. Int. Conf. Inf. Commun. Technol. (ICICTM)*, May 2016, pp. 65–70.
[163] *Encryption Mandate: Clear and Present Danger for the Energy Sector*, Infosecur. Mag., London, U.K., 2017. [Online]. Available: https://www.infosecurity-magazine.com/white-papers/encryption-mandate/
[164] *Ciphertex Data Security*. Accessed: Feb. 13, 2020. [Online]. Available: https://www.ciphertex.com/oil-gas
[165] *Thales eSecurity*. Accessed: Feb. 13, 2020. [Online]. Available: https://www.thalesesecurity.com/
[166] J. Askar. *Keeping Data Secure in the Oil and Gas Industry*. Accessed: Feb. 13, 2020. [Online]. Available: https://www.infosecurity-magazine.com/opinions/secure-oil-gas/
[167] L. Musthaler. (2015). *Are you Overlooking Tokenization as a Data Security Measure? Tokenization is Another Way to Obscure Data and Protect it in the Event of a Breach*. [Online]. Available: https://www.networkworld.com/article/3002307/are-you-overlooking-tokeni zation-as-a-data-security-measure.html
[168] M. Zborowski, "As oil and gas data multiply, so do the cybersecurity threats," *J. Petroleum Technol.* Sep. 2019. [Online]. Available: https://pubs.spe.org/en/jpt/jpt-article-detail/?art=5992
[169] K. Wong, C. Dillabaugh, N. Seddigh, and B. Nandy, "Enhancing suricata intrusion detection system for cyber security in SCADA networks," in *Proc. IEEE 30th Can. Conf. Electr. Comput. Eng. (CCECE)*, Apr. 2017, pp. 1–5.
[170] J. Luswata, P. Zavarsky, B. Swar, and D. Zvabva, "Analysis of SCADA security using penetration testing: A case study on modbus TCP protocol," in *Proc. 29th Biennial Symp. Commun. (BSC)*, Jun. 2018, pp. 1–5.
[171] *Dragos Industrial Cybersecurity Platform*. Accessed: Feb. 13, 2020. [Online]. Available: https://dragos.com/platform/
[172] *ICS CyberVision: Why Sentryo?*. Accessed: Feb. 13, 2020. [Online]. Available: https://www.sentryo.net/why-sentryo/
[173] *CyberX for Oil and Gas*. Accessed: Feb. 13, 2020. [Online]. Available: https://cyberx-labs.com/oil-and-gas/
[174] (Aug. 2019). *Assessing the Threats, Risks, and Activity Groups Affecting the Global Oil and Gas Industry*. [Online]. Available: https://dragos.com/wp-content/uploads/Dragos-Oil-and-Gas-Threat-Perspec tive-2019.pdf

TRUNG NGUYEN received the B.Eng. degree from the Ho Chi Minh City University of Technology (HCMUT), Vietnam, and the M.Eng. and Ph.D. degrees from MUN, Canada. He is currently a Postdoctoral Fellow at Memorial University of Newfoundland (MUN). His current research interests include artificial intelligence/machine learning, autonomous vehicles, mechatronics, and digitalization of the oil and gas industry.

RAYMOND G. GOSINE received the bachelor's degree from Memorial University and the Ph.D. from Cambridge University. Subsequently, he held teaching and research positions at Cambridge University, The University of British Columbia, and Memorial University. These appointments included an NSERC Chair in Industrial Automation at UBC (supported by BC Packers) and the J.I. Clark Chair of Intelligent Systems for Operations in Harsh Environments at Memorial University (supported by C-CORE). He is currently an Associate Vice-President (Research) at Memorial University of Newfoundland. His main research is in the areas of intelligent systems, robotics, and automation with a particular interest in the applications of these technologies to natural resource industries. He is also interested in the broader implications of advanced technologies, and recently served as the Chair of a Public Review Panel that considered the scientific, socio-economic, public policy, regulatory, environmental, and public health issues associated with unconventional oil and gas development in Western Newfoundland.

PETER WARRIAN is currently a Senior Research Fellow at the Munk School of Global Affairs and Public Policy, University of Toronto. He is Canada's leading academic expert on the steel industry. He was formerly the Research Director of the United Steelworkers of America and Chief Economist of the Province of Ontario. His current researches are on knowledge networks, supply chains, and digital manufacturing. As a member of the Innovation Systems Research Network (ISRN), funded by the Social Sciences and Humanities Research Council of Canada, he has worked on the interface between the steel industry and the auto industry, particularly in the area of lightweight materials and the interaction of software and advanced materials.