A glance about the Big Data Analytics in the Oil&Gas industry

DX.DOI.ORG/10.19199/2020.2.1121-9041.036

A glance about the Big Data Analytics in the Oil&Gas industry

Alberto Manuel Garcia Navarro*
Alfonso Capozzoli**
Vera Rocca*
Raffaele Romagnoli*

* Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino
** Department of Energy "Galileo Ferraris", Politecnico di Torino, Italy

Corresponding author: Vera Rocca

This article aims to provide a glance about the current and future applications of Big Data Analytics (BDA) within the Oil&Gas industry, which is searching for an improvement in data managing equal to the improvement in techniques and technologies required by current and future hydrocarbon production. The connection and compatibility of BDA with the hydrocarbon sector is analyzed, highlighting its multiple applications, most of them already present in an early stage. Special emphasis is made on the description of case studies (related to prospection, drilling and reservoir production optimization) looking at the new challenges the industry needs to overcome in order to get the maximum of this tool, extract value and increasing profits. New trends and forecast of possible areas to work on are also discussed.
***Keywords**: Big Data, Big Data Analytics, machine learning, data management, Oil&Gas upstream industry.*

***Applicazioni dei Big Data Analytics nel settore Oil&Gas**. Questo articolo ha lo scopo di fornire una panoramica sulle applicazioni attuali e future dei Big Data Analytics (BDA) nel settore Oil&Gas, costantemente alla ricerca di una gestione dei dati ottimizzata per le nuove tecniche e tecnologie adottate per la produzione di idrocarburi. Viene analizzata la connessione e la compatibilità dei BDA con il settore degli idrocarburi, evidenziandone le molteplici applicazioni, molte delle quali già presenti in fase prototipale. Particolare enfasi è posta sulla descrizione di casi studio (relativi alla prospezione, alla perforazione e all'ottimizzazione della produzione degli idrocarburi) che mettono in evidenza le nuove sfide che l'industria petrolifera deve affrontare per massimizzare le applicazioni dei BDA. Vengono inoltre discusse le nuove tendenze e le possibili future aree di applicazione.*
***Parole chiave**: Big Data, Big Data Analytics, machine learning, data management, settore upstream dell'Oil&Gas.*

1. Introduction

Digital era faces a constantly increasing mole of data never available before. Almost 20% of all current worldwide available data was generated only in 2019, and an exponential increase is expected in the near future (O'Dea, 2020). The enormous intrinsic value of data is strongly dependent on the capacity of gather, store, process and analyze them for extracting high-quality information. Big Data Analytics (BDA) in conjunction with High Performance Computing (HPC) have been offering an effective response considering that only in 2019 the global big data market reached 49 billion U.S. dollars and by 2027 it is expected to twofold up to 103 (Holst, 2020). It comes to no surprise when talking about Big Data (BD) and BDA in the Oil&Gas sector, which accounted alone in 2019 around 3.8% of the global economy GDP (Gross Domestic Product) (Investopedia, 2020).

Data acquisition and interpretation have always been the core of different disciplines involved in Oil&Gas upstream sectors such as exploration, drilling, reservoir characterization and production optimization. Petroleum industry is not new to the challenges of managing massive data, both in space and over time. Such as example, seismic acquisition for prospection can involve hundreds or thousands square kilometers of investigated surface; real-time data acquisition (with a sampling frequency in the order of milli seconds) and analysis are keys factors for enhancing the efficiency and safety of well drilling operations. Even so, exploration and production from even more complex scenarios (i.e. operating in deep and ultra-deep waters or in harsh conditions, drilling and completing complicate wellbore profiles, increasing the recovery factor in mature assets) have asked in the last decades for advances in techniques and technologies capable of handling the increase amount and variety of gathered data. At the same time, the environmental regulations currently enforced in most countries have imposed more restrictive practices and standards, with consequent improvement in data acquisition technology: such an example, the monitoring of ground movement for subsidence studies is moving from GPS to satellite images. This latter approach is based on synthetic aperture radar (SAR) data acquisition and provides millimeter-scale movements of large zones (Codegone et al, 2016, Giani et al, 2017; Giani et al, 2018; Coti et al, 2018). These new technological solutions could be fully exploited if associated with an adequate improvement in data storage, processing, analysis, application and visualization approaches. Therefore, the oil industry has been approaching BD and BDA as a potential solution to the abovementioned challenges, calling for the same breakthrough effects that this branch of science has been gushing in other domains of applications. In fact, BDA have shown very interesting and promising capacity in uncovering hidden patterns and models, boosting or boosted by the always-invoked multidisciplinary approach.

Massive investment of the oil industry in BD and BDA sector accounts for 3.8 billion dollars on 2018 and it is expected to pose a CAGR (Compound Annual Growth Rate) of over 15% till 2023 (Technavio, 2019). The major players from both the technology providers and the oil industry are involved, as reported by the Oil&Gas Intelligence Center of Global Data on January 2020: Alphabet (Google), Amazon, Arista Networks, Cisco, Cloudera, HP, IBM, Intel, Microsoft, Nvidia and Oracle from one side and ADNOC, BP, Chevron, Conoco Phillips, Equinor, ExxonMobil Gazprom, Rosneft, Repsol, Shell and Woodside from oil industry side. Worthy of pointing out are the efforts of services companies like Schlumberger and Baker Hughes in developing and providing industry-tailored BD solutions. Moreover, from the academic side, numerous high-level educational programs are devoted to the application of BD and BDA approaches in the Oil&Gas sectors, often in collaboration with major oil companies. Among the others, we quote the MSc Petroleum Data Management at the University of Aberdeen in collaboration with Shell, Total and Chevron, and the Petroleum Data Management master at the Institut Français du Pétrole (IFP) in collaboration with Statoil, Total, Schlumberger, Teradata, CVA Engineering, etc.

The scope of the present paper is overviewing the past and current role of BD and BDA in the Oil&Gas industry and forecasting their possible and most promising future applications.

2. Big Data and Big Data Analytics

Big Data encompasses huge data sets that cannot be managed or studied with current regular technologies but, instead, require more robust frameworks based on data analytics to be developed with the help of very fast and accurate processors. Big Data includes unstructured (not organized and text-heavy) and multi-structured data (including different data formats resulting from people/machines interactions) (Mircea and Mihaela, 2014). Big Data Analytics is on the other hand the sets of strategies and tools to deal with them.

BD meets the next five V's criteria (Ishwarappa and Anuradha, 2015): *Volume* as a reference of the enormous quantity of data or information. The data can come from any sensor or data recording tool and they are challenging to be handled due to storage, sustainability, and analysis issues. *Velocity* as a characteristic of the speed of data generation. *Variety* refers to the various types of data, which are generated, stored, and analyzed according to the recording devices or sensors. The formats can be text, image, audio or video. The classification can be done more technically as structured, semi-structured, and unstructured data. *Veracity* as a measure of the quality and usefulness of the available data for analysis and decision-making. *Value* as the importance of the returned value of investments for BD infrastructures. Analyzing huge data sets looks toward revealing the underlying trends and helps forecasting the potential issues or investing opportunities.

BDA has been defined as "*a holistic approach to managing, processing and analyzing the 5V data-related dimensions (i.e. volume, variety, velocity, veracity and value) to create actionable insights in order to deliver sustained value, measure performance and establish competitive advantages*" (Fosso *et al.*, 2015). It is an extension of the processes performed by Business Intelligence Systems, described by H.P. Luhn for the first time at the IBM JOURNAL on 1958: "Business Intelligence (BI) represents the applications and processes to analyze that data using tools such as analytics and data mining".

Traditionally BI was used to abstract (on volume) and clean (only structured) data to perform analysis and correlate variables through models. BDA expands its capacities by including nonlinear methods and unstructured data from massive volumes of information generated at higher velocities, processing them to extract value by identifying hidden patterns, unknown correlations, market trends and customer preferences allowing faster and more precise responses on the Decision Making Process (DMP) as (Searchbusinessanalytics, 2020).

BDA, as part of advanced BI strategies, aims towards answering four major questions about a set of data related to the studied event: *what*, *why*, *if* and *how*. The correspondent analytics areas are descriptive, diagnostic, predictive and prescriptive (Michigan State University, 2019). Common tasks taking part on these processes include: text and data mining; supervised and unsupervised Machine Learning (ML) methods and deep learning algorithms; statistical, stochastic and deterministic analysis (Searchbusinessanalytics, 2020). Figure 1 resumes these 4 approaches, stating from their main objectives (as questions on the right side), and outputs and analysis instruments (on the left side).

Beside the well-known applications for major corporations on improving process efficiencies or managing data, examples of applications of BDA include the banking sector where ML and AI (Artificial Intelligence) could help to detect fraudulent activities by analyzing customers behaviors. Clustering techniques can identify new branch locations according to the demand. In the medical sector the storage and efficient access of patient records are increased by cloud storage and data can be captured and analyzed in real time from wearable devices. Within the disaster management area, BDA can provide tools for evaluate weather parameters, measured at local stations, and correlate them with satellite and radar information to forecast dangerous situations allowing authorities to design emergency plans (Dutta, 2019).

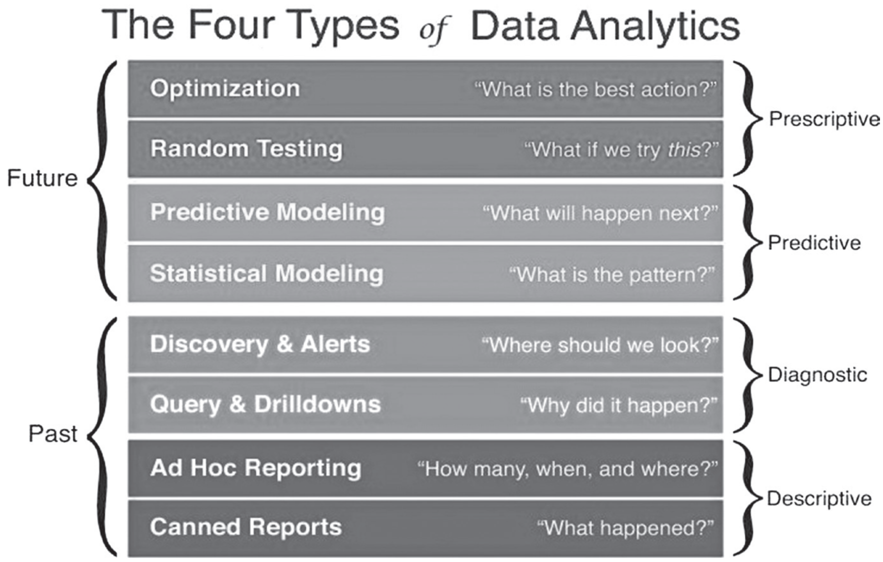

Figure 1. The Four Types of Data Analytics (https://exagoinc.com, 2017).
Le quattro tipologie di Data Analytics.

3. Overview of Big Data Analytics in the O&G Upstream Industry

3.1. Seismic survey

One of the key exploration and characterization technique is the seismic survey, whose interpretation requires great computational capacities along with proper visualization tools. Seismic acquisition and interpretation allow potential targets to be identified during the prospection phase. A proper interpretation of data provide structural (fault systems), stratigraphic (surfaces), petrophysical (such as porosity) and mechanical (elastic properties) information about the investigated formations. Furthermore, the 4D seismic approach opens to more efficient and accurate interpretation opportunities together with a strong increase of data storage, interpretation and visualization. 4D seismic data are 3D data collected at different times over the same area to assess changes in a producing hydrocarbon reservoir with time. ML tools can represent a valid approach for managing and also integrating seismic data with other structured and unstructured data also in real-time from different sources (e.g., drilling and logging operations, geology or lab results) for an interdisciplinary data analysis approach.

Researches conducted by Roden (2016), incorporated Principal Component Analysis – PCA ("a procedure for reducing the dimensionality of the variable space by representing it with a few orthogonal (uncorrelated) variables that capture most of its variability") (ScienceDirect, 2020) with Self-Organizing Maps – SOM ("an unsupervised neural network that reduces the input dimensionality in order to represent its distribution as a map") (ScienceDirect, 2020) to carry out a multi-component seismic analysis for an offshore Gulf of Mexico location, aiding interpreters to effectively discriminate and better comprehend the ad-hoc more relevant data. In a 5-stage process two 2D-SOM were generated incorporating 18 instantaneous seismic attributes from a 3D survey in a PCA analysis, with the first principal component being attenuation and identifying for it five prominent ones (envelope second derivative, envelope modulated phase, trace envelope, average energy and sweetness). The second principal component was flat spots and another four prominent attributes were identified: instantaneous frequency, thin bed indicator, acceleration of phase, dominant frequency. With a 5 X 5 neuron matrix for both SOM, one potential new drilling location (previously undetected) and fluid contacts (subsequently confirmed at wellbore) where clearly highlighted.

Olneva *et al.* (2018) clustered 1D, 2D, and 3D geological maps for West Siberian Petroleum Basin with seismic data (emphasizing the Achimov play with 207 fields discovered), developing two compatible approaches for analysis. One approach, defined as "*from general to particulars*" (i.e. top-down), incorporated drilling data coming from 5000 existing wells with regional geology structures and tectonic elements, and available paleographic charts. The second one, defined as "*from particulars to* *generals*" (i.e. bottom-up), started from regional 2D seismic and geological data, aiming to create a training sample based on seismic geological patterns, counting on a regional database including more than 40.000 km^2 of 3D data. The top-down approach, involving the clustering algorithm K-means, allowed the update and the improvement of the regional clustering model of the play, in agreement with the adopted mathematical model and the local geology. The bottom-up approach, based on pattern recognitions, revealed its usefulness on generating a library of typical seismic images for specific events in the Achimov sequence, mostly associated with turbidites.

3.2. Drilling

The decrease of easy recoverable oil has been accompanied by an increasing complexity of drilling scenarios, i.e. ultra-deep waters, harsh conditions, shale oil and gas reservoirs. A massive adoption of new techniques and tools, such as Visual Analytics and Machine Learning approach, is expected to contribute in the drilling sector for an immediate and effective improvement of both operation safety and efficiency, and consequently cost-reduction (Crooks, 2018). Solving the problem of data distribution and heterogeneity is the first step for integrating in real-time data from logging-wile-drilling logs, drilling mud properties, circulation pressure, drill bit and string tensions, micro seismicity and even lithology. This approach can provide a real improvement in terms of drilling program design and real time operations, to prevent failures or fast-track the schedule.

Johnston and Guichard (2015) demonstrated how the role of BDA can become essential when dealing with structured and unstructured data from different formats, sources and properties. They studied how to reduce the risks associated with drilling operations combining well loggings, drilling and geological formation data from about 350 oil and gas wells in the North UK. The data set contained information about a large geographical area, whose analysis with conventional analysis techniques would have been difficult, or even impossible, for handling and interpreting it in its entirety. The adopted iterative BDA approach allowed the discovery of unexpected correlation between borehole quality and drilling parameters (function of stratigraphy and region) and the identification of predictive optimal drilling parameters to avoid bad hole conditions.

Maidla *et al.* (2018) analyzed the dangers of working with wrong tagged "big data" measurements. The drilling performance improvement (in terms of NPT, Non-Productive Time, and ILT, Invisible Lost Time) was analyzed by combining data from 4 hour morning reports, electronic drilling recorder, cross-plots of weight on bit and differential pressure by implementing a BDA process. Nevertheless, no definitive conclusions were obtained in absence of both high quality controlled data and proper correlations with the physics phenomena occurring during drilling operations. In that case the authors emphasized the aspects of the adopted approach that must be improved: the need of accurate and proper sampling time in data recording, the stringent conditions to be defined as "big data" (at least volume and velocity) and the need to correctly correlate the physics of the drilling process with the corresponding performance analysis.

Duffy *et al.* (2017) improved the drilling rig efficiency by implementing best-safe-practice initiatives which were identified through an Automated Drilling State Detection of Monitoring Service (ADSDMS) based on a BDA process. According to their results, more than 11 days of drilling operation were saved on a single pad of nine wells drilled by the same rig, while the need of performing non-destructive test to locate surface-breaking defects on pipes was reduced by 45%. The results were replicated by Yin *et al.* (2018) with the aim of finding and optimizing invisible non-production time by exploiting collected real-time logging data, mathematical statistics, simulated data and cloud computing.

Machine Learning models, for instance for drill bit wear, are now being integrated into real-time data science workflows as the "Streaming DataOps workflows" developed by TIBCO, allowing cost/risk reductions and improving HSE performances (Vendetti, 2019). Hutchinson *et al.* (2018) worked with data obtained from downhole vibration sensors to characterize the drill string dynamics. The combination of real and simulated data resulted in the development of a drilling automation application. The developed model reduced the risks of drilling failures and the drilling development costs.

3.3. Reservoir production optimization

Production area is not new to BA and BDA (Subrahmanya et al, 2014; Honorio et al, 2015), especially regarding the production optimization. In order to estimate the hydrocarbon reserves and to forecast the reservoir production due to different strategies, a 3D numerical simulation approach is commonly adopted. Reservoir studies requires a multidisciplinary approach to integrate a large number of data related to different disciples: geological, geophysical, petrophysical information, fluid characteristics together with production data (position and completion of wells and produced liquid volumes, pressure and temperature for each well). Models are updated during all reservoir production life via a calibration phase to continuously increase their reliability in forecasting. Furthermore, production forecasting is carried via stochastic risk analysis approaches with further increase of data to be managed (Dake, 1998). Consequently, a massive and systematic adoption of BDA can improve drastically the efficiency of data management, calibration and forecast phases.

A case study of production optimization of unconventional oil and gas reservoirs is reported by Popa *et al.* (2015). Heavy oil reservoirs in San Joaquin Valley (US) were produced via Steam Assisted Gravity Drainage (SAGD) and cyclic steam operations. The authors collected and elaborated different structured and unstructured, static and dynamic types of data from near to 14,200 wells. The proposed BD approach integrated DTS (Distributed Temperature Sensing) data with well completions and geological information. The major obtained improvement was in terms of data integrity checks and monitoring pressure-temperature status of the system. The advantages related to additional high-quality information in real time were shown in terms of well integrity and optimization of SAGD operations and surface facilities reliability.

Chelmis *et al.* (2012) worked also in the area of Artificial Intelligence and Semantic Web techniques to manual curation of data in smart oil fields – a philosophy that encompass stream of data from digital sensors, processing, integration and analysis in an ad-hoc facility or office (Whaley, 2015). In order to deal with the large volumes of unstructured data generated by users, currently stored with no proper metadata, the authors proposed a semi-automatic approach for creating and maintaining a File Naming Convention Ontology, facilitating the data curation process. This ontology was employed to carry out automated and interactive annotation of filenames, generating semantically enriched metadata: ontology and metadata were simultaneously generated and updated from the filenames where all user-supplied terms were processed and annotated. Their approach made use of recent advancements in Semantic Web, Linguistic Processing, Named Entity Recognition and Classification techniques. As the authors stated, the same approach can be applied to other types of datasets (i.e. seismic image processing, interpretation and analysis workflows) by including additional domain specific metadata attributes in the target dataset.

Several case studies refer to the adoption of Automated Decline Analysis (ADA) based on the identification of patterns in production data and the forecast of real-time production performance, while training models on a range of pattern-recognition techniques (Padmanabhan, 2014). Seemann *et al.* (2013) developed a smart forecast and flow method to perform ADA, while Rollins *et al.* (2017) adopted Hadoop as processing tool and Visual Analytics for data presentation based on public data. Furthermore, economics and engineering aspects were simultaneously matched on development stages by Ockree *et al.* (2018). They developed a reservoir optimization via the superposition of production type curves obtained by coupling Artificial Intelligence and economic analysis.

BDA strategies turn out to be effective in improving hydraulic fracturing modeling as shown by Lin (2014). The authors combined the traditional Physical related Computational Operators (PCO) based on mathematical and physical description of the phenomena, with Analytics-related Computational Operators (ACO) used if strong physical models are on their way of definition, such as the model of fracture propagation into the reservoir. ACO aimed to identify trends or rules in data, while PCO solved physical equations. This ambivalent approach proved to be effective in dealing with data uncertainty. Udegbe *et al.* (2017) worked with pattern-recognition tools to detect trends associated with favorable and unfavorable shale-gas well re-stimulation candidates. Texture-based attributes called Haar-like features were adapted and adopted: it is a binary classification framework from real-time face detection. It was used to characterize different rate/production profiles for the candidate wells simulated via a dual-permeability model by modifying fracture, reservoir, and operational parameters. Using a cascaded AdaBoost classifier algorithm, the proposed approach was able to recognize refracturing candidates in the presence of rate erratic fluctuations. The results were validated versus public production data of 17 hydraulically fractured gas wells in Texas, in similar and coherent shale formations. In addition, Joshi *et al.* (2018) used BDA to analyze the micro-seismic data sets and model the fracture propagation patterns during hydraulic fracturing. The authors used the Hadoop platform to manage the massive datasets generated by micro-seismic tools, being able to integrate datasets from exploration, drilling and production operations.

4. BD and DBA: the way ahead

The most promising fields for BDA application starting from the nearest future seem to be related to the development of unconventionals and to the extension of the economic production life of brown fields. Among the unconventionals, both drilling and production from oil or gas shale reservoirs can benefit from a massive application of BDA. Shale minerals are particularly sensitive to time-dependent stability degrade processes occurring within and around borehole because of interaction with drilling fluids. The capability to collect, store, organize, analyze a vast amount of data coming from different sources and in real-time mode can improve operation safety and efficiency. Furthermore, due to their low permeability, shale reservoir production asks for hydraulic fracturing operations: modelling and planning fracturing operation remain a challenging task due to the involved coupled fluid-flow and mechanical phenomena and the complex nature of the formation (Shahid, 2016). Moreover, the modeling of induced fracture propagation and their possible intersection with the natural ones is currently an open issue. Patterns recognition methodologies have showed to be effective tools for facing these issues.

Brown fields are hydrocarbon reservoirs matured to a production plateau or even progressed to a stage of declining production. The extension of their economic production life has the advantage of a deep knowledge of the reservoirs and the involved phenomena due to long-time monitoring, together with all existing facilities. The amount of data (related to the produced volumes of fluids, pressures and temperatures at wells, well logs, lab data, well completion data, etc.) gathered during decades of production history (and so from all-fashion sheet format to the digitalized one) represents also one of the most challenging task to be addressed by means of BDA to increase efficiency in terms of storage, organization and harmonization.

As a general comment, the new generation of sensors is enabling an explosion of gathered data, both considering improved technology fiber-optic sensors or low-cost and so high-density ones. They are and will be adopted by all Oil&Gas sectors (well drilling, reservoir characterization and production, etc.) and will require a constant and continuous technology upgrade of data sorting, organizing, analysis.

The major oil and service companies are moving toward the BD and BDA self-own infrastructures but many difficulties must be overtaken, such as high cost also in terms of time (one for all, trading algorithms is an intensive time-consuming process) and the need of new generation data analysists. Moreover, since most technologies are still in their R&D phase, no dominant design in terms of DB architecture or user interface has been established, even though in the BDA approach some criteria are already settled down. The solution at the moment has been patched by outsourcing but with a big concern about data and systems security, as demonstrated by the case of Saudi Aramco in 2012 and in 2016 (the Shamoon and Shamoon 2 viruses).

5. Conclusions

The increasing complexity of hydrocarbon reserves, in terms of exploration, characterization, drilling and production, has asked for improvement in techniques and technologies with consequent increase of amount and variety of gathered data. Therefore, the oil industry, which is not new in dealing with massive data of different formats and nature, has been approaching BDA as a potential solution for a further improvement in data management. In different branches of application, BD and DBA have been showing their potentiality in terms of data acquisition, storage, processing, analysis and visualization. They are adopted for facing the problem of data distribution and heterogeneity, providing integration of data from different sources, formats and properties to enable effective analysis. They help in data cleaning to obtain high quality data. They uncover hidden patterns and unknown correlations form massive data and between different typology of information not routinely compared. Furthermore, the new visualization approaches intuitively reflect massive data via graphs or tables, conveying information more clearly and effectively.

Successful application of BD and BDA are reported in numerous Oil&Gas upstream sectors, such as well drilling, production optimization, seismic survey. The most promising fields for BA/BDA application starting from the nearest future seem to be related to the extension of the economic production life of brown or mature fields and to the development of unconventional reserves. Among the latter, hydrocarbon shale reservoirs are still related to several open issues about safety and efficiency of well drilling operations and about reliability and efficiency of the hydraulic fracturing modelling. Moreover, the new generation sensors adopted by the majority of Oil&Gas sectors are associated with an explosion of gathered data, both considering improved technology of fiber-optic sensors or low-cost and so high-density ones.

At the moment, oil companies rely on outsourcing and strategic alliances with providers, but a movement toward self-own infrastructures is evident, also induced by security and confidentiality issues.

The future of BA/BDA in Oil&-Gas industry seems to be bright.

Nevertheless, several issues must be considered: to favour multi-disciplinarity, the need of new generation data analysis specialists, the customization and standardization of BA and BDA approaches and criteria for the Oil&Gas industry needs, and finally the usual long-term research and investment time frame for targeting business properly.

References

Chelmis, C., Zhao, J., Sorathia, V.S., Agarwal, S., & Prasanna, V., 2012. *Semiautomatic, semantic assistance to manual curation of data in smart oil fields*. Paper SPE-153271-MS, presented at the SPE Western Regional Meeting, Bakersfield, California 21-23 March 2012. doi: https://doi.org/10.2118/153271-MS

Codegone, G., Rocca, V., Verga, F., Coti, C., 2016. *Subsidence Modeling Validation Through Back Analysis for an Italian Gas Storage Field*. Geotechnical And Geological Engineering, vol. 34 (6), pp. 1749-1763. ISSN 0960-3182, doi: 10.1007/s10706-016-9986-9

Coti, C., Rocca, V. Sacchi, Q., 2018. *Pseudo-elastic response of gas bearing clastic formations: An Italian case study*. EnergiesOpen Access, Volume 11, Issue 9, September, Article number en11092488. ISSN: 19961073, DOI: 10.3390/en11092488

Crooks, E., 2018. *Drillers turn to big data in the hunt for more, cheaper oil*. Financial Times. Retrieved from: https://www.ft.com/content/19234982-0cbb-11e8-8eb7-42f857ea9f09. February 12. Accessed March 20, 2020.

Dake L.P., 1998. *Fundamentals of reservoir engineering*. ISBN 0-444-41830-X. Elsevier science B.V.

Duffy, W., Rigg, J., & Maidla, E., 2017. *Efficiency Improvement in the Bakken Realized through Drilling Data Processing Automation and the Recognition and Standardization of Best Safe Practices*. Paper SPE-184724-MS, presented at the SPE/IADC Drilling Conference and Exhibition, The Hague, The Netherlands, 14-16 March 2017. DOI: https://doi.org/10.2118/184724-MS

Dutta, P., 2019. *Top 20 Best Big Data Applications & Examples in Today's World*. Retrieved at: https://www.ubuntupit.com/best-big-data-applications-in-todays-world. August 14. Accessed May 1, 2020.

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., and Gnanzou, D., 2015. *How 'Big Data' Can Make Big Impact: Findings from a Systematic Review and a Longitudinal Case Study*. International Journal of Production Economics. doi: 10.1016/j.ijpe.2014.12.031

Giani, G., Orsatti, S., Peter, C., Rocca V., 2018. *A coupled fluid flow-geomechanical approach for subsidence numerical simulation*. EnergiesOpen Access, Volume 11, Issue 7, ISSN: 19961073, doi: 10.3390/en11071804

Giani, G.P., Gotta, A., Marzano, F., Rocca, V., 2017. *How to Address Subsidence Evaluation for a Fractured Carbonate Gas Reservoir Through a Multi-disciplinary Approach*. Geotechnical and Geological Engineering. Volume 35, Issue 6, December, Pages 2977-2989. ISSN: 09603182. doi: 10.1007/s10706-017-0296-7

Globaldata, 2020. *Sheer volume of data being created by oil and gas companies driving adoption of big data, says GlobalData*. Retrieved from: https://www.globaldata.com/sheer-volume-data-created-oil-gas-companies-driving-adoption-big-data-says-globaldata, January 21. Accessed March 15, 2020.

Holst, A. (2020, March 2). Big data market size revenue forecast worldwide from 2011 to 2027. Retrieved from: https://www.statista.com/statistics/254266/global-big-data-market-forecast. Accessed March 15, 2020.

Honorio, J., Chen, C., Gao, G., Du, K., & Jaakkola, T., 2015. *Integration of PCA with a Novel Machine Learning Method for Reparameterization and Assisted History Matching Geologically Complex Reservoirs*. Paper SPE-175038-MS. Presented at the SPE Annual Technical Conference and Exhibition, Houston, Texas, 28-30 September 2015. doi: https://doi.org/10.2118/175038-MS

Hutchinson, M., Thornton, B., Theys, P., & Bolt, H., 2018. *Optimizing drilling by simulation and automation with Big data*. Paper SPE-191427-MS, presented at SPE Annual Technical Conference and Exhibition, 24-26 September, Dallas, Texas 2018. doi: https://doi.org/10.2118/191427-MS

Investopedia, 2020. *What Percentage of the Global Economy Is the Oil and Gas Drilling Sector?* Retrieved from: https://www.investopedia.com/ask/answers/030915/what-percentage-global-economy-comprised-oil-gas-drilling-sector.asp., February 15. Accessed March 15, 2020.

Ishwarappa, J., Anuradha, J., 2015. *A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology*. In: Procedia Computer Science vol. 48, pp. 319-324, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.04.188

Johnston, J., & Guichard, A., 2015. *New findings in drilling and wells using Big data analytics*. Paper OTC-26021-MS, presented at the Offshore Technology Conference, Houston, Texas, 04-07 May 2015. doi: https://doi.org/10.4043/26021-MS

Joshi, P., Thapliyal, R., Chittambakkam, A.A., Ghosh, R., Bhowmick, S., & Khan, S.N., 2018. *Big Data Analytics for Micro-seismic Monitoring*. Paper OTC-28381-MS, presented at the Offshore Technology Conference Asia, Kuala Lumpur, Malaysia, 20-23 March 2018. doi: https://doi.org/10.4043/28381-MS

Lin, A., 2014. *Principles of Big Data Algorithms and Application for Unconventional Oil Introduction: Insufficient Resources, (ISR) Computing and BD*. Paper SPE-172982-MS, presented at the Large Scale Computing and Big Data Challenges in Reservoir Simulation Conference and Exhibition, Istanbul, Turkey, 15-17

September 2014. doi: https://doi.org/10.2118/172982-MS

Luhn, H.P., 1958. A Business Intelligence System, in: IBM Journal of Research and Development, October, pp. 314-319.

Maidla, E., Maidla, W., Rigg, J., Crumrine, M., & Wolf-Zoellner, P., 2018. *Drilling analysis using Big data has been misused and abused*. Paper SPE-189583-MS, presented at the IADC/SPE Drilling Conference and Exhibition, Fort Worth, Texas, 6-8 March 2018. doi: https://doi.org/10.2118/189583-MS

Michigan State University, 2019. *4 Types of Data Analytics and How to Apply Them*. Retrieved from: https://www.michiganstateuniversityonline.com/resources/business-analytics/types-of-data-analytics-and-how-to-apply-them, 8 October. Accessed May 1, 2020.

Mircea, T., Mihaela, I. (2014). Big Data: Present and Future, in: Database Systems Journal, vol.V, no. 1, pp. 32-41

Ockree, M., Brown, K.G., Frantz, J., Deasy, M., & John, R., 2018. *Integrating Big data analytics into development planning optimization*. Paper SPE-191796-18ERM-MS, presented at the SPE/AAPG Eastern Regional Meeting, Pittsburgh, Pennsylvania, 7-11 October 2018. doi: https://doi.org/10.2118/ 191796-18ERM-MS

O'Dea, S., 2020. *Volume of data/information created worldwide from 2010 to 2025*. Retrieved from: https://www.statista.com/statistics/871513/worldwide-data-created, February 28. Accessed March 15, 2020.

Olneva, T., Kuzmin, D., Rasskazova, S., & Timirgalin, A., 2018. *Big data approach for geological study of the Big region West Siberia*. Paper SPE-191726-MS, presented at the SPE Annual Technical Conference and Exhibition, Dallas, Texas, 24-26 Semptember 2018. doi: https://doi.org/10.2118/191726-MS

Padmanabhan, V., 2014. *Big Data analytics in oil and gas, converting the promise into value*. Bain & Company. Retrieved from: https://www.bain.com/insights/big-data-analytics-in-oil-and-gas/, March 26.

Popa, A.S., Grijalva, E., Cassidy, S., Medel, J., & Cover, A., 2015. *Intelligent Use of Big Data for Heavy Oil Reservoir Management*. Paper SPE-174912-MS, presented at the SPE Annual Technical Conference and Exhibition, Houston, Texas, 28-30 September 2015. doi: https://doi.org/10.2118/174912-MS

Roden, R., 2016. *Seismic Interpretation in the Age of Big Data*. Geophysical Insights, pp. 4911-4915. doi:10.1190/segam2016-13612308.1

Rollins, B.T., Broussard, A., Cummins, B., Smiley, A., & Dobbs, N., 2017. *Continental production allocation and analysis through Big data*. Paper URTEC-2678296-MS, presented at the USPE/AAPG/SEG Unconventional Resources Technology Conference, Austin, Texas, 24-26 July 2017. doi: https://doi.org/10.15530/urtec2017-2678296

Seemann, D., Williamson, M., & Hasan, S., 2013. *Improving Reservoir Management through Big Data Technologies*. Paper SPE-167482-MS, presented at the SPE Middle East Intelligent Energy Conference and Exhibition, Manama, Bahrain, 28-30 October 2013. doi: https://doi.org/10.2118/167482-MS

Shahid, A.S.A., Fokker, P.A., Rocca, V., 2016. *A review of numerical simulation strategies for hydraulic fracturing, natural fracture reactivation and induced microseismicity prediction*. Open Petroleum Engineering Journal, Volume 9, June, Pages 72-91. doi: 10.2174/1874834101609010072

Subrahmanya, N., Xu, P., El-Bakry, A., & Reynolds, C., 2014. *Advanced Machine Learning Methods for Production Data Pattern Recognition*. Paper SPE-167839-MS, presented at the SPE Intelligent Energy Conference & Exhibition, Utrecht, The Netherlands, 1-3 April 2014. doi: 10.2118/167839-M

Technavio, 2019. *Global Big Data Market in the Oil and Gas Sector 2019-2023*. Use of Big Data by AI and ML Tool to Boost Growth. Retrieved from: https://www.businesswire.com/news/home/20190522005315/en, May 22. Accessed March 15, 2020.

Udegbe, E., Morgan, E., & Srinivasan, S., 2017. *Face Detection to Fractured Reservoir Characterization: Big Data Analytics for Restimulation Candidate Selection*. Paper SPE-187328-MS, presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 9-11 October 2017. doi: https://doi.org/10.2118/187328-MS

Vendetti, J., 2019. *Big Dollars for Big Oil with Big Data – Analytics Saves Billions in Upstream Exploration and Production*. Retrieved from: https://www.tibco.com/blog/2019/10/10/big-dollars-for-big-oil-with-big-data, October 10. Accessed March 20, 2020.

Whaley, J., 2015. *What is a Digital Oil Field?* GEOExPro, Vol 12, No 2. Retrieved from: https://www.geoexpro.com/articles/2015/12/what-is-a-digital-oil-field

Yin, Q., Yang, J., Zhou, B., Jiang, M., Chen, X., Fu, C., Liu, Z, *et al.*, 2018. *Improve the Drilling Operations Efficiency by the Big Data Mining of Real-time Logging*. Paper SPE-189330-MS, presented at the SPE/IADC Middle East Drilling Technology Conference and Exhibition, Abu Dhabi, UAE, 29-31 January 2018. doi: https://doi.org/10.2118/189330-MS

Electronic References

https://www.geoexpro.com/articles/2015/12/what-is-a-digital-oil-field (accessed April 5, 2020).

https://www.sciencedirect.com/topics/engineering/self-organizing-map (accessed April 5, 2020).

https://www.sciencedirect.com/topics/medicine-and-dentistry/principal-component-analysis (accessed April 5, 2020).

https://www.searchbusinessanalytics.techtarget.com/definition/big-data-analytics (accessed May 1, 2020.