Energy Policy 81 (2015) 117–121

Contents lists available at ScienceDirect

Energy Policy

journal homepage: www.elsevier.com/locate/enpol

Viewpoint

Data as an asset: What the oil and gas sector can learn from other industries about "Big Data"

Robert K. Perrons [a,b,*], Jesse W. Jensen [c]

[a] Queensland University of Technology, GPO Box 2434, Brisbane, Qld 4001, Australia
[b] Centre for Strategy and Performance, University of Cambridge, United Kingdom
[c] Intel Corporation, 2200 Mission College Blvd., Santa Clara, CA 95054, USA

HIGHLIGHTS

- Upstream oil and gas industry frequently discards or ignores the data it collects
- The sector tends to view data as descriptive information about the state of assets
- Leaders in Big Data, by stark contrast, regard data as an asset in and of itself
- Industry should use Big Data tools to extract more value from digital information

ARTICLE INFO

Article history:
Received 9 October 2014
Received in revised form
15 January 2015
Accepted 22 February 2015
Available online 3 March 2015

Keywords:
Big data
Oil and gas
Information technologies
Data

ABSTRACT

The upstream oil and gas industry has been contending with massive data sets and monolithic files for many years, but "Big Data" is a relatively new concept that has the potential to significantly re-shape the industry. Despite the impressive amount of value that is being realized by Big Data technologies in other parts of the marketplace, however, much of the data collected within the oil and gas sector tends to be discarded, ignored, or analyzed in a very cursory way. This viewpoint examines existing data management practices in the upstream oil and gas industry, and compares them to practices and philosophies that have emerged in organizations that are leading the way in Big Data. The comparison shows that, in companies that are widely considered to be leaders in Big Data analytics, data is regarded as a valuable asset—but this is usually not true within the oil and gas industry insofar as data is frequently regarded there as descriptive information about a physical asset rather than something that is valuable in and of itself. The paper then discusses how the industry could potentially extract more value from data, and concludes with a series of policy-related questions to this end.

1. Introduction

"Big Data"[1] is a rather vague term that describes the application of new tools and techniques to digital information on a size and scale well beyond what was possible with traditional approaches (Lohr, 2012b), typically involving data sets that are so large and complex that they require advanced data storage, management, analysis, and visualization technologies (Chen, et al., 2012). Like many industries, the upstream oil and gas sector has seen a flurry of initiatives and high-profile publications (e.g., Anand, 2013; Beckwith, 2011; Holdaway, 2014) about this topic, which have in turn translated into significant discussion within industry conferences (e.g., Feblowitz, 2013) and among practitioners. Critics of Big Data caution that the transformational potential of these analytical capabilities may be somewhat oversold and misunderstood (Harford, 2014; Lohr, 2012b),[2] but the oil and gas sector has already been noticeably impacted by several of the technologies underpinning these changes.

* Corresponding author at: Queensland University of Technology, GPO Box 2434, Brisbane, Qld 4001, Australia.
E-mail address: perrons@alum.mit.edu (R.K. Perrons).

[1] The term "Big Data" is believed to have been coined by the astronomy and genomics communities in the 2000s (Mayer-Schönberger and Cukier, 2013, p. 6), but the concept has been used more widely since then.

[2] Gary King, the director of Harvard University's Institute for Quantitative Social Science, goes as far as predicting that there "is no area that is going to be untouched" by Big Data (Lohr, 2012a).

http://dx.doi.org/10.1016/j.enpol.2015.02.020
0301-4215/

By most accounts, the oil and gas industry's data is already “big.” Modern oil and gas seismic data centers can easily contain as much as 20 petabytes[3] of information, which is roughly equivalent to 926 times the size of the U.S. Library of Congress (Beckwith, 2011). If this amount of information was copied into books and put on a single continuous bookshelf, it would go around the Earth's equator approximately six times (Beckwith, 2011). And while seismic data sets are notoriously large and cumbersome, many other aspects of the oil and gas industry are also generating significantly more data than they used to (Perrons, 2010b). What is more, there is every reason to believe that this trend towards more digital information is just getting warmed up. Current estimates suggest that the total amount of digital data in the world—including things like books, images, e-mails, music, and video—is doubling every 2–3 years (Lohr, 2012a; Mayer-Schönberger and Cukier, 2013).

Big Data tools and techniques have most famously delivered value in the social media sector and retail industries (Harford, 2014; Mayer-Schönberger and Cukier, 2013), but have also led to major breakthroughs in a diverse range of other contexts, including scientific research (Frankel and Reid, 2008; Kluger, 2014; Lynch, 2008), healthcare (Chen, et al., 2012), heavy equipment manufacturers (Mehta, 2013), and professional sports (Leahey, 2013).[4]

2. What has made Big Data possible?

Big Data is not a result of a single silver-bullet technology, but rather the coming together of several innovations and novel ideas in a highly complementary way. Four of these technological developments are particularly noteworthy:

(1) A precipitous decline in data storage costs. As shown in Fig. 1, the cost of storing digital information has been falling at an exponential rate for a long time. Several years ago, it was standard practice in many industries to discard significant collections of data when their initial use had passed, as there was a real economic expense associated with archiving the data afterward (e.g., Feblowitz, 2013). This is much less true today (Komorowski, 2014).

(2) Continued growth in the processing speeds of computing devices. Moore's Law, which states that the number of transistors on integrated circuits doubles approximately every two years, has been continuing unabated since the 1970s (Fig. 2).[5] It therefore follows that the amount of computing power offered in commercially available devices has been increasing at a similarly impressive rate (Ball, 2000).

(3) Breakthroughs in relevant areas of mathematics. Whereas traditional data sets have historically needed to be fairly structured, orderly, and static, digital information in the era of Big Data is frequently noisy, messy, raw, unstructured, and dynamic (Ouellette, 2013). Recent developments in mathematics—most notably, geometry—have significantly helped Big Data practitioners see through the messiness of these new data sets to find useful information and relationships. Considerable progress has been made in representing massive data sets as networks of geometrical nodes and edges so that the data can be rationalized using a suite of mathematical tools known as topological data analysis (TDA). Simply put, TDA is a way of getting structured information out of unstructured data so that machine-learning algorithms can be applied to it (Carlsson, 2009; Ouellette, 2013).

(4) The development of software platforms such as Google's MapReduce or its open-source rival Apache™ Hadoop®.[6] These tools make it possible to break large data sets into smaller chunks that can be delegated to several computing devices. The results of the calculations arising from each of the smaller chunks can then be re-integrated at the end of the process. This approach frequently uses cloud computing infrastructure as a platform for transferring these data chunks to different computing devices, and then bringing back the results.

3. How is Big Data different from what was done previously?

Despite the hype that has been generated around the topic of Big Data, the overarching objective is something that organizations in the energy sector and everywhere else have been aspiring to for a long time: to make better decisions (Regalado, 2014). What is changing, however, are the specific mechanisms by which these decisions are made. First, Big Data differs from traditional approaches on account of the “three Vs”: volume, velocity, and variety (McAfee and Brynjolfsson, 2012). The steadily decreasing costs associated with collecting and storing data have resulted in a fundamental shift in thinking about data quality and volume. Historically, data collection was predicated on sampling from a subset of an overall population, and trying to make the collected data from that sample as accurate as possible. By contrast, the move towards Big Data has led to a much greater tolerance for messiness and imprecision. This more relaxed approach to vagueness has been compensated, however, by much larger volumes of data. Underlying this change of philosophy is the belief that “more trumps better” (Mayer-Schönberger and Cukier, 2013, p. 33).

These much larger volumes of data are now able to move with ever-increasing velocities such that a bewildering number of system variables can be monitored in nearly real-time. Moreover, this data is coming from a wider variety of sources and in an increasingly broad array of formats. Some data is “born digital,” meaning that it was created specifically for digital use by a computer or data processing system (e.g., e-mail, web browsing, GPS locations); other data is “born analog,” meaning that it comes from the physical world, but can increasingly be converted into a digital format (e.g., voice or visual information captured by phones, cameras, or video recorders, or data collected from wearable devices). The rising capability of “data fusion” makes it more and more possible to bring together disparate sources of data to glean fresh insights that nobody predicted (White House, 2014).

4. How Big Data is unfolding differently in the oil and gas sector

In light of this sweeping global trend, it is hard to imagine a future in which the oil and gas industry is not collecting

[3] This statistic is less impressive when you consider that Walmart, the U.S. retail giant, collects more than 2.5 petabytes of data every hour from customer transactions (McAfee and Brynjolfsson, 2012).

[4] The 2011 movie *Moneyball*, starring Brad Pitt, chronicles how the low-budget Oakland A's started using historical performance data and arcane baseball statistics to spot undervalued players. Intensive data analysis has since become commonplace not only in baseball but also in other sports, including English soccer (Lohr, 2012a).

[5] Note that even though this figure's timeline stops at the year 2000, the microprocessor industry has more or less continued to deliver new products that have kept pace with Moore's Law until the present day.

[6] Using Hadoop, Visa was able to reduce the processing time for two years' worth of test records—which translates to approximately 73 billion transactions—from one month to only 13 min (Mayer-Schönberger and Cukier, 2013, p. 46).