Business Horizons (2017) xxx, xxx—xxx

Available online at www.sciencedirect.com

ScienceDirect

KELLEY SCHOOL OF BUSINESS
INDIANA UNIVERSITY

www.elsevier.com/locate/bushor

Addressing barriers to big data

Abdulkhaliq Alharthi [a], Vlad Krotov [b,*], Michael Bowman [b]

[a] College of Business, Abu Dhabi University, P.O. Box 59911, Abu Dhabi, U.A.E.
[b] Murray State University, 109 Business Building, Murray, KY 42071, U.S.A.

KEYWORDS
Big data;
Big data infrastructure;
Data privacy;
Organizational culture;
Competitive advantage;
Hadoop

Abstract Increasingly, big data is viewed as the most strategic resource of the 21st century, similar in importance to that of gold and oil. While sitting on these vast pools of data, many organizations are simply not ready to take advantage of this new strategic resource. Embracing big data requires addressing a number of barriers that fall into the domains of technology, people, and organization. A holistic, socio-technical approach is required to overcome these barriers. This article introduces the specific tactics we recommend for addressing big data barriers, which involve changes to technology infrastructure, a focus on privacy, promotion of big data and analytic skills development, and the creation of a clear organizational vision related to big data.

1. The era of big data

Recent advancements in information and communications technologies (ICT), as well as the ever-increasing affordability and ubiquity of networks and electronic devices, have resulted in a massive volume of data from various sources and in different formats. The volume of data available today is measured in zettabytes (ZB)—a measure equal to one trillion gigabytes (GB) and equivalent to the data storage capacity of about 250 billion DVDs. The world wide web alone was estimated to contain 0.5 ZB of data in 2009 (Fan & Bifet, 2013). This amount of data is available from more than one trillion web pages currently accessible on the web. The total amount of digital data in the whole world reached 1.8 ZB in 2011 and is predicted to grow to approximately 90 ZB by 2020 (Jeon, 2012). Currently, about 90% of the digital data available was created in the last 2 years (Gobble, 2013). These massive amounts of recently created digital data are often referred to as *big data*.

Participants of the 2012 World Economic Forum in Davos, Switzerland, declared that big data has become a strategic economic resource, similar in significance and liquidity to currency and gold (Johnson, 2012). While organizations today across the globe realize that big data analytics will be an

* Corresponding author
E-mail addresses: alharthi2000@yahoo.com (A. Alharthi), vkrotov@murraystate.edu (V. Krotov), michael.bowman@murraystate.edu (M. Bowman)

0007-6813/$ — see front matter © 2017 Kelley School of Business, Indiana University. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bushor.2017.01.002

important source of competitive advantage in the future, a number of barriers prevent many organizations from fully exploiting the opportunities big data can offer. These barriers include outdated IT infrastructure, the inherent complexity and messiness of big data, lack of data science skills within organizations, privacy concerns, and organizational cultures that are not conducive to data-driven operations or data-driven decision making. This article proposes specific tactics for removing these barriers, namely: (1) the utilization of commodity hardware and specialized big data software, (2) collaboration with educational institutions, (3) installation of policies and processes that would support individual privacy, and (4) development of a clear organizational vision in relation to big data.

2. What is big data?

The term big data is used to describe the massive volume of digital data produced by human activity that is very difficult to manage using conventional data analysis tools. While scholars and practitioners may have different views on what precisely is meant by big data, the Gartner Group's definition is widely used. According to the definition (Gartner, 2017), big data is characterized by the 3 Vs:

- *Volume* refers to the vast quantity of structured and unstructured data that is hard to collect, manage, and analyze with the existing IT infrastructure and tools; thus, these massive data sets require new and innovative tools and approaches for capturing, storing, and analyzing data.
- *Variety* refers to the fact that the data comes from various sources such as spreadsheets, traditional databases, text documents, and digital data streams.
- *Velocity* refers to the fact that these big data sets are often comprised of and continuously expanded by real-time data streams.

The term big data refers not only to data, but also to the tools and practices for analyzing, processing, and managing these massive, complex, and rapidly evolving data sets. Therefore, the terms big data and big data analytics are used in this article interchangeably.

Big data analytics is similar to the concept of business intelligence (BI). Both concepts are used to denote usage of data management technologies and computer-based analytical tools for discovering actionable business knowledge and facilitating decision making based on organizational data. Yet these two concepts have three major differences related to the type of data used by BI and big data. These differences are closely related to the 3 Vs discussed previously (see Table 1). BI architecture is often conceptualized as an organizational data architecture that relies on a finite set of highly structured data sources (e.g., various internal and external databases) that are accessed in an offline mode (e.g., via a specialized data mart extracted from an organizational data warehouse). In contrast, big data movement aims to develop data management technologies and analytical tools that can handle an infinite number of data sets comprised of highly complex and 'messy' data in virtually real-time formats.

Table 1. Big data vs. business intelligence

Characteristics	Big data	Business intelligence
Volume	Infinite	Finite
Velocity	Real time	Offline
Variety	Unstructured	Structured

3. Big data opportunities

Organizations across the globe increasingly have come to the realization that the ability to analyze and use big and complex data sets will be the most important source of competitive advantage in the 21st century. Big data has the potential to deliver better customer experience, enhance internal efficiency, and, ultimately, improve profitability and competitiveness of organizations across all industries. Organizations can use big data to get smarter and innovative in ways that were not possible before the advent of the 'zettabyte era' (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). Some examples of successful big data initiatives are discussed in the next sections.

3.1. Etihad Airways

Based in Abu Dhabi, United Arab Emirates, Etihad Airways flies to more than 89 destinations across the globe and carries 10 million passengers annually. Since the company's founding in 2003, Etihad began using hundreds of sensors on every plane to generate data about its fleet. This arrangement is similar to the Internet of Things (IoT) concept, in that a