Business Horizons (2016) **59**, 673—688

Available online at www.sciencedirect.com

ScienceDirect

KELLEY SCHOOL OF BUSINESS
INDIANA UNIVERSITY

www.elsevier.com/locate/bushor

Managerial work in the realm of the digital universe: The role of the data triad

Vijay Khatri

Kelley School of Business, Indiana University, 1309 E. Tenth Street, Bloomington, IN 47405-1701 U.S.A.

KEYWORDS
Analytics;
Big data;
Managerial decision making;
Managerial work;
Digital universe

Abstract With the explosion of the digital universe, it is becoming increasingly important to understand how organizational decision making (i.e., the business-oriented perspective) is intertwined with an understanding of enterprise data assets (i.e., the data-oriented perspective). This article first compares the business- and data-oriented perspectives to describe how the two views mesh with each other. It then presents three elements in the data-oriented perspective that are collectively referred to as the data triad: (1) use, (2) design and storage, and (3) processes and people. In describing the data triad, this article highlights practices, architectural techniques, and example tools that are used to manage, access, analyze, and deliver data. By presenting different elements of the data-oriented perspective, this article broadly and concretely describes the data triad and how it can play a role in the redefined scope of work for data-driven business managers.

1. Uncovering the hidden potential of the digital universe

The digital universe refers to the digital data that is created, replicated, and consumed. It includes digital movies playing on high-definition televisions, banking data that is transmitted when credit cards are swiped at stores, airport security footage, music that is listened to via smartphones, social media interactions, and fitness information as gauged by smart wearables. Much like the physical universe, the digital universe is enormous and diverse. Both the rise of social media and the growth of internet business transactions—the latter of which are expected to increase to 450 billion a day by 2020 (Gantz & Reinsel, 2010)—are fueling the expansion of the digital universe. Indeed, the digital universe is doubling in size every two years and is projected to explode from 4.4 trillion gigabytes in 2013 to around 44 trillion gigabytes in 2020 (Turner, Gantz, Reinsel, & Minton, 2014). Tens of billions of connected devices (e.g., PCs, laptops, smartwatches, televisions) generate a deluge of diverse data. The digital universe is comprised of not only structured but also unstructured data—estimated at around 90% of enterprise data (Gantz & Reinsel, 2011)—that includes, for example, human-generated

E-mail address: vkhatri@indiana.edu

0007-6813/$ — see front matter © 2016 Kelley School of Business, Indiana University. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bushor.2016.06.001

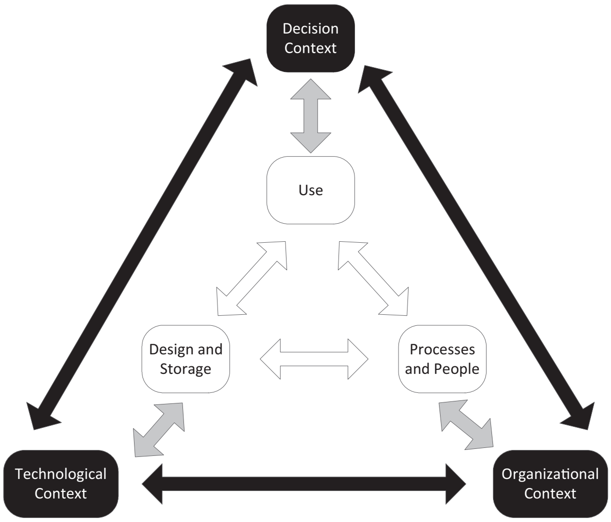
textual data from research reports, product descriptions and customer reviews, machine-generated web log files, and sensor-based data.

The latent potential of the digital universe is recognized in every sector, be it government, financial services, healthcare, not-for-profit, retail, services, or manufacturing. Data is taking prediction capabilities to new heights. For example, based on 10 years of 61,000+ 'trouble tickets' and details of the 21,000 miles of aging power cables, some dating back to the 1880s, manhole explosions in New York have been accurately predicted (Ehrenberg, 2010). Likewise, based on their purchase histories, customers' life stages (e.g., pregnancy; Duhigg, 2012) and the timing of their future purchases (Bensinger, 2014) have been accurately predicted. With a plethora of data, not only can the accuracy of existing decision making be enhanced, but also new and creative ways can be developed to enable decision making that was not possible even a few years ago. For example, Facebook 'likes' can be used to predict an individual's personality, sometimes more accurately than as described by friends or family members (Youyou, Kosinski, & Stillwell, 2015). A McKinsey report projects that service providers may find the potential value of global personal location data to be as high as $100 billion in revenue over the next 10 years (Manyika et al., 2011); the same report estimates the potential annual value of data in US healthcare and the European public sector at $300 billion and €250 billion, respectively. Given its ability to help identify and create new business opportunities, it is little wonder that data is now referred to as the new oil of the digital economy (Toonders, 2014).

Being data-driven—that is, making better decisions up and down the organization chart—is now acknowledged as a worthwhile organizational endeavor (Redman, 2013). While organizations that embrace data-driven decision making perform better on objective measures of financial and operational results (McAfee & Brynjolfsson, 2012), the sheer size, diversity, and vibrancy of the digital universe continues to present them with a challenge. In addition to the inherent complexity associated with better employing the digital universe for decision making, there is a distinct shift toward self-service analytics—also referred to as the consumerization of analytics—wherein the line-of-business professionals are empowered and encouraged to quickly blend, analyze, and visualize data on their own with minimal support from technology specialists. The consumerization of analytics implies that the demand (i.e., 'user') side of leveraging data and the supply (i.e., 'provider') side of managing the data infrastructure are increasingly coalescing. In the context of consumerization of analytics that is based on ever-expanding and diverse data, business managers need to address many questions. For example: How will managerial decision making be influenced by assumptions associated with data processing and analysis? How will immediacy in managerial decision making influence the need for real-time data access? How will the need for real-time data access influence technology choices (e.g., those related to data storage)? And how will the organization's beliefs and culture influence the processes that are designed to enable desirable uses of data?

As indicated by the preceding questions, being data-driven implies the need to grasp how business decision making is intertwined with an understanding of enterprise data assets. While the *business-oriented perspective* (black boxes and arrows, Figure 1) focuses on how the business goals of revenue enhancement, cost reduction, and/or risk management are reflected in managerial decision making, the *data-oriented perspective* (white boxes and arrows, Figure 1) focuses on the processes of and practices for processing and analyzing enterprise data to generate insights that support managerial decision making. This article suggests that these perspectives are two sides of the same coin, with the former signifying the user side of leveraging the digital universe and the latter indicating the provider side of managing the data assets in that universe. Figure 1 highlights three key points about the two perspectives. First, one must be cognizant of three key contexts that are central to the business-oriented perspective: *decision*, *technological*, and *organizational*. Second, one must understand the three elements that comprise the data-oriented perspective: data *use*, *design and storage*, and the

Figure 1. Business- and data-oriented perspectives



associated *processes and people*. These elements are collectively referred to as the *data triad*. Third, one must recognize the linkages between the two perspectives (grey arrows, Figure 1).

The value of data lies in its utility. From the business-oriented perspective, utility refers to decision making that influences the business imperative. From the data-oriented perspective, utility refers to realization of business uses that support decision making; in turn, the requisite utility both drives and is driven by technology and organization. While business managers primarily focus on the outer, business-oriented triangle depicted in Figure 1, this article suggests that data-driven business managers need to better recognize (1) the significance of the inner, data-oriented triad; and (2) how the two perspectives intertwine with each other.

In the context of two concomitant, significant shifts in the nature of work of business managers—(1) the need to be more data-driven and (2) consumerization of analytics—this article seeks to integrate the business- and data-oriented perspectives and concretely outline practices, architectural techniques, and example tools that better delineate the data-oriented perspective.

The remainder of this article first compares the business- and data-oriented perspectives to describe how the two views intertwine with each other. It then provides a holistic data-oriented perspective for business managers who need to understand (1) use, (2) design and storage, and (3) processes and people associated with the inner data triad. By presenting different elements of the data-oriented perspective, this article broadly and concretely describes the data triad and how it can play a role in the redefined scope of work for a data-driven business manager.

2. The business- and data-oriented perspectives

To illustrate the complementary nature of the business- and data-oriented perspectives, Table 1 lists sample business-oriented questions and consonant data-oriented ones. As depicted, decisions that are aligned with enterprise objectives are central to the business-oriented perspective, while data that drives decision making is pivotal to the data-oriented perspective. From the business-oriented perspective, the utility of data lies in the decision making data enables and how that decision making influences the business in terms of its impact on revenue and cost, and the management of risks (Fisher, 2009). Similarly, the data-oriented perspective focuses on the uses of data, along with the assumptions underlying different types of uses. For example, one of the uses of historical sales data may be to forecast future demand so that the firm can decide whether or not to add production or service capacity. The need to enhance the effectiveness of decision making in turn may drive, for instance, how the forecast of future sales is rendered graphically to help with capacity decisions, which may also have cost and revenue implications.

The desired value or utility shapes and is shaped by the technological and organizational contexts. The *technological context* refers to the how, when, and where associated with the decisions. For example, the need for centralized or decentralized decision making is intertwined with how scattered or hidden the data may be. The immediacy in decision making that is based on a time-series forecast (a business-oriented perspective) establishes the need for real-time access to data (a data-oriented perspective). Finally, the tools that support clarification of assumptions in decision making influence the technologies that are employed in managing the lifecycle of data. The data-oriented perspective also delves into investigating technologies and practices that are used to capture, clean, extract, and analyze data.

The *organizational context* focuses on the decision maker (i.e., who?), along with the decision-making processes. As the key constituents in decision making are recognized, the requisite skills and training needed to better use the data are identified. Finally, from the viewpoint of the organizational context, the business-oriented perspective focuses on the strategy and organizational beliefs and culture that are contextual to decision making. For its part, the data-oriented perspective emphasizes data-related processes for managing the lifecycle of data in its entirety. It also focuses on issues related to data ownership.

This article provides a lens through which to view and understand the data-oriented perspective holistically. The data-oriented perspective suggests that three aspects—use, design and storage, and processes and people—are themselves intertwined. As a business manager identifies new business uses of data, she/he also needs to realize that use is inextricably intertwined with how the database is designed and stored. For example, if doctors' addresses are known to have an accuracy of 85%, this data may be acceptable for some purposes, such as running a marketing campaign that involves mailing low-cost informational flyers about new products. However, this threshold of data quality may not be appropriate for other uses, such as notifying the doctors about a product recall. Thus, different practices for managing data quality—an aspect

Table 1. The scope of business- and data-oriented perspectives

Business-oriented perspective	Data-oriented perspective
Decision Context (What?)	*Use*
● How can business goals be quantified? What decisions need to be made?	● What are the different uses of data for decision making?
● What is the impact of decision on revenue enhancement, cost reduction, and risk management?	● What are the assumptions associated with different uses of data? What types of structured and unstructured data should be focused on?
● What needs to be done to enhance the effectiveness of decision making?	● How can insights be rendered graphically to support decision making?
Technological Context (How? When? Where?)	*Design and Storage*
● Does the technology for decision making need to be centralized or decentralized?	● How scattered and hidden is the organizational data?
● What technology is required that would support the requisite immediacy in decision making?	● What data requires real-time access? In designing different types of systems, what are the storage and use tradeoffs?
● How do different tools support clarification of assumptions in decision making?	● Based on different uses, what technologies and practices can be used to capture, clean, extract, and analyze data during its lifecycle?
Organizational Context (Who?)	*Processes and People*
● Who are the key constituents in decision making?	● Who holds decision rights for various data decision domains? What skills are required to better use data? What are the training-related requirements that would enable desirable uses of data?
● What are the beliefs and culture that can influence decision making?	● What processes need to be designed to enable desirable uses of data? Who owns the data?
● How do regulations influence decision making?	● What are the acceptable processes of data uses? How to ensure and monitor data-related regulations? How is data inventoried?

important to design and storage—may be apt for different types of uses. Given that processes and people also influence use, a business manager must ask questions, such as: What is the process of production of doctor-related data? How is that data processed? Who owns doctor-related data? How old is that data? Having described the data-oriented perspective in the context of the business-oriented perspective, three elements of the data triad are outlined next.

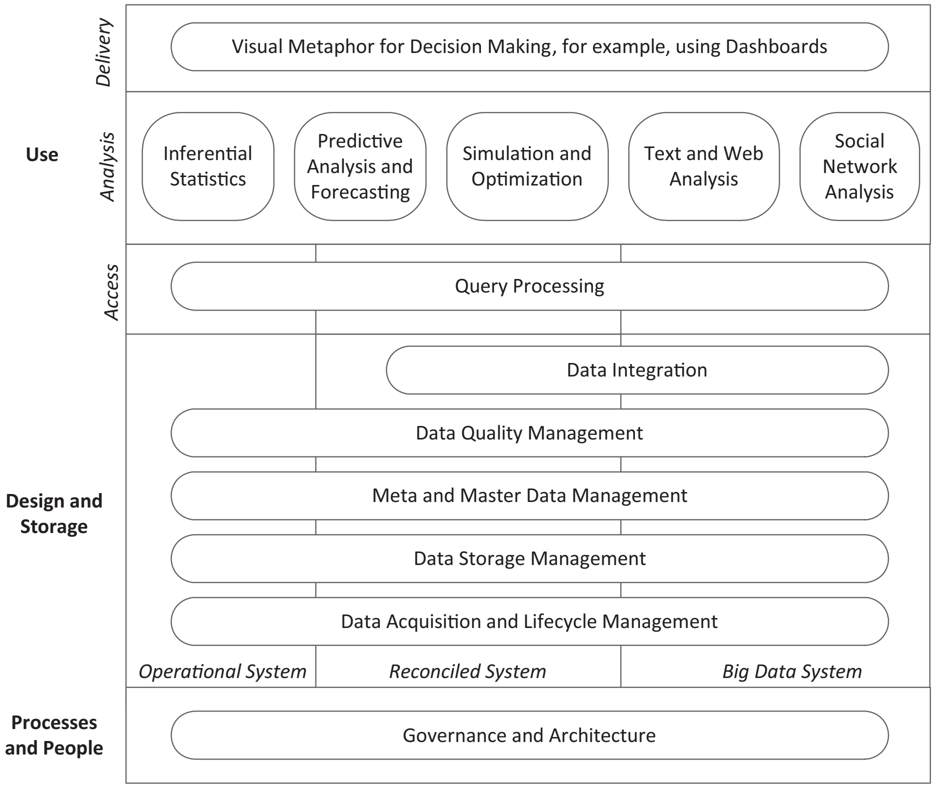
3. Three elements of the data triad

Figure 2 portrays and expounds upon three key elements of the data-oriented perspective, which are referred to as the data triad: (1) use, (2) design and storage, and (3) processes and people. While this section provides an overview of the three elements, subsections 3.1., 3.2., and 3.3. detail the individual components within each element.

With respect to *use* in Figure 2, data is often delivered (*delivery*) to users via graphical metaphors (e.g., using dashboards). Employing data for business use (i.e., decision making) may require data *analysis* using techniques such as predictive analysis and forecasting, simulation and optimization, text and web analysis, and social network analysis. The analysis of data is contingent on data *access* (using *query processing*) to different types of systems: (1) operational systems typically are accessed using Structured Query Language, or SQL; (2) corporate-wide integrated data—sometimes referred to as the reconciled system—may be accessed using another query language, such as Multidimensional Expressions; and (3) the big data system may be accessed using a query language, such as HiveQL. Section 3.2. describes these different types of uses in detail.

With respect to *design and storage*, there are several use-related considerations—for example, requirements related to query response time, the ability to handle a wide variety of data, and

Figure 2. The data triad



the associated cost. These considerations prompt the need for different types of systems: operational, reconciled, and big data. The *operational system*, which refers to an architecture that supports online transaction processing (OLTP), focuses on individual transactions; for example, an operational system that is used to manage the stock of books at a given store. These operational systems are typically optimized for inserting or updating rows in the database, or retrieving a small number of rows (Jensen, Pedersen, & Thomsen, 2010). A *reconciled system*, which refers to an architecture that supports online analytical processing (OLAP), employs data structures that can handle massive queries based on potentially distributed and disparate operational systems. For example, a reconciled system could be employed to track all historical sales of books that were sold online and in-store, each managed by potentially different operational systems. Reconciled systems may duplicate data from the operational systems, do not delete or update their data and can add data periodically, and are optimized for retrieving and summarizing large numbers of rows, thereby helping in what-if analysis.

There is a recent trend of amalgamation of these two architectures, operational and reconciled. Hybrid Transaction/Analytical Processing (HTAP), a term coined in early 2014 by Gartner, describes a new generation of in-memory data platforms that can perform both OLTP and OLAP without requiring data duplication. HTAP systems are commercially available from major vendors such as SAP Business Suite 4 HANA (S/4 HANA). While such architectures can support low latency data reporting and analysis—for example, forecasting and simulations—they may not be appropriate for processes that do not require real-time planning, forecasting, and what-if analysis (Pezzini, Feinberg, Rayner, & Edjlali, 2014).

Big data refers to data that is too big, too slow, too hard, and too expensive for existing tools to process (Madden, 2012). 'Too big' means that data is of petabyte scale as based on clickstream, transaction histories, sensors, etc. 'Too slow' implies that data needs to be processed quickly (e.g., for fraud detection at the point of sale). 'Too hard' is a catchall for data that does not fit the processing and analysis that may be required. Finally, 'too expensive' refers to the need to scale as necessary in a cost-effective manner (e.g., using commodity hardware). This description maps with that of Gartner: The high volume, velocity, variety, and/or value data assets that require new and innovative forms of processing for business insights are referred to as big data (Beyer & Laney, 2012). An architecture that supports storage, management, and retrieval of big data is referred to as a *big data* *system*. While traditional database systems can handle multipetabyte-sized databases, they lag in other requirements as compared to big data systems. Interestingly, while the hype around big data is high, big data systems have not yet widely been realized as a resource, and may not be apt for every application. (See Table 2 for a comparison of applications of traditional and big data systems.) According to a 2014 Gartner survey, while 74% of respondents indicated that they had invested or were planning to invest in big data systems within 24 months, only 14% reported deploying a big data project (Heudecker & Kart, 2014).

In the context of the previously mentioned database systems, five classes of practices, architectural techniques, and tools are used to manage the design and storage of data, which in turn influences data use. First, data needs to be acquired from internal (e.g., Enterprise Resource Planning (ERP) systems) or external sources (e.g., the data brokerage market) (*data acquisition and lifecycle management*). Second, data needs to be stored so that it can be easily accessed (*data storage management*). Third, the ease of access of stored data needs to be enhanced by clear definitions of data (*metadata management*); additionally, core entities within the business—such as customers and products—need to be defined consistently (*master data management*). Fourth, the quality of data needs to be tracked and managed (*data quality management*). Finally, data that resides in multiple autonomous heterogeneous data sources may need to be combined (*data integration*). Section 3.1. presents these five components, which are central to design and storage of the digital universe.

Table 2. Traditional vs. big data systems

As technology platforms, traditional and big data systems are designed for different scenarios. Traditional database systems focus on reducing processing time. Their objective could be summarized as follows: "Given a set of machines [i.e., cost in dollars], try to minimize the response time of each request" (Florescu & Kossmann, 2009, p. 43). With the advent of big data, as the amount of data rose exponentially, it was not feasible to take the number of machines—or, cost in dollars—as given, and what is being optimized lately is as follows (Florescu & Kossmann, 2009, p. 43): "Given a response time goal for each request, try to minimize the number of machines (i.e., cost in $)." The two types of systems also provide different levels of transactional support. In the tradeoff between consistency and availability, big data systems strive for availability and eventual consistency, and not strict consistency or consistent views of changing data; that is, updates and replicas are expected to be consistent over a long period of time. The lack of strict consistency has acted as a barrier to big database systems being employed for mission-critical tasks in large enterprises. That is why 90% of reconciled systems are projected to remain and not be replaced by big data implementations (Laney et al., 2014). Finally, while the traditional database system mandates well-architected deductive development of information stores, big data initiatives emphasize an opportunity-oriented inductive approach (Buytendijk & Laney, 2013). In the latter approach, analytics, programming, and data integration are done in an ad-hoc manner, usually by the same person—a lofty expectation for an individual. Such unrealistic expectations may have made their mark; indeed, 57% of respondents of a Gartner survey indicated that a skills gap is influencing the adoption of big data systems (Heudecker & Adrian, 2015).

The third element of the data triad, *processes and people*, serves as the glue that connects models, policies, and standards governing how data is stored, integrated, and put to use. *Data governance* refers to the assignment of the locus of control for different decision domains (e.g., that related to data acquisition and lifecycle management) that ensure effective management and use of data (Khatri & Brown, 2010). The *data architecture* is the specification artifact that defines strategic data requirements and guides integration and control of data assets to ensure that data investments are aligned with the enterprise strategy (Mosley, 2010). Section 3.3. describes processes and people, outlined above, in greater detail.

From a manager's point of view, the first element of the data triad—use—is the most important. However, the story of data begins with its inception. This occurs within the element of the data triad referred to as design and storage, which is described next.

3.1. Design and storage

As an enterprise asset, data must be:

1. Acquired and managed throughout its lifecycle;
2. Stored effectively;
3. Identified, defined, and understood;
4. Quality controlled; and
5. Consistently accessible even if the individual data stores may have been designed independently.

These five requirements are supported by five different practices, as described in the following

Table 3. Sample business considerations and tools/technologies for managing storage and design

Design and Storage	*Sample Business Considerations*	*Example Tools/Technologies*
Data Acquisition and Lifecycle Management	● What is the program for data definition, production, retention, and retirement for different types of data, both internal and external? ● How do compliance issues related to legislation affect data retention and archiving? ● In view of pervasive sensing, what data should be archived vs. discarded? ● How do data-brokered products match the business uses of data? ● How can the visibility of data lineage be improved at the point of consumption?	● IBM's Infosphere Optim Archive ● Oracle's ILM Assistant
Data Storage Management	● Does the application rely on immediate consistency and real-time analytics? ● What are the different types of structured and unstructured data that need to be stored? ● For different types of data, how fast is the anticipated growth of data?	● Database management and data warehouse systems from, for example, Oracle, Teradata, IBM, SAP, Microsoft ● HDFS for storage, NoSQL database, along with the programming model such as MapReduce
Metadata Management	● How will data be consistently defined and modeled so that it is interpretable? ● What is the plan to keep different types of metadata up-to-date? ● Is the business meaning of data easy to understand? ● Is data simple and quick to find? ● What computing resources are needed to find data?	● MetaDex by Compact Solutions ● MetaCenter by DataAdvantageGroup ● InfoSphere Information Governance Catalog by IBM ● Oracle Enterprise Metadata Management by Oracle
Master Data Management	● How important is the shared use of terminology across multiple data sources? ● How does the shared use of business terminology relate to consistency of non-transactional data (e.g., customer, product, location)? ● Is there a need for a consolidated 360-degree view about important non-transactional data (e.g., customers)? ● How can shared use of business terminology influence compliance with regulations?	● Product Master Data, for example: ❖ ProductHub EBS by Oracle ❖ Informatica MDM Product 360 ● Customer Master Data, for example: ❖ Informatica ❖ Siebel UCM
Data Quality Management	● How inaccurate, incomplete, and unreasonable is the data? ● How does data quality influence the business imperatives (e.g., those related to agility and performance)? ● How does current data quality influence compliance and transparency? ● What are the standards and metrics for data quality with respect to accuracy, timeliness, completeness, and credibility? ● What is the program for establishing and communicating data quality? ● How will data profiling be conducted against current business rules? ● How standardized is the terminology? ● How can responsibility for certified data sources be formalized?	● IBM InfoSphere Information Server ● Spectrum Technology Platform by Pitney Bowes ● Talend Open Studio for Data Quality ● Informatica's Data Quality and Rev

Table 3 (*Continued*)

Design and Storage	*Sample Business Considerations*	*Example Tools/Technologies*
Data Integration	• How is business integration impacted by data integration? • What data—structured vs. unstructured—needs to be integrated? • How important is a 360-degree view of, for example, customer data? • How is access response time influenced by data integration? • How is data quality influenced by data integration?	• Informatica platform and IBM InfoSphere Information Server Enterprise Edition • Apache Flume collects, aggregates, and transfers data from disparate sources to a centralized store; Apache Sqoop allows import and export of data among structured data stores and Hadoop

subsections and illustrated in Figure 2. Table 3 outlines example business considerations and sample tools/technologies that can automate these practices.

3.1.1. Data acquisition and lifecycle management

Data acquisition and lifecycle management refers to a set of practices for acquiring data and managing it along its lifecycle. Data acquisition can entail the procurement of data from, for example, enterprise-wide order fulfillment systems, sensor-based data collection systems, or external data brokers. ERP systems are a major source of data acquisition for most organizations.

Sensor-based data collection systems, which use sensors to measure data such as pressure and proximity, can be thought of as cyber-physical systems as they convert physical features to digital signals. Such cyber-physical systems form the core of the Internet of Things (IoT), which is predicted to number around 25 billion units by 2020 and generate incremental revenue exceeding $300 billion (Lopez & Cantara, 2014).

By 2019, 75% of analytics solutions are projected to include 10 or more data sources from third-party data providers referred to as *data brokers*. Data brokers are businesses that aggregate, cleanse, and enrich information from a variety of sources with the intention of licensing it to other organizations (Faria, Linden, & Laney, 2016). The report presents a taxonomy of data brokerage that includes, for example, data about individuals (addresses, phone numbers, demographics, credit scores—as provided by data brokers such as Acxiom, comScore, Harte Hanks, Equifax, and Experian); organizations (information as provided by data brokers such as Bloomberg and Dow Jones); real estate (sales rates, prices, comparables—as provided by data brokers such as Zillow); and products/brands/reviews (information as provided by data brokers such as Zagat and BrightPlanet).

Data lifecycle management refers to policies, practices, and tools that are used to align the business value of data with the most appropriate and cost-effective storage infrastructure, from the point of data conception to data retirement. In the context of decreasing storage costs and potential risk of non-compliance to regulatory requirements (e.g., Sarbanes-Oxley, HIPAA), there is a tendency to err on the side of caution and retain all data, forever. Data lifecycle management seeks to meet the need of reduced storage cost while concurrently ensuring regulatory compliance and auditability by moving data between different types of storage, changing access permissions, and backing up data periodically.

3.1.2. Data storage management

The data in the digital universe needs to be stored in such a way that it is easy to analyze. According to an EMC report, only 22% of the digital universe was a candidate for analysis in 2012; less than 5% was actually analyzed (IDC, 2014). *Data storage management* refers to the organization of collected data in a format that facilitates and eases analysis. Operational systems employ the relational model, which uses tables—composed of columns and rows—as building blocks (i.e., data structures) to store and access data. Relational databases supporting business transactions are so pervasive that it is hard to imagine an organization without one. A limitation of relational databases is the way in which related data is spread around the database. For example, an order processing system for books may have a table each for books, order headers, order lines, and customers. While organizing the data logically in multiple tables (i.e., normalization) helps toward maintaining consistency in the database, it does not support easy or fast access. That is why reconciled systems that are designed to enhance data access typically employ another data structure, such as multidimensional cubes.

As compared to traditional database systems, big data systems can employ (1) a scalable distributed file system (e.g., the Google File System or one that is based on it, such as the Hadoop Distributed File System [HDFS]); (2) a NoSQL database (e.g., Amazon's Dynamo or Google's BigTable) that supports easy replication, eventual consistency, and processing capacity for huge amounts of data; (3) a programming model (e.g., MapReduce or Dryad) that facilitates data analysis applications. For example, Hadoop is a popular open-source infrastructure that is used as an alternative to store and process petabyte-scale data sets in a Hadoop Distributed File System, using the MapReduce framework on commodity hardware. In the context of big data, a *data lake* refers to an emerging data storage architectural concept, not tools and technologies. The data lake represents an environment for storing data in its native or near-native format, without the assurances and benefits associated with semantic consistency, governance, and security (Heudecker, Beyer, & Randall, 2015). While data lakes provide the flexibility and agility usually associated with data of unknown scope and use, they are unable to balance this with the auditability, stability, and performance typical of an enterprise data warehouse.

3.1.3. Metadata and master data management

Metadata is data that describes other data, thereby augmenting the usability of that data over its lifecycle (Beyer, Thompson, Lapkin, Gall, & Simoni, 2011). Metadata accomplishes this by making it easier to understand the meaning or the semantics of data by simplifying and speeding the search for data, thus improving the efficiency and effectiveness of the associated data use. In the context of big data, better inventorying data assets can be pivotal in the discovery of new uses for data (Simoni, Judah, & Zaidi, 2015). For example, as part of a marketing campaign involving social, digital, and in-store touchpoints, sentiment analysis employing big data (e.g., number of comments/posts/likes on Facebook) can be even more effective when combined with sales in the traditional reconciled system; this, in turn, requires metadata management that spans across multiple systems. Thus, enterprise metadata management tools are evolving from standalone vendors to vendors in a multivendor environment that support end-to-end data lineage (Simoni et al., 2015). Examples are provided in Table 3.

While metadata refers to the description of the meaning and the business rules associated with data, *master data* refers to an authoritative, business-relevant repository associated with non-transactional data (master data will have metadata associated with it). Master data refers to the consistent set of attributes that (1) are used across multiple business processes in an organization, and (2) depict the core entities (White, 2010). Examples of core entities are parties (e.g., customers, prospects, citizens, employees, vendors, suppliers), places (e.g., locations, offices), and things (e.g., assets, policies, products, services). Creating a common set of attributes—for example, for customers, products, and suppliers enterprise-wide—can ensure that the information assets are reused and that they provide maximal value to the business.

Master data management refers to the technology-enabled discipline that ensures uniformity of description among and consistency across an enterprise's shared master data assets (Spruit & Pietzka, 2015). The tools that support master data management facilitate a common use of business terminology for non-transactional data (Loshin, 2008), a consolidated 360-degree view of customers (Pula, Stone, & Foss, 2003), compliance with regulations (Delbaere & Ferreira, 2007), and improved reporting and analytics (Radcliffe, 2011). Master data management of product data—for example, using Oracle's Product Hub EBS—helps synchronize product information across heterogeneous data sources and enables a single product view for various business initiatives, thereby supporting customer service, returns, and logistics (White, O'Kane, Palanca, & Moran, 2015). Master data management of customer data—for example, using Siebel UCM (O'Kane & Judah, 2015)—can synchronize customer data across multiple sources to support customer-facing sales teams and processes. Note that master data management solutions can cannibalize customer master database service providers such as Acxiom, Austin-Tetra, and Experian, as they deliver data as a service (Zornes, 2010).

3.1.4. Data quality management

The process of managing the creation, transformation, and transmission of data such that it meets the use-related requirements of all users is referred to as *data quality management* (Mosley, 2010). Often, data quality is conceptualized in terms of accuracy alone; that is, how the fidelity of the data in the 'machine' world maps to the fidelity of the data in the 'real' world. However, data quality has many other dimensions besides accuracy. These additional dimensions include timeliness (i.e., meeting the time expectation of accessibility), completeness (i.e., all expected attributes have values), consistency (i.e., data in multiple datasets have consistent values), and currency (i.e., the degree to which data is current). The practice of data quality management hinges on five elements, the first of which entails identifying the acceptable threshold of data quality as based on anticipated and current uses of data (Mosley, 2010). The second element involves data quality-related awareness, to ensure the necessary buy-in from organizational stakeholders. The third element entails the identification of anomalies and the definition of business rules that ensure the requisite data quality. The fourth element encompasses the institutionalization of control processes for conformance with data quality rules. The fifth and final element involves monitoring compliance with defined data quality-related SLAs. Data quality management tools typically support data profiling, parsing and standardization, and cleansing and monitoring.

3.1.5. Data integration

Data integration refers to the practices, architectural techniques, and tools that facilitate consistent access and delivery of data across an organization's data subject areas toward the end of meeting different uses of data across a wide variety of business processes (Friedman & Zaidi, 2014). A key aspect of data integration platforms is these tools enable data that was designed independently to be leveraged together. The data integration platform generically follows three steps: (1) *extraction*, which involves connecting to source systems and selecting and collecting the requisite data; (2) *transformation*, which encompasses conversion of data to a standard format; and (3) *loading*, which includes importing the extracted and transformed data to target systems.

Data integration tools are useful in the context of organizations that provide data to, and receive data from, external trading partners such as customers and suppliers. These tools are also increasingly required in the support of data delivery to, and access of data from, platforms typically associated with big data initiatives. Given the recent advent of cloud-based services, the roles of data integration platforms are being rethought to include scenarios such as cloud-to-cloud, on-premises and cloud, and multienterprises with cloud configurations (Thoo, 2012).

3.2. Use

Webster's Dictionary defines *use* as "the act or practice of employing something." From the data-oriented perspective, use refers to the application of data to a particular end; in the context of this article, that end is generating insights for decision making. Based on the extent of data processing needed and comprehension of underlying data structures required, this article differentiates use as access, analysis, and delivery (see Figure 2). *Access*—or, specifically, data access—entails employing a query language (e.g., SQL) to retrieve data from a database. It requires user understanding of both data structures/storage and the associated query language. *Analysis*—which employs techniques whose origins may lie in diverse fields such as statistics, machine learning, econometrics, linguistics, or operations research—is a tool of conceptualization, not just data retrieval. It requires capturing essential elements of a problem and judgment in the interpretation of results. While the insights based on data access call for data retrieval (also colloquially referred to as *reporting*), those of data analysis are associated with algorithmic data processing. *Delivery*, in the context of data-driven decision making, refers to the ability to explore business questions and uncover insights, usually using graphical metaphors. For example, a dashboard using SAS Visual Analytics, Oracle Discoverer, and Tableau does not require the user to be cognizant of all the ways data might be stored and structured. Data delivery may involve data access/analysis and does not require knowledge of a query language. Table 4 outlines example business uses and sample tools/technologies that automate these uses. Next, the three types of uses are described in detail.

3.2.1. Access

The query language employed for data access is dependent on the system that it is queried upon. The scope of SQL, the lingua franca for operational systems, includes the definition of data structures, primarily in terms of tables or relations, and operations on the stored data. Multidimensional Expressions (MDX) is often used to query data that is conceptualized as multidimensional cubes. Big data systems employ several different query languages—for example, HiveQL and HBase. Developed by Facebook in 2007 and an Apache open-source project since 2008, Hive is an open-source petabyte data warehousing framework based on Hadoop. HiveQL is a SQL-like language that is used for running batch processes on Hadoop. Apache HBase is a NoSQL key/value store that runs on top of the Hadoop Distributed File System. Unlike HiveQL, HBase operations run in real-time on its database rather than batch-oriented MapReduce jobs.

Another technology that is being employed to enhance query processing is in-memory computing. In-memory computing relies on data retention in a server's main memory, or Random Access Memory (RAM), as a means of processing at faster speed. This enables high-performance, response

Table 4. Example of different types of uses and the associated tools/technologies

Use	*Example Business Uses*	*Example Tools/Technologies*
Access		
Operational	● Determine the average rating of books. ● Find the customer whose last name is "Jones."	● SQL (Structured Query Language) is the lingua franca for relational database management systems (RDBMSs), with a market size valued at $24 billion worldwide in 2012 (Fontecchio, 2012).
Reconciled	● List the average rating of different types of books, sold in-store and online, published in the last decade.	● Multidimensional Expressions (MDX).
Big Data	● Reporting that assesses web-based purchasing and digital retail.	● Apache Hive allows for querying petabyte-scale data stored in HDFS (Hadoop Data File System) for analysis using HiveQL, a SQL-like language. It runs batch processes on Hadoop. ● Apache HBase is NoSQL key/value store that runs in real-time on its database.
Analysis		
Predictive Analysis	● List the reasons for churning of customers. ● List the drivers of insurance fraud.	● Closed source tools (e.g., SAS Enterprise Miner, IBM SPSS Modeler) ● Open-source tools (e.g., R, KNIME) ● Mahout in the Hadoop framework
Forecasting	● Determine the sales in the next quarter. ● Ascertain the number of customer visits next week.	● SAS Forecast server ● Excel-based forecasting
Simulation	● Predict the number of fire companies dispatched on alarm. ● Understand bottlenecks in a system.	● Simio ● Arena ● ProModel
Optimization	● Determine transport route planning and loan approvals (operational). ● Allocate budget or other limited resources (tactical). ● Decide which company to acquire (strategic).	● Optimization modeling platforms include general purpose software tools (e.g., AIMMS, AMPL, FICO, IBM, Portfolio Decisions, River Logic, GAMS, LINDO, SAS). ● Optimization platforms are used with optimization solvers—such as Conopt, FICO Xpress, Gurobi, IBM CPLEX, SAS/OR, and Knitro—to build optimization applications (Kart & Zaidi, 2015).
Text Analysis	● Determine what customers are saying on social media. ● Improve product quality by analyzing repair notes.	● IBM SPSS Modeler ● Teradata Aster Analytics ● RapidMiner Studio
Web Analysis	● Measure the performance of a website. ● Evaluate return on online advertising. ● Optimize online content.	● Adobe Analytics ● Google Analytics ● IBM Digital Analytics ● WebTrends
Social Network Analysis	● In consumer market, identify the influencers. ● Detect online community and its evolution.	● DNA-7 ● Maven-7 ● Optimice ● Syndio

Table 4 (*Continued*)

Use	*Example Business Uses*	*Example Tools/Technologies*
Delivery		
Dashboard	- Show sales using geographic metaphor of a map.	- Tableau - Qlikview - Spotfire by Tibco

time-critical, and low-latency use cases, such as real-time customer interaction and fraud prevention (Edjlali, Feinberg, & Jain, 2014). SAP Hana and IBM DB2 with Blu Acceleration are examples of in-memory computing tools, which are now also available in the form of cloud services. In the context of big data systems, Apache Spark's multistage in-memory primitives provide performance up to 100 times faster than Hadoop's MapReduce paradigm. For example, Pinterest utilizes Spark Streaming with other memory-optimized technologies, like MemSQL, to measure and understand user engagement trends in real-time (Vanian, 2015).

3.2.2. Analysis

Different techniques for analysis are sometimes categorized as descriptive, predictive, and prescriptive. *Descriptive* techniques (e.g., inferential statistics) help understand and analyze business performance. They address the question: What has happened? *Predictive* techniques help provide explanatory or predictive relationships between predictors and target variables using, for example, data mining or predictive analysis, forecasting, and Monte Carlo simulation. They address the question: What will happen? Finally, *prescription* techniques seek to computationally determine high-value alternative decisions—using optimization, scheduling, queueing, and like methods—given objectives and a constrained set of conditions and resources. They address the question: What should happen? Table 4 lists several different techniques for analysis, including predictive analysis, forecasting, simulation, optimization, text analysis, web analysis, and social network analysis.

Predictive analysis refers to the ability to predict future behavior or unknown outcomes by employing a range of techniques, such as regression, decision trees, neural net, and support vector machines. Predictive analysis is used in predicting churn (or customer attrition), insurance fraud, crime, and even health outcomes. For example, credit risk models—which use data from a loan application to predict the risk of taking a loss—employ predictive analysis and are indispensable in credit decisions. *Forecasting* is a specific type of prediction using time-series methods that predict the value of an outcome at a specified time (e.g., number of visits on a website in the next week). Forecasting can be done using several different techniques, each with their own associated assumptions; these techniques include the moving average, exponential smoothing, and the regression model (Chambers, Mullick, & Smith, 1971). While predictive analysis focuses on generating a score for whatever is being analyzed (e.g., specific customer churning, a certain part failing), forecasting provides an overall aggregate estimate (e.g., sales in the next quarter).

Under the *Monte Carlo simulation* technique, a system model is built to experiment with or study how performance metrics might be affected by changes to the system design. Simulation models of business processes can help examine business practices in order to improve them. By focusing on interdependencies of resources and animating their interactions, bottlenecks can be pinpointed and new techniques tried in order to see if small changes can make a big impact. Instances of applications for simulation include analyzing patient flows in a hospital to allocate beds across wards, analyzing the manufacturing process in a refinery to improve throughput, and analyzing how CO_2 emissions might change in response to a cap-and-trade agreement.

Optimization—an approach that involves many alternatives, constraints, and tradeoffs—entails the use of mathematical algorithms to choose the best alternatives from a set of feasible solutions based on specified objectives and constraints. The objective is often formulated as maximization or minimization of, for example, profit or loss in a business setting or expected return in the scenario of investment. Optimization is finding different types of use cases including operational (e.g., dynamic allocation of seats on a flight among several fare classes, planning the timing and depth of markdowns in retail, providing sales people in the field with targeted prices for negotiations, scheduling airline crews), tactical (e.g., choosing retail assortments to be offered across several stores of a chain, allocating inventory across several nodes in a distribution network), and strategic (e.g., designing a distribution network—which type of facilities to build or lease, and where; designing a global supply network) (Kart & Zaidi, 2015).

Text analysis automates the processing of text data—including documents, discussion forums, emails, and surveys—from inside or outside the enterprise (Yuen, 2012). Input may be culled from emails, webpages, customer reviews, and social media content; corresponding output can be keywords, sentiments, and other structured information. The field of text analysis is at the intersection of information retrieval, machine learning, statistics, linguistics, and data mining. Text analysis has a wide variety of applications. Among others, it can help determine what customers are saying about a product or service on call center logs and social media streams like blogs, tweets, forum posts, and newsfeeds. It can clarify warranty claims based on warranty analysis. It can identify product quality issues by analyzing repair notes. It can lower cost of claims by scrutinizing customer service records. It can also help examine medical records and patient notes for the purposes of predictive epidemiology. A recent example of employment of text analysis involves Facebook's Topic Data. Topic Data was designed for and is aimed at businesses that want to better understand what people think about topics related to their business.[1] It allows marketers on Facebook to drill down into data of what over 1.5 billion users are saying about events, brands, subjects, and activities, thereby enabling businesses to make better decisions about how they market their products on Facebook.

Web analysis is performed via specialized analytic applications that are used to understand the web actions of users so as to enhance their online experience. Web analysis considers information from web server access logs, browser logs, user profiles, user sessions and transactions, cookies, user queries, mouse clicks, and other data generated by interactions of users on the web. Web analysis is employed to measure the performance of a website (e.g., number of visits, time spent on pages), understand who the visitors are (e.g., geographical information), gauge return on online advertising, and enable content optimization (e.g., time spent on different types of pages, transaction vs. informative). Leading web analysis platforms now provide support for different channels (e.g., mobile web) and have added advanced analysis—for example, automatically identifying valuable customer segments and the attributes that define them.

Social network analysis is used to examine patterns of relationships among people. Linkage-based social network analysis generally highlights link prediction, community detection, social network evolution, and social influence. For example, social network analysis is used in investigative applications, such as fraud detection or crime prevention. In the consumer market space, it is used to identify influencers. In telecommunications, social network analysis has employed call distribution records—such as number dialed, incoming caller number, call count, and type of call—to find out about calling circles of individual consumers (Rozwell & Aggarwal, 2015). Recently, vendors such as Jive and Zimbra have automated data collection and have embedded features—for example, people to follow or content to filter—thus enhancing the practical uses of social network analysis (Rozwell & Aggarwal, 2015).

3.2.3. Delivery

Dashboards—user interfaces that depict graphical presentations of trends in an organization—are often employed for the continuous monitoring of business performance as dictated and measured by metrics or key performance indicators (KPIs). The metrics often are based on a holistic scorecard—for example, the balanced scorecard (Kaplan & Norton, 1992). Dashboards that are used to monitor strategy are referred to as *performance management systems* (Quinn, 2010). Dashboards provide a mechanism for data delivery wherein users are not required to be cognizant of the underlying data structures of the data assets.

3.3. Processes and people

The third element of the data triad, processes and people, serves as the glue for the other two elements. A 2006 survey of 359 North American organizations with business intelligence and analytics systems revealed that a program for the governance of data was reported to be one of five success practices for deriving business value from data (see Khatri & Brown, 2010). Data governance has four elements (Logan, Dayley, & Childs, 2013):

1. Establishing who holds the decision rights and is held accountable for an organization's policies about capabilities for the creation, capture, storage, usage, control, access, archival, and deletion of data assets (Khatri & Brown, 2010; Tallon, Ramirez, & Short, 2013);

2. Harmonizing investments in accordance with policies;

3. Establishing measures to monitor adherence to policies; and

4. Ensuring processes are in accordance with the policies within tolerance to support decisions (i.e., risk management).

[1] Source: https://www.facebook.com/business/news/topic-data

In summary, the goal of data governance is twofold (Tallon et al., 2013). The first aim is to maximize the value of data so that reliable data is effectively employed for decision making. The second objective is to protect data so that its value is not diminished by inappropriate use.

To capitalize on the increasingly large and diverse digital universe, the role of Chief Data Officer (CDO) has recently come into prominence. The CDO's role is to manage, govern, and utilize data as an organizational asset. Gartner reports that by 2019, 90% of large organizations will have a CDO (Logan, Popkin, & Faria, 2016). The same study indicates that the most effective reporting relationship for CDOs falls outside the IT group, as data management is a business function rather than a technical one.

A wide variety of data assets requiring management presents with an increased level of complexity, and a data architecture provides a systematic arrangement of component elements (i.e., blueprint) so as to enhance planning data governance and portfolio management. The enterprise data architecture typically includes three components (Mosley, 2010):

1. An enterprise data that identifies subject areas, business entities—along with the associated properties of those entities—and business rules between business entities.

2. Information value chain analysis that aligns business entities of the data model with business processes and other architecture components such as the goals, strategies, projects, and technology platforms.

3. Data delivery architecture including data integration architecture, enterprise taxonomy for content management, metadata architecture, etc.

Data architecture management refers to the integrated process of defining, managing, and maintaining the specification artifacts or master blueprints that (1) support common business vocabulary, (2) define strategic data requirements, (3) outline high-level designs that meet the requirements, and (4) align with enterprise strategy (Mosley, 2010).

4. Discussion

The digital universe encompasses not only the virtual realm (i.e., one that has no physical equivalent, like social media) but also digital reflection of the physical realm (i.e., data sensed by actions humans or machines are already taking). With the exponential growth of the digital universe, data-driven decision making requires (1) identification of new data sources that support novel ways of representing problems, (2) identification of novel techniques for data analysis, and (3) discovery of creative applications of analytical insights for creating business value. New ways of representing problems can enable innovative algorithms or heuristics for solving the problems. Given that novel ways of representing or encoding problems can enhance problem solving, managers need to identify new data sources and the types of representations they enable.

Given the volume, variety, and vibrancy of the digital universe, we need to think of novel ways of representing problems and new ways of data analysis. Innovative data analysis can, however, have pitfalls. Consider the case of the now-defunct Google Flu Trends (GFT). A study published in *Nature* magazine famously declared the ability to predict—two weeks earlier than the Centers for Disease Control (CDC)—the prevalence of flu in geographic areas, based solely on Google searches (Ginsberg et al., 2009). In 2013, when GFT missed the peak of the 2013 flu season by 140%, the failure was ascribed to foundational issues related to validity of the measurement procedure (*construct validity*) and stability of the measures (*reliability*) (Butler, 2013; Lazer, Kennedy, King, & Vespignani, 2014).

To create value for the business, organizations must link data management, access, analysis, and delivery with decision making. Decision making in turn needs to dovetail with the business imperative of enhancing revenue, decreasing costs, and managing risks.

This article suggests that data-focused business managers can act as drivers for innovative organizations that are constantly employing new sources of data, identifying new ways of generating insights, and employing those insights for decision making. It is projected that by 2018 in the United States, an additional 1.5 million data-savvy managers will be needed to derive value from the huge and vibrant digital universe (Manyika et al., 2011).

5. Conclusion

Given the distinct shift in focus from data production to data consumption, business managers need to better understand how innovative uses of data can influence their work. As a marketing commentator blogged, value creation during data consumption is onerous (Palmer, 2006):

> Data is just like crude [oil]. It's valuable, but [left] unrefined it cannot really be used. It has

> to be changed into gas, plastic, chemicals, etc. to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.

While data undeniably has value, deriving value requires both broad and deep understanding of the digital universe. With respect to creating value for businesses, this article suggests that one must first link data analysis with the business imperatives of enhancing revenue, decreasing costs, and managing risks. While business managers are primarily keen on using data to generate insights, this article proposes that they also need to understand the roles of design and storage and processes and people in data management. These three elements—use, design and storage, and processes and people—comprise the data triad of the digital universe.

Acknowledgment

The author is grateful to Goker Aydin and Binny Samuel for their thoughtful comments on earlier versions of this article. He is indebted to the editor, Marc Dollinger, for his feedback and insight during the review process. Finally, he is thankful to Lisa Faye Miller for her very thoughtful copy editing of this article.

References

Bensinger, G. (2014, January 17). Amazon wants to ship your package before you buy it. *Wall Street Journal*. Retrieved from http://blogs.wsj.com/digits/01/17/amazon-wants-to-ship-your-package-before-you-buy-it/

Beyer, M. A., & Laney, D. (2012, June 21). The importance of 'big data': A definition. *Gartner.* Available at https://www.gartner.com/doc/2057415/importance-big-data-definition

Beyer, M. A., Thompson, J., Lapkin, A., Gall, N., & Simoni, G. D. (2011). Gartner clarifies the definition of metadata. *Gartner.* Available at https://www.gartner.com/doc/1424022/gartner-clarifies-definition-metadata

Butler, D. (2013). When Google got flu wrong. *Nature, 494*(7436), 155—156.

Buytendijk, F., & Laney, D. (2013, September 12). Big data business benefits are hampered by 'culture clash'. *Gartner.* Available at https://www.gartner.com/doc/2588415/big-data-business-benefits-hampered

Chambers, J. C., Mullick, S. K., & Smith, D. D. (1971). How to choose the right forecasting technique. *Harvard Business Review, 49*(4), 45—70.

Delbaere, M., & Ferreira, R. (2007). Addressing the data aspects of compliance with industry models. *IBM Systems Journal, 46*(2), 319—334.

Duhigg, C. (2012, February 16). How companies learn your secrets. *New York Times*. Retrieved from http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

Edjlali, R., Feinberg, D., & Jain, A. (2014). Market guide for in-memory DBMS, 2015. *Gartner.* Available at https://www.gartner.com/doc/2940217/market-guide-inmemory-dbms

Ehrenberg, R. (2010, July 7). Predicting the next deadly manhole explosion. *Wired*. Retrieved from http://www.wired.com/2010/07/manhole-explosions/

Faria, M., Linden, A., & Laney, D. (2016). Understand the data brokerage market before choosing a provider. *Gartner.* Available at https://www.gartner.com/doc/3183418/understand-data-brokerage-market-choosing

Fisher, T. (2009). *The data asset: How smart companies govern their data for business success*. Hoboken, NJ: John Wiley & Sons.

Florescu, D., & Kossmann, D. (2009). Rethinking cost and performance of database systems. *SIGMOD Record, 38*(1), 43—48.

Fontecchio, M. (2012, April 12). *Oracle the clear leader in $24 billion RDBMS market*. Retrieved from http://itknowledgeexchange.techtarget.com/eye-on-oracle/oracle-the-clear-leader-in-24-billion-rdbms-market/

Friedman, T., & Zaidi, E. (2014, December 15). Research library: Fundamentals for data integration initiatives. *Gartner.* Available at https://www.gartner.com/doc/2945018/research-library-fundamentals-data-integration

Gantz, J., & Reinsel, D. (2010, May). The digital universe decade—Are you ready? *IDC*. Retrieved from https://www.emc.com/collateral/analyst-reports/idc-digital-universe-are-you-ready.pdf

Gantz, J., & Reinsel, D. (2011, June). Extracting value from chaos. *IDC*. Retrieved from www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature, 457*(7232), 1012—1014.

Heudecker, N., & Adrian, M. (2015, May 12). Survey analysis: Hadoop adoption drivers and challenges. *Gartner.* Available at https://www.gartner.com/doc/3051617/survey-analysis-hadoop-adoption-drivers

Heudecker, N., Beyer, M. A., & Randall, L. (2015). Defining the data lake. *Gartner.* Available at https://www.gartner.com/doc/3053217/defining-data-lake

Heudecker, N., & Kart, L. (2014, September 9). Survey analysis: Big data investment grows but deployments remain scarce in 2014. *Gartner.* Available at https://www.gartner.com/doc/2841519/survey-analysis-big-data-investment

IDC. (2014, April). *The digital universe of opportunities: Rich data and the increasing value of the Internet of Things*. Retrieved from http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm

Jensen, C. S., Pedersen, T. B., & Thomsen, C. (2010). *Multidimensional databases and data warehousing*. San Rafael, CA: Morgan & Claypool.

Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard—Measures that drive performance. *Harvard Business Review, 70*(1), 71—79.

Kart, L., & Zaidi, E. (2015, May 28). Market guide for optimization solutions. *Gartner.* Available at https://www.gartner.com/doc/3064419/market-guide-optimization-solutions

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM, 53*(1), 148—152.

Laney, D., Linden, A., Buytendijk, F., White, A., Beyer, M. A., Chandler, N., et al. (2014). Answering big data's 10 biggest vision and strategy questions. *Gartner.* Available at https://www.gartner.com/doc/2822220/answering-big-datas-biggest

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science, 343*(6176), 1203–1205.

Logan, D., Dayley, A., & Childs, S. (2013). Information governance best practice: Adopt a use case approach. *Gartner.* Available at https://www.gartner.com/doc/2630023/information-governance-best-practice-adopt

Logan, D., Popkin, J., & Faria, M. (2016). First Gartner CDO survey: Governance and analytics will be top priorities in. *Gartner.* Available at https://www.gartner.com/doc/3183117/gartner-cdo-survey-governance-analytics

Lopez, J., & Cantara, M. (2014, November 21). 2015 predicts: Digital business disrupts paradigms — Financial, military, and process. *Gartner.* Available at https://www.gartner.com/doc/2920318/-predicts-digital-business-disrupts

Loshin, D. (2008). *Master data management*. Burlington, MA: Morgan Kaufmann.

Madden, S. (2012). From databases to big data. *IEEE Internet Computing, 16*(3), 4–6.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey & Company.* Retrieved from http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10), 61–68.

Mosley, M. (2010). *The DAMA guide to the data management body of knowledge*. Bradley Beach, NJ: Technics Publications, LLC.

O'Kane, B., & Judah, S. (2015, November 11). Magic quadrant for master data management of customer data solutions. *Gartner.* Available at https://www.gartner.com/doc/3166220/magic-quadrant-master-data-management

Palmer, M. (2006, November 3). Data is the new oil. *ANA Marketers.* Retrieved from http://ana.blogs.com/maestros/2006/11/data_is_the_new.html

Pezzini, M., Feinberg, D., Rayner, N., & Edjlali, R. (2014). Hybrid transaction/analytical processing will foster opportunities for dramatic business innovation. *Gartner.* Available at https://www.gartner.com/doc/2657815/hybrid-transactionanalytical-processing-foster-opportunities

Pula, E. N., Stone, M., & Foss, B. (2003). Customer data management in practice: An insurance case study. *Journal of Database Marketing, 10*(4), 327–341.

Quinn, K. (2010). How business intelligence makes performance management work. *Business Intelligence Journal, 15*(1), 8–16.

Radcliffe, J. (2011, February 28). MDM in 2011: Who's interested in MDM and why? *Gartner.* Available at https://www.gartner.com/doc/1565114/mdm-whos-interested-mdm-

Redman, T. C. (2013, July 11). Are you data driven?. Take a hard look in the mirror. *Harvard Business Review.* Retrieved from https://hbr.org/2013/07/are-you-data-driven-take-a-har

Rozwell, C., & Aggarwal, A. (2015, July 31). Market guide for social network analysis. *Gartner.* Available at https://www.gartner.com/doc/3104139/market-guide-social-network-analysis

Simoni, G. D., Judah, S., & Zaidi, E. (2015). Market guide for metadata management solutions. *Gartner.* Available at https://www.gartner.com/doc/3092921/market-guide-metadata-management-solutions

Spruit, M., & Pietzka, K. (2015). MD3M: The master data management maturity model. *Computers in Human Behavior, 51*(Part B), 1068–1076.

Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013). The information artifact in IT governance: Toward a theory of information governance. *Journal of Management Information Systems, 30*(3), 141–178.

Thoo, E. (2012, August 9). Data in the cloud: Harness the changing nature of data integration. *Gartner.* Available at https://www.gartner.com/doc/2113515/data-in-the-cloud-

Toonders, J. (2014). Data is the new oil of the digital economy. *Wired.* Retrieved from http://www.wired.com/2014/07/data-new-oil-digital-economy/

Turner, V., Gantz, J. F., Reinsel, D., & Minton, S. (2014). The digital universe of opportunities: Rich data and the increasing value of the internet of things. *IDC.* Retrieved from http://www.emc.com/leadership/digital-universe/2014iview/index.htm?.cmp=micro-big_data-general-emc

Vanian, J. (2015, February 18). Pinterest is experimenting with MemSQL for real-time data analytics. *GIGAOM.* Retrieved from https://gigaom.com/2015/02/18/pinterest-is-experimenting-with-memsql-for-real-time-data-analytics/

White, A. (2010, September 21). MDM 'primer': How to define master data and related data in your organization. *Gartner.* Available at https://www.gartner.com/doc/1438416/mdm-primer-define-master-data

White, A., O'Kane, B., Palanca, T., & Moran, M. P. (2015). Magic quadrant for master data management of product data solutions. *Gartner.* Available at https://www.gartner.com/doc/3166917/magic-quadrant-master-data-management

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences, 112*(4), 1036-1040.

Yuen, D. (2012, March 21). How BI leaders can get started with text analytics. *Gartner.* Available at https://www.gartner.com/doc/1956821/bi-leaders-started-text-analytics

Zornes, A. (2010). 10 key trends in MDM. *Information Management, 20*(3), 27–30.