International Journal of Data Science and Big Data Analytics

Publisher's Home Page: https://www.svedbergopen.com/

SvedbergOpen
DISSEMINATION OF KNOWLEDGE

Research Paper

Open Access

Trends in Data Warehousing Technology

Bola Abimbola[1*]

[1]Universidad de Oviedo, Escuela Universitaria Jovellanos, Francisco Tomas y Valiente 33201 Gijon, Oviedo, 33006, Spain.
E-mail: UO285018@uniovi.es

Abstract

The research looks at the concepts associated with data warehousing. These include: cloud computing and big data analytics. By analyzing the growth of these technologies, it is possible to tell the trend in which data will take in the coming future. A lot of data is in use at the moment in very large packets and volumes. The technologies that are associated with data use are associated with data use are immense and greatly spread across all computing gadgets in use. Cell phones, microcomputers, and other computing gadgets all take advantage of data warehousing technology. The research presents a case for the proliferated use of big data technologies across the world. Data use, and data users are separated using the concept of 'abstraction.' As users continue to use big data analytics and warehousing more often, they are less aware of the dangers and concerns in the use and reliance of virtual computing facilities. The study seeks to establish the influence of current technologies on data warehousing and how they lead to the growth of data storage units across the globe. The survey on data warehousing trends is focused on software technologies that highly rely on the data warehousing facilities. Social media is acknowledged as one of the leading factors behind the growth of data warehousing facilities. This research seeks to establish the use of data in social media and other applications. The concept of data analytics is stressed as a key issue in data warehousing. Being that the data in warehouses is highly sophisticated and voluminous; the need for specialized software to undertake sorting and searching purposes is reviewed. The study seeks to establish that the use of big data is an important concept as well. This concept is evaluated and well explained. The concerns with data warehousing are also evaluated in an effort to realize more need to secure data. The research concludes with an appeal for user awareness on data warehouse capabilities. Users are expected to be more aware of the purpose and capability of data and in essence, use these facilities from a professional point of view; not as novices.

Keywords: Data warehouse, Big Data, Virtual storage, Cloud storage

1. Introduction

This Data warehousing technologies have been in existence since the advent of cloud computing. For data to exist within a cloud setup, it ought to be stored in a storage location. This location should be abstract to the cloud user but in essence, exist physically. The aspects of the location that would make it suitable as a secure storage for data are considered in the setting up of the facility (Thusoo *etal.,* 2010). This facility is referred to as a data warehouse. It is

often the case that a warehouse will comprise tens to hundreds of data storage units referred to as data banks. A data bank is a ubiquitous unit of data aggregation that stores records relating to a particular file system. A data bank makes it possible for the user to retrieve millions of bits in an ordered manner without taking long to reference specific files within a database. In modern data computing technologies, the concept of data mining is applied to fish out specific data from a large storage unit. In essence, data mining takes the form of an algorithm that sequences data in a storage facility using searching data structures. A search can peruse through millions of files in a minute to ease the work of the user. This is the same technology applied in mining tweets on Twitter and comments on Facebook (Halter *et al.,* 2016). Data mining has become essential in the modern world due to the proliferation of social media and the ability of micro computing units such as the smartphones and the palmtops to store larger volumes of data. Cell phones nowadays even have the technological capacity to search using voice commands. As algorithms get more complicated to handle the growing hardware and software needs in modern smartphones, so does the complexity of the data in the storage units advance (Gonzalez *et al.,* 2010)? Every piece of data in a file system is stored within a data file. A file will often comprise a few kilobytes to even Gigabytes of data. Every file is denoted using a file name and a file extension. The extension indicates the type of file stored in the storage unit or device. The common file extensions in use are:.docx,. rft and .doc (for word processors), .xls (for spreadsheets), .gif, .jpeg, and .jpg (for images) as well as .txt (for common text files). Data storage facilities are based on the extended use of database management software to handle very large data units (Thusoo *et al.,* 2010). The software in use is a more advanced form of structured query language that handles large volumes of data. Such volumes are referred to as 'big data.' The software used to handle big data is referred to as 'analytical software for big data' and the technology used is the 'big data analytics.' It is not practical to separate the notion of data warehousing from big data analytics trends and technologies (Inmon *et al.,* 2010). Trends in data warehousing technology seek to address new challenges in handling data. These challenges include; handling large volumes of data, securing data from unauthorized access, preventing the tapping of data during transmission and the guarantee of data safety and perpetuity even when not in use (Halter *et al.,* 2016). Various companies maintain data warehouses and offer services, such as, data storage and backups, data recovery and error handling as well as database administration. As more people get hold of the personalized digital assistants (smartphones), there is need to consider the reality that data handling will be done on a more personalized level in the future. This is why trends in data management technology are taking the shape of analytical software engines that require little human assistance (Takecian *et al.,* 2013).

2. Data Warehousing Concepts

2.1. The Concept of Big Data

A bit is the smallest unit of data. It is a string comprising of eight characters in an array format. This array contains data packets of very menial sizes. A byte comprises 1024 bits arranged in a longer, more complex array. As the sizes increase, the array becomes more complicated (Krishnan, 2013). The next unit of data is the kilobyte. A kilobyte consists of a thousand bytes but in computing terms; this is exactly estimated to 1024 (28). A thousand Kilobytes make up a megabyte (the size of most files used in computing is in megabytes). A thousand Megabytes give rise to a large unit referred to as Gigabytes, which then lead to Terabytes. The term 'big data' often alludes to Terabytes. It denotes data that cannot be stored or sorted in a single computing unit. This is why there is a need for a data warehouses to store the data. The processing is done using special processors as is the maintenance of the data (Erickson, 2013). Big data is in a sense the technology of exponential data use. Data cannot easily be quantified. This is because, bits and bytes are generated and reproduced each day. It is quite difficult to get professional data figures in real time. However, the expedited use of data can be managed using monitoring facilities within the big data analytics software. The software can determine data usage and advice on the need to expand the facility. However, since users are sending packets and data streams each day, it is quite difficult to ensure that the use of the data is common to all people accessing the cloud. Data and its use are separate entities. However, there is anxiety that the use of the facility is not always simultaneous among users. Queuing concerns make it difficult for users to access facilities when there is a deadlock occasioned by similar requests over long periods of time. Big data analytics software must thus be optimized to guarantee that it does not interfere with queuing of requests (Abellò *et al.,* 2013). Big data is a phenomenon that is as popular as cloud computing and mobile computing. However, the intricacies of big data are not well known. Most users are unaware that while using cloud computing products and services, they interact with big data (Krishnan, 2013). Big data has made cloud computing possible. The number of activities undertaken by cloud servers often generates a lot of data and processed information. Such information is channeled to big data server and

facilities for storage. Mobile computing platforms utilize big data to store aspects of user identity in specialized drives. For instance, Google's Drive on smart phones is used as a repository for all social media applications that run on the Android platform. Social media platforms and applications such as WhatsApp, Facebook and Twitter gather a lot of personal data from the user. It is thus crucial that the user back up such information so as to store special moments about their lives. Big data technologies make this possible (Haq, 2016).

2.2. Big Data Analytics

Big data requires specialized software for analysis, storage and sorting purposes. Such software is referred to as analytical software for big data. Examples of analytical software for big data include: Enterprise Resource Planning Software, Business Intelligent software, Supply Chain Management software and Space Exploration software (Abelloˊ et al., 2013). The data applied in the analysis is often too large to be processed by a single processor thus requires millions of processors grouped together. The speed involved in big data analytics software is similar to what is applied in the supercomputer processors. The technology is referred to as 'floating point' and the speed is in floating points per second (FLOPS). This is larger than the instructions per second processor applied in the ordinary microprocessor. The fastest speeds range from 10-50 FLOPS. Such systems comprise of over 5 million processors (Haq, 2016). Every organization processes data. Data and information are significant concepts in the field of information technology (Erickson, 2013). Where a business' role is to get input from users and process it into information of some kind; the user is expected to develop a critical sense of reasoning and understanding on the processes involved. Users need to process a lot of information to offer clients the necessary data output they require. With business intelligent systems, the user needs to understand complex formulae and algorithms (Revels and Nussbaumer, 2013). The system is developed to process a significant number of computations concerning the functioning of the business. It is also critical that the business realizes profits based on a model that is within reach of its goals. Business systems develop and actualize models for the user (Sremack, 2015). They make it possible to run on little resources and achieve great impact without straining to meet certain labor and capital requirements. Productivity is thus made easier. Business intelligence systems do not just evaluate the performance of the organization; they predict trends and promote growth patterns. The systems used enable companies to minimize waste by reducing operational costs. Business intelligence systems are preferred in organizations that make very critical decisions about small aspects of data (Gersil, 2016). Such organizations utilize big data systems, such as, data banks and cloud storage facilities to maintain accurate repositories on user and organization information. An intelligence system is able to determine where the weak link in an organization is. This is done using parameters, such as, income statements, productivity, client feedback systems and ordinary messaging. It is critical that information on clients such as public relations and account management changes that need to be done are captured. The system is thus able to provide an accurate assessment of an issue and offer an amicable solution to solve an imminent problem (Sremack, 2015). The challenge with business analytical systems is that they are quite expensive and highly optimized for specialty. It is thus not easy to implement some of these systems from one organization to the next (Krishnan, 2013). It is often critical for an organization to invest in a system that is specific and tailored to their needs. This enables the organization to beat competition and offer solutions that are innovative. More importantly, software of this nature is bound by copyright laws. Such laws govern the working of users and discourage infringement of patent rights. For instance, an organization dealing with security systems may want analytical software that explores the security loopholes in their own organization and offers solutions to handle the concern. Software of such nature cannot be sold to any other party as it may compromise their internal security arrangements (Takecian, et al., 2013). They would thus have to meet the cost of software copyrights and non-disclosure on the part of the developers. The trends in database management and security imminent in society demand for advanced features and mechanisms to manage software. It is crucial for an organization to restrict the authorization for use of data analytical software to few users (Halter et al., 2016). Most of the information that proceeds from such systems is of a valuable nature. It may jeopardize not only the trade secrets of the organization but specific employee and client information as well. Security risks of such nature can only be minimized where the number of users is limited as well. Regardless, there are exceptional cases where technical or expert intervention is required. Such cases include: database migration, virus attacks and system malfunctions (Revels and Nussbaumer, 2013). Where such cases arise, there should be a specific pass key allocated to the administrator of the system to restrict the database view to a particular extent. No external users need to be allowed full disclosure where it comes to business intelligence systems. Such an action places an organization under threat of intruder attacks. Indeed, many companies that have witnessed colossal data theft suffered such attacks while undertaking maintenance and review of the system.

3. Trends in Data Warehousing

3.1. Growth and Proliferation of Data Warehousing Technology

A data warehouse is a very useful facility. Above all, it comprises the entire storage unit of data used in operation, such as, the management of large network systems, such as, global Internet storage facilities, institutional big data analytics, such as, weather forecasting, space exploration, and research, as well as handling large volumes of data used in social networking. Data warehouses ensure the conditioning of computer parts. A computer requires sufficient cooling and prevention from exposure to strong magnetic fields, ultraviolet and gamma rays as well as ionic liquids (Abello´ *et al.*, 2013) .

The warehouse ensures that the facility is kept in a cool, dry and secure place. Many of these facilities include guards, security cameras and electric fences to bar intrusion. The goal is often to ensure that the fidelity of the components within the databanks is guaranteed. A warehouse can have a hundred to even thousands of databanks, each with over a thousand powerful processors. This technology is referred to as compartmentalized large scale integration of processors. A warehouse helps to secure this very expensive facility. Data warehousing technology use has been on the rise since the advent of cloud networks. Data in clouds is often stored in a databank for safe keeping and ease of access. However, in the modern era; where social networking technologies are on the rise; there is a lot of insistent on data warehouses for managing the network and growing the client data without interfering with the storage capacity of the system (Sremack, 2015). Essentially, social media networks do not limit the size of storage space a network or user can have. They keep expanding to meet growing user demands as they continue to offer more storage capacity to marketers and users alike. Regardless of the number of users, social media sites can handle data traffic and guarantee fidelity of data stored on the networks. This calls for the social network administration to maintain cloud networks on data warehouses they set up and propagate (Inmon *et al.,* 2010). Many organizations are weighing the costs and benefits of having a data center. Data warehouses have the basic advantage of minimizing the number of servers requests loop onto in a hypertext mark-up universal resource locator page request. When a facility is closer to the users, the requests made to the databank servers are processed faster than when the resource is further away (Sremack, 2015). It is important to however consider that while this may be necessary for some organizations, it is not always necessary for others. There are organizations that have very few requests to the server. They thus may not need to have data warehouses or banks. Essentially, there is a realistic concern about the cost of maintaining data warehouses. It would suffice if setting up the facility would only encompass hardware and software costs. However, overheads such as security and housing costs make the accrued expenses too high for many companies (Abello´, 2013).

3.2. Data Reporting

Data reporting is a growing concern in the cloud business world. Information from the source to the user is often needed fast and increasingly at an alarming rate. Handling of such information calls for special software that keeps track of the changing user needs. This software is referred to as the analytical software for big data. Software of such nature exists in the market as demand increases (Bennett, 2016). The issue is never about processing but reporting of information. Where data integrity issues are concerns, it is not only the accuracy but the timeliness of data that matters to the user. Data presented needs to be within the scope of time that makes it useful. To make this possible, it is critical that the user employs software that prompts outputs whenever they are desired. Such software is referred to as data reporting software. It takes advantage of database facilities such as macro, modules, forms and tables to generate accurate reports where and when needed (Watson *et al.*, 2015). Data reporting software has to be very accurate. The nature of information it can take as input also needs to be vetted using a strict algorithm to ensure validity (Watson *et al.*, 2015). Data validation rules and procedures that make this possible are within the basic scripting languages. However, to every organization; especially those using analytical software, more validation procedures are desired. It is thus critical to ensure that these procedures are put in place before the data reporting software is used. Where such software is used without proper data verification, there exists a situation of false positives. Such a name is given to inaccurate data that has been processed using accurate tools (Bennett, 2016). An example where this happens includes; analysis of census data collected by unauthorized census officials can lead to misreporting hence false positives.

4. The Fusion Between Cloud Networks

4.1. Existing Cloud Storage Facilities

Cloud storage facilities are offered by service providers with significant resources to develop a sound infrastructure that encompasses all the services managed within the cloud. These services are then offered to the user at a subscription fee. The major vendors that offer these services include: Amazon Cloud Services, the Microsoft Cloud, the Google

Cloud Service and the IBM Cloud (Sremack, 2015). Other companies that offer one or two of the cloud services include; VMware, Citrix and Oracle. These networks interlink the World Wide Web servers using special networking protocols. Essentially, these protocols make it possible for cloud users subscribed to any vendor to access their files though a different cloud provider at a small fee. The use of cloud networks is thus universal and interlinked. Among the existing virtual drives, infrastructure is the most valued of the services offered. The ability of a vendor to capture more users depends on the number of warehouses they have and the rates they offer for storage space. Many thus offer at least 2 Gigabytes of free data to entice users to the infrastructure (Miller and Han, 2009). A virtual drive setup can be as simple as a set of two computers sharing data over a cloud. The need for a data facility within a locality that is abstract to both users is the only major impediment to such a setup. The cloud ensures that abstraction is not only guaranteed, but it is also encouraged. Abstraction of user data needs to be in a manner that does not allow users to be informed of the location of any facility they use (Miller and Han, 2009). This keeps these facilities away from intruders. IP spoofing is one major threat to computing resources. Where an intruder gathers information about IP addresses to various data equipment, they can launch any attacks on the network. Cloud service providers must thus be discrete about the location of warehouses and the use of cloud facilities to ensure that they can relocate at will without informing the user; for the user's own safety (Miller and Han, 2009).

4.2. Virtual Storage Concerns

The perpetuity of cloud data is based on the use of data management systems. Storage is important and should never be tampered with. Regardless, many data storage systems are not fail-safe. Cloud networks need to have proper administration in order to guarantee proper functioning and validity of all the sectors of data storage drives within the data banks. Regardless of the number of databanks or data warehouses in use, there is need for a vendor to guarantee the user that they will always have the ability to store and retrieve data at will (Sremack, 2005). Such guarantee is managed by ensuring that there is a redundant system that powers the storage system (data warehouse) as well as functioning analytics software that is compatible with a database and does not fail in its function to store sequentially arranged data. Storage in the computer cloud can be hampered by the available storage space. The vendor thus needs to evaluate the storage space available from time to time in order to guarantee cloud users that they can expand their data storage capacities at will. All these measures are based on effective research so that the vendor maintains a cost-efficient and profitable cloud (Aji *et al.*, 2013). Cloud network infrastructure comprises the hardware and software components that interlink various computer clouds. The concept of parallel computing best explains how the cloud network infrastructure works. Since users are expected to share a common pool of resources, they have different time slices to access the resources they need. However, due to the capabilities and superior speeds guaranteed by the cloud, these time slices can be calibrated in 1/1000 of a second (March and Hevner, 2007). The use of the resource thus does not seem shared. The infrastructure's ability to abstract its capabilities from the user is based on the number of users accessing the resource at the time. If all users decided to upload a Facebook video at the same time, the Facebook file server would definitely be constrained for resources. However, based on the usages of the network, the service provider has an algorithm that tracks the amount of traffic in the network. This keeps the social network growing at a rate that matches the user demands. Virtual storage involves various cryptic systems. The cloud is expected to offer practically any service desired by the user. However, these services are based on some subscription fee that has to be paid from time to time by the user (Kutemperor, 2015). While many users can afford the considerable fees imposed by the service provider, there are issues about the safety information on clouds that service providers have to deal with on a daily basis. As more users enlist for their services, cloud service providers have to worry about expanding their facilities to accommodate user needs. Users also face potential data loss from their own negligence. When such losses occur, they are often borne by the service provider out of good will. The insurance sector has yet grown to accommodate data assurance in most parts of the world. Data thus remains a vital resource yet users do not put in place pro-active measures to guarantee the fidelity of this data (Miller and Han, 2009).

4.3. Virtual Storage Capabilities

Cloud storage capacities are often abstract phenomena that the user is not aware of. The use of cloud resources is abstract to the user so that they get the feeling that they are privy to a large storage and processing power that is often marketed as 'infinite.' Storage on the cloud is never a quantity that can be easily determined by the vendor as well. When a warehouse facility is set up, it is often several millions of terabytes in size (Halter *et al.*, 2016). The facility also include storage capabilities for more data banks. Since each data bank can comprise hundreds of thousands of processors, it is practical in a sense to say that the cloud storage capability is infinite. Nevertheless, the more the processors and hard drives on these units, the more difficult it is to maintain them. They consume a lot of power and cost a fortune to install. It is thus advisable that a vendor invests in research to find out the necessary storage

capacity they require before engaging Engineers to develop the data warehouses (Chen *et al.,* 2012). The storage disks used for cloud storage are magnetic in nature. They require to be kept in regions where activity is minimized and potential for damage is minimal. However, the limitation of hard drives in a data warehouse is based on cost effectiveness and the ability of such facilities to host sensitive data (Revels and Nussbaumer, 2013). Most often, an organization will have more than one facility to cater for the different sources of data. There are sources that require very secure facilities. Communications such as government cables need to be safe guarded and thus such a client would get preferential treatment. Nevertheless, organizations expand their facilities as users make economically viable requests that they can pay for. It is often about the client's ability to foot the bill. However, due to the colossal financial investment required to set up warehouses, organizations that have these facilities are very few. The capability issue is based on the viability and not capacity to store data. Where capacity lacks, organization outsource data storage and management services to more endowed organizations at a fee (Kutemperor, 2015).

5. The Future of Virtual Storage Facilities

The onset of the 'dot com' generation in the year 2000 did not anticipate the nature of computing power experienced currently. Computers have grown to continue to offer services and products that meet user needs from time to time. Virtual storage facilities were initially known to be mail and file servers belonging to large organizations (Revels and Nussbaumer, 2013). Currently, more organizations are adopting virtual storage for practically all their data needs. The storage of data in data warehousing is becoming a common trend. Many companies and individuals are adopting the storage of data in drives, the use of software platforms on virtual storage locations and the access to infrastructure resources that they do not own. Such trends will be common in the future as the growth of infrastructure used in virtual storage systems continues to be more sophisticated. It is thus essential that as more people embrace technology more, the reality of the dangers they expose themselves to in the virtual world is made more aware to them (Gersil, 2016). Currently, the coverage of internet globally has not reached the global population of 7 billion people. The same challenges have been realized with the global accessibility of smartphones and computing devices. Even as efforts are made to convince more people to embrace technology, it is difficult to get everyone on the internet radar and in consistent use of smartphones (Kutemperor, 2015). Challenges that make smartphone coverage difficult include; lack of electricity and little network reception for network wave forms and primitivism. Efforts such as the global digital migration will make changes to the current situation on warehousing technology adoption. As more users demand of digital services, so will the vendors make them available to these users. Growth in technology is thus highly dependent on market changes and government digital adoption policies. While data storage facilities continue to advance with the cloud, there is a significant trend towards expansion of user disk spaces. The mobile computing devices users use in the modern world contains internal storage spaces and removable storage units such as memory cards and flash drives (Sremack, 2015). Users can back up data in more than one way. However, the issue is in the safety of these devices. Information in external storage drives can easily be used to taint the owner's public image. Where public image is concerned, the user needs to secure their storage devices with all cryptosystems at their disposal. It is necessary for users to consider insuring their storage facilities as well. Even as they back up their data on drives and cloud storage facilities, the safety of their computing devices is still at stake if they have no insurance cover (Revels and Nussbaumer, 2013). Insuring smart phones and computers offers the user the satisfaction that they will not only have their data backed up on the cloud but alternative mobile computing devices to use in case of accidental loss or damages.

6. Conclusion

Data warehousing technologies across the world keep improving to make the user experience in data management better. More users are in tune with the latest technologies in use at the moment. Some of the basic skills needed for managing a data warehouse are already handled by the service providers. There is thus no need to bother about how to handle or deal with big data. In essence, even as society continues to embrace technology and the use of data warehouses, they are less able to understand the technology itself. Abstraction makes it less common for users to understand the underlying technologies applied in data warehousing technology than they understand the functionality of the technology. Data warehousing and big data technologies make data across the globe appear to be centralized. This is because, the use of the data is indiscriminate and independent of platform chosen. Users can thus choose any vendor to manage their data while retaining fidelity of information stored at all time. As big data and data warehousing technologies continue to gain traction, it is important to consider the security and usability issues in big data. Such considerations will enable the technology to attract more users in the future.

References

Abello´, A., Darmont, J., Etcheverry, L., Golfarelli, M., Mazo´nLo´pez, J. N., Naumann, F., and Vossen, G. (2013). Fusion Cubes: Towards Self-Service Business Intelligence. *International Journal of Data Warehousing and Mining*, 9(2), 66-88.

Aji, A., Wang, F., Vo, H., Lee, R., Liu, Q., Zhang, X., and Saltz, J. (2013). Hadoop GIS: A High Performance Spatial Data Warehousing System Over Map-Reduce. Proceedings of the VLDB Endowment, 6(11), 1009-1020.

Bennett, M. (2016). An Industry Ontology for Risk Data Aggregation Reporting. *Journal of Securities Operations Custody*, 8(2), 132-145.

Chen, H., Chiang, R.H., and Storey, V.C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165-1188.

Erickson, P. (2013). Data Warehouse: 82 Most Asked Questions: What You Need to Know. Emereo Publishing. Brisbane, Australia.

Gersil, M. (2016). Data Mining Practice in the Scope of CRM: Churn Management in the Telecommunication Sector. *International Journal of Business Management Economic Research*, 7(2), 560-567.

Gonzalez, H., Halevy, A., Jensen, C. S., Langen, A., Madhavan, J., Shapley, R., and Shen, W. (2010). Google Fusion Tables: Data Management, Integration and Collaboration in the Cloud. In Proceedings of the 1st ACM Symposium on Cloud Computing (pp. 175-180). ACM.

Halter, O., Kromer, M., Kutemperor, N., and Soares, D. (2016). Dipping a Toe into Data Lakes. *Business Intelligence Journal*, 21(2), 40-46.

Haq, Q.U. (2016). *Data Mapping for Data Warehouse Design.* Amsterdam: 1st Edition - December 8,

2015. Author: Qamar Shahbaz, eBook ISBN: 9780128053355. Paperback ISBN: 9780128051856.

Inmon, W. H., Strauss, D., Neushloss, G. (2010). DW 2.0: The Architecture for the Next Generation of Data Warehousing. 1st Edition, Morgan Kaufmann Publishers. USA.

Krishnan, K. (2013). Data Warehousing in the Age of Big Data. 1st Edition, Morgan Kaufmann Publishers. USA.

Kutemperor, N.T. (2015). Price/Performance Considerations in Building Data Warehouses for Big Data. *Business Intelligence Journal*, 20(2), 8-13.

March, S.T., and Hevner, A.R. (2007). Integrated Decision Support Systems: A Data Warehousing Perspective. *Decision Support Systems*, 43(3), 1031-1043.

Miller, H.J., and Han, J. (Eds.).(2009). *Geographic Data Mining and Knowledge Discovery*. CRC Press.

Revels, M., and Nussbaumer, H. (2013). Data Mining and Data Warehousing in the Airline Industry. *Academy of Business Research Journal,* 369-82.

Sremack, J. (2015). Responding to a Legal Data Request: The Role of the Data Warehouse Team. *Business Intelligence Journal*, 20(1), 39-43.

Takecian, P. L., Oikawa, M.K., Braghetto, K.R., Rocha, P., Lucena, F., Kavounis, K., ... and Ferreira, J.E. (2013). Methodological Guidelines for Reducing the Complexity of Data Warehouse Development for Transactional Blood Bank Systems. *Decision Support Systems*, 55728-739. doi:10.1016/j.dss.2013.02.008

Thusoo, A., Shao, Z., Anthony, S., Borthakur, D., Jain, N., SenSarma, J., ... and Liu, H. (2010). Data Warehousing and Analytics Infrastructure at Facebook. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data (pp. 1013-1020). ACM.

Watson, A., Watson, B., and Vallmuur, K. (2015). Estimating Under Reporting of Road Crash Injuries to Police Using Multiple Linked Data Collections. *Accident Analysis and Prevention,* 8(3), 18-25. doi:10.1016/j.aap.2015.06.011