

Big Data Analytics in Healthcare: Opportunities and Challenges

Norah Dhafer Al-Qahtani

Faculty of Computing & Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Shatha Ali Al-Asem

Faculty of Computing & Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract

Healthcare is a vast and data rich industry. The administrative details stored into database hold a large number of transactions for each treated patient. The expansion and the adoption of the electronic health have been increasing the overall amount of the data that is available rather exponentially. In spite of all the efforts that have been made in the healthcare industry it has still been unable to rather leverage the vast data that can help in improving the care and the healthcare operations simultaneously. Along with this, the rapidly growing field of the big data analytics in the healthcare has started playing a major or the integral role in evolution and the development of the healthcare practice and research. Big data analytics have also provided the tool that shall help in accumulating, managing, assimilating, analyzing, the structured and the unstructured data. The concept of dig data analytics has been applied with the aim of increasing the efforts of the care delivery and the disease exploration simultaneously. In spite of all the efforts the adaptation rate and the research development has been hindered with the major problems that inherent within the big data analytics model. The paper basically discusses the concept of big data analytics in healthcare. Moreover, the paper also looks to explore the advantages of the concept. Similarly, the paper has also explored the opportunities and challenges that are associated with the idea of big data analytics in healthcare. At last the paper is closing with the general examination of the idea and the exploration.

1. Introduction

Historically, healthcare industry produced a huge amount of data called big data, causing to have a huge number of patients' records. The present trend in healthcare data storage is to rapidly digitizing the big data instead of hard copy. The volume of healthcare data is growing during the last a few years, in addition to the changes of healthcare compensation models, emerging new critical factors into healthcare environment. Big data refers to the large amount of data complexity, rapidity, and variability which require new technologies to enable capturing, storing, distributing, managing, and analyzing data. Current big data analytics are applied for the reason of the vast amount of patients' data regarding to their medical and health data. Similarly, big data in healthcare is relatively associated with data analytics in order to describe big data from three basic perspectives: variety, velocity, and volume (Raghupathi & Raghupathi, 2014).

The challenges of healthcare environment limit the healthcare sector to be extended. The history of patients, clinical data, progressing comments, effects of drugs, tests results etc. should be kept into database. The concept of data analytics involves the optimality of hospital resources usage that improve the service provided to patients, using dataset's combination in order to effectively distribute costs, process data innovatively, and support decision making. Several researches claimed that using some tools such as warehousing, decision support systems, or data mining techniques is optimized to overcome the critical problems and issues of traditional resources usage. Moreover, the protection of data privacy of the patients under data sharing pressure between organizations represents more challenging issue to healthcare. These private and sensitive data might be shared between insurance companies, hospitals, and clinics that minimize the level of data protection. On the other hand, the most important question should ask about the mechanisms used for big data analysis in healthcare in order to efficiently improve healthcare delivery. Predictive modeling, quality measurement, and population health are the major applications used for data analytics and they are rapidly progress (White, 2014).

The healthcare market for quite a while has been creating furthermore assembling extensive measure of data which is extricated and driven by record keeping, administrative necessities, patient care and consistence. Most of the data is by and large put away in a printed copy frame, the most recent or the present day pattern is towards the quick digitalization of the a lot of data (F. E. Dewey, 2014). Since it is driven the compulsory and mandatory necessities and the possibility to enhance the general nature of healthcare conveyance which has diminished the expense, these gigantic and substantial amounts of data is called as the enormous data which holds the guarantee of rather supporting the more extensive scope of medicinal furthermore the healthcare capacities, which likewise incorporate the other clinical choices, populace health administration, ailment observation (Manyika J, 2011).

Huge data in healthcare essentially is all the electronic healthcare data sets so expansive furthermore complex that they are now and then difficult to deal with the officially created programming and equipment

which can't be overseen effortlessly with the normal data administration instruments furthermore through the conventional techniques.

The organization of this paper will be partitioned into five sections: big data analytics in healthcare, big data analytics opportunities for Healthcare, challenges, literature review, and conclusion. First section discusses some identification of big data, big data analytics, and big data analytics in healthcare. Second section, big data analytics in healthcare, is partitioned into two subsections: benefits to healthcare and market factors driving healthcare. Third section called challenges discusses the potential challenges may face big data analytics in healthcare. Fourth section includes a set of studies and researches related to this topic from previous periods. Finally, some conclusions and results are stated at the end of this paper.

2. Big data analytics in healthcare

The healthcare information volume is all normal to develop and extend itself in the coming years. Alongside this, the healthcare repayment models have been changing where significant pay for performance have been rising as the basic new considers the healthcare environment. Since the benefits is not by any means the only inspiration, it is fairly essential vital for the healthcare associations to really hope to get the distinctive instruments, systems, foundation to influence the enormous information adequately or even danger of losing the potential a huge number of dollars in the income and benefits (White, 2014).

The objective of big data analytics is to help decision makers and to improve the performance of the healthcare personnel and the patients. The three aspects of the big data analytics are the medical image analysis, genomic data processing and the physiological signal processing. The overall exponential growth has been a major factor that has helped the patients with the innovative solutions to the process. The trend of the adaptation of the computational systems has been processing the research and it has also been practicing the medical professionals (L. A. Celi, 2013). Big data analytics in healthcare industry minimally should provide the primary functionalities of data processing. These functions form the general platform of data analytics which should include ease of use, quality assurance, availability, scalability, privacy, security, continuity, and the ability of dissimilar levels manipulation. The key requirement in healthcare is the real-time big data analytics, needing to address the lag between collecting and processing data (Raghupathi & Raghupathi, 2014). The improvement of healthcare data analytics indicates to the motivation of creating new innovations and technologies. Potentially, the outcomes of improvements include the analysis of big data and reduce costs in healthcare. The big data analytics is considered a valuable asset for healthcare by driving more clever treatment decisions (Berg, 2015).

Big data analytics in healthcare industry minimally should provide the primary functionalities of data processing. These functions form the general platform of data analytics which should include ease of use, quality assurance, availability, scalability, privacy, security, continuity, and the ability of dissimilar levels manipulation. The key requirement in healthcare is the real-time big data analytics, needing to address the lag between collecting and processing data (Raghupathi & Raghupathi, 2014).

2.1 Advantages to healthcare

With the combination of digitalizing, combing and also effectively using the dig data, it helps the healthcare firms which are starting from single-doctor offices to the multi-provider groups to the large hospitals. All of them have been benefitting from the big data (Figure 1). The major or the potential benefits of big data includes the detection of disease at the first or the initial stage where the disease can be cured effectively and efficiently. Along with this, it also helps in managing the individual and the population health which can overcome the fraud issues also. Along with this, the advantage of big data analytics in healthcare is that it helps in enabling more than \$300 billion in savings each year in the United States healthcare only. The clinical operations and the R&D have been the two largest areas for the potential savings with \$165 billion and \$109 billion in waste only.

According to research, the specific areas which have enhanced the use of data and analytics have been able to gauge in the highest results which have also included the accurate identification of patients who have the biggest consumer of the resources of health and who have also been at the biggest risk for the adverse results where the individuals are provided with the required information they actually need. It is also helpful in managing the health situation of the patients who can use the data and consult the doctors who are specialist in the respective fields. The big data even helps the patients with the effective and timely treatment of their health concerns (D. S. Liebeskind, 2015). It has been quite an issue where the patient data has not been saved and kept along; with the use of big data analytics this issue has been resolved to an extent. Big data analytics also help in identifying the treatment, tracking healthier behaviors, managing the health concerns and other issues simultaneously (Dutta, 2014).

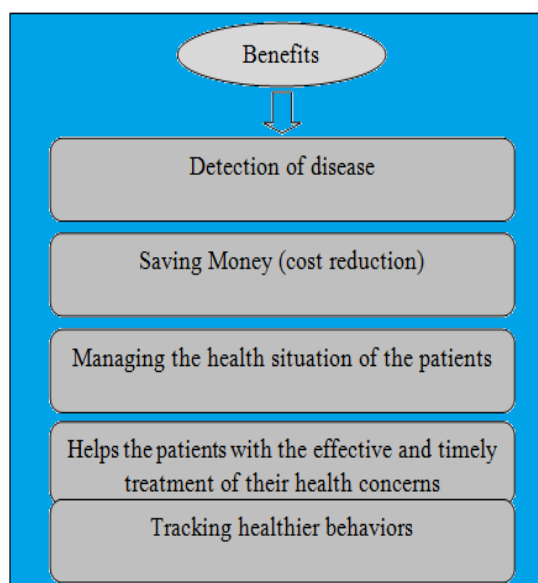


Figure 2. Big Data Analytics Advantages to Healthcare

2.2 Market Factors in Healthcare

There are some market factors that effecting healthcare. The huge rise in population, growing the problems of health and chronic illness lead to open the door for a lot of opportunities of healthcare (Figure 2). The competition between consumers for providing the best healthcare solution increasing with increase in patients and this allow competition innovate in this field to improve quality of solutions that provided to patients and also to reduce cost and time that is required for these solutions (P. Raj, 2015).

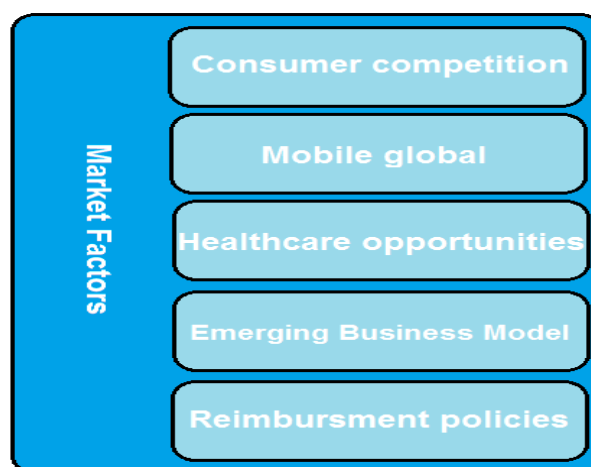


Figure 2. Market factors in healthcare

3. Big Data Analytics Opportunities for Healthcare

The opportunities that is available for the big data analytics is that is that it is used to improve the efficiency and the delivery of care. The current healthcare big data analytics is used to actually solve quite a few clinical and operational issues even if it is not properly executed. The application of data analytics in healthcare has been helping in improving health, predictive modeling and the quality measurement along the way. The first major opportunity of big data analytics in healthcare is that it helped in using the CMS and also other different healthcare payers for the prevention of fraud. Predictive modeling has been using statistical techniques and also the different historical data to help in the estimation of probability in the future. CMS contractors have been using this technique to help in determining the claims that are likely to be fraudulent prior to the overall payment of the services (White SE, 2011).

Along with the detection of CMS for the fraud in Medicare has been claiming the performing of the post payment reviews through the contractors such as the recovery audit contractors. This pre-payment auditing

has been quite similar to the review of the activities that are used by the credit card companies. The CMS model is a helpful tool in the big analytics for the healthcare. The idea of the big data analytics is that it helps in overcoming the fraud that has been happening quite regularly (White SE, 2011). Predictive modeling can also be a helpful tool in determining the patient that is most likely to benefit from a healthcare management plan. The care management plan has been used to prevent the hospitalizations of the patients that have been facing the chronic conditions such as the diabetes, pulmonary disorder or the asthma. Along with this, it often includes the contact with a healthcare professional to rather ensure that the patients does not given any other treatment that does not overcome the issue. Therefore, the one major opportunity of big data analytics is that it shall identify the high cost risk drivers and it shall also allow the early intervention and the patient management (Howe, 2007).

Another opportunity is the tracking and the identification of the patients with the type 2 diabetes is also discussed in the big data analytics. The research has suggested that by using a two-step process to identify the subsets of the patients can have similar primary diagnosis and it can also help in the statistical clustering method that can further divide the subset (Dutta, 2014). The method is generally used to administrative the datasets, which can determine the treatment of patient. Hence, it is a method where patient data is available over a period of time and across the providers. Along with this, the big data analytics also provide an opportunity to measure and also report the quality of healthcare. The health quality alliance has helped in the patient initiatives of different patients. The nigh data analytics have been underway in more than 26 states across the United States that has been committed to serve the high quality service (F. E. Dewey, 2014).

All the different commercial insurance companies have been using some level of big data analytics to overcome the fraudulent issues. Some of them have been implementing the pay for performance that helps in seeking the unnecessary services and also help in overcoming the highest quality care (Tannen RL, 2009). Along with this, the big data analytics have also been quite challenging in the processing of the physiological data to deal with, the big data analytics therefore helps in each step of the process and it provides the opportunity for the systematic improvement within the healthcare research and the different practice communities (Dutta, 2014).

Similarly, big data analytics also provides the opportunity to overcome the data wrangling, harmonizing and aggregating and discrete medical data formats which help in the understanding of the patient needs and requirements. Similarly the electronic medical records data has also been a part of the big data analytics that have been providing the opportunity to study the drug efficacy. The researchers have been quite committed to assume that big data analytics in the healthcare help in controlling the covariates and they also propose the methods of innovation to help in adjusting the data along with the patient needs and wants (Cousins MS, 2002). The potential use of the data driven discoveries in the therapeutics has helped in the improved services. With the personalized medical treatments and their protocols it might be identified through the mining of large clinical database.

4. Big Data Analytics Challenge for Healthcare

The major challenge of big data analytics in healthcare is that it sometimes makes it difficult to use the healthcare data to its fullest extent. First of all, the data in many of the healthcare provider, especially the hospitals have become soloed and segmented. The data has also been erased and has been tempered in some occasions which challenge the credibility of the data (Zhong, T., 2013). Along with this, the administrative data such as the reimbursement, the claims and the cost information is sometimes stored and also used by the financial and operational management teams for other purposes also which can affect the data. The data is also used to carry the business side of the healthcare but in general it is not used to inform the patient care or the treatment protocols.

The clinical data such as the vital signs, the patient history, results of the diagnostics, progress notes are also stored in the EHR. The clinical data is also maintained and accessed by the nurses, the physicians and also the other frontline clinical staff and it also used to track the patient care and communicate the treatment plans that are carried throughout the team of the clinical providing care to the different sets of patients (Plaisant, C., 1999). The outcomes and the quality of data such as the surgical site, rate of return to surgery, the surgical infections, patient falls and the Centers for the Medicare and the Medicaid Services (CMS) which are the value based prochain measures are also in the domain of the quality or the risk management departments.

Another major challenge that is associated with the big data analytics is that is typically used to make the retrospective measurement of the performance of the provider. The research conducted by the Clinical Informatics Survey in the year 2011 confirmed that the survey has confirmed that more than 43% respondents have complained that the data being kept in the silos throughout the organizations has become barrier to rather analyzing the clinical data (Tannen RL, 2009). The survey has also confirmed that the respondents have been professionals who actually challenge the credibility of the big data analytics sometimes. The issue that has been popped out has expanded beyond the providers throughout the health care industry (Marchal, S., et al. 2014).

The big data analytics also involves the optimal use of the hospital resources and the improving patient outcome which can be achieved only by the combination of the datasets which helps in meeting the second portion of the Gartner definition, the effective of cost, new and innovative forms of the processing of information

to enhance ideas and the decision-making (Dutta, 2014). Along with this, most of the providers in the big data analytics have been working hard enough to overcome the issues and the challenges through different methods and tools such as the decision support database and the data warehouse that actually allows the researcher and the analysts to combine the data from the traditional segmented sources. Similarly, a big data analytics platform in the healthcare is open source platforms (L. A. Celi, 2013).

Big data analytics has another challenge which is that it should be and it actually needs to be acknowledged in the healthcare data analytics where it is most of the times used as the source of secondary data. For example, administrative data that is collected primarily for the purpose of the accounting services which are rendered and the overall collection of the payment. EHR data has been primarily collected to track the performance and the progress of the patient the treatment and the payments. EHR has been collecting the tract of the progress of each patient, their cliental status and their treatments respectively. When the overall data is used to measure the quality and the outcomes, the original use of the data helps in the overcoming of the validity and the reliability of the results models. The comprehensive and the complete data and all the information of the governance shall be used in order to address the challenges within and across the providers. Along with this data governance program shall also be used to use the appropriate data sources and the data fields also.

The success of the big data analytics in the healthcare therefore becomes dependent on the data that has been retrieved and it sometimes can be tempered. Along with this, the big data analytics have the issue of ownership and the governance and the standards have to also be considered. The big data analytics in the social insurance is once in a while institutionalized which now and again is divided or produced in legacy in the legacy of the IT frameworks with the contradictory configurations (Zhong, T., 2013). There are another challenges of big data analytics for healthcare illustrated in (Figure 3).

5. Literature Review

Big data therefore is a high velocity, variable data, complex that actually requires the different techniques and the technologies to help in capturing, storing, managing and distributing the information. The current techniques that have been present in the industry have been rather patient related in the healthcare and the medical data that shall be used to further understand the outcomes which can be applied at different points. The population data and the individual data shall be helpful in informing the patient during the process of decision making process (White, 2014).



Figure 3. Another challenge of Big Data Analytics (source: Raj et al., 2015)

Huge data in the medicinal services has been a staggering thought in view of its huge volume furthermore as a result of the assorted qualities of the distinctive data sorts and the velocity at which it has been overseen. The totality of the data has been identified with the patient healthcare and the healthcare of the enormous data in the healthcare business. It also includes the clinical data from the CPOE and the different client decision support systems which includes the details such as the prescription of the doctor, the medical imaging, the pharmacy, the insurance, the laboratory and all the different administrative data which also includes the patient data in the electronic records, social media, Facebook, Twitter, blogs, status update, web pages and also the journals in the medical journals (Rubin JC, 2014).

Reports have been recommending that the healthcare system in the United States alone has come to 150 Exabyte's according to the data of 2011. At the present development rate, the huge data for the United States healthcare is good to go to achieve the historic point of zettabyte scale and soon it might even reach to the yottabyte. As indicated by the California based healthcare system, which has around 9 million people, is genuinely acknowledged to have between the 26.5 and 44 petabytes of rich data that has been expelled from the EHRs that in like manner join the photos and annotations (Manyika J, 2011).

By actually understanding the association and also the understanding patterns within the data, the concept of big data analytics have become the potential to improve care, lower the cost and also help in saving lives. The big data analytics in the healthcare has been taking advantage of the explosion in the overall data that helps in extracting the secret information that is used to better inform others and also help in decision making. When big data is analyzed and synthesized and all the patterns, the associations, the trends are revealed the different stakeholders in the healthcare delivery system can help in developing the more insightful treatments that shall result in the diagnosis of an improved and a healthier result (Raghupathi & Raghupathi, 2014).

The likely induction of the big data analytics in the healthcare has even helped in the improved outcome that has been prevalent across the different scenarios by analyzing the characteristics of patient and also the overall outcome and the cost of care that helps in the identification of the most clinical and the most cost effective treatment which helps in the assessment of the patient profile. It also helps in the rather proactive identification of all the different predictive support and event that have been helping to best identify the value and also help in checking the analytics systems for the fraud detection (Tannen RL, 2009). Later on, big data analytics in social insurance is good to go to become quickly by actualizing the utilization of the big data over the distinctive medicinal services associations in the human services industry. As the examination demonstrates, big data analytics have been helping in conquering the difficulties talked about above (Marchal, S., et al. 2014).

With the progression of time as the big data analytics turn out to be more standard the issue of ensuring protection, building up standard and administration, shield security is good to go to acquire consideration. Big data analytics and the applications in the social insurance right now is building stage and with the utilization of quick improvement in the stage and instruments can likewise quicken the developing procedure. Along with this, big data analytics has been leveraging a vast host of disparate, structured and the unstructured data sources which shall play an important role in the way healthcare shall be performed in the coming times (Rubin JC, 2014). The objective of big data analytics have been helping in the decision making and it has also improved the performance of the healthcare personnel and the patients. The three aspects of the big data analytics are the medical image analysis, genomic data processing and the physiological signal processing. The overall exponential growth has been a major factor that has helped the patients with the innovative solutions to the process. The trend of the adaptation of the computational systems has been processing the research and it has also been practicing the medical professionals (L. A. Celi, 2013).

The big data analytics in healthcare have been helping the overall situation of the medical industry. It has been helping and supporting the decision making process which helps in curing the health issues of the patients. The big data analytics has been quite helpful in meeting the definition of the big data. In spite of the different challenges that have been surrounding the full aggregation and the use of healthcare data are not quite insurmountable. To meet all these challenges shall require the cultural shift in the healthcare to provide the care for both the internal to providers and also between the providers and the other portions of the industry (Marchal, S., and et al. 2014).

The major challenge however of the big data analytics in healthcare is to determine the proper balance between the protecting the information of the patient and also maintaining the integrity and the usability of that specific data. Robust and the strong information and the data governance programs shall help in addressing quite a few of those challenges. The overall sharing of the data is between the organizations shall be helpful in addressing before the full potential of the big data in the healthcare may also be unlocked. The overall concept of the Learning Health System shall be serving the core values which shall serve as the guiding concept for the advancing of all the efforts to actually help in creating a collection of the healthcare data that shall help in realizing the different opportunities to serve the patients (L. A. Celi, 2013).

6. Conclusion

Big data analytics can possibly really change the way the medicinal services suppliers have been utilizing the recently presented modern advancements that help in picking up the knowledge from the clinical furthermore the other data vaults and it has additionally helped in settling on the educated choices.

As per the definition by Gartner “high velocity, high volume and the high variety information assets that have been demanding the innovation forms of the processing of information, the cost effective visions and the decision making. Finally, it can be said that the big data analytics in the healthcare has the ability to actually transform the way in which the providers of healthcare use the sophisticated technologies to help gain the insight from their respective clinical and other details which has helped the healthcare industry become more customized and aligned with the needs and wants of the patients. Healthcare data surely is applicable with the concept of big data. It is not impossible to aggregate all healthcare data challenges and issues, but it is difficult to meet them which require different cultural movements internally and externally. In addition, the need of addressing these challenges is increasing over the time.

References

- Berg, G. (2015). 3 ways big data is improving healthcare analytics. The potential to improve outcomes and contain costs from the analyzing big data in healthcare are well big
- Cousins MS, Shickle LM, Bander JA. An introduction to predictive modeling for disease management risk stratification. *Dis Manag.* 2002;5(3):157–167
- D. S. Liebeskind and E. Feldman, “Imaging of cerebrovascular disorders: precision medicine and the collaterome,” *Annals of the New York Academy of Sciences*, 2015.
- Dutta, R., et al.: Recommending Environmental Big Data Using Semantic Machine Learning. In: *CRC Book on Future Trend on Big Data Analytics*, pp. 463–494. CRC Press, Taylor & Francis Group (2014)
- F. E. Dewey, M. E. Grove, C. Pan, et al., “Clinical interpretation and implications of whole-genome sequencing,” *JAMA*, vol. 311, no. 10, pp. 1035–1045, 2014.
- Howe, Rufus, and Christopher Spence. *Population health management: Healthways' Pop Works*. HCT Project 2004-07-17, volume 2, chapter 5, pages 291-297.
- L. A. Celi, R. G. Mark, D. J. Stone, and R. A. Montgomery, “‘Big data’ in the intensive care unit: closing the data loop,” *American Journal of Respiratory and Critical Care Medicine*, vol. 187, no. 11, pp. 1157–1160, 2013.
- Manyika J, Chui M, Brown B, Buhin J, Dobbs R, Roxburgh C, Byers AH. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. USA: McKinsey Global Institute; 2011.
- Marchal, S., et al.: A Big Data Architecture for Large Scale Security Monitoring. In: *2014 IEEE International Congress on Big Data (Big Data Congress)*. IEEE (2014)
- Plaisant, C., et al.: Interface and data architecture for query preview in networked information systems. *ACM Transactions on Information Systems (TOIS)* 17(3), 320–341 (1999).
- Raghupathi, W. & Raghupathi, V. (2014). *Big data analytics in healthcare: promise and potential health information science and systems*. Health Information Science and Systems.
- Raj, P., Raman, A., Nagaraj, D. and Duggirala, S. (2015). *High-Performance Big-Data Analytics*. Cham: Springer International Publishing.
- Rubin JC, Friedman CP. Weaving together a healthcare improvement tapestry. Learning health system brings together health data stakeholders to share knowledge and improve health. *J AHIMA*. 2014;85(5):38–43.
- Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomized controlled trial findings. *BMJ*. 2009;338:b81
- White SE. Predictive modeling 101. How CMS’s newest fraud prevention tool works and what it means for providers. *J AHIMA*. 2011;82(9):46–47.
- White, SE. (2014). A review of big data in health care: challenges and opportunities. Health Information Management and Systems Division. The Ohio State University, Columbus, OH, USA. Volume 2014:6 Pages 13–18.
- Zhong, T., et al.: On mixing high-speed updates and in-memory queries: big-data architecture for real-time analytics. In: *2013 IEEE International Conference on Big Data*. IEEE(2013)