Towards Big Business Process Mining

Badr Omair, Ahmed Emam
Information Systems
King Saud University, KSU
Riyadh, Saudi Arabia
e-mail:balomair@gmail.com, aemam@ksu.edu.sa

***Abstract*—Nowadays, the topic of Big Data has received much attention from researchers. Because Big Data provides excellent analytics on a broad spectrum of data, data science is currently emerging as an interesting scientific discipline as a field of study comprehensively identified with the accumulation, administration, and examination of data. Business process mining is a process-centric technique focused on the mining of data and, for this reason, Big Data can be a big help for business process mining. In this paper, we will review Big Data and business process mining and present a model for mapping between Big Data characteristics and business process mining techniques. The mapping model has discovered that Big Data can help business process mining in different areas and open the door for more help.**

Keywords-business process mining; Big Data; business process; smart business processes.

I. INTRODUCTION

Business process mining techniques are very useful for any organization in any field. The power of these techniques comes from the fact that they are based on fact-based data [1]. Most of the time, it is very difficult to know all the facts of a given situation. One of the most important reasons for this is the inability of current technology to host huge volumes of data [2]. In some fields, like health care, seismology, astronomy and finance, archive data is regularly deleted to save on storage space. This data usually contains much important information and so business process techniques are unable to work well without it. Even in fields that do not generate huge data volumes, they will, in the long run, have increasingly large event logs, which will eventually create performance problems. Machines, sensors, and surveillance devices generate huge amounts of data that are also frequently deleted due to the incapability of saving such a large volume of information. We are now approaching the Internet of things era. Business process mining techniques must be able to handle the huge amount of data logs these devices create in order to be able to manage the facts and correlate them efficiently, or they will not be able to work successfully. In addition to the volume problem, currently, data is restricted to the structured data type only in process mining techniques [3], but there are also lots of facts available in semi and unstructured data types. For these reasons, some efforts are being made to solve these issues by using emerging Big Data technologies. These efforts are scattered, disorganized and do not cover comprehensive views, and so they do not take full advantage of Big Data technologies for process mining techniques. In this paper, we will present a model that maps every characteristic of Big Data to process mining techniques. This model proves that Big Data can contribute a great deal to business process mining. Comprehensive discovery, accurate prediction abilities, visibility, efficiency, and flexibility represent the main advantages of the mapping model. This mapping model will open the door for using Big Data at different levels for business process mining and will maximize its use in business process mining. Future efforts should now be aimed at investigating each help track in detail to ascertain a practical implementation, especially as Big Data technologies are beginning to mature and become more available. The rest of this paper is organized as follows: Section II describes Big Data definitions and characteristics, Big Data versus business intelligence, the Big Data life cycle, Big Data opportunities and challenges, and Big Data architecture, business process and business process mining; Section III describes Big Data and business process mining tools; Section IV reviews the previous work in using Big Data for business process mining; Section V will present the mapping model; and Section VI shows the conclusion and future work.

II. BACKGROUND

A. Big Data definitions and characteristics

Due to the boom of generated data, Big Data has become a strong concern for many organizations. Current technologies, such as Relational Database Management System (RDBMS), data warehouses, and business intelligence, do not have the capabilities to support the Big Data business goal, which is to enable organizations to create actionable business insights in a rapidly changing environment [4].

In a book by Hurwitz et al. [5], Big Data is defined as the ability to deal with an enormous volume of divergent data, at the speed, and inside the time span to permit continuous investigation and response.

In Savitz's [6] gave a more detailed definition as: "Big Data are high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization".

Watson [2] provided the most comprehensive definition of Big Data as: "Big data is a term that is used to describe data that is high volume, high velocity, and/or high variety; requires new technologies and techniques to capture, store, and analyze it; and is used to enhance decision making, provide insight and discovery, and support and optimize processes". Big data is considered as the most recent era in the development of decision support data management [7].

Thus, from these definitions of Big Data, we can see that the main characteristics of Big Data are: volume, velocity, and variety, also known as the 3Vs. The term volume alludes to the huge size of the data set, velocity demonstrates the speed of data movement, and variety refers to the different data types [8].

B. Big Data versus business intelligence

Chen et al. [9] stated that business intelligence and analytics fall into three categories: first, Database Management System (DBMS) structured content, utilizing traditional analytic tools via data warehousing, Extract, Transform, and Load (ETL), Online Analytical Processing (OLAP) and data mining; second, web-based and unstructured content, utilizing tools in information retrieval, opinion mining, question answering, web analytics, social media analytics, social network analysis, and spatial-temporal analysis; and third, mobile and sensor-based content, utilizing tools in location-awareness analysis, person-centered analysis, context-relevant analysis, and mobile visualization.

The main differences between Big Data and a data warehouse are: Big Data is stored in a distributed file system rather than on a central server; the Big Data format can be unstructured or semi-structured rather than only structured data; Big Data includes real-time data as well as offline data [10]; Big Data comes from a variety of sources, including new data sources such as web data, social media, device logs and mobile data; and finally, Big Data is mainly used for predictive analysis. Hu et al. [11] made a comparison between big data and traditional data as depicted in TABLE I.

C. Big Data life cycle

There are some applied Big Data life cycles, but Hu et al. [11] stated that the most common Big Data life cycle consists of four phases, the first of which is data generation. . Second is data acquisition, which refers to the process of obtaining information and is subdivided into data collection, data transmission, and data pre-processing. Third is data storage, which refers to the persistent storage and management of large-scale datasets. Fourth is data analysis, which leverages analytical methods or tools to inspect, transform, and model data to extract value. Figure 1 depicts these phases with exemplar technologies.

D. Big Data opportunities and challenges

Big Data has a wide variety of sources, from traditional transactional processing to processes that incorporate Internet data (e.g., clickstream and social networking), research data (e.g., reviews and industry reports), area information (e.g., cell phone information and geospatial information), pictures (e.g., observation and satellites), store network information (e.g., Electronic Data Interchange (EDI) and merchant inventories), and device data (e.g., sensors and Radio Frequency Identification (RFID) gadgets) which allow for many Big Data applications [12]. For example, by integrating a customer's profile with his habits, location and interests, which can be obtained from the Internet (e.g., Google, Twitter, Facebook, LinkedIn and other social media), we can build tailored products, conduct customer sentiment analysis and provide targeted services [13]. Integrating the huge amount of patients' historical data with medicine manufacturing data can help to provide personalized medication and gain insights into genetic and environmental causes of diseases [14]. Support companies can provide better services and improved troubleshooting, cyber security and uptime for their customers by monitoring recorded data from smart meters and machine logs [10].

In addition to retail, manufacturing, banking and finance, and healthcare, C. L. Philip Chen & Zhang [15], stated that Big Data applications also lie in many scientific disciplines, such as astronomy, atmospheric science, medicine, genomics, biology, biogeochemistry and other complex and interdisciplinary scientific researches. They also conducted a survey about Big Data benefits in the business sector, as depicted in Figure 2.

Big Data security and privacy are a big challenge. In addition, inconsistency, incompleteness, scalability and timeliness of the data are also challenges [16][17]. In a predictive analytics study in 2013 at The Data Warehousing Institute (TDWI) [18], a survey was made regarding the challenges of Big Data. It found that data integration complexity, architecting Big Data systems, lack of skills or staff, dealing with real-time data, poor data quality, and data privacy are the most important challenges.

Watson [2] stated that the clear business need, strong management support, the alignment of analytics strategies with business need, skilled people, the right analytics tools and fact-based decision making are the keys to Big Data analytics success.

E. Big Data architecture

The most famous software framework for processing Big Data is Apache Hadoop. For this reason, we will take it as an example to demonstrate Big Data architecture. The Apache Hadoop software library is a massive computing framework consisting of several modules including Hadoop Distributed File System (HDFS), Hadoop MapReduce, HBase, and Chukwa [11] as depicted in Figure 3. TABLE will briefly describe these modules. Chan [4] illustrated architecture for Big Data analytics and investigated Big Data innovations that incorporated Not only SQL (NoSQL) databases, HDFS and MapReduce. He examined running batch and real-time analytics using Hadoop. Its bidirectional association with conventional data warehouses and data mining analytics is depicted in Figure 4.

F. Business process

"A business process instance represents a concrete case in the operational business of a company, consisting of activity instances" [19]. Business processes have become progressively vital in numerous enterprises because they define the method for developing value and distributing it to customers [20]. Business processes are the key drivers behind three critical success factors—cost, quality, and time [21]. There are some well-known business processes methodologies, such as Six Sigma, Lean, BP Trends, Hammer and Rummler–Brache. The main elements of these methodologies are: (1) management and leadership, for describing how the processes will be managed, (2) process improvement, for describing the improvement process steps, (3) measurement, for describing how the processes will be measured, (4) learning, for describing training needs, (5) alignment with organizational priorities, for prioritizing process improvement projects, (6) continuous improvement, for determine how are

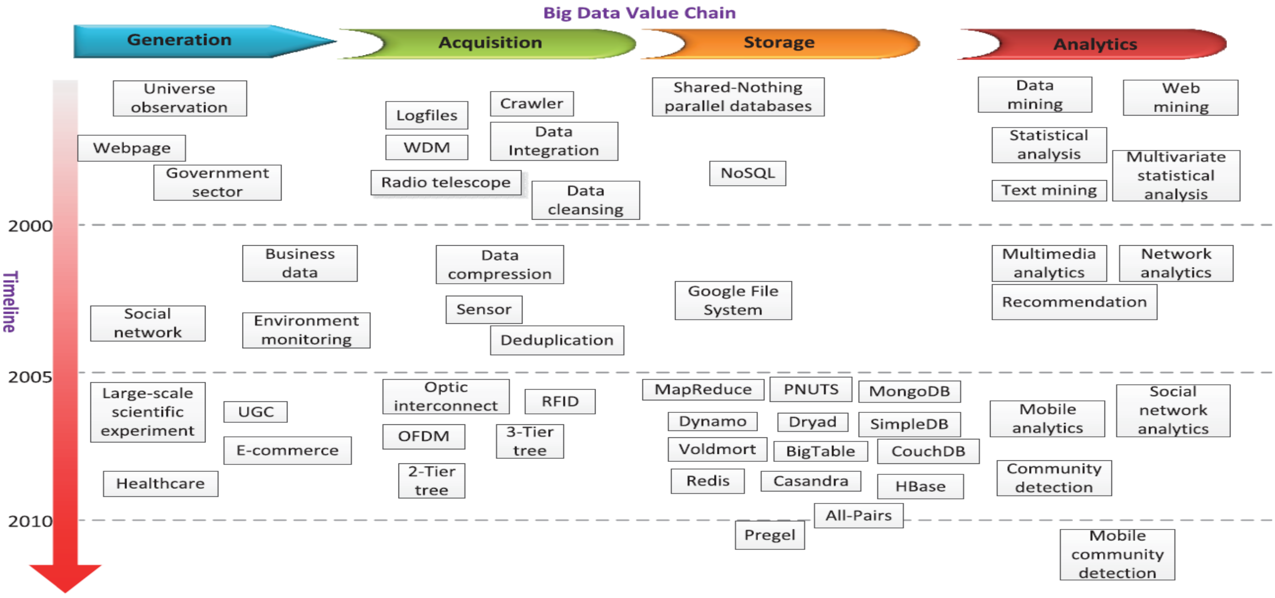

Figure 1. Big Data life cycle technology map [11].

TABLE I. BIG DATA VERSUS TRADITIONAL DATA [11].

	Traditional Data	Big Data
Volume	gigabyte	Constantly updated (terabyte or petabyte currently)
Generated rate	per hour/day	more rapid
Structure	structured	semi-structured or unstructured
Data source	centralized	fully distributed
Data integration	easy	difficult
Data store	RDBMS	HDFS, NoSQL
Access	interactive	batch or near-time

processes be monitored and by whom, (7) technology, for specifying Business Process Management (BPM) tools, (8) common practices, for specifying the organization standards, (9) change management, for describing the change management process for the business processes [22]. BPM covers how we study, distinguish, change and screen business methodologies to guarantee that they run easily and can be enhanced over the long term. It involves a constant assessment of existing processes and identification of approaches to improve them, in order to bring about global organizational enhancement [23].

G. *Business process mining*

W. M. P. van der Aalst [24] stated that there is currently a missing link between business processes and the real processes with information systems. Process mining has arisen as a new scientific discipline to provide a link between process models and event data [25]. Simeonova [26] defined process mining as techniques that help to find, screen and enhance genuine procedures by concentrating learning from event logs. Data is gathered from assorted types of systems and examined to identify deviations from standard processes and see where the bottlenecks are. Process mining is based on fact-based data and starts with an analysis of data, followed by the creation of a process model. This differs from the classic way of first modeling a control flow and then adding data to it. For example, in navigation devices, there is a link between the current reality and the models; they are not a static map, but a dynamic one that we use every day for understanding the locations of traffic jams, listening to directions, and estimating the remaining journey time. For this reason, using process mining together with traditional process documenting techniques will give more accurate results, as shown in Figure 5. Companies often use process intelligence, mining or analytics, and apply a variety of statistical and artificial intelligence techniques to measure and analyze process-related data [27]. The three types of Business Process Analysis (BPA) are validation, verification, and performance, which all require collecting and storing large volumes of process and event data [28]. In the following section, we will describe the well-known tools of Big Data and process mining.

III. TOOLS OF BIG DATA AND BUSINESS PROCESS MINING

A. *Big Data tools*

Watson [2] stated that the criteria to choose the right Big Data platform are: the applications that use the platform; the capabilities of processing the volume, velocity and variety of data; real-time or batch processing; people skills; and finally the implementation cost. As previously mentioned, Apache Hadoop is the most famous software framework for processing big data. It has the capabilities to process large amounts of data across potentially massively parallel clusters of servers (for example, Yahoo! has over 42,000 servers in its Hadoop installation) [2]. Apache Hadoop consists of client machines and clusters of loosely-coupled commodity servers. Hadoop has two main components: first, HDFS, which is the storage system and distributes data files over large server clusters; and second, MapReduce is the distributed processing framework for parallel processing of large data sets that distribute computing jobs to each server in the cluster and collects the results. There are three major categories of machines in HDFS: client machines, master nodes and slave nodes [29]. The client machines load data

 ISBN: 978-1-61208-445-9

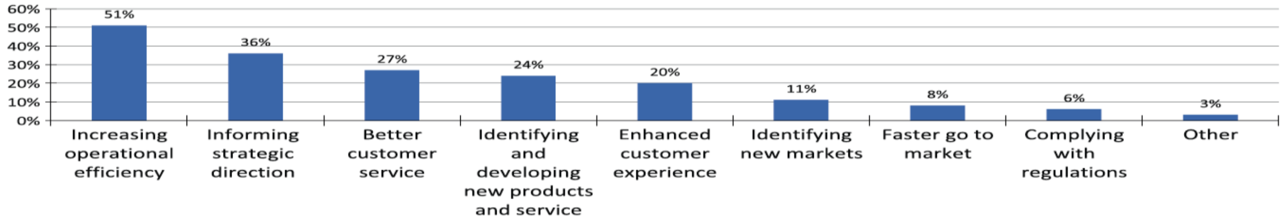

Figure 2. Operational Big Data opportunities [15].

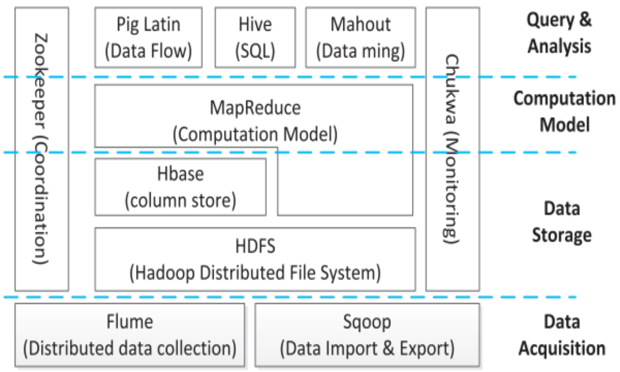

Figure 3. Hadoop architecture [11].

TABLE II. SUMMARY OF A HADOOP MODULE [11].

Function	Module	Description
Data Acquisition	Flume	Data collection from disparate sources to a centralized store
	Sqoop	Data import and export between structured stores and Hadoop
Data Storage	HDFS	Distributed file system
	HBase	Column-based data store
Computation	MapReduce	Group-aggregation computation framework
Query & Analysis	Pig Latin	SQL-like language for data flow tasks
	Hive	SQL-like language for data query
	Mahout	Data mining library
Management	Zokeeper	Service configuration, synchronization, etc.
	Chukwa	System monitoring

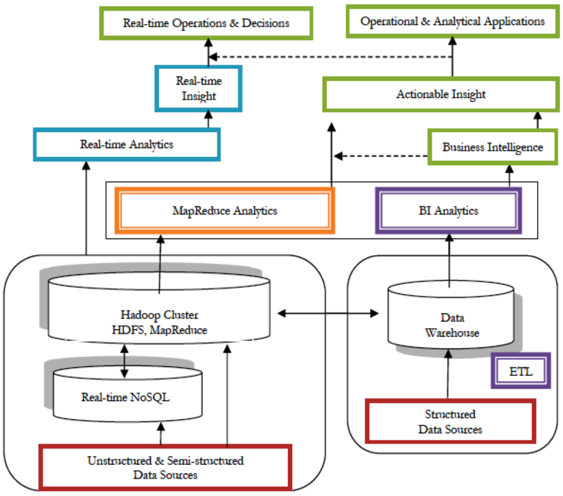

Figure 4. Big Data analytics architecture [4].

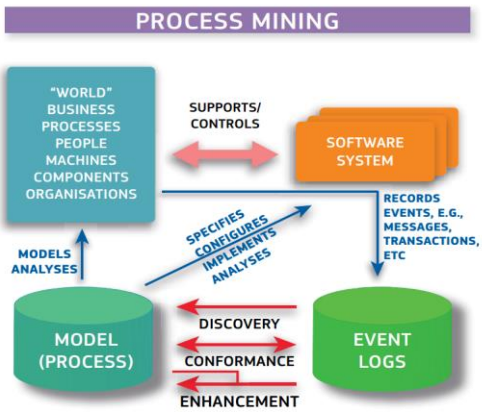

Figure 5. Process mining [25].

into servers and retrieve results. The master nodes have two types: HDFS nodes (name nodes), which are responsible for keeping the directory of all files in the HDFS file system, and MapReduce (job tracks) nodes, which are responsible for assigning MapReduce tasks to slave nodes.

B. *Business process mining tools*

W. M. P. van der Aalst [1] stated that there are three main techniques in BPM, as depicted in Figure 6. The first is the discovery technique, which takes an event log and produces a model without using a priori information. The second is the conformance technique, where an existing process model is compared to an event log of the same process. The third is the enhancement technique, which covers or improves an existing process model using information about the actual process recorded in an event log.

The most commonly used tools for process mining are ProM and Disco. ProM is a generic open-source framework for implementing process. It is very powerful but complicated. On the other hand, Disco is a commercial product and is very easy to use, but it lacks some of the process mining techniques. In the following section, we will show the most important effort made for using Big Data with business process mining.

IV. PREVIOUS WORK OF USING BIG DATA TO CONDUCT BUSINESS PROCESS MINING

Traditional Business Intelligence (BI) and Decision Support System (DSS) tools require something more than

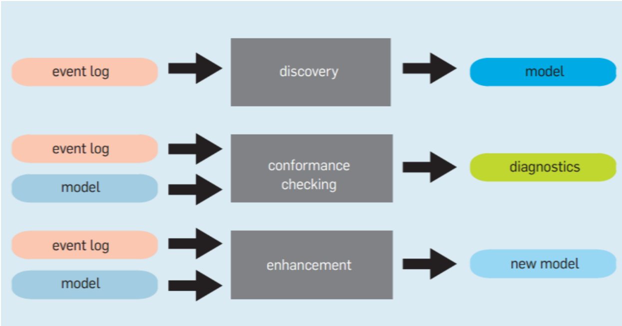

Figure 6. Types of process mining techniques in terms of input and output [1].

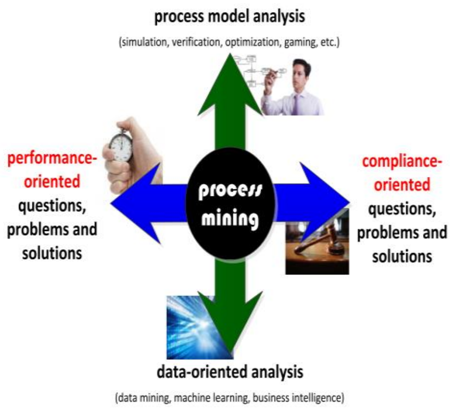

Figure 7. Process mining positions [31].

the use of mere historical data and rudimentary analysis tools to be able to predict future actions, identify trends or discover new business opportunities [30]. Real-time, low latency monitoring and analysis of business events for decision-making are required, but difficult to achieve [20]. It is a challenge to turn lots of event data ("Big Data") into valuable insights related to performance and compliance. Process mining with Big Data can do more than automate process discovery based on event data; it can also be used to answer a wide variety of performance and compliance questions in a unified and integrated manner, as shown in Figure 7. Transaction data and sensor data will enable new process mining applications to replace traditional analysis based on hand-made models [25]. Therefore, we should focus on a wide variety of different event data for mining. Process mining with Big Data will enable us to develop business processes that follow the right path in each situation.

Vera-Baquero et al. [32] introduced a business process improvement methodology for overcoming processing time and data size limitations by integrating process improvement with Big Data-based DSSs. The methodology is explained in Figure 8, and consists of five phases. The first phase is the definition phase, and it intends to not only identify and represent the business process that has a significant value to the organization, but also to have clear insight into the strategic management of the enterprise and a good understanding of the business goals being pursued. The definition phase involves four steps: (1) determine the scope and boundaries of the global business process (cross-organizational); (2) identify operational flaws within each single organization, including interactions between operational units (interdepartmental); (3) identify the level of detail that the global business process will be broken down into (level of sub-activities); and (4) develop process and activity tables. The configuration phase comes next and intends to prepare the analytical environment for receiving structural event data from the operational systems that will feed the DSS for later analysis. It includes identifying software boundaries and interdepartmental processes within business nodes, the selection of an event data format, the determination of instance correlations, and implementation of software listeners, along with a selection of metric and their threshold values. Next, the execution phase starts to capture the operational data and send business event data to the DSS. Finally, the control phase monitors and analyzes the outcomes of the DSS, and the diagnosis phase identifies deficiencies and weaknesses in the business processes identified in the definition phase. We use visualization to identify hotspots, or re-run event streams in a simulation mode in order to perform root cause analysis, among others. In the following section, we will present the mapping model, which map the Big Data characteristics to business process techniques.

V. MAPPING BIG DATA CHARACTERISTICS TO BUSINESS PROCESS TECHNIQUES

In Figure 9, we have presented the 3Vs characteristics of Big Data and showed how each of them can affect process mining techniques. The 3Vs characteristics of Big Data are volume, variety, and velocity, and they are considered the most important characteristics of Big Data as previously explained in this paper. These characteristics are mapped to process mining techniques of process discovery, process enhancement, and conformance checking; they are also explained previously in this paper. This mapping is very useful to discover how Big Data can help in process mining. We will explain this mapping in detail by describing each help track as follows:

1) Adequateness for discovery: in some application domains like the healthcare and finance sectors, it is impossible to preserve data for more than a year due to the huge volume of data generated from their systems. The same thing happens for data logs that are generated by machines and sensors. For this reason, this data is frequently deleted, even it is very important, in order to save on storage space. However, with existing Big Data technologies, we can save this data and take advantage of saving it for the long term. In process discovery techniques, we depend on event logs to discover the business process model. So, a bigger event log will ensure that we cover almost all the cases that may happen in the system. The data volume will increase the accuracy of the process discovery technique.

2) Adequateness for prediction: there are very useful probability techniques that can applied to the event logs for prediction, for example, what-if analysis, and the decision tree that can be applied to predict the path of a business process. For this reason, the more samples we have, the more we can improve the probability techniques. In addition, prediction reports will be improved with the volume of data, such as estimating the completion date.

 ISBN: 978-1-61208-445-9

3) Visibility: Event logs cannot include all the desired facts for the business process model. Big Data technologies can help us to cover all the facts by looking at extra data that was previously impossible to analyze. Big Data can help us to look for semi and unstructured data, like social media, images, emails and competitors' web sites, to feed the business process model with useful data.

4) Flexibility for efficiency: with increasing numbers of devices that are connected to the Internet, and the advent of the Internet of things era, it is very important to take advantage of these devices as much as we can. Because Big Data technologies can give us the ability analyze data in real-time by handling the huge data logs of these devices, this will help us to use these devices to increase the efficiency of the business process model. These devices can be used to extract the data from inside or outside the organization. For example, the business process model path can be changed in real-time as a quick response to use the logs of some meters like workload and inventory balance.

5) Flexibility for conformance: similar to flexibility for efficiency, flexibility for conformance can be achieved by enabling the business process model path to be changed automatically in real-time as a quick response to generated data logs from conformance devices. Temperature meters and surveillance are examples of such devices.

VI. Conclusion and Future Work

As we have seen in this paper, there are some un-coordinated efforts existing for discovering Big Data applications for business process mining, but they are not yet sufficient for covering the whole picture. For this reason, we have provided a model for mapping the 3Vs characteristics of Big Data to process mining techniques. The model will help us to bring all the applicable benefits of Big Data into business process mining. Adequateness for discovery, adequateness for prediction, visibility, flexibility for efficiency, and flexibility for conformance are the main help tracks shown in the model. In future work, these tracks should be studied in detail to ascertain practical help for business process mining using Big Data technologies.

Acknowledgment

I would like to thank KSU-CCIS Ph.D. committee for their unlimited support.

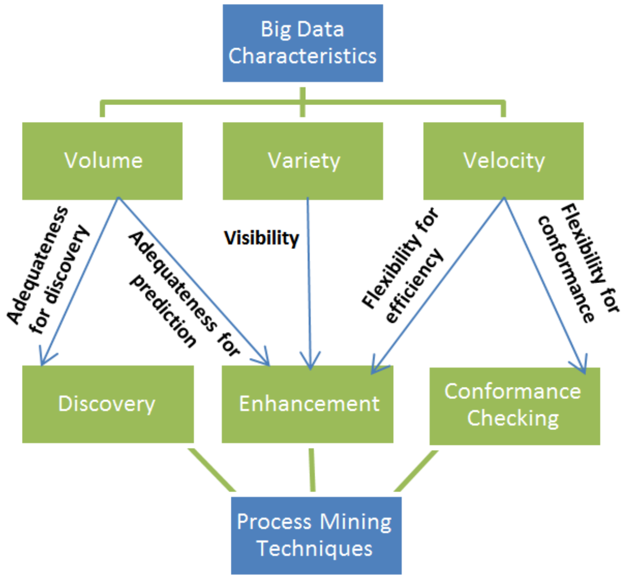

Figure 9. Mapping between Big Data characteristics and business process mining techniques.

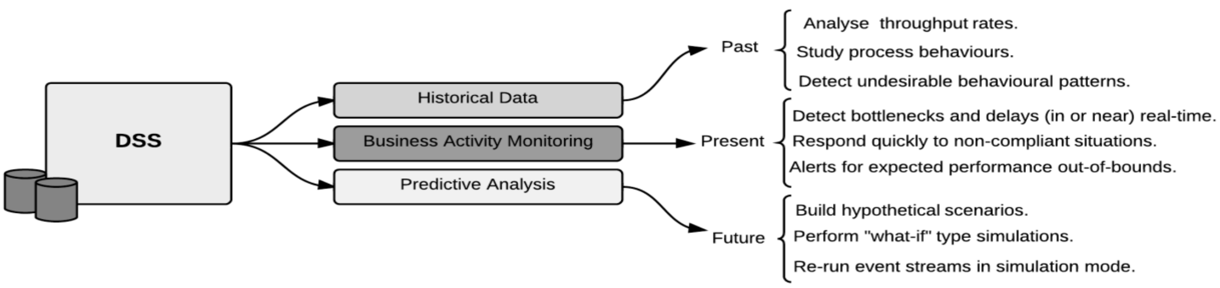

Figure 8. Business Process Analytics for Different Dimensions [32].

References

[1] W. M. P. van der Aalst, "Process Mining," *Communications of the ACM*, Springer Berlin Heidelberg, 2012.

[2] H. J. Watson, "Tutorial: Big data analytics: Concepts, technologies, and applications," *Commun. Assoc. Inf. Syst.*, vol. 34, 2014, pp. 1247–1268.

[3] A. Rozinat, "How Big Data Relates to Process Mining – And How It Doesn't — Flux Capacitor," 2014. [Online]. Available: http://fluxicon.com/blog/2011/12/how-big-data-relates-to-process-mining-and-how-it-doesnt/. [Accessed: 13-Dec-2014].

[4] J. Chan, "An architecture for Big Data analytics," *Commun. IIMA*, vol. 13, 2013, pp. 1.

[5] J. Hurwitz, A. Nugent, F. Halper, and M. Kaufman, *Big data for dummies*. For Dummies, 2013, pp. 15.

[6] E. Savitz, “Gartner: 10 Critical Tech Trends For The Next Five Years,” 2012. [Online]. Available: http://www.forbes.com/sites/ericsavitz/2012/10/22/gartner-10-critical-tech-trends-for-the-next-five-years/. [Accessed: 29-Dec-2014].

[7] H. J. Watson and O. Marjanovic, “Big Data: The Fourth Data Management Generation,” *Bus. Intell. J.*, vol. 18, no. 3, 2013, pp. 4–8.

[8] C. L. Philip Chen and C.-Y. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data,” *Inf. Sci. (Ny).*, vol. 275, 2014, pp. 314–347.

[9] H. Chen, R. H. L. Chiang, and V. C. Storey, “Business intelligence and analytics: from Big Data to big impact,” *MIS Q.*, vol. 36, 2012, pp. 1165–1188.

[10] N. Sawant and H. Shah, *Big Data Application Architecture Q&A: A Problem-Solution Approach*. Apress, 2013, pp. 2.

[11] H. Hu, Y. Wen, T. Chua, and X. Li, “Towards Scalable Systems for Big Data Analytics: A Technology Tutorial,” *IEEE Access*, 2014.

[12] M. Minelli, M. Chambers, and A. Dhiraj, *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today’s Businesses*. 2013, pp. 5–8.

[13] K. V. N. Rajesh, “Big Data Analytics: Applications and Benefits.,” *IUP J. Inf. Technol.*, vol. 9, no. 4, 2013, pp. 41–51.

[14] P. Chandarana and M. Vijayalakshmi, “Big Data analytics frameworks,” in *Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014 International Conference on*, 2014, pp. 430–434.

[15] C. L. Philip Chen and C. Y. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data,” *Inf. Sci. (Ny).*, vol. 275, Aug. 2014, pp. 314–347.

[16] D. Agrawal et al., “Challenges and Opportunities with Big Data 2011-1,” *Purdue e-Pubs*, 2011.

[17] R. T. Kouzes, G. A. Anderson, S. T. Elbert, I. Gorton, and D. K. Gracio, “The Changing Paradigm of Data-Intensive Computing.,” *IEEE Comput.*, vol. 42, no. 1, 2009, pp. 26–34.

[18] “TDWI predictive analytics study,” *TDWI*, 2013. [Online]. Available: http://www.business2community.com/big-data/drive-real-time-revenue-world-big-data-01109279. [Accessed: 31-Dec-2014].

[19] M. Weske, *Business process management: concepts, languages, architectures*. Springer Science & Business Media, 2012, pp. 7.

[20] A. Vera-Baquero, R. Colomo-Palacios, and O. Molloy, “Business process analytics using a big data approach,” *IT Prof.*, vol. 15, 2013, pp. 29–35.

[21] S. Adam, N. Riegel, J. Doerr, O. Uenalan, and D. Kerkow, “From business processes to software services and vice versa - An improved transition through service-oriented requirements engineering,” in *Journal of software: Evolution and Process*, 2012, vol. 24, pp. 237–258.

[22] S. Sweet, “Which BPM Methodology is Best for Us?,” *bpminstitute*. [Online]. Available: http://www.bpminstitute.org/resources/articles/which-bpm-methodology-best-us. [Accessed: 17-Feb-2015].

[23] “AIIM - What is Business Process Management?,” *AIIM*. [Online]. Available: http://www.aiim.org/What-is-BPM-Business-Process-Management. [Accessed: 30-Dec-2014].

[24] W. M. P. van der Aalst, “Business Process Management: A Comprehensive Survey,” *ISRN Softw. Eng.*, 2013, pp. 1–37.

[25] W. M. P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, vol. 136. 2011, pp. 80.

[26] D. Simeonova, “BIG DATA AND PROCESS MINING,” *DG DIGIT*, 2014. [Online]. Available: https://ec.europa.eu/digit-ict/sites/digit-ict/files/ictinterview.pdf. [Accessed: 13-Dec-2014].

[27] C. Janiesch, M. Matzner, and O. Müller, “Beyond process monitoring: a proof-of-concept of event-driven business activity management,” *Business Process Management Journal*, vol. 18. 2012, pp. 625–643.

[28] M. P. van der A. Wil, W. Mathias, and W. Guido, “Advanced Topics In Workflow Management: Issues, Requirements, And Solutions,” *J. Integr. Des. Process Sci.*, vol. 7, 2003, pp. 49–77.

[29] B. Hedlund, “Understanding Hadoop Clusters and the Network,” 2011. [Online]. Available: http://bradhedlund.com/2011/09/10/understanding-hadoop-clusters-and-the-network/. [Accessed: 29-Dec-2014].

[30] A. V. Baquero and O. Molloy, “A Framework to Support Business Process Analytics.,” in *KMIS*, 2012, pp. 321–332.

[31] W. van der Aalst, “‘ Mine your own business ’: using process mining to turn big data into real value,” *Proc. Eur. Conf. Inf. Syst.*, 2013, pp. 1–9.

[32] A. Vera-Baquero, R. Colomo-Palacios, and O. Molloy, “Towards a Process to Guide Big Data Based Decision Support Systems for Business Processes,” *Procedia Technol.*, vol. 16, 2014, pp. 11–21.