INSIST Vol. 4 No. 2, October 2019 (236–239)
http://insist.unila.ac.id/
DOI:10.23960/ins.v4i2.236
eSSN: 2502-8588
Received: 14/06/2019
Accepted: 19/08/2019
Published online: 20/10/2019

Empirical Study on Big Data Analysis for Supply Chain Management

Subhra Prosun Paul[1*], Shormee Saha[1**]

Abstract—with globalization, outsourcing is reaching beyond continents. Design is done in one part of the world, manufacturing in another low-cost country and distribution to other countries in the world. Procurement is illuming as a central focus that requires to be synchronized with all other business functions. As a matter of course, a sizeable amount of a firm's revenue goes for its supply chain that interprets the significance of the supply chain lays in a firm's bottom-line. So, the supply chain has a tremendous opportunity to get used of data. Nowadays, the supply chain is attracting much and more attention because in terms of analytics it is behind other functions of a firm. Specifically, this paper will (1) redefine, by research on scientific work, what BDA means in the context of Supply Chain Management, and how it differs and has evolved from analytics technologies; (2) evolve taxonomy of Big Data within SCM that identifies and classifies the different sources and types of data arising in modern supply chains and (3) suggest some applications of BDA and show the potential high value of this technology offers to solve intricate SCM challenges. This research tries to explore how the behavior of Big Data can succor procurement and SCM in greater decision making. Big data can be a lightening of a resilient environment while managing suppliers in global SCM is a challenging task. Another studied aspect is having access to a greater pool of data and what kind of potential data can render benefit SCM. SCM professionals were interviewed to understand what they expect from their logistics, procurement and marketing systems and how Big Data can contribute to that. What type of transparency is needed? What requires to be automated? What delineation of data is useful? Furthermore, how Big Data can help with SCM risk management.

Keywords—Big Data, Supply Chain Management, Business analytics, Big Data analytics, Big Data Architecture.

I. Introduction

During the last few years, due to a large amount of data available and the meteoric increase in data creation, the phenomenon has been generically called Big Data. It mainly emerged from the combination of ubiquitous internet access, an increasing number of consumer products and services, powerful computing power and large data warehouses. This pool of data is now used in both the public realm and the business world. Specifically, in the business realm, Big Data is regarded as a tool that provides companies with more information, faster than ever before.

In this study, the focus has been particularly set on potential benefits that procurement and sourcing can reap from Big Data analytics implementation but first, we need to understand what Big Data is. There are many definitions for Big Data. The Big Data term comes from a massive volume of accessible data, that has a high velocity in the generation of data and transferring that generates/comes from different kinds of sources. These sources of data range from omnipresent smart devices of consumers to manufacturing and SCM/logistics operational processes that can be captured thanks to IoT technology. It is unclear albeit Big Data is not considered a black box among firms but still in operation.

[1] Department of Computer Science & Engineering, Royal University of Dhaka, Dhaka, Bangladesh.
Corresponding author: *su_p_pl@yahoo.com,**165shormee@gmail.com

II. Background Study

A. Business Analytics and Big Data Analytics

Holsapple et al (2014) try to define business analytics in the form of a framework with some building blocks. The first building block is movement. In any entity such as an organization or supply chain, the movement is similar to a working culture that welcomes the usage of analytics. It is grounded in identifying and solving problems through evidence.

Another building block is a collection of practices and technologies. This part can exist without the previous block, the movement but it might not be effective enough. In this part lies techniques and technologies that are used for an evidence-based operation to derive knowledge for decision making. Business analytics is usually viewed by this part, techniques, and practices.

The next building block is a transformational process. At this stage, the practices and techniques turn evidence into understanding and action. Business analytics has a set of capabilities. These capabilities are the usage of statistical and techniques and working with descriptive, predictive and prescriptive analytics that provides evidence for decision making.

The final important building block is the decisional paradigm. Business analytics is considered as an approach in decision making alongside other approaches. These approaches can be in line with each other or can be conflicting. The issue that is raised here is assessing whether this decisional paradigm is appropriate for an organization considering analytics movement and the presence of required capabilities.

It should be noted that owning Big Data analytics on its own does not bring a competitive advantage to a firm. It is merely a technology that facilitates decision making and its performance is largely dependent on the quality of data and algorithms in use. In order to move in direction of gaining competitive advantage informed decision making should be accelerated.

B. Big Data Analytics in Supply Chain Management

From upstream to downstream of supply chains, massive data is generated in each operation and supplier in the chain. The provision of accurate and real-time insights helps supply chain professionals to continually identify and respond to supply chain issues. These data are in the form of structured, semi-structured and unstructured data.

INSIST Vol. 4 No. 2, October 2019 (236–239)
http://insist.unila.ac.id/
DOI:10.23960/ins.v4i2.236

Received: 14/06/2019
Accepted: 19/08/2019
Published online: 20/10/2019

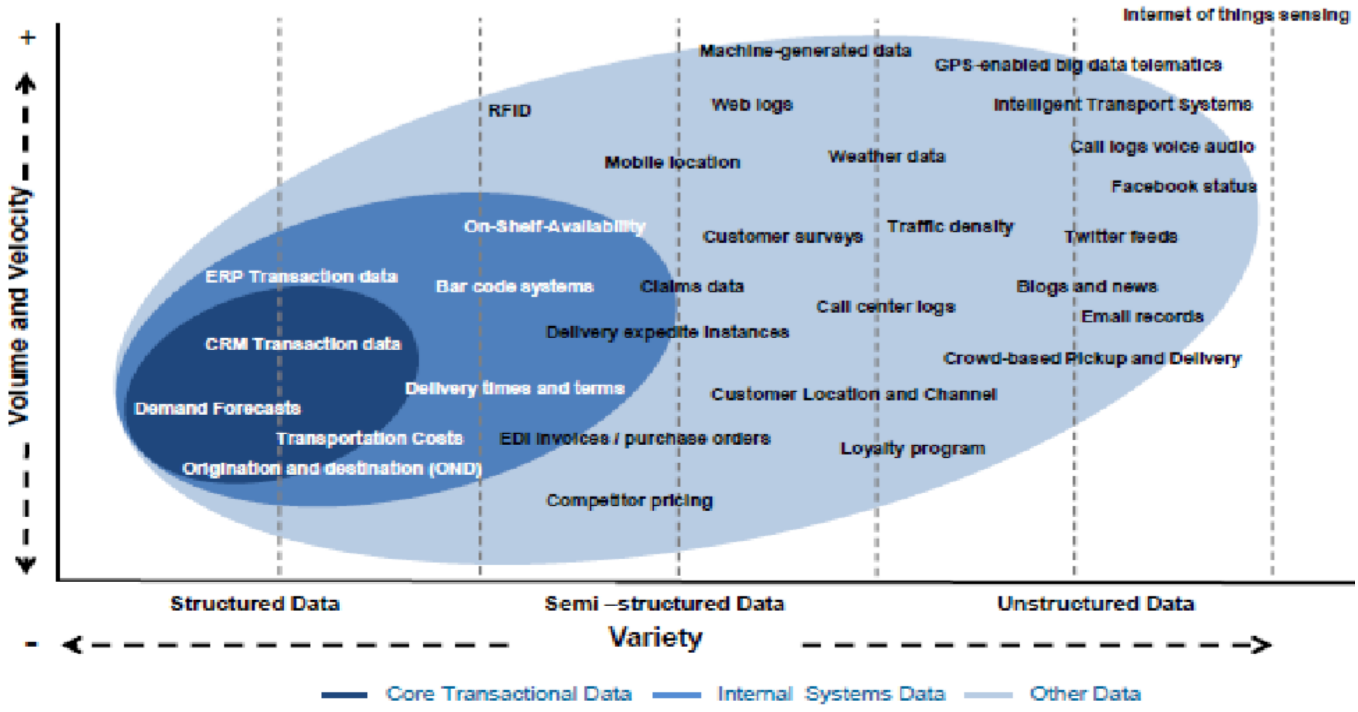

Fig.e 1: Taxonomy of data in supply chain (Rozados et al, 2014)

Figure 1 shows the taxonomy of data in the supply chain. It is evident that structured data is mostly present in internal systems and has lower volume and velocity. As we go further in horizontal axis variety increases as data becomes more unstructured. At the same time volume and velocity surge. Even though enterprise resource planning (ERP) and customer relationship management (CRM) data are core data, they only compromise a small fraction of whole data sources available to use.

Gunasekaran et al (2017) argue that big data analytics can aid the supply chain through improving visibility, resilience, and robustness. It generates cost-saving and contributes to the competitiveness of a firm. Additionally, firms can use it to beat the competition by increased transparency, effective decision making and customer segmentation provided by data capabilities.

III. A Typical Big Data Architecture

Chan (2013) proposed architecture for Big Data analytics identifying the nature, characteristics and potential applications of Big Data. In essence, the architecture is based on the client-server protocol.

On the client-side, the author proposes an architecture that consists of NoSQL databases, distributed file systems, and a distributed processing framework. A NoSQL database is essentially a non-relational, non-SQL-based database. However, it stores records in key-value pairs and works very efficiently with unrelated data which is similar to a traditional relational database. NoSQL databases are adaptable to distributed systems and they are highly scalable that makes these databases ideal for Big Data applications. In the client-side, below the NoSQL layer, Chan (2013) proposes a scalable distributed file system that is capable of handling a large volume of data, and a distributed processing framework which is responsible to distribute the computations over large server clusters.

IV. A Proposed Big Data Architecture For SCM

We propose a Big Data architecture and an analytics framework for SCM applications following the previous researches. The various input data sources in the supply chain are represented by the entities at the bottom-most layer which are suppliers, manufacturers, warehouses, distributors/ retailers and the customers. These entities provide the input data in both structured and unstructured formats. While data generated and retrieved from traditional databases such as relational databases are structured data, the input data received from various sensors, RFID tags, etc. are all unstructured in nature. The agglomeration of these humongous volumes of data results in the generation of Big Data in the system. The Big Data is then supplied to the Big Data architecture as input. While the structured data is extracted by ETL mechanisms and is populated into a data warehouse, the unstructured counterpart is taken care of by the HDFS and Map Reduce systems of the Hadoop cluster and is also stored in the NoSQL database. After the ETL operation, an operational data store (ODS) is structured for the purpose of the structured data inputs before they are gathered into the data warehouse. The Operational Data Store database is capable of integrating data from several sources so that different additional operations can be carried out on the data. In the ODS, data can be pre-processed, filtered, resolved for repetition and checked for integrity and compliance with the corporate rules. Structured input data used in the current

INSIST Vol. 4 No. 2, October 2019 (236–239)
http://insist.unila.ac.id/
DOI:10.23960/ins.v4i2.236

Received: 14/06/2019
Accepted: 19/08/2019
Published online: 20/10/2019

operation can be housed in the ODS before it is transferred to the data warehouse for persistent storage and archiving.

A real-time intelligence (RTI) system then access data in the data warehouse. RTI is an approach to data analytics that is used to achieve real-time data by directly accessing operational systems from a real-time data warehouse or feeding business transactions into and business intelligence (BI) system. The technologies that enable real-time RTI include data virtualization, data federation, enterprise information integration (EII), enterprise application integration (EAI) and service-oriented architecture (SOA).

RTI supports instant decision-making since it uses complex event processing tools to analyze data streams in real-time and either triggers automated actions or alerts the users to patterns and trends. The output of the RTI may be directly introduced into the analytics applications so that the user can visualize the results of analytics in real-time. However, the output of RTI may be fed into a dimensional data store (DDS), for non-real-time analytics. A DDS is a database that stores the data warehouse data or the output of the RTI module in a different form than the format of OLTP in traditional relational database systems. The reasons for getting the data from the source data warehouse into the DDS and then querying the DDS rather than querying the source data warehouse directly is that the arrangement of data in a DDS is in dimensional format which is more appropriate for analytics engine. In addition, when the ETL system loads the data into the DDS, various data quality checks are carried out by the data quality rules. Low-quality data is fed back into the data quality (DQ) database in order to be corrected by the source data warehouse.

With the help of a control system that is based on the sequence, rules, and logic stored in the metadata of the data warehouse, the ETL system is thus processed and orchestrated. The metadata is a database containing the summary of the data in the data warehouse and includes information such as the data structure, the data quality rules, the data meaning, the data usage, and different information about the data. The outputs of both the RTI module and the DDS module are fed into the Data Mining module that is responsible for finding patterns and relationships in the data that is of interest to the analytics engine. The output of the analytics engine is then suitably presented to the user by rich visualization techniques in the form of reports, charts, graphs, etc. In some situations, the output of the DDS module may be directly fed into the Analytics engine like: for patterns and alerts that do not require complex data mining algorithms to identify them.

V. Protocols For Security in Big Data Systems

With the advent of Big Data, a massive amount of information in a system is available to its user. However, some of the data and information may be private and sensitive in nature and it may not be desirable that all users in the system should be able to get access to that information. Establishing security and ensuring the privacy of sensitive data is a vital requirement for any Big Data system.

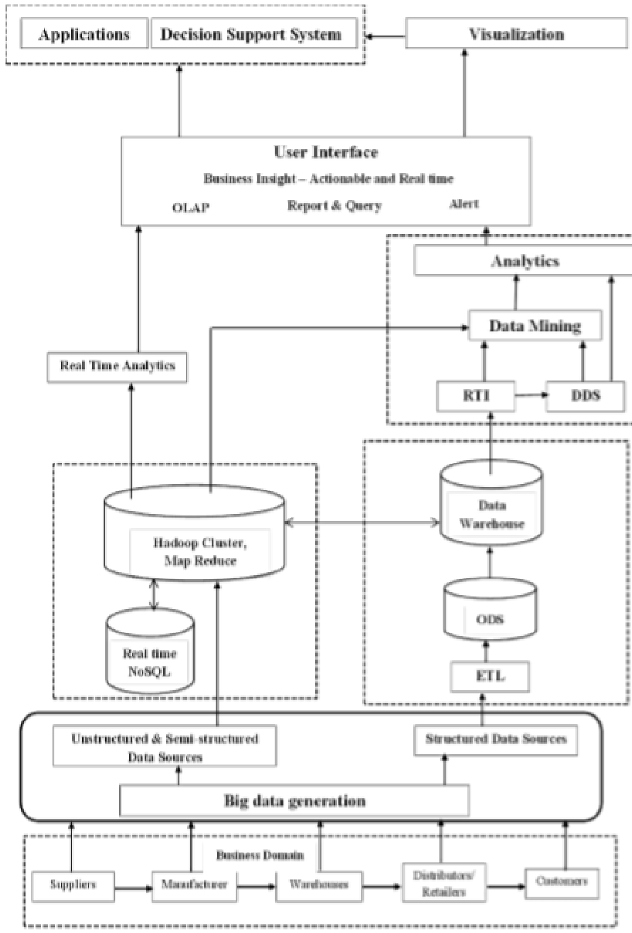

Fig. 2. A proposed architecture of Big Data analytics for SCM

In the generic SDI Architecture model for e-science (e-SDI), Layer 5 has been termed as Federated Access and delivery Infrastructure (FADI). Infrastructure components, including policy and collaborative user groups, support functionality are included in FADI. The main component of the FADI standard includes infrastructure components to support inter-cloud federations services such as Cloud Service Brokers, Trust Brokers, and Federated Identity Provider. For ensuring security in the system and privacy in data communication, each service/cloud domain contains an identity provider (IDP), authentication, authorization, and accounting (AAA) service agent. These entities typically communicate with their peers in other domains via service gateways.

The privacy of sensitive data is ensured by the use of a data-centric access control mechanism. Based on the data type and format, two basic types of access control mechanisms are deployed: (i) resource-based access control, (ii) document-based access control. For document-based access control, the eXtensible Access Control Markup Language (XACML) policy language is perfectly matched. Native access control mechanisms can be used for the purpose of resource-based access control.

In the case of defining policies for fine granular access control, XACML language is extremely efficient. Based on the request context attributes such as subject/user, data identifiers, actions, actions or lifetimes and also on the

Received: 14/06/2019
Accepted: 19/08/2019
Published online: 20/10/2019

structured data contents, fine granular access controls are defined. However, there is a disadvantage of XACML. XACML policies are used to lead in significant computational overhead for large documents or complex data structures.

Most of the commercial NoSQL databases for structured data storage have inbuilt security and access control mechanisms. A major part of them provides coarse-grain authorization features, both on user management and on protected data granularity which is similar to table-level or row-level security.

Even after having access controls in the databases, data-at-rest in remote machines may remain still unprotected. With the usage of encryption, and enhanced access control policies data in remote machines can be protected. Encryption enhanced access control mechanisms use attribute-based encryption and allows data decryption only for the targeted subject or attribute owner. These kinds of mechanisms are constructive in Big Data use cases for healthcare or targeted broadcast of streaming data in a distributed wireless sensor networks.

A dynamic infrastructure trust bootstrapping protocol (DITBP) is recommended that deploys trusted computing group reference architecture (TCGRA) and trusted platform module (TPM) for establishing trust among the computing entities. TPM infrastructure generates a key pair level in the machine hardware in order to utilize security at the hardware level. The private key is never revealed and it is used for decryption only when the machine a known and is in a trusted state. The (public, private) key pair is then used for authenticating the machine and then to decrypt the data payload.

VI. Benefits of Big Data in Supply Chain Management

Benefits of Big Data in Supply Chain Management are divided into 4 groups: visibility and accuracy of the information, operational efficiencies, higher service quality, and new business models and better prediction. Regarding visibility it enables product and service visibility, identification of problematic suppliers and problems for suppliers, giving early warnings and reduction of inventory and supply chain risk. Regarding operational efficiencies, it aids in expediting decision making, real-time vendor management and real-time view of demand and sale for expediting sourcing process. Service quality is improved through close interactions with the customer. New business models lead to the emergence of companies that are information-driven and act as intermediaries. A large scale survey of supply professionals has been conducted to understand the benefits of predictive analytics in SCM. Additionally, some more benefits are included. For example, greater power in relationship with suppliers and customers, agility in response to changing environment and better bargaining position in negotiations.

VII. Challenges of Big Data in Supply Chain Management

Regardless of the general challenges of Big Data that are data management capabilities, privacy, data ownership and deriving meaningful insights from a large volume of data, implementing it in each domain has its own unique challenges and requires domain knowledge. Usually, supply chains are not owned by one specific company, there are multiple numbers of players, each with their own fragmented system which requires standardization efforts. Many stakeholders do not understand the value that can be gained by Big Data. Even though sharing information creates value, firms can be hesitant in this regard especially if they are concerned about losing competitive advantage. Finally shared data requires a repository accessible for all partners and such repository currently do not exist.

VIII. Conclusion

This study was conducted to elicit the needs and expectations of supply and purchasing professionals from Big Data analytics. The focus of this study was placed on the procurement function and how it can benefit from Big Data. Big Data is an evolving concept that is under investigation and there is no concrete definition for it in the minds of people. Challenges regarding internal data were identified, it was not only limited to internal data of procurement but other departments that procurement relies on them as well. This study was a qualitative explorative study trying to understand the needs of supply and purchasing professionals regarding Big Data. Consequently, the results of this study touch the surface of the topic and give a holistic view of what the expectations from Big Data are in supply chain management.

References

[1] Assunção, M.D., Calheiros, R.N., Bianchi, S., Netto, M.A. and Buyya, R., (2015). "Big Data computing and clouds: Trends and future directions". Journal of Parallel and Distributed Computing, 79, pp.3-15.

[2] Baxter, P. and Jack, S., (2008). "Qualitative case study methodology: Study design and implementation for novice researchers". The qualitative report, 13(4), pp.544-559.

[3] Chen, H., Chiang, R.H. and Storey, V.C., 2012. Business intelligence and analytics: from big data to big impact. MIS quarterly, pp.1165-1188.

[4] Chen, D.Q., Preston, D.S. and Swink, M., (2015). "How the use of big data analytics affects value creation in supply chain management". Journal of Management Information Systems, 32(4), pp.4-39.

[5] Chithur, D. (2014). "Driving Strategic Sourcing Effectively with Supply Market Intelligence". Tata Consultancy Services (TCS).

[6] Chowdhary, P., Ettl, M., Dhurandhar, A., Ghosh, S., Maniachari, G., Graves, B., Schaefer, B. and Tang, Y., (2011), "Managing procurement spend using advanced compliance analytics". In e-Business Engineering (ICEBE), 2011 IEEE 8th International Conference on pp. 139-144. IEEE.

[7] Demchenko, Y., Grosso, P., De Laat, C. and Membrey, P., (2013), "Addressing big data issues in scientific data infrastructure". In Collaboration Technologies and Systems (CTS), 2013 International Conference on pp. 48-55. IEEE.

[8] Fan, Y., Heilig, L. and Voß, S., (2015). "Supply chain risk management in the era of big data". In International Conference of Design, User Experience, and Usability, pp. 283-294. Springer, Cham.

[9] Gandomi, A. and Haider, M., (2015). "Beyond the hype: Big data concepts, methods, and analytics". International Journal of Information Management, 35(2), pp.137-144.

[10] Gao, J., Koronios, A. and Selle, S., (2015). "Towards a process view on critical success factors in big data analytics projects".

[11] Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S.F., Childe, S.J., Hazen, B. and Akter, S., (2017). "Big data and predictive analytics for supply chain and organizational performance". Journal of Business Research, 70, pp.308-317.