Big Data Usage in Retail Industry

PhD candidate Anna Timofeeva
University of Economics - Varna, Varna, Bulgaria
ansergeeva@ue-varna.bg

Abstract
This article is a review of the essence and application of Big Data and Analytics in the retail industry, including e-commerce. The vast and complex data generated nowadays is in the scope of the Big Data technologies. The computer-based automation of data management and analysis enables business organizations to discover hidden models and useful knowledge which refer to the business processes. The article highlights data-driven and analysis-based approaches to commerce and identifies the leading software solutions and their capabilities. The main aim is to bring out the business benefits of using Big Data and Analytics technologies in the retail industry. In the digital era the speed and breadth of knowledge turnover within the economy increases and the advantages become more accessible for the industries.

Keywords: Big Data, Artificial Intelligence, Analytics, Retail Industry

Introduction
In the era of digitalization, more and more of the actions the Internet users take generate data that is able to be registered, processed and stored. The collection of this data as such however, is just the first step in the unveiling of its potentials. It needs to be converted and organized into a database, and then into a knowledge. This complex and comprehensive process for processing and analysis of a huge amount of information is a subject of the developing technology Big Data. This technology is a powerful source for the creation of a useful awareness of how people behave and what they prefer as users. The Big Data technology can be used in many areas of economy. This research examines into the potential Big Data has in the retail industry, aiming at disclosing the advantages from its use and analysis. Finding the user behavior pattern is essential for the commercial organizations. The quality information makes it possible to coordinate and optimize the supply of goods and services in compliance with the attitudes and needs of the users, and to increase the proceeds of sale by a more effective and target-oriented management of the commercial organization.

1. Definition of Big Data
The term Big Data refers to data sets that are too large and complex to be dealt with by the traditional data processing and data analytic methods, being essential due to the implicit information they contain. Big Data may provide clearer insight into the hidden models, tendencies and associations, especially in the context of the need to make good management decisions.

Big Data was originally determined as „high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation". (Gartner Inc.) This *attributive* definition known as the 3Vs of Big Data reveals the three dimensions of the Big Data technology. (Figure 1) Oracle goes on to add the *comparative* definition: „big data is larger, more complex data sets, especially from new data sources. These data sets are so voluminous that traditional data processing software just can't manage them. But these massive volumes of data can be used to address business problems you wouldn't have been able to tackle before." (Oracle Corp.) According to Thomas H. Davenport's general definition Big Data "is the collection and interpretation of massive data sets, made possible by vast computing power that monitors a variety of digital streams – such as sensors, marketplace interactions and social information exchanges – and analyses them using "smart" algorithms. In short, big data searches сomb massive amounts of digital information looking for useful correlations." (Davenport, 2014) The complex nature of Big Data causes many interpretations.

NIST[1] gives an *architectural* definition: „Big Data consists of extensive datasets – primarily in the characteristics of volume, variety, velocity, and/or variability – that require a scalable architecture for efficient storage, manipulation and analys". (NIST, 2019)
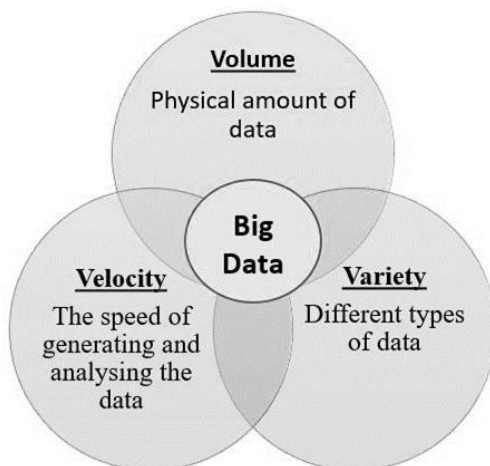

Figure 1. The 3Vs definition of Big Data.

Due to the more thorough study and exploration of the Big Data concept, its attributive definition has been improved. IBM went deeper into the Big Data definition by adding one more V– for Veracity, which refers to the accuracy and trustworthiness of data. Only part of the large amount of data is true and able to be used for solid decision making. IBM's infographic (Figure 2) gives a more detailed notion of the understanding of Big Data by certain facts connected with the 4Vs.
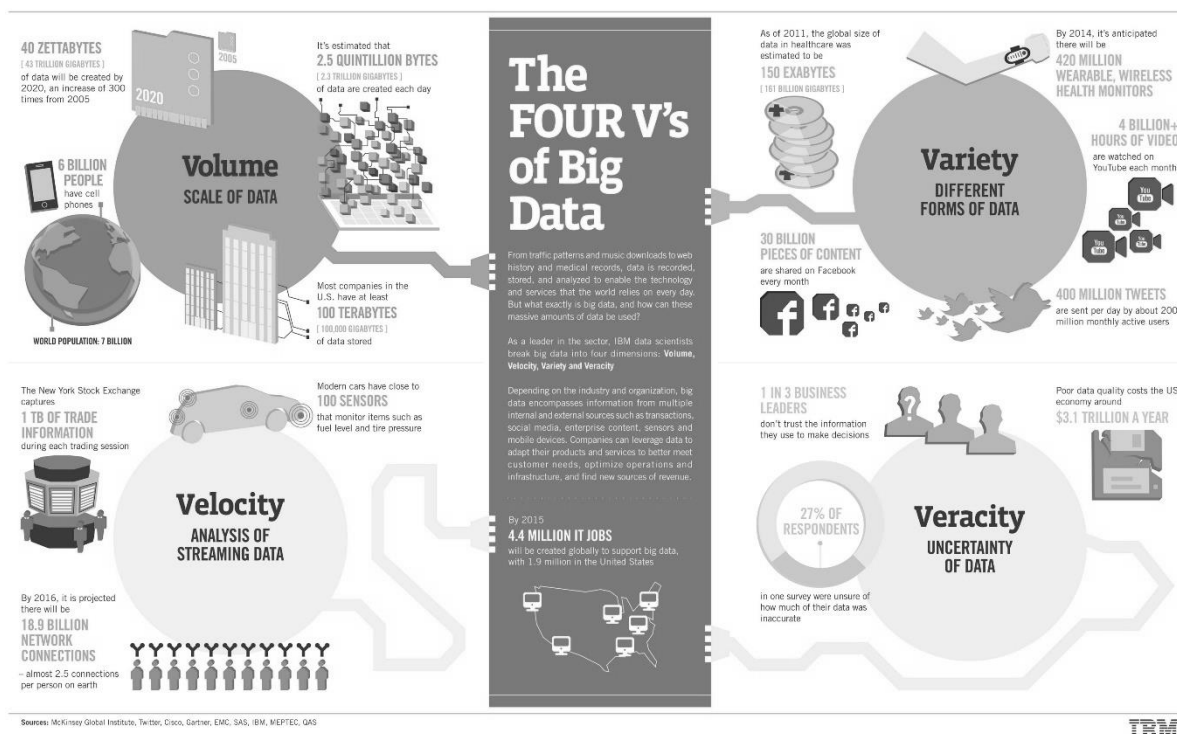

Figure 2. Infographic: IBM's 4Vs definition of Big Data. (IBM, 2019)

The evolution of information over the last years is obvious in respect of the volume, variety,

---

[1] NIST – (National Institute of Standards and Technology) is an agency of the U.S. Commerce Department

speed of generation and importance. At the same time new dimensions of the Big Data concept have been distinguished and defined to expand to 8Vs – volume, variety, velocity, veracity, variability, value, viscosity and virality. These characteristics are given a more detailed study in Table 1 below. (Rouse, 2019; Spaulding, 2019)

Table 1. 8Vs Definition of Big Data.

Attribute	Specific meaning
Volume	The amount of data from myriad sources
Variety	The types of data: structured, semi-structured, unstructured
Velocity	The speed at which data is generated
Veracity	The degree to which data can be trusted (accuracy of the data)
Variability	The way in which Big Data can be used and formatted
Value	The business value of the collected data
Viscosity	The way data matches business needs and triggers action
Virality	The capacity to aggregate a message that can be shared with a large audience

The complex nature of Big Data affords a wide range of opportunities for analysis through the use of the Artificial Intelligence (AI) mechanisms. The introduction of technologies conformable to the needs of the organization is an opportunity for mastering the business challenges. The extraction of information that is good for the business organizations contributes to the business process optimization. The insights based on the complex analysis are essential as they serve for improvement of the business practices.

2. Big Data Usage in Retail

With the development of information technologies and business process digitalization Big Data are seen to be used extensively in economy. The organizations realize that in order to be able to react adequately to the challenges in a highly competitive environment and make use of the advantages the innovations offer, they need to implement new technologies. The significant growth in the amount of data generated over the last few years mediates the adoption of the Big Data concept and the application of specialized technologies in the various economic sectors and industries. The storage, processing and analysis of information are decisive for the sustainable development of the organizations. The diagram below (Figure 3) which is based on a research study conducted by Dresner Advisory Services in 2018 (Columbus, 2018), shows the extent to which Big Data has been adopted among the industries. It's obvious that in the retail industry the adoption of the technology is mid-level, showing potential for development in this sector over the next years.

Both in the conventional retail industry and in the e-commerce the information comes from a variety of sources – suppliers, shoppers, the market (including competitors), and part of the information is generated in the commercial organization itself. After a process of transformation and integration, the information may be brought to a state suitable for analysis. By so doing, the information is sifted to pick out that part of it that is important and provides a basis for management decision making in respect of on-shelf availability, promotional effectiveness, budget and planning, new item introductions, category planning, competitive awareness, shoppers' experience, holistic business planning. (Figure 4)
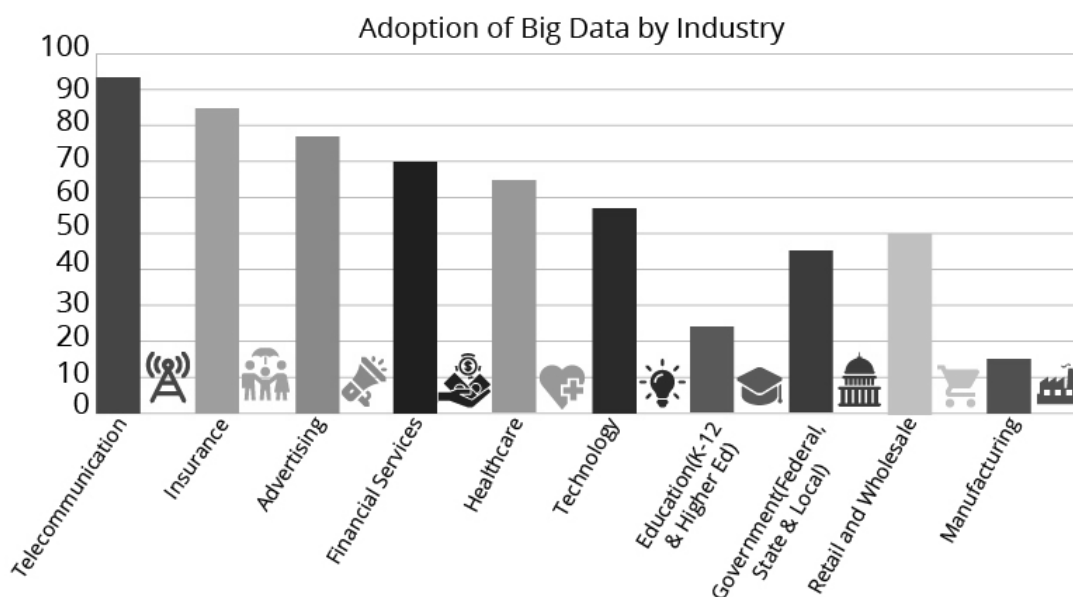
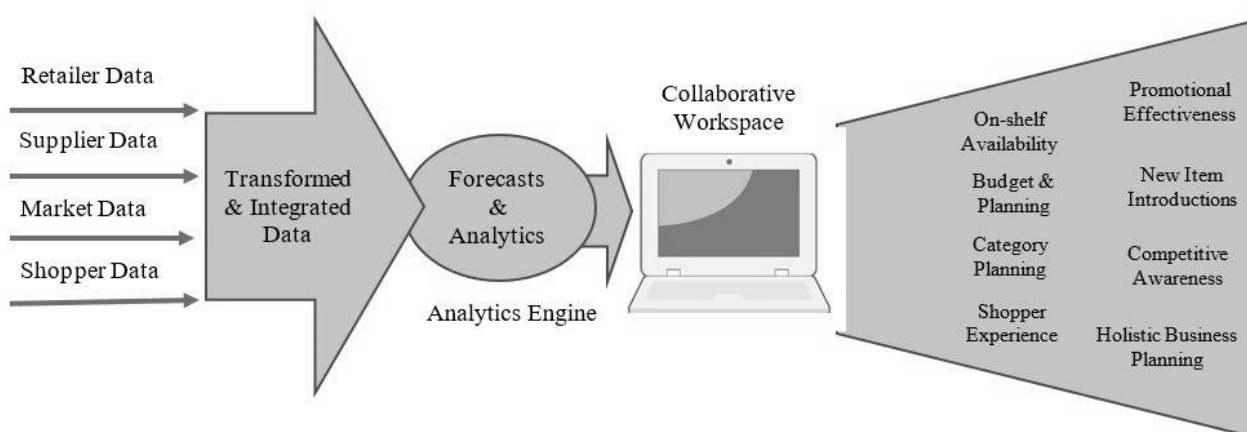Figure 3. Adoption of Big Data by Industry. (Agrawal, 2019)

Figure 4. Big Data in Retail Industry. (Ben-Zvi, 2017)

Due to the huge amount of data circulating in the retail industry it is important to pick out the useful information. In this regard, Big Data is used together with AI – to convert data to a resultant knowledge. The information passes through the following stages of AI analysis:

- descriptive analytics – helps to understand what happened (in the context of the business processes and environment)
- diagnostic analytics – tries to understand why it happened
- predictive analytics – tries to forecast what could happen
- prescriptive analytics – tries to figure out what should be done
- cognitive analytics – causes something to happen

Thanks to the analytical capabilities of AI there is formed a knowledge which can reveal a more accurate view of hidden patterns and trends and can be used to make better decisions in the future and can also provide additional insight. (Figure 5) This knowledge can be effectively used for the purpose of strategic, tactical and operational management of the retail organizations.

The technologies intended for Big Data manipulation may be classified into two types: Operational Big Data Technologies and Analytical Big Data Technologies. The first type processes the incoming data, while the second type is more complex and deal with the data that is transformed

and carries essential information. Big Data Technologies may be classified into 4 groups: Data Storage, Data Mining, Data Analytics and Data Visualization. (Kiran, 2019; SoftwareTestingHelp, 2018)
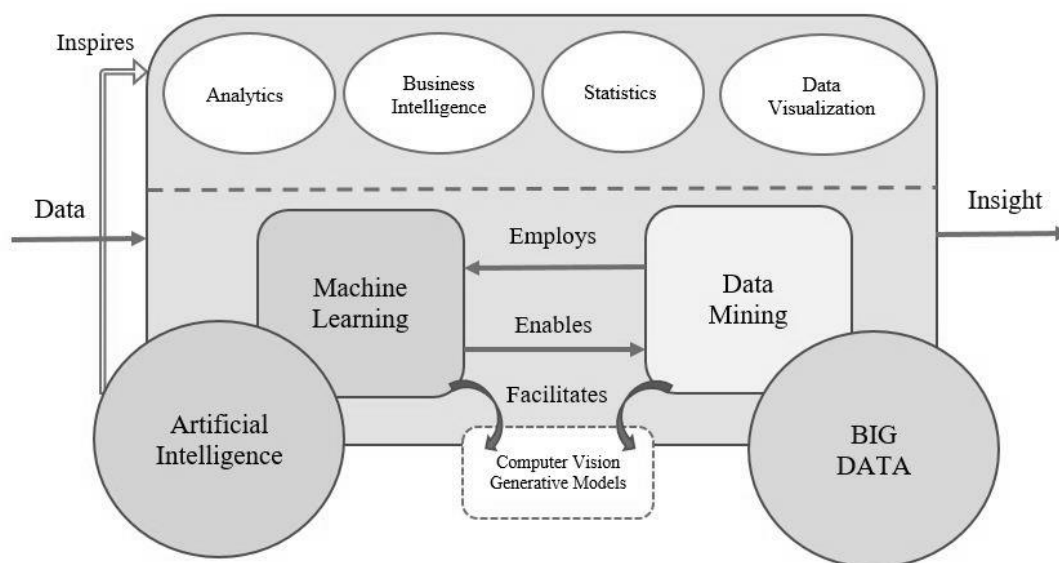

Figure 5. AI and Big Data Collaboration.

Here are the most used technologies from the first group:

- Apache Hadoop is an open source framework which is compatible with all platforms. It automatically stores, processes and analyzes data in a distributed data processing environment. It can deal with single servers or thousands of machines, each providing local computation and storage. The technology ensures fault-tolerance and high availability. Hadoop is one of the most commonly used Big Data software products and many solutions are based on it. Moreover, many vendors offer cloud computing services and maintenance that support this technology.
- MongoDB is a NoSQL document database. It is an open source and free to use software that supports different operating systems. The technology is a flexible alternative to the relational databases and it is easy to handle.
- RainStor is a database management system for structured data which analyzes and archives data. It supports SQL data access. The technology is designed with advanced data security and control features like encryption, authentication, masking, access policies and rules, data duplication.
- Hive is a cross-platform data warehouse software. It is designed for data summarization, query and analysis of structured and semi-structured data.
- Apache Cassandra is an open-source NoSQL database management system. It is designed to quickly operate with huge volumes of data stored on different servers. It provides high accessibility and interaction with the database. The technology has a log-structured storage and automated replication and it is free of charge.
  The tools which provide data mining are:
- RapidMiner is a centralized cross-platform IT solution which has a powerful graphical user interface (GUI) and is well integrated with APIs and cloud. It is designed for machine learning and predictive analytics, and it supports various scripting languages.

- Elasticsearch is an open-source, cross-platform, multitenant-capable, full-text search engine used for analysis. It handles with unstructured and semi-structured data from various sources.
- Presto is an open source distributed SQL query engine. It is designed for interactive analytic queries that can unite data from multiple sources.
  The preferred technologies for data analytics are:
- Apache Kafka is a publish-subscribe messaging IT solution. The system provides messaging to a particular category. The system is rapidly operating, flexible and scalable.
- Apache Spark is a cluster computing framework for real-time processing. It supports variety of applications written on several programming languages. The system is fast, handy and operates with multiple types of data.
- Splunk system is designed to collect, index and link real-time data in a data warehouse. It provides reports, graphs, alerts, dashboards and data visualizations. It is used for business and web analytics.
- KNIME is an open-source, cross-platform tool. It provides reports, researches, CRM, data mining, text mining, data analytics and business intelligence. It is well integrated with other technologies and languages.
- R is an open-source, free dynamic software environment which is used for statistical analytics, charts and graphics.
  Good solutions for data visualization are:
- Tableau is a system for analytics, business intelligence and data visualization. It can handle with huge data and it can provide flexible interactive real-time customized dashboards.
- Plotly is designed to build up quickly and efficiently interactive graphs. It analyzes data by using stats tools.
- Qubole data service is good for cloud-native performance. It uses structured, semi-structured (Fisher & Raman, 2018) and unstructured data for generating reports, dashboards and visualizations. The service improves workload efficiency.

In the retail industry to improve customer relations and to maintain customer loyalty is a core target. In this regard, the use of the above technologies may bring a number of benefits to the retailers and their customers. Among the most essential and beneficial effects are: (McDaniel, 2019; Mbufung, 2018; Fisher & Raman, 2018; Kolassa, 2019; Snowden, 2019)

➢ Enhanced shopping experience – predicting future consumer behavior based on the average number of products bought for a certain period of time, preferences for certain items, reactions to bonuses and subscription programs;

➢ Dynamic customer service – improving speed and quality of customer service, dealing with negative feedbacks and even making them positive, improving customer communications in general;

➢ More effective personalization – improving customer service by way of segmentation and satisfaction of the individual needs, making recommendations at the time of purchase as accurate as possible, offering the most recent products and services for each individual buyer, enhancing buyer's loyalty and satisfaction;

➢ Improved security – identifying credit card frauds in the event of online payments, product returns frauds or identity theft, reduction of losses by preventing theft of merchandise;

➢ Clearer supply chain – opportunity for order tracking while the purchased items are on their way, their status and location update, supply chain monitoring;

➢ Optimized dynamic pricing – connected with a better management of the pricing models, more adequate response to the demand and supply fluctuations, more specific competitive advantages in pricing, better differentiation of the retailer and the goods he offers, more targeted marketing campaigns and improvement of supply strategies;

➢ Increased sales, maximized profit – increasing sales and revenue, cost optimization that makes for profit maximization and prosperity of the commercial organization;

➢ Real-time analytics – the real-time dataflow and real-time analysis allow timely assessment of events and conditions and immediate decision making, supply update, enhancement of the supply chain effectiveness;

➢ New products and services development – focusing on new products and services in demand, entering new market segments;

➢ Reduced investment – opportunity for creation of strategies to optimize commercial processes, as well as increasing the timeframe for return on investment;

➢ Trend prediction, demand forecast – predicting and developing new consumer behavior management strategies, reducing the time spent on purchasing, forecasting stock needs in view of changes in seasonal patterns or in economy, maintaining the development of favorable tendencies in the development of the commercial organizations.

The number of commercial organizations able to benefit from the above advantages of their use has grown with the development of technologies. The improvement of the business opportunities and business processes in the era of digitalization is becoming increasingly possible for both the conventional trade and the e-commerce.

Conclusion

There are a variety of opportunities for introduction of the Big Data technologies in the retail industry. Notwithstanding certain challenges that may arise, the trade organizations make good use of the opportunities offered by the digital transformation. The potential for introduction and wider use of Big Data and AI in retail industry is substantial, as well as the opportunity for improvement depending on the type of business – traditional retail or e-commerce. By effective extraction of data and text and by application of various analysis methods and data visualization tools, the Big Data analytics allows for making the best business decisions. New technological advancements emerge every day, and as it takes place they convert the huge amount of data to useful real-time knowledge and insight, which is the basis for making informed decisions and undertaking adequate actions with regard to customers, suppliers and competitors.

References

1. Agrawal, V., 2019. *Big data in a nutshell.* [Online] Available at: https://jaxenter.com/big-data-nutshell-159112.html [Accessed 8 10 2019].
2. Ben-Zvi, G., 2017. *Big Data Helps Retail Revive.* [Online] Available at: https://sqream.com/dead-or-alive-and-kicking-big-data-helps-retail-revive/ [Accessed 8 10 2019.
3. Columbus, L., 2018. *Big Data Analytics Adoption Soared In The Enterprise In 2018.* [Online] Available at: https://www.forbes.com/sites/louiscolumbus/2018/12/23/big-data-analytics-adoption-soared-in-the-enterprise-in-2018/#2e98a39f332f [Accessed 8 10 2019].
4. Davenport, T. H., 2014. How strategists use "big data" to support internal business decisions, discovery and production. *Strategy & Leadership,* 42(4), pp. 45-50.
5. Fisher, M. & Raman, A., 2018. Using Data and Big Data in Retailing. *Production & Operations Management,* p. 1665–1669.
6. Gartner, Inc., n.d. *Gartner Glossary.* [Online] Available at: https://www.gartner.com/en/information-technology/glossary/big-data [Accessed 21 9 2019].
7. IBM, 2019. *The Four V's of Big Data.* [Online] Available at: https://www.ibmbigdatahub.com/infographic/four-vs-big-data
8. Kiran, R., 2019. *Top Big Data Technologies that you Need to know.* [Online] Available at: https://www.edureka.co/blog/top-big-data-technologies/

9. Kolassa, S., 2019. Forecasting the Future of Retail Forecasting. *Foresight: The International Journal of Applied Forecasting,* pp. 11-19.

10. Mbufung, C., 2018. *5 Ways Big Data And Analytics Will Impact E-Commerce In 2019.* [Online] Available at: https://medium.com/@constantinenmbufung/5-ways-big-data-and-analytics-will-impact-e-commerce-in-2019-e127d53ac13c

11. McDaniel, S., 2019. *7 Ways Big Data Will Change E-Commerce Business In 2019.* [Online] Available at: https://www.talend.com/resources/big-data-ecommerce/

12. NIST, 2019. *NIST BIG DATA INTEROPERABILITY FRAMEWORK: VOLUME 1, DEFINITION.* [Online] Available at: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1r2.pdf [Accessed 13 10 2019].

13. Oracle, C., n.d. *The Definition of Big Data.* [Online] Available at: https://www.oracle.com/big-data/guide/what-is-big-data.html [Accessed 21 9 2019].

14. Rouse, M., 2019. *What is Big Data and Why is it Important?.* [Online] Available at: https://searchdatamanagement.techtarget.com/definition/big-data [Accessed 8 10 2019].

15. Snowden, E., 2019. Retailers offer insights on how they are leveraging big data and technology to implement store-centric approaches to visual merchandising. *Retail Merchandiser,* pp. 88-89.

16. SoftwareTestingHelp, 2018. *Top 15 Big Data Tools (Big Data Analytics Tools) In 2019.* [Online] Available at: https://www.softwaretestinghelp.com/big-data-tools/

17. Spaulding, N., 2019. *Big Data: Making It "BIG" For E-Commerce Trends.* [Online] Available at: http://www.instantshift.com/2019/10/18/big-data-for-ecommerce/ [Accessed 8 10 2019].