**STATE OF ART REVIEW AND APPLICATION OF BIG DATA ANALYTICS IN THE CONTEXT OF LOGISTICS AND TRANSPORTATION**

**Abstract**

This review will briefly cover big data and how it is used in analytics. It will touch on the relevant technologies and techniques used in big data analytics, as well as the benefits of big data analytics across various sectors. The review will also present some of the challenges and barriers faced by those who use big data analytic tools and try to determine if the results of the analytics are worth the costs of overcoming these challenges. In the end, the review will make recommendations for the future.

**1 Introduction**

In today's world, big data is created through a variety of means. Social media, online transactions, clickstream data, and sensor data are just a few of the sources that generated massive amounts of structured, unstructured, and semi-structured data, and this heterogeneous data is referred to as big data(Kumar & Singh, 2018). The volume of data being generated today is constantly increasing. However, it is generally agreed that the amount of data being generated today is extremely large and continues to grow at an exponential rate. Big data is a term for data sets that are too large or complex for traditional data-processing application software to adequately deal with. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy, and data source. However, the process of analyzing large data sets to discover patterns, trends, and correlations is known as big data analytics(Osman, 2019). Big data analytics can help you make better decisions, optimize business processes, and gain a competitive advantage.

Big data is characterized primarily by these aspects of complexity known as the 4Vs (Volume, Velocity, Variety, and Veracity). Data that is enormous in size, beginning with a terabyte or more, is considered volume. Velocity is the speed at which data is collected. Variety is the wide range of data types and formats that must be stored and analyzed together, and veracity is the uncertainty of data(Cui et al., 2020).

Big data analytics can be used for tasks such as machine learning, computer vision, and web statistics. There are many potential applications for big data analytics. Some examples include: Improving healthcare outcomes by analysing large data sets of patient information(Li et al., 2021), Detecting fraud and financial crimes (Omolara et al., 2018),Improving marketing campaigns by understanding customer behaviour (Surendro, 2019), Big data analytics is a key part of the logistic and transportation sector. Big data analytics is used to manage and track the movement of goods and information(Yudhistyra et al., 2020).

**Technology Review**

The technology of Big Data Analytics has been advancing at a rapid pace in recent years. There are numerous approaches to Big Data Analytics (BDA), and the technology used varies based on the application. BDA technology, overall, includes tools for data collection, storage, processing, and analysis. Sensors, social media feeds, and transaction records are examples of data collection methods. Hadoop, NoSQL databases, and cloud computing platforms are examples of storage and processing tools(Ahmed et al., 2021). Data mining, machine learning algorithms, data visualisation tools, and statistical methods. For example, in the tyre industry, statistical inference is used to forecast sales using big data. Scalability, another big data analytics technique, is used to understand brand perceptions using Twitter social network data(Araz et al., 2020). There are a few steps that can be taken to address the big data issue. Using big data management platforms is one of the most common solutions. These platforms provide a centralised location for managing large amounts of data and can help to reduce the time and cost associated with data collection and analysis. And computational solutions like data mining, machine learning and data visualisation have become more efficient and effective. There are numerous technical solutions to big data issues. However, the fundamental question is how we can put together what we really need to choose the best solution.

**Data Mining and Machine Learning**

One of the most important aspects of big data analytics is predictive analytics, which uses statistical methods like data mining and machine learning to examine current and historical data in order to predict future trends. Data mining is the process of extracting valuable information from large data sets. It involves sorting through data to find patterns and trends. Data mining can be used to predict future trends and behaviors. Machine learning is very similar to data mining in that both processes scan for trends and patterns. Unlike data mining methods, which extract data based on human understanding, machine learning uses that data to improve the program's understanding. Machine learning recognizes data patterns and modifies program functions appropriately(Kumar & Singh, 2018).

**Distributed File System (DFS)**

The Distributed File System (DFS) is a set of services that enables file contents to be distributed across a network of nodes. Location transparency, replica management, fault tolerance, data rebuild, and error detection and correction are the primary features of the DFS. The metadata and its management are at the heart of the DFS.

**Google File System (GFS)**

The Google File System (GFS) is a specialised distributed file system created by Google purpose of providing efficient, dependable data access on large clusters of commodity hardware(Savaliya & Saxena, 2019).

**Hadoop DFS (HDFS)**

The Hadoop Distributed File System (HDFS) is a storage and processing system created to handle large amounts of distributed unstructured data. It reliably stores data and provides fast, scalable access to information with fault tolerance. It is used in conjunction with mapreduce, which allows for the processing of massive amounts of data. Hadoop's major elements are HDFS and mapreduce. Hadoop can process data from various sources, such as relational databases, text files, and streaming data(Dwivedi & Dubey, 2014).

**Architecture of HDFS**

HDFS has two elements NameNode and DataNode. NameNode serves as the master node it manages the name space of the system. When a person tries to save a large data file, the file is divided into blocks and stored in DataNodes. NameNode performs block mapping on various data nodes and stores meta data information. when the NameNode give instruction to the DataNode it  perform  creation, replication, and deletion of blocks(Dwivedi & Dubey, 2014).

**APPLICATION REVIEW**

This section of this paper focuses on the application of big data analytics in logistics and transportation. Six research papers were selected and reviewed through a rigorous process.

The ship transport sector tracks container delivery using unique identifier codes, but manual reading can be error prone. Because of the low code/container contrast and the wide range of sizes, colours, positions, inter-spaces, and alignment of these codes, container-code detection and recognition present more challenges. (Ayed et al., 2015)through their study they proposed a Hadoop big data analytics system can be used to detect and recognize container code. They have chosen a texture-based text detection approach using Haarwavelet transform and Support Vector Machine (SVM) to characterize these features into text and non-text classes. To reduce the computational time of text recognition and detection technologies for ship transportation industry applications, they used Hadoop MapReduce. Container code is managed to capture using surveillance cameras or mobile devices and saved to a Hadoop distributed system file (HDFS). The container code is then recognized by using Optical Character Recognition (OCR) to combine individual characters.

**Graphical user interface, treemap chart

Description automatically generated**

Figure Examples of container codes

(Neilson et al., 2019) did a Systematic Review of Literature on Big Data in Transportation their literature review that holds a broad view of transportation in general and cross-references with BDA techniques in transportation. The researchers are trying to find ways to make better use of big data in transportation. They reviewed many big data architectures. The first proposed big data architecture consists of six layers Data sources, Data normalization, data brokering, data storage, and data visualization then broadly categorize these layers as data collection, data pre-processing, analytics, and decision making. These layers are expected to fill and incorporate data from different sources and show the results of the analysis in graphical forms to assist users in making informed decisions. The second proposed big data architecture is called Hut. Hut is a real-time data management platform that was used in Madrid to manage and process data from traffic sensors. The framework has three layers: the speed layer, the batch layer, and the serving layer. If users require it, the serving layer supports both real-time and historical data access, either separately or concurrently.

(Yan et al., 2019) Small and medium-sized businesses will struggle to develop big data capabilities on their own due to a lack of resources to incorporate large-scale Internet of Things terminals and logistics information systems. To equip themself with big data capabilities, SMEs must use technology companies as a platform. Platform-based companies in the field of logistics and internalization have exploded in recent years, but there is still a significant gap in the industry's big data necessities. According to the findings of the study, SMEs are not adequately open or aware of how to enhance themself through these platforms. Large enterprises, on the other hand, can incorporate Internet of Things terminals and logistics information systems to collect logistics big data if they have adequate resources and technological expertise.

**UPS Case study**

Over 46,000 UPS delivery vehicles are equipped with telematics sensors, which generate massive amounts of data such as delivery truck speed, location, and total vehicle condition (Davenport & Dyché, 2013). To incorporate the data and tools, they used Hadoop components such as HDFS for storage, Map Reduce for processing, Kafka streaming, Sqoop for ingestion, Hive and Pig for querying unstructured data, and cluster monitoring equipment. big data storage and computing power and used analytics across 300 million+ data points to deliver tens of thousands of real-time route optimizations per minute Apart from the financial advantages, this resulted in a reduction of approximately 100 million delivery miles and, as a result, 100,000 metric tonnes of carbon emissions(Davenport & Dyché, 2013).

**Etihad Airways case study**

Etihad Airways flies to 89 worldwide destinations and transports 10 million passengers per year. Etihad Airways' collection of data includes data about aircraft, passengers, air quality, fuel efficiency, and passenger behavior. This information is used to manage the whole fleet of planes and make sure that passengers have a pleasant flight. Etihad Airways uses IoT to generate data that it analyses in order to improve its plane fleet. The company constantly monitors the status and location of its planes in order to anticipate problems. As a result, the company saves a lot of money(Alharthi et al., 2017).

<https://brightspace.hud.ac.uk/d2l/le/content/226185/viewContent/1787279/View>

References

Ahmed, S. T., Basha, S. M., Arumugam, S. R., & Patil, K. K. (2021). *Big Data Analytics and Cloud Computing: A Beginner's Guide*. MileStone Research Publications.

Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. *Business Horizons*, *60*(3), 285-292.

Araz, O. M., Choi, T. M., Olson, D. L., & Salman, F. S. (2020). Role of analytics for operational risk management in the era of big data. *Decision Sciences*, *51*(6), 1320-1346.

Ayed, A. B., Halima, M. B., & Alimi, A. M. (2015). Big data analytics for logistics and transportation. 2015 4th international conference on advanced logistics and transport (ICALT),

Cui, Y., Kara, S., & Chan, K. C. (2020). Manufacturing big data ecosystem: A systematic literature review. *Robotics and computer-integrated Manufacturing*, *62*, 101861.

Davenport, T. H., & Dyché, J. (2013). Big data in big companies. *International Institute for Analytics*, *3*(1-31).

Dwivedi, K., & Dubey, S. K. (2014). Analytical review on Hadoop Distributed file system. 2014 5th International Conference-Confluence The Next Generation Information Technology Summit (Confluence),

Kumar, S., & Singh, M. (2018). Big data analytics for healthcare industry: impact, applications, and tools. *Big data mining and analytics*, *2*(1), 48-57.

Li, W., Chai, Y., Khan, F., Jan, S. R. U., Verma, S., Menon, V. G., & Li, X. (2021). A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. *Mobile Networks and Applications*, *26*(1), 234-252.

Neilson, A., Daniel, B., & Tjandra, S. (2019). Systematic review of the literature on big data in the transportation domain: Concepts and applications. *Big Data Research*, *17*, 35-44.

Omolara, A. E., Jantan, A., Abiodun, O. I., Singh, M. M., Anbar, M., & Kemi, D. (2018). State-of-the-art in big data application techniques to financial crime: a survey. *International Journal of Computer Science and Network Security*, *18*(7), 6-16.

Osman, A. M. S. (2019). A novel big data analytics framework for smart cities. *Future Generation Computer Systems*, *91*, 620-633.

Savaliya, R. R., & Saxena, A. (2019). A Study of Comparatively Analysis for HDFS and Google File System towards to Handle Big Data.

Surendro, K. (2019). Predictive analytics for predicting customer behavior. 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT),

Yan, Z., Ismail, H., Chen, L., Zhao, X., & Wang, L. (2019). The application of big data analytics in optimizing logistics: a developmental perspective review. *Journal of Data, Information and Management*, *1*(1), 33-43.

Yudhistyra, W., Risal, E., Raungratanaamporn, I.-s., & Ratanavaraha, V. (2020). Exploring big data research: a review of published articles from 2010 to 2018 related to logistics and supply chains. *Operations and Supply Chain Management: An International Journal*, *13*(2), 134-149.