International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Bigdata Analytics in E-Commerce

Sarita Baburao Tembugade

CDOE, Mumbai University

1.Abstract

The exponential growth of digital transactions and consumer interaction has positioned big data analytics as a core enabler of innovation in the e-commerce industry. As of 2025 e- commerce platforms are increasingly adopting real time analytics, machine learning and hybrid data processing architectures to improve decision making, personalization and operational efficiency. This paper explores the current landscape of big data analytics in e- commerce, examining modern tools, emerging trends and challenges. This study explores how big data analytics is transforming e-commerce. It highlights how modern technologies like real time processing, machine learning, and data Lakehouse architectures enable businesses to understand consumer behaviour, optimize operations and deliver personalized experiences. The paper also introduces an original framework designed for dynamic data processing and privacy conscious analysis in online retail environments.

Keywords: Big data analytics, E-commerce, Data Analytics, Machine Learning, Customer Personalization, Supply chain Optimization.

Introduction: -

The e-commerce sector has experienced remarkable growth over the past decade, driven by increased internet access, mobile commerce and digital marketing innovations. With this expansion comes a surge in data generation – from user clicks and transactions to reviews, product and social media mentions. In 2025, leveraging this data effectively is no longer optional; it is essential for survival and competitive differentiation. Big data analytics enables e-commerce platforms to convert vast, complex datasets into actionable insights. These insights power real time recommendation systems, predictive inventory management, fraud detection and customer sentiment analysis. The shift from traditional business intelligence to intelligent, real time data processing architectures has been accelerated by developments in distributed computing, machine learning and cloud native platforms.

This paper aims to provide a current, comprehensive view of how big data analytics is transforming e-commerce. We explore key technologies such as Apache Spark, Kafka, Delta Lake and vector databases and highlight their applications in real world business scenarios.

Furthermore, we propose a novel model for customer behaviour analytics using multi source integration and discuss the implications of data privacy regulations on big data strategy.

Literature survey

Big Data Analytics: Impacting Business in Big way:

In today's business environment, data is generated abundantly through clickstreams on authentic websites, social media interactions, mobile device location information and machine-generated abundantly data. The central components driving this transformation include Intelligent Connected Machines equipped with Internet Connectivity and advanced sensors for data capture, automation controls and software application. Data hasemerged as a game-changer in the e-commerce landscape. Studies emphasize its ability to enhance decision-making, personalize experiences and improver business operation.

Big data analytics and its application in E-commerce:

It demonstrated how the internet and technological progress have produced huge quantities of data that can give firms insightful information. Both small and large e-commerce companies are utilizing big data analytics to gain a competitive edge by deeply understanding their markets, products and client.

1. Personalized Customer Experience

E-commerce businesses use big data to understand customer behaviour, preferences, and purchase history. This allows them to tailor product recommendations, offers, and content to individual customers, leading to increased engagement and conversions. For example, Amazon uses big data recommend products based on

browsing history and past purchases.

2. **Dynamic pricing**

Real-time price adjustments based on demand, competitor pricing, inventory, and customer profile. For Example, Airlines and online retailers like Flipkart and Alibaba use this.

3. **Customer Sentiment Analysis** Analyses reviews, feedback and social media to understand public perception. Help in product improvement and brand management.

4. **Inventory and Supply chain Management**

Predicts demand patterns to optimize stock levels. Improves delivery logistics by analysing traffic, weather and route data.

5. **fraud Detection and Prevention**

Monitors transactions for unusual patterns or suspicious activity. Uses predictive modelling to flag fraudulent behaviors in real-time.

6. **Marketing Optimization**

Tracks campaign performance and customer segmentation. Enables real-time bidding for ads and targeted content delivery

Identification of problems in Bigdata Application E-commerce:

1. **Data Collection Challenges**

• **Volume and Variety**: E-commerce paltforms collect data from various sources -websites, mobile apps, social media, customer support chats etc. The massive volume and variety of data make it difficult to integrate and manage.

• **Inconsistent Formats:** Data may come in structured, semi-structured or unstructured formats (e.g. logs, images, texts), which makes standardization a challenge.

2. **Data Storage and Management**

• Scalability Issues: As customer bases grow, so does the data. Many e-commerce businesses struggle to scale their data infrastructure efficiently.

• **Cost:** High storage costs for maintaining massive volumes of raw and processed data, especially for startups and medium-sizes enterprises.

3. **Data Security and Privacy**

• **Sensitive Information Exposure:** E- commerce platforms handle sensitive customer data (credit card info, addresses, login, credentials), which are prime targets for cyberattacks.

4. **Integration with Legacy Systems**

• Many older e-commerce systems are not designed to handle modern big data technologies. Integration with Hadoop, Spark or cloud-based analytics tools is difficult and expensive.

Analytics:

The data analytics state is the stage at which the value of the big e-commerce data platform is determined. The primary objectives of this stage ate to gather pertinent data insights, support decision making and effectively advance the e- commerce industry. At this stage, it is necessary to perform statistical, data mining, knowledge discovery, and visual analytics applications. In our platform, we consider novice users, which are users without data science skills but have deep knowledge of the e-commerce domain. To make this stage in our platform accessible to novice users in the e-commerce sector and allow them to perform analytically tools and the cluster of machines is established through a master node(s). The connections were established via Open Database Connectivity (ODBC) drivers, which are open standard Application Programming interfaces (APIs) for accessing data sources [24], which permit access to the data-querying components Spark, and further, the data lake can be reached. The following analytical tools were used in this implementation of e-commerce applications:

- MATLAB [5] used tall arrays to work with data backed by a distributed data store.

- Python and PySpark [6] for analytics libraries that include data analysis, natural language processing, and image processing packages.

- Tableau [7] for visual analytics and dashboards. It should be connected to other analytical tools for a variety of applications

Data Processing:

Our platform organizes the stored data into “data lake zones”, as described above. In the proposed platform, the processing is accomplished on a cluster of machines that follows a worker paradigm. In this paradigm, applications interact with the leader node to deploy a data processing job. The job is further

decomposed into tasks that can be executed concurrently by the worker nodes. The proposed platform is designed to process diverse types of data. These data can be in their original format (e.g., log records and social network feeds.). Alternatively, the data can be stored as traditional relational data structured as rows and columns with fixed schemas.
Moreover, e-commerce requires analytical tasks to be performed in batches and real-time (stream).for this ,the platform uses a distributed system architecture namely, the lambda architecture, which is data deployment model for data processing that consists of batch-based data pipeline and a fast streaming based data pipeline for handling real time data [20,21], to process real time streams of e-commerce data as well as baches of data on a cluster of commodity machines in distributed and parallel fashion.

Data Analytics Tasks:

The core objective of performing data analytics tasks is to gain insights and support the decision-making process that effectively endorses e-commerce optimization operations. This includes:

• Find interesting e-commerce data correlations within the data lake.

- Predicting events before they occur using big data.

• Analysed data to understand customers and optimize products and processes to meet customer expectations.

• Study the effects of social media campaigns on e-commerce, including product sales.

- Provision of personalized service and customized products.

• Rral time data analytics provides personalized services with unique content and makes relevant promotional offering.

- Dynamic pricing systems with monitoring competing prices and alerts.

- Identify credit card fraud, product returns, and identify theft.

• Performing fraud detection in real time by analysing data from multiple sources, such as transactional data, customer purchase history, social media feeds and location data.

- Analysing customer churn and making decisions to avoid it.

Furthermore, the presented big e-commerce data platform includes a feedback cycle to perceive the outcomes of diction's made about e-commerce firms. This is potentially beneficial in the monitoring of sales of items after promotions or the monitoring of a cluster of customers after personalized promotions are offered.

Generally, applications that include tasks such as data mining, knowledge discovery and visual analytics can be performed at this stage to optimize the e-commerce firm and the outcomes are monitored for corrective action.

data pipeline for handling real time data [20,21], to process real time streams of e-commerce data as well as baches of data on a cluster of commodity machines in distributed and parallel fashion.

Findings:

1. Enhanced Customer Personalization
2. Improved Operational Efficiency
3. Real-Time Fraud Detection
4. Customer Segmentation and Targeting
5. Challenges in Data Integration
6. Concerns Around Data Privacy
7. Emergence of Predictive and AI-Based Analytics

Conclusion:

Big data analytics (BDA)has emerged as the new frontier of innovation and competition in the wide spectrum of the e-commerce landscape due to the challenges and opportunities created by the information revolution. Big data analytics (BDA) increasingly provides value to e- commerce firms by the dynamics of people, processes, and technologies to transform data into insights for robust decision making and solutions to business problems. This is a holistic process which deals with data, sources, skills and systems in order to create a competitive **advantage. Leading e-commerce firms such as** Google, Amazon, ASOS, Netflix and Facebook have already embraced BDA and experienced enormous growth.

The study presents and approach for encapsulating all the best practices that build and shape BDA capabilities. In addition, the study of big data is well understood; and challenges are properly addressed, the BDA application will maximize business value through facilitating the pervasive usage and speedy delivery of insights across organizations.

References:

1. Arya, K., Kumar, T.; Jain M.K. Big data analytics on global e-commerce organisations: A study, survey and analysis.Int.J.Sci.Eng.res.2016,7,82-84[Google Scholar]

2. Vinodhini, M.; Appukuttan, M. A survey on big data analytics in e-commerce. *IJEST ISSN* **2016**, *1*, 61–64. [Google Scholar]

3. Weiqing, Z.; Wang, M.C.; Nakamoto, I.; Jiang, M. Big Data Analytics in E-commerce for the U.S. and China Through Literature Reviewing. *J. Syst. Sci. Inf.* **2021**, *9*, 16–44. [Google Scholar] [CrossRef]Cloudera ODBC Driver for Impala. Cloudera,
Palo Alto, CA, USA. 2016. Available online: http://www.cloudera.com/documentation/other/connectors/impala- odbc/latest/Cloudera-ODBC-Driver-for- Impala-Install-Guide.pdf (accessed on 1 September 2023.).5.MATLAB. Tall Arrays.
Available online: https://www.mathworks.com/help/matlab/import-export/tall-arrays.html (accessed on 1 September 2023)6.PySpark Available online: https://spark.apache.org/docs/latest/api/python/ (accessed on 1 September 2023).

7. Tableau Software. Tableau. Available online: https://www.tableau.com (accessed on 1 September 2023)

8. Cheng, Y.; Liu, F.; Jing, S.; Xu, W.; Chau, D. Building Big Data Processing and Visualization Pipeline through Apache Zeppelin. In Proceedings of the Practice and Experience on Advanced Research Computing, New Orleans, LO, USA, 9–13 July 2017; pp. 1–7. [Google Scholar] [CrossRef]

9. Akter, S., & Wamba, S. F. *"Big data analytics in E-commerce: a systematic review and agenda for future research."* *Electronic Markets*, 26 (2016) — comprehensive review of data types (transactional, clickstreams, voice, video) and business value in e-commerce link.springer.comresearchgate.net.

10. Edosio, U. Z. *"Big Data Analytics and its Application in E-Commerce."* E-Commerce Technologies Conf., University of Bradford, 2014 – explores paradigm shift, analytical technologies, and case studies (Amazon, Walmart, Adidas)

11. Sharma, N., Sawai, D., & Surve, G. *"Big data analytics: Impacting business in big way."* ICDMAI 2017, Pune – highlights enterprise- level impact worldscientific.com.

12. Wamba, S. F., et al. *"How 'Big Data' can make big impact: Findings from a systematic review and longitudinal case study."* *International Journal of Production Economics*, 165 (2015)