International Journal of Science, Strategic Management and Technology
Volume **02** Issue **06** June-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

A Survey on E-commerce Customer Behavior Analytics: Challenges, Insights and Tools

Ashok Kumar Sahoo
Department of Master of Computer Applications
GIFT Autonomous, Bhubaneswar, Odisha, India, ashok2024@gift.edu.in

Smruti Ranjan Swain
Head of Department Master of Computer Applications
GIFT Autonomous, Bhubaneswar, Odisha, India, hodmca@gift.edu.in

https://doi.org/10.55041/ijsmt.v2i6.060

Cite this Article: Sahoo, A. K. (2026). A Survey on E-commerce Customer Behavior Analytics: Challenges, Insights and Tools. International Journal of Science, Strategic Management and Technology, *02*(6). https://doi.org/10.55041/ijsmt.v2i6.060

***Abstract*—Large volumes of customer data are generated every day from modern e-commerce platforms and digital technologies such as online shopping systems, cloud computing, and business intelligence applications. Analysis of these massive datasets requires advanced analytical techniques and scalable platforms to extract meaningful insights for business decision making. Therefore, customer behavior analytics in e-commerce has become an important area of research and development. The basic objective of this paper is to explore the impact of customer behavior analytics, challenges in e-commerce data processing, open research issues, and various analytical tools associated with it. This paper also discusses the role of Databricks in handling large-scale customer datasets and generating interactive business insights through visualization techniques. As a result, this study provides a platform to understand customer purchasing patterns, product preferences, and sales trends at different stages of e-commerce analytics. Additionally, it opens new opportunities for researchers and organizations to develop intelligent and data-driven solutions for improving customer experience, operational efficiency, and business profitability.**

***Keywords*—E-commerce Analytics; Customer Behavior; Databricks; Big Data; Business Intelligence; Data Visualization; Customer Insights; Sales Analytics**

I. INTRODUCTION

The rapid growth of digital commerce platforms has significantly changed the way customers interact with businesses and purchase products online. Modern e-commerce systems continuously generate large volumes of customer-related data through online transactions, browsing activities, product searches, payment systems, customer reviews, and digital interactions. This continuous generation of data has increased the importance of customer behavior analytics in modern business environments. Organizations are increasingly focusing on understanding customer preferences, purchasing patterns, and customer engagement in order to improve user experience and enhance business profitability. As online shopping platforms continue to expand globally, customer analytics has become an essential component of strategic decision making in the e-commerce industry [1], [5].

E-commerce customer analytics involves collecting, processing, and analyzing customer datasets to extract meaningful business insights. These datasets include customer profiles, browsing histories, transaction records, product preferences, and purchasing activities generated from multiple digital platforms. Managing such large-scale customer data requires efficient analytical systems capable of handling high-volume and real-time data processing.

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **06** June-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Important characteristics of customer analytics data include volume, velocity, variety, and veracity, which directly influence the efficiency and reliability of business intelligence systems. Proper data management and processing techniques are essential for generating accurate analytical results and improving organizational performance [1], [7].

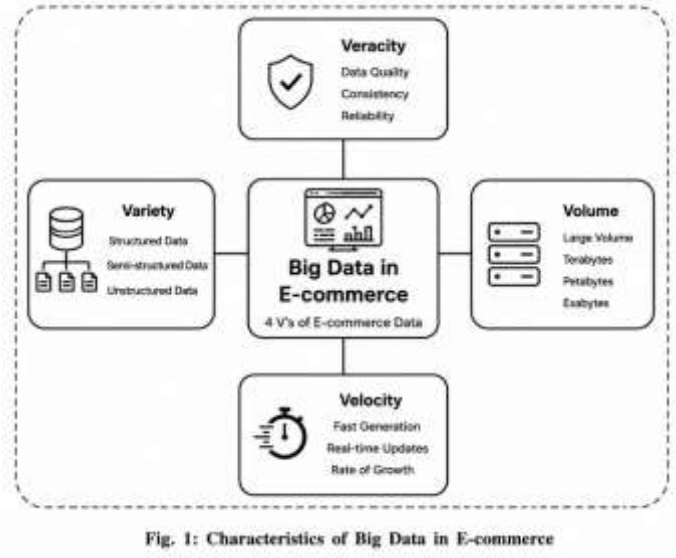

Fig. 1: Characteristics of Big Data in E-commerce

Modern organizations use scalable analytics platforms and cloud-based technologies to process customer data and generate actionable insights. In this project, Databricks is used as the primary platform for customer data processing, transformation, and dashboard visualization. Databricks provides distributed computing capabilities, scalable data processing, and integration with Apache Spark for handling large-scale customer datasets efficiently [2], [3]. The platform also supports data engineering workflows such as ETL processing, medallion architecture, and real-time analytical operations. Using these capabilities, the proposed dashboard performs customer segmentation, sales analysis, trend identification, and visualization of important business metrics including customer acquisition trends, product popularity, and customer retention rates [3], [14].

Interactive dashboards and visualization techniques help business analysts understand customer behavior patterns and identify important business trends effectively. The dashboard developed in this project provides graphical insights into purchasing behavior, sales performance, customer engagement, and regional sales distribution. These insights help organizations improve marketing strategies, optimize product recommendations, and enhance overall customer satisfaction. However, processing continuously growing customer datasets also introduces challenges related to scalability, data quality, security, and real-time processing [4], [8]. Therefore, efficient analytical platforms and intelligent data processing techniques are required to support effective business intelligence and data-driven decision making in modern e-commerce systems [3], [15].

II. CHALLENGES IN E-COMMERCE CUSTOMER ANALYTICS

In recent years, the e-commerce industry has experienced rapid growth in online shopping platforms, digital payment systems, mobile commerce applications, and customer interaction channels. Modern e-commerce platforms continuously generate large volumes of customer-related data from browsing activities, purchase transactions, product searches, customer reviews, and online engagement records. The continuous growth of customer information creates new opportunities for business intelligence, customer behavior analysis, and data-driven decision making.

However, processing and analyzing large-scale customer datasets also introduces several technical and operational challenges. E-commerce customer analytics requires scalable storage systems, efficient processing frameworks, interactive dashboard visualization, and secure data management solutions for handling large volumes of customer information effectively. The proposed E-commerce Customer Behavior Dashboard addresses these challenges using the Databricks platform and Medallion Architecture for structured customer data processing and analytics [1], [3].

A. Data Storage and Processing

One of the major challenges in e-commerce customer analytics is handling the continuously growing volume of customer and transaction data generated from multiple online platforms. Traditional database systems often face limitations in storing and processing large-scale customer datasets efficiently. In addition, e-commerce data exists in different formats such as structured transaction records, semi-structured activity logs, and unstructured customer feedback, making data integration and analytical processing more complex. Another important issue is maintaining data quality, consistency, and reliability before performing analytical operations.

To overcome these challenges, the proposed system utilizes Bronze, Silver, and Gold layers within the Databricks Medallion Architecture for scalable data ingestion, cleaning,

transformation, and analytical processing. The Bronze layer stores raw customer and transaction data collected from e-commerce sources, while the Silver layer is responsible for data cleaning, preprocessing, and standardization. The Gold layer contains transformed analytical datasets used for dashboard reporting and business intelligence operations. This architecture improves processing efficiency, data organization, and scalability while supporting reliable dashboard visualization and efficient customer analytics [3], [4].

B. Data Cleaning and Analytical Complexity

E-commerce datasets often contain duplicate records, missing values, inconsistent formats, and invalid transaction details that directly affect analytical accuracy. Processing such large and inconsistent datasets requires efficient data transformation and validation techniques. Traditional analytical methods may not perform efficiently when dealing with continuously growing customer datasets due to computational complexity and processing limitations [5],

Fig. 2: Challenges in E-commerce Data Analytics

[11].

The proposed dashboard addresses these challenges through scalable ETL operations and data transformation workflows implemented within Databricks. The Silver layer is mainly used for cleaning, validating, and transforming raw customer data into analysis-ready datasets. Advanced analytical tables and business metrics are then generated in the Gold layer to simplify dashboard reporting, customer segmentation, and sales trend analysis. These processing techniques help improve analytical accuracy and generate meaningful customer insights for business decision making [2], [3].

C. Scalability and Dashboard Visualization

Scalability is another significant challenge in e-commerce customer analytics because online shopping platforms continuously generate large volumes of real-time customer data. Analytical systems must be capable of handling increasing datasets efficiently without affecting performance. Customer interactions such as product searches, browsing behavior, payment activities, and purchase transactions require scalable processing systems and optimized analytical workflows for efficient business intelligence operations [4], [7].

In addition, effective dashboard visualization is essential for understanding customer behavior and business trends. Poor visualization techniques may create difficulties in interpreting analytical insights and reduce decision-making efficiency. The developed dashboard integrates multiple visualization techniques such as bar charts, pie charts, line charts, KPI indicators, and sales trend reports to provide meaningful business insights. These visualizations help business analysts understand customer acquisition trends, product popularity, customer retention rates, and regional sales distribution effectively while supporting data-driven business decision making [6], [15].

D. Data Security and Privacy

Customer data contains sensitive information such as payment details, browsing activities, transaction histories, and personal customer records, making security and privacy critical challenges in e-commerce analytics systems. Unauthorized access, data leakage, and privacy violations may significantly affect both organizations and customers. Therefore, secure data management and controlled access mechanisms are necessary during customer data processing and dashboard analytics operations [8], [13].

The proposed system focuses on maintaining secure analytical processing within the Databricks environment using structured data management and controlled access policies. Proper data handling practices and organized processing layers help improve reliability, consistency, and privacy protection during customer analytics operations. In addition, authentication mechanisms, encryption techniques, and role-based access control help prevent cyber threats and unauthorized access to customer information. Continuous monitoring and secure backup systems are also important for maintaining customer data availability, integrity, and long-term protection against security threats and system failures [3], [8].

III. OPEN RESEARCH ISSUES IN E-COMMERCE CUSTOMER ANALYTICS

E-commerce customer analytics has become an important research area in both industry and academia due to the rapid growth of digital commerce platforms and large-scale customer datasets. Modern e-commerce organizations continuously generate data from online transactions, browsing activities, customer reviews, payment systems, mobile applications, and product interaction records. Extracting meaningful business insights from these datasets requires scalable analytical platforms, efficient data processing techniques, and interactive visualization systems. The increasing complexity of customer data creates several open research challenges related to real-time analytics, cloud-based data processing, predictive customer analysis, personalized recommendation systems, and secure customer data management. This section discusses major research areas associated with e-commerce customer analytics and modern data engineering technologies [1], [5].

A. Real-Time Customer Analytics

Modern e-commerce platforms continuously generate real-time customer data through product searches, browsing activities, payment transactions, and customer interactions. Managing and analyzing these continuously generated data streams remains a major challenge in customer analytics systems. Organizations require scalable analytical frameworks capable of processing real-time customer activities for sales monitoring, customer engagement analysis, and trend identification [2], [7].

Real-time analytical systems help organizations identify customer purchasing patterns, monitor product demand, and generate faster business insights. In the proposed dashboard, Databricks supports scalable processing of continuously generated customer datasets and enables real-time analytical reporting through interactive dashboard visualization. However, maintaining low-latency processing and high analytical performance for large-scale customer datasets still remains an important research challenge in modern e-commerce analytics systems [3], [4].

B. Cloud-Based Customer Analytics

Cloud computing has become an important technology for handling large-scale customer datasets because of its scalability, flexibility, and distributed processing capabilities. E-commerce organizations require scalable infrastructures for storing, processing, and analyzing massive customer datasets generated from multiple online platforms. Cloud-based analytical systems provide on-demand computing resources for customer behavior analysis, sales trend identification, and dashboard reporting operations [3], [4].

The integration of cloud-based platforms such as Databricks with Apache Spark enables efficient customer data processing, distributed computation, and scalable dashboard visualization. Cloud-based systems also improve operational efficiency by supporting collaborative analytical workflows and centralized data management. However, research challenges related to cloud security, data synchronization, processing optimization, and system reliability still require further improvement in customer analytics environments [4], [13].

C. Predictive Customer Behavior Analysis

Predictive analytics techniques are increasingly used in e-commerce systems for customer segmentation, purchasing behavior analysis, product recommendation, and sales forecasting. Analytical models can help organizations identify customer interests, predict purchasing patterns, and improve personalized marketing strategies. Predictive customer analytics also helps businesses improve customer retention and optimize operational performance through data-driven decision making [6], [10].
However, developing highly accurate predictive models using continuously growing customer datasets remains a significant research challenge. Customer behavior changes dynamically based on market trends, seasonal demand, and user interests, making predictive analysis more complex. In addition, large-scale customer datasets often contain inconsistent or incomplete information that may affect analytical accuracy. Therefore, future customer analytics systems require more scalable analytical models and efficient data processing frameworks for improving predictive business intelligence operations [5], [11].

D. Customer Knowledge Discovery Systems

Customer analytics systems not only focus on data processing but also emphasize business knowledge discovery and intelligent decision support. Knowledge discovery systems help organizations extract meaningful business insights from customer data and apply them in real-world business operations. These systems involve multiple

stages such as customer data collection, data transformation, business insight generation, dashboard visualization, and strategic decision making [1], [15].

In the proposed E-commerce Customer Behavior Dashboard, analytical reports and interactive visualizations help business analysts identify customer acquisition trends, product popularity, customer retention patterns, and regional sales distribution. These knowledge exploration systems improve business intelligence and support organizations in developing targeted marketing strategies and improving customer engagement. Future research in customer knowledge discovery systems will mainly focus on improving dashboard intelligence, automated reporting systems, and real-time business insight generation using scalable analytical platforms such as Databricks [3], [15].

IV. TOOLS AND FRAMEWORKS FOR E-COMMERCE CUSTOMER ANALYTICS

Modern e-commerce platforms continuously generate large volumes of customer-related data from online transactions, browsing activities, payment systems, product searches, and customer interactions. Processing these datasets requires scalable analytical frameworks and distributed computing technologies capable of handling large-scale customer analytics efficiently. Modern customer analytics systems use various tools and platforms for data ingestion, transformation, storage, visualization, and real-time analytical reporting. The proposed E-commerce Customer Behavior Dashboard utilizes Databricks, Apache Spark, cloud-based analytics, and dashboard visualization frameworks to process and analyze customer datasets using Medallion Architecture. This section discusses the major technologies and frameworks used for e-commerce customer analytics and dashboard development [1], [3].

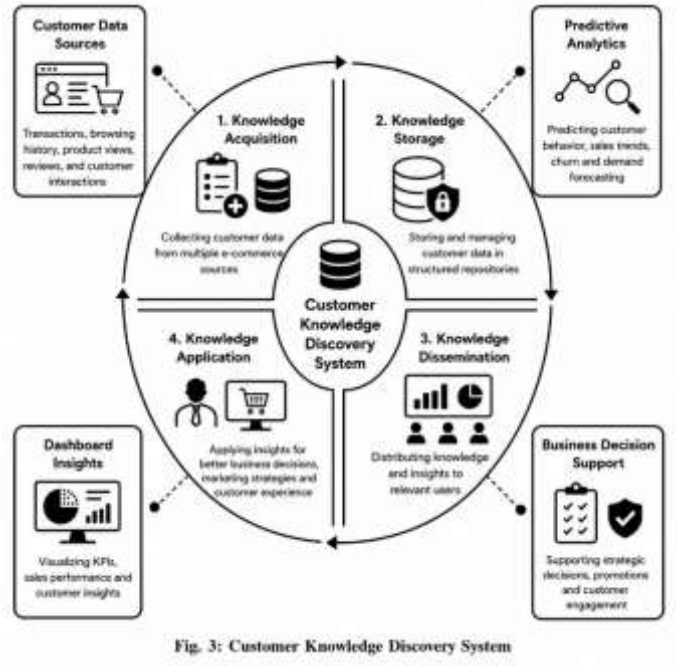

Fig. 3: Customer Knowledge Discovery System

A. Apache Spark for Customer Analytics

Apache Spark is a distributed big data processing framework designed for scalable and high-speed analytical computation. It supports in-memory processing, distributed execution, SQL analytics, streaming operations, and interactive dashboard reporting. In the proposed customer analytics system, Apache Spark is integrated with the Databricks platform to process large-scale customer datasets efficiently. Spark helps perform customer data ingestion, transformation, aggregation, and analytical processing within the Bronze, Silver, and Gold layers of the Medallion Architecture [2], [14].

The Spark architecture consists of a driver program, cluster manager, worker nodes, executors, and distributed tasks. The driver program manages analytical execution, while worker nodes process customer datasets in parallel to improve scalability and computational performance. Apache Spark significantly reduces processing time for customer analytics operations and supports scalable dashboard generation for customer behavior reporting and visualization.

Apache Spark also provides strong support for scalable data engineering by enabling parallel processing across distributed computing environments. In the proposed dashboard, Spark SQL and Data Frame operations are used to process customer transaction records, browsing activities, and sales data efficiently within the Databricks environment. The integration of Spark with Medallion Architecture improves data reliability by supporting structured transformation, filtering, aggregation, and analytical computation across Bronze, Silver, and Gold layers. These distributed processing capabilities help improve dashboard performance and support efficient business intelligence operations [2], [3].

In addition, Apache Spark supports advanced analytical applications such as real-time data processing, scalable analytics, and customer behavior analysis. The framework processes large volumes of customer and transaction data while maintaining computational efficiency and scalability. Within the proposed system, Spark enables efficient

customer analytics and supports interactive dashboard visualization for customer acquisition trends, product popularity, regional sales analysis, and purchasing behavior reporting. Spark also provides fault tolerance, distributed memory management, and high-speed processing capabilities that improve the reliability and performance of

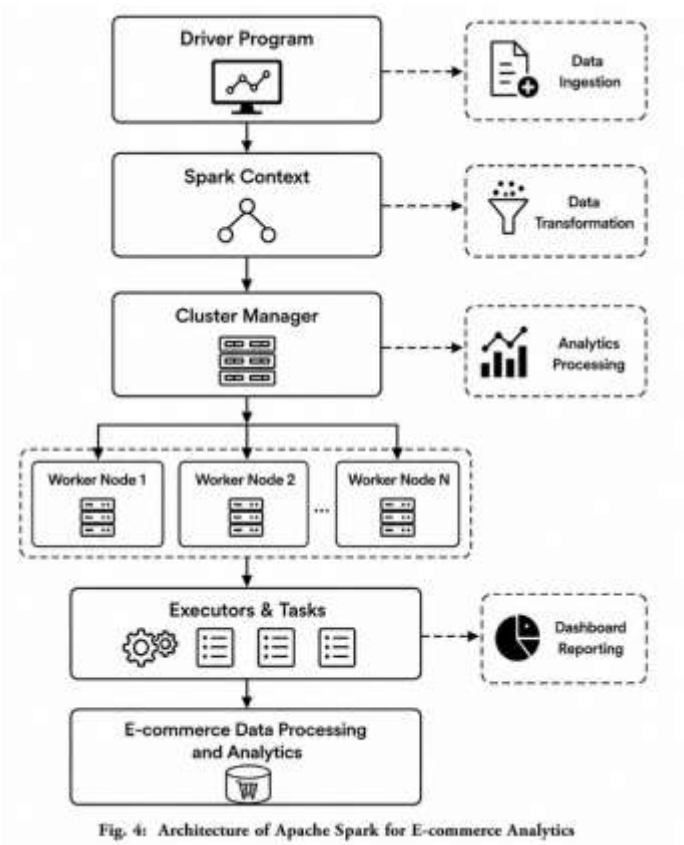

Fig. 4: Architecture of Apache Spark for E-commerce Analytics

analytical operations in modern e-commerce environments [2], [7].

B. Databricks Platform

Databricks is the primary analytical platform used in the proposed E-commerce Customer Behavior Dashboard. The platform combines scalable data engineering, distributed processing, SQL analytics, and dashboard visualization within a unified cloud-based environment. Databricks is built on Apache Spark architecture and provides efficient support for ETL workflows, analytical processing, and business intelligence operations [2], [3].

In this project, Databricks is used to process customer datasets through Bronze, Silver, and Gold layers for structured customer analytics. The Bronze layer stores raw customer and transaction data collected from e-commerce platforms, while the Silver layer performs data cleaning, preprocessing, and transformation. The Gold layer contains business-level analytical datasets used for dashboard reporting and visualization. This layered architecture improves scalability, processing efficiency, collaboration, and data reliability for customer analytics systems [3], [4].

Databricks also supports notebook environments, SQL-based analytics, real-time processing, and collaborative dashboard

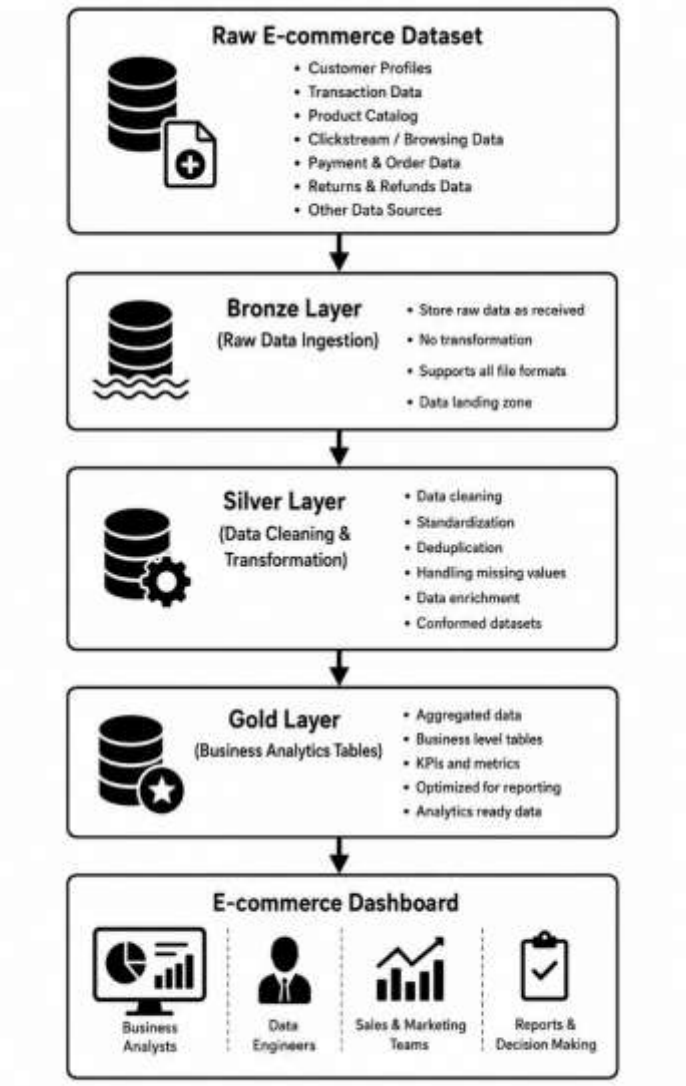

Fig. 5: Medallion Architecture for E-commerce Analytics

development for data engineers and business analysts. These capabilities help organizations improve customer behavior analysis, sales monitoring, and operational decision making through scalable and reliable customer analytics solutions. In addition, Databricks provides automated workflow management, secure cloud integration, and distributed processing capabilities that improve analytical performance and support efficient dashboard visualization systems [3], [14]

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **06** June-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

C. Cloud Computing for Customer Data Processing

Cloud computing provides scalable and flexible infrastructure for storing and processing large customer datasets generated from modern e-commerce systems. E-commerce organizations require cloud-based platforms to manage continuously growing customer information and analytical workloads efficiently. Cloud technologies support distributed customer analytics, real-time processing, and centralized dashboard reporting systems [4], [13].

The integration of Databricks with cloud-based infrastructure improves resource scalability, processing efficiency, data accessibility, and collaborative analytical operations. Cloud computing also reduces infrastructure management complexity and supports efficient customer data processing for large-scale business intelligence applications. These capabilities help organizations process customer interactions and sales activities more effectively within modern e-commerce environments [3], [4].

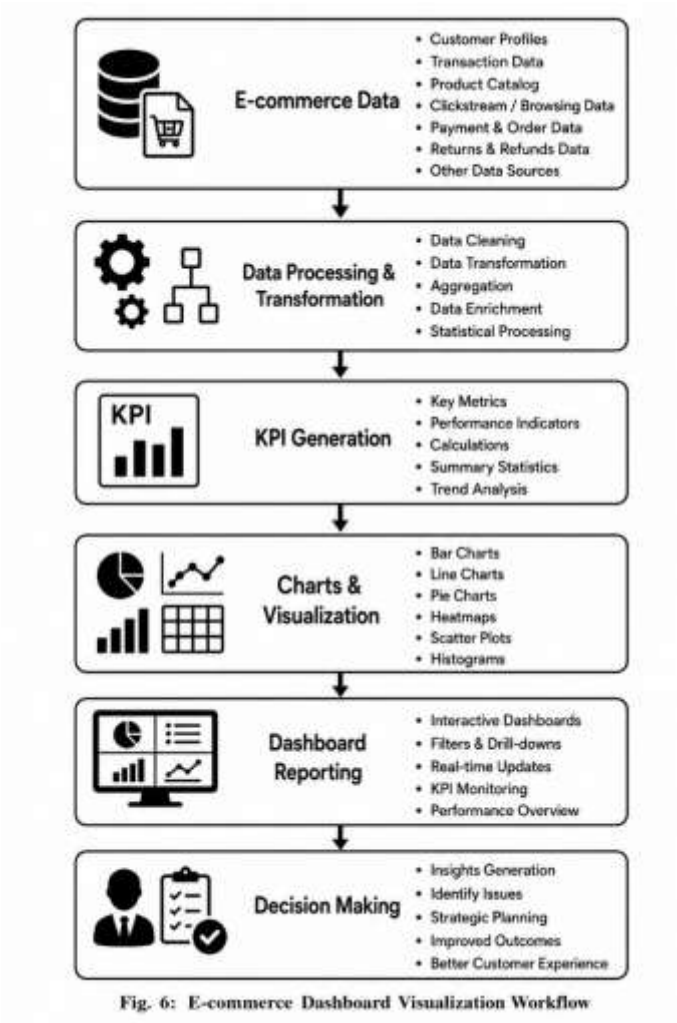

Fig. 6: E-commerce Dashboard Visualization Workflow

D. Dashboard Visualization and Reporting

Data visualization plays a major role in customer analytics because business analysts require meaningful graphical insights for decision making. The proposed dashboard uses multiple visualization techniques including bar charts, line charts, pie charts, KPI indicators, heatmaps, and sales trend reports to analyze customer behavior and business performance [15].

Interactive dashboard reporting helps organizations monitor customer acquisition trends, product demand, customer retention rates, purchasing behavior, and regional sales distribution. Effective visualization improves business intelligence and supports strategic decision making within e-commerce organizations. The dashboard developed in Databricks transforms complex customer datasets into understandable business insights through interactive analytical reporting and visualization systems [6], [15].

E. Real-Time Customer Data Processing

Modern e-commerce systems continuously generate real-time customer data through browsing activities, product searches, transactions, and payment interactions. Processing these continuously generated data streams remains a major challenge due to the high velocity of customer information [2], [7].

Distributed analytical frameworks such as Apache Spark and Databricks enable scalable real-time customer analytics and support continuous monitoring of customer activities and sales performance. Real-time processing capabilities improve dashboard responsiveness and help organizations generate faster business insights for customer engagement and operational decision making [3], [4].

F. Data Security and Privacy Management

Customer analytics systems handle sensitive information such as payment details, transaction histories, browsing records, and personal customer information. Maintaining data confidentiality, integrity, and privacy is an important requirement in e-commerce analytics systems [8], [13].

The proposed customer analytics framework focuses on secure customer data handling and structured analytical processing within the Databricks environment. Authentication mechanisms, encryption techniques,

controlled access policies, and secure cloud integration help improve data security and reliability during customer analytics operations. Proper data management practices also support secure dashboard reporting and long-term protection of customer information within modern e-commerce analytics platforms [3], [8].

G. Scalable Customer Analytics Framework

Scalability is one of the most important requirements for customer analytics because e-commerce datasets continuously grow over time. Analytical systems must efficiently process increasing customer records and transaction activities without affecting system performance and reliability [4], [7].

The proposed E-commerce Customer Behavior Dashboard uses distributed processing frameworks and Medallion Architecture to support scalable customer analytics and business-level dashboard reporting. The system demonstrates how modern big data technologies and cloud-based analytical platforms can improve customer data management, business intelligence, and customer behavior analysis within modern e-commerce environments [2], [3].

V. SUGGESTIONS FOR FUTURE WORK

The rapid growth of customer data generated from e-commerce platforms, online transactions, mobile applications, and digital payment systems has increased the importance of scalable customer analytics systems. Although the proposed E-commerce Customer Behavior Dashboard provides an efficient framework for customer data processing and visualization using Databricks and Medallion Architecture, several research opportunities still exist for improving large-scale customer analytics systems. Future analytical platforms may focus on developing more intelligent, scalable, and real-time customer analytics solutions capable of handling continuously increasing customer datasets with improved analytical accuracy and operational efficiency [1], [3].

One of the major future research areas is the integration of advanced predictive analytics and intelligent recommendation systems within customer analytics platforms. Predictive models can help organizations identify customer purchasing behavior, forecast product demand, improve customer retention, and optimize marketing strategies. Intelligent analytical systems may also support automated customer segmentation, personalized product recommendations, and trend prediction for improving customer engagement. However, handling inconsistent customer datasets, missing transaction records, and continuously changing customer behavior still remains a significant research challenge in predictive customer analytics systems [5], [6].

Another important future direction is real-time customer analytics using streaming customer data generated from online shopping platforms and digital commerce applications. Modern e-commerce systems continuously generate customer interaction data through browsing activities, payment transactions, product searches, and customer engagement records. Processing these real-time customer streams efficiently requires distributed processing frameworks and scalable cloud-based analytical platforms. Future research may focus on integrating Apache Spark Streaming, Databricks workflows, and real-time dashboard systems to improve customer behavior monitoring and business intelligence operations [2], [7].

Scalability and customer data security are also important research challenges for future customer analytics systems. As customer datasets continue to grow rapidly, analytical frameworks must efficiently process large-scale transaction records and customer activities without affecting performance and reliability. In addition, customer analytics systems contain sensitive information such as payment details, browsing histories, and transaction records, requiring secure data management and controlled access mechanisms. Future analytical frameworks may incorporate advanced encryption methods, secure cloud computing models, and privacy-aware analytical techniques for secure customer data processing and business reporting [4], [8].

Another promising research direction involves enhancing dashboard systems using intelligent visualization and automated reporting techniques. Future dashboards may include advanced analytical features such as real-time KPI monitoring, interactive drill-down analysis, automated business reporting, and intelligent customer insight generation. The integration of scalable cloud-based analytics, distributed processing frameworks, and advanced visualization techniques can significantly improve business intelligence and operational decision making within e-commerce organizations [3], [15].

Furthermore, future research may focus on integrating intelligent analytical assistants and conversational

dashboard systems within customer analytics platforms for improved user interaction and automated business insights. These intelligent systems may help business analysts interpret customer behavior data more efficiently and support faster data-driven decision making. Therefore, the continuous development of scalable data engineering technologies, distributed analytical frameworks, and intelligent dashboard systems will play a significant role in the future of e-commerce customer analytics and digital business intelligence [2], [6].

VI. CONCLUSION

The rapid growth of customer data generated from e-commerce platforms, online transactions, mobile applications, and digital payment systems has increased the importance of scalable customer analytics systems. Processing and analyzing these large and complex customer datasets using traditional analytical approaches is a challenging task. This paper presented an E-commerce Customer Behavior Dashboard developed using the Databricks platform and Medallion Architecture for scalable customer data processing, transformation, and visualization. The proposed framework efficiently processes customer datasets through Bronze, Silver, and Gold layers to improve data quality, analytical consistency, and business-level reporting [1], [3].

The developed customer analytics system provides meaningful insights into customer acquisition trends, purchasing behavior, product popularity, regional sales distribution, and customer retention through interactive dashboard visualization. Multiple analytical and visualization techniques such as bar charts, line charts, pie charts, KPI indicators, and sales trend reports were implemented to support customer behavior analysis and business decision making. The integration of Apache Spark and Databricks improved distributed processing efficiency, scalability, and analytical performance for handling large-scale customer datasets [2], [14].

This work also highlighted several important challenges and research opportunities associated with e-commerce customer analytics, including real-time customer data processing, predictive analytics, dashboard scalability, data security, and intelligent visualization systems. The proposed framework demonstrates how modern big data technologies and cloud-based analytical platforms can be effectively utilized for customer analytics and business intelligence systems. Furthermore, Medallion Architecture provides a structured and reliable approach for customer data engineering, ETL processing, and dashboard reporting operations [3], [4].

In the future, customer analytics systems may integrate advanced predictive models, real-time customer monitoring, intelligent recommendation systems, and automated dashboard reporting for enhanced business intelligence and customer engagement. The proposed E-commerce Customer Behavior Dashboard can serve as a scalable foundation for future customer analytics systems and demonstrates the practical application of modern data engineering and visualization technologies within the e-commerce industry [6], [15].

REFERENCES

[1] A. Gandomi and M. Haider, "Beyond the Hype: Big Data Concepts, Methods, and Analytics," *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.

[2] M. Zaharia et al., "Apache Spark: A Unified Engine for Big Data Processing," *Communications of the ACM*, vol. 59, no. 11, pp. 56–65, 2016.

[3] Databricks Inc., "The Databricks Lakehouse Platform," Databricks Documentation, 2024. Available: https://www.databricks.com/

[4] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The Rise of Big Data on Cloud Computing: Review and Open Research Issues," *Information Systems*, vol. 47, pp. 98–115, 2015.

[5] X. Jin, B. W. Wah, X. Cheng, and Y. Wang, "Significance and Challenges of Big Data Research," *Big Data Research*, vol. 2, no. 2, pp. 59–64, 2015.

[6] O. Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, G. K. Karagiannidis, and K. Taha, "Efficient Machine Learning for Big Data: A Review," *Big Data Research*, vol. 2, no. 3, pp. 87–93, 2015.

[7] K. Kambatla, G. Kollias, V. Kumar, and A. Grama, "Trends in Big Data Analytics," *Journal of Parallel and Distributed Computing*, vol. 74, no. 7, pp. 2561–2573, 2014.

[8] A. Jacobs, "The Pathologies of Big Data," *Communications of the ACM*, vol. 52, no. 8, pp. 36–44, 2009.

[9] N. Mishra, C. Lin, and H. Chang, "A Cognitive Adopted Framework for IoT Big Data Management and Knowledge Discovery Perspective," *International Journal of Distributed Sensor Networks*, vol. 2015, pp. 1–13, 2015.

[10] D. P. Acharjya, S. Dehuri, and S. Sanyal, *Computational Intelligence for Big Data Analysis*. Springer International Publishing, Switzerland, 2015.

[11] M. Herland, T. M. Khoshgoftaar, and R. Wald, "A Review of Data Mining Using Big Data in Health Informatics," *Journal of Big Data*, vol. 1, no. 2, pp. 1–35, 2014.

[12] I. Merelli, H. Perez-Sanchez, S. Gesing, and D. D'Agostino, "Managing, Analysing, and Integrating Big Data in Medical Bioinformatics: Open Problems and Future Perspectives," *BioMed Research International*, vol. 2014, pp. 1–13, 2014.

[13] T. Huang, L. Lan, X. Fang, P. An, J. Min, and F. Wang, "Promises and Challenges of Big Data Computing in Health Sciences," *Big Data Research*, vol. 2, no. 1, pp. 2–11, 2015.

[14] Apache Spark Documentation, "Spark SQL, Machine Learning and Streaming Analytics," Apache Software Foundation, 2024. Available: https://spark.apache.org/

[15] Microsoft, "Business Intelligence and Data Visualization Solutions," Microsoft Learn Documentation, 2024. Available: https://learn.microsoft.com/