International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 06 June-2026 | Impact Factor: 3.5

A Medallion Architecture-Based Healthcare Patient Analytics Dashboard Using Databricks

Biswa Ranjan Behera

Department of Master of Computer Applications

GIFT Autonomous, Bhubaneswar, Odisha, India, bbehera2024@gift.edu.in

Smruti Ranjan Swain

Head of Department Master of Computer Applications

GIFT Autonomous, Bhubaneswar, Odisha, India, hodmca@gift.edu.in

How to Cite this Article:

Behera, B. R. (2026). A Medallion Architecture-Based Healthcare Patient Analytics Dashboard Using Databricks. International Journal of Creative and Open Research in Engineering and Management, 02(6). https://doi.org/10.55041/ijcope.v2i6.082

https://doi.org/10.55041/ijcope.v2i6.082

Abstract—Healthcare organizations generate a large volume of patient-related data every day through hospital systems, medical records, laboratory reports, and treatment activities. Analyzing these large datasets is important for improving patient care, hospital performance, and healthcare decision-making. This paper presents a Healthcare Patient Analytics Dashboard developed using the Databricks Lakehouse Platform and Medallion Architecture for scalable healthcare data processing and analysis. The proposed system processes healthcare datasets through Bronze, Silver, and Gold layers to improve data quality, perform transformation, and generate analytical insights. Interactive dashboards are used to visualize patient demographics, disease distribution, admission trends, treatment outcomes, and hospital performance using various charts and analytical reports. The developed system helps healthcare administrators and medical professionals identify healthcare patterns, monitor operational efficiency, and support data-driven decision-making. Furthermore, this work demonstrates how modern big data technologies and visualization tools can be effectively utilized in the healthcare sector for intelligent analytics and healthcare management.

Keywords—Healthcare Analytics; Databricks; Medallion Architecture; Big Data Analytics; Patient Dashboard; Data Visualization; Lakehouse Platform; Healthcare Management

I. INTRODUCTION

In the modern healthcare environment, a huge amount of patient-related data is generated daily from hospital management systems, electronic medical records, laboratory reports, diagnostic centers, and healthcare applications. The rapid growth of digital technologies and cloud-based healthcare systems has significantly increased the volume and complexity of healthcare data. Managing and analyzing these datasets using traditional data processing techniques has become a major challenge for healthcare organizations. Healthcare data generally exists in structured, semi- structured, and unstructured formats and requires efficient processing methods to generate meaningful insights for decision-making. Big data technologies and modern analytical platforms provide scalable solutions for processing large healthcare datasets and improving operational efficiency.

The concept of big data in healthcare is commonly associated with characteristics such as volume, velocity, variety, and veracity. Volume represents the continuously increasing amount of healthcare data generated from multiple sources, while velocity refers to the speed at which this data is produced and processed. Variety indicates the different forms of healthcare data including patient records, prescriptions, medical imaging reports, and admission details. Veracity focuses on data quality, consistency, and reliability for accurate healthcare analysis. Efficient healthcare analytics helps hospitals improve patient care, monitor disease trends, optimize treatment planning, and support data-driven medical decisions.

This paper presents a Healthcare Patient Analytics Dashboard developed using the Databricks Lakehouse Platform and Medallion Architecture for scalable healthcare data processing and visualization. The proposed system processes healthcare data through Bronze, Silver, and Gold layers to perform data ingestion, cleaning, transformation, and business-level analytics. Interactive dashboards are created to analyze patient demographics, disease distribution, hospital performance, treatment outcomes, and admission trends using multiple visualization techniques. The developed framework demonstrates how modern big data technologies and healthcare analytics can be integrated to support intelligent healthcare management and operational decision-making.

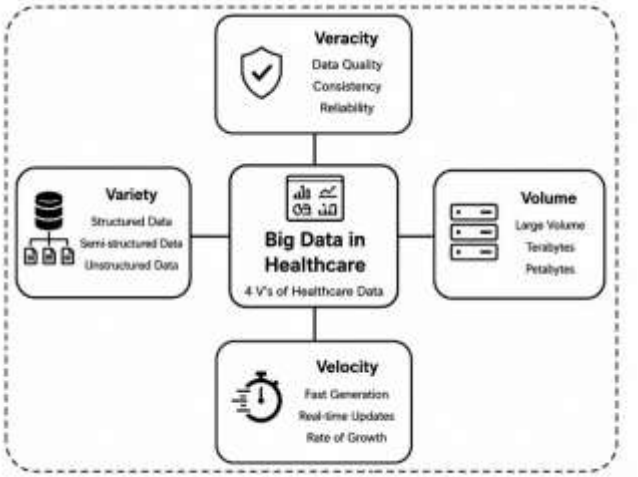

Fig. 1: Characteristics of Big Data in Healthcare

II. CHALLENGES IN HEALTHCARE DATA ANALYTICS

In recent years, the healthcare sector has experienced a rapid increase in the generation of digital data from hospital management systems, electronic health records, laboratory reports, diagnostic centers, wearable devices, and healthcare applications. The continuous growth of healthcare information creates new opportunities for data-driven analysis and intelligent decision-making. However, processing and analyzing large-scale healthcare datasets also introduces several technical and operational challenges. Healthcare data analytics requires efficient storage systems, scalable processing frameworks, data visualization techniques, and secure data management solutions to handle large volumes of patient information effectively. The proposed Healthcare Patient Analytics Dashboard addresses these challenges using the Databricks Lakehouse Platform and Medallion Architecture for structured healthcare data processing and analytics.

A. Data Storage and Processing

One of the major challenges in healthcare analytics is handling the continuously growing volume of patient data generated from multiple healthcare sources. Traditional database systems often face limitations in storing and processing large-scale healthcare datasets efficiently. In addition, healthcare data exists in different formats such as structured patient records, semi-structured reports, and unstructured clinical notes, making data integration and analysis more complex. Another important issue is maintaining data quality, consistency, and accuracy before performing analytical operations. To overcome these challenges, the proposed system utilizes Bronze, Silver, and Gold layers for scalable data ingestion, cleaning, transformation, and analytical processing. The Medallion Architecture helps

organize healthcare data systematically and improves processing efficiency for analytical reporting while supporting reliable dashboard visualization, faster analytical computation, improved healthcare decision-making, and scalable management of large healthcare datasets efficiently.

B. Data Cleaning and Analytical Complexity

Healthcare datasets often contain duplicate records, missing values, inconsistent formats, and invalid information 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 large-scale healthcare data due to computational complexity and processing limitations. The proposed system addresses these issues through the Silver layer, where healthcare data is cleaned, standardized, and transformed into analysis-ready datasets. Advanced analytical tables are then generated in the Gold layer to simplify dashboard reporting and business-level analytics.

C. Scalability and Data Visualization

Scalability is another significant challenge in healthcare data analytics because healthcare organizations continuously generate large amounts of real-time patient information. Analytical systems must be capable of handling increasing datasets without affecting performance. In addition, effective visualization of healthcare data is essential for understanding patient trends, disease distribution, hospital performance, and treatment outcomes. Poor visualization techniques may lead to difficulty in interpreting analytical results. The developed dashboard integrates multiple visualization techniques such as bar charts, pie charts, line charts, heatmaps, scatter plots, and histograms to provide meaningful healthcare insights. These visualizations help healthcare administrators and medical professionals analyze operational performance and support data-driven decision-making.

D. Data Security and Privacy

Healthcare data contains highly sensitive patient information, making security and privacy a critical challenge in healthcare analytics systems. Unauthorized access, data leakage, and privacy violations may affect both healthcare organizations and patients. Therefore, secure data management and controlled access mechanisms are necessary during healthcare data processing and analysis. The proposed system focuses on maintaining data integrity and secure analytical processing within the Databricks environment. Proper data handling practices and structured processing layers help improve reliability, consistency, and privacy protection during healthcare analytics operations while supporting secure healthcare reporting, controlled user access, and efficient regulatory compliance management. In addition, encryption techniques, authentication mechanisms, and role-based access control are essential for preventing cyber threats, maintaining patient confidentiality, and ensuring secure healthcare data sharing across analytical and cloud-based healthcare platforms. Furthermore, continuous security monitoring and secure backup systems are important for maintaining healthcare data availability, integrity, and long-term protection against system failures and cyberattacks.

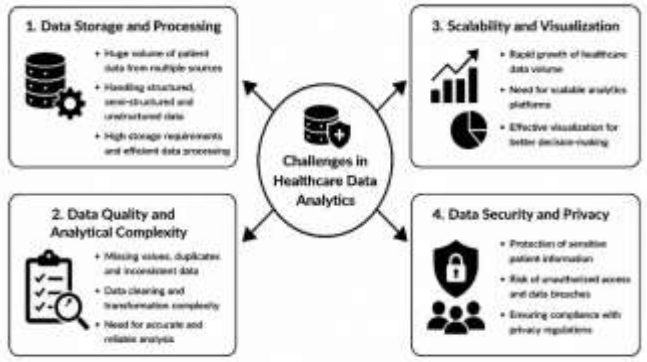

Fig. 2: Challenges in Healthcare Data Analytics

III. OPEN RESEARCH ISSUES IN HEALTHCARE DATA ANALYTICS

Healthcare data analytics has become an important research area in both industry and academia due to the rapid growth of digital healthcare systems and large-scale patient data. Modern healthcare organizations continuously generate data from hospital management systems, wearable devices, medical imaging systems, laboratory reports, and electronic health records. Extracting meaningful knowledge from these datasets requires scalable analytical platforms, intelligent data processing techniques, and efficient visualization systems. The increasing complexity of healthcare data creates several open research challenges related to real-time analytics, cloud-based healthcare systems, machine learning integration, predictive healthcare analysis, and secure healthcare data management. This section discusses major open research areas associated with healthcare data analytics and modern big data technologies.

A. IoT and Real-Time Healthcare Analytics

The integration of Internet of Things (IoT) devices in healthcare systems has significantly increased the generation of real-time patient data. Smart healthcare devices such as wearable sensors, smart monitors, and remote patient tracking systems continuously produce large volumes of healthcare information. Managing and analyzing these continuous data streams remains a major challenge in healthcare analytics. Efficient analytical frameworks are required to process real-time healthcare data for disease monitoring, patient tracking, and emergency response systems. Machine learning and predictive analytics techniques can help extract meaningful healthcare insights from IoT-generated patient data. Figure 3 illustrates the flow of healthcare data collection and analytical processing using IoT-based healthcare systems.

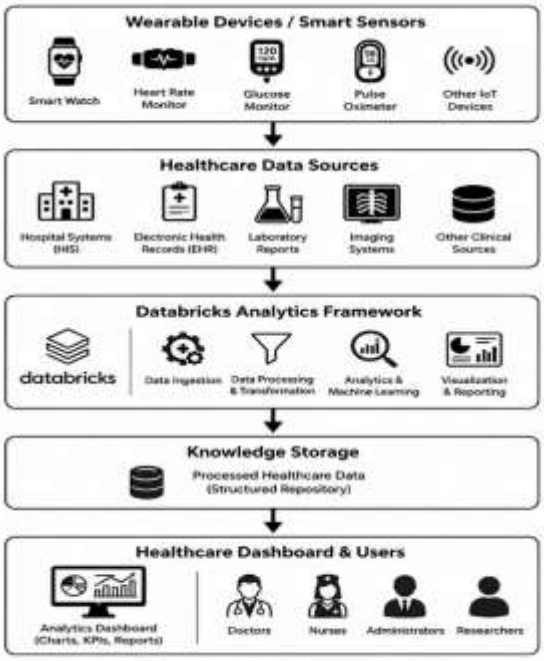

Fig. 3: IoT-Based Healthcare Data Analytics Framework

B. Cloud Computing for Healthcare Analytics

Cloud computing has become an important technology for handling large-scale healthcare datasets due to its scalability, flexibility, and cost efficiency. Healthcare organizations require scalable infrastructure for storing, processing, and analyzing massive patient datasets generated from multiple healthcare systems. Cloud-based healthcare analytics platforms provide on-demand resources for big data processing and advanced analytical operations. The integration of cloud computing with platforms such as Databricks enables efficient healthcare data processing, distributed computation, and scalable visualization systems. However, challenges related to data privacy, system reliability, and secure healthcare data sharing still require further research and development.

C. Artificial Intelligence and Predictive Healthcare Systems

Artificial Intelligence and machine learning techniques are increasingly used in healthcare analytics for disease prediction, treatment recommendation, patient risk analysis, and healthcare automation. Predictive healthcare systems can help hospitals identify critical patient conditions and improve healthcare decision-making. However, developing highly accurate predictive models using large-scale healthcare datasets remains a complex research challenge due to data inconsistency, imbalance, and privacy concerns. Future healthcare analytical systems may combine big data platforms, machine learning algorithms, and intelligent visualization systems to support advanced healthcare management and predictive medical analysis.

D. Healthcare Knowledge Exploration Systems

Healthcare analytics systems not only focus on data processing but also emphasize knowledge discovery and knowledge dissemination for healthcare decision-making. Knowledge exploration systems help healthcare professionals extract meaningful insights from analytical results and apply them in real-world healthcare operations. These systems involve multiple stages such as healthcare knowledge acquisition, knowledge storage, knowledge dissemination, and knowledge application. Figure 4 presents a conceptual healthcare knowledge exploration system for healthcare analytics and intelligent decision support.

Fig. 4: Healthcare Knowledge Exploration System

IV. TOOLS AND FRAMEWORKS FOR HEALTHCARE BIG DATA ANALYTICS

Large volumes of healthcare data are continuously generated from hospital systems, laboratory reports, wearable devices, electronic health records, and healthcare applications. Processing these datasets requires scalable analytical frameworks and distributed computing technologies capable of handling healthcare big data efficiently. Modern healthcare analytics systems use various tools and platforms for data ingestion, storage, processing, visualization, and real-time analytical reporting. The proposed Healthcare Patient Analytics Dashboard utilizes Databricks, Apache Spark, cloud-based analytics, and visualization frameworks to process and analyze healthcare datasets through the Medallion Architecture. This section discusses the major technologies and frameworks used for healthcare big data analytics.

A. Apache Spark for Healthcare Analytics

Apache Spark is a distributed big data processing framework designed for fast and scalable analytical computation. It supports in-memory processing, distributed execution, machine learning, streaming analytics, and interactive data visualization. In the proposed healthcare analytics system, Apache Spark is integrated with the Databricks platform to process large-scale healthcare datasets efficiently. Spark helps perform healthcare data ingestion, transformation, aggregation, and analytical processing within the Bronze, Silver, and Gold layers of the Medallion Architecture.

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 healthcare datasets in parallel to improve scalability and processing performance. Apache Spark significantly reduces processing time for healthcare analytics and supports scalable dashboard generation for healthcare reporting and visualization.

Apache Spark also provides strong support for scalable healthcare data engineering by enabling parallel data processing across multiple computing nodes. In the proposed Healthcare Patient Analytics Dashboard, Spark SQL and DataFrame operations are used to process large healthcare datasets efficiently within the Databricks environment. The integration of Spark with the Medallion Architecture improves data reliability by supporting structured data transformation, aggregation, filtering, and analytical computation across Bronze, Silver, and Gold layers. This distributed processing capability helps reduce execution time and improves the overall performance of healthcare analytics and dashboard reporting systems.

In addition, Apache Spark supports advanced analytical applications such as machine learning, real-time stream processing, and predictive analytics, which are highly beneficial in modern healthcare systems. The framework can process large volumes of patient records and operational healthcare data while maintaining scalability and computational efficiency. In the proposed system, Spark enables efficient healthcare data analysis and supports interactive dashboard visualization for patient demographics, disease trends, treatment outcomes, and hospital performance analysis. The use of Apache Spark within the Databricks platform demonstrates how distributed computing technologies can enhance healthcare analytics and support intelligent healthcare decision-making systems. Furthermore, Spark provides fault tolerance, distributed memory management, and high-speed parallel processing capabilities that improve the reliability and performance of healthcare analytical operations. The framework also supports integration with cloud-based platforms and large healthcare databases, enabling healthcare organizations to perform scalable analytics, real-time monitoring, and advanced reporting efficiently. These capabilities help improve healthcare operational management, optimize resource utilization, and support data-driven clinical and administrative decision-making processes. Additionally, Apache Spark enables flexible healthcare workflow management, faster data transformation, and efficient handling of complex healthcare analytical tasks across distributed computing environments and modern cloud-based healthcare infrastructure systems.

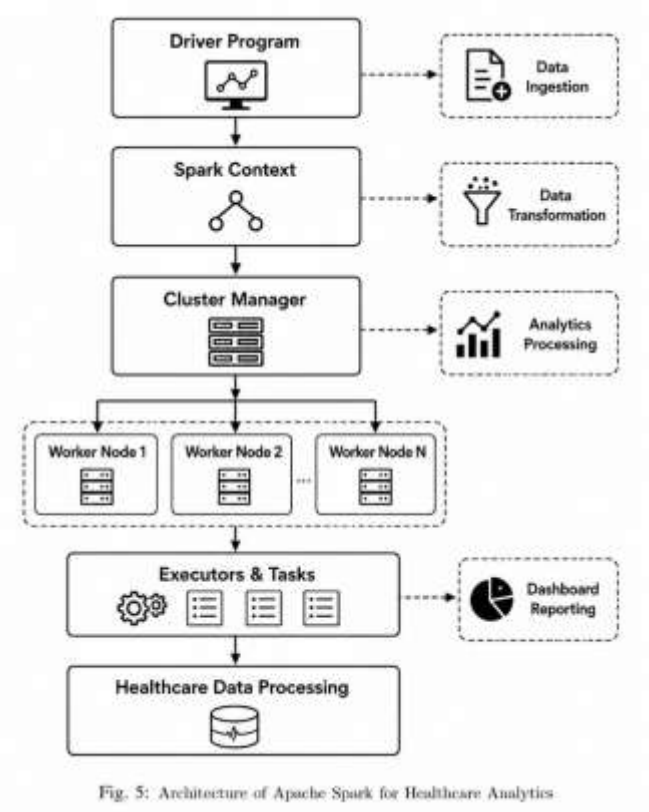

Fig. 5: Architecture of Apache Spark for Healthcare Analytics

B. Databricks Lakehouse Platform

The Databricks Lakehouse Platform combines the capabilities of data lakes and data warehouses for scalable healthcare analytics. It provides an integrated environment for data engineering, machine learning, SQL analytics, and dashboard visualization. In the proposed system, Databricks is used to process healthcare datasets through Bronze, Silver, and Gold layers for structured healthcare analytics. The platform supports distributed data processing and enables healthcare organizations to manage large-scale patient information efficiently while improving analytical performance and operational scalability.

The Bronze layer stores raw healthcare datasets, the Silver layer performs data cleaning and transformation, and the Gold layer generates business-level analytical tables for dashboard visualization. The Databricks platform improves scalability, collaboration, and distributed processing efficiency for healthcare analytical systems. In addition, the platform supports real-time data processing, cloud-based analytics, and collaborative healthcare reporting for data engineers and analysts. These capabilities help improve healthcare decision-making, optimize resource utilization, and support intelligent healthcare management through scalable and reliable healthcare analytics solutions.

Furthermore, Databricks provides automated workflow management, secure cloud integration, and scalable analytical processing capabilities that enhance healthcare data management, improve collaboration among healthcare professionals, and support efficient real-time healthcare reporting and dashboard visualization systems.

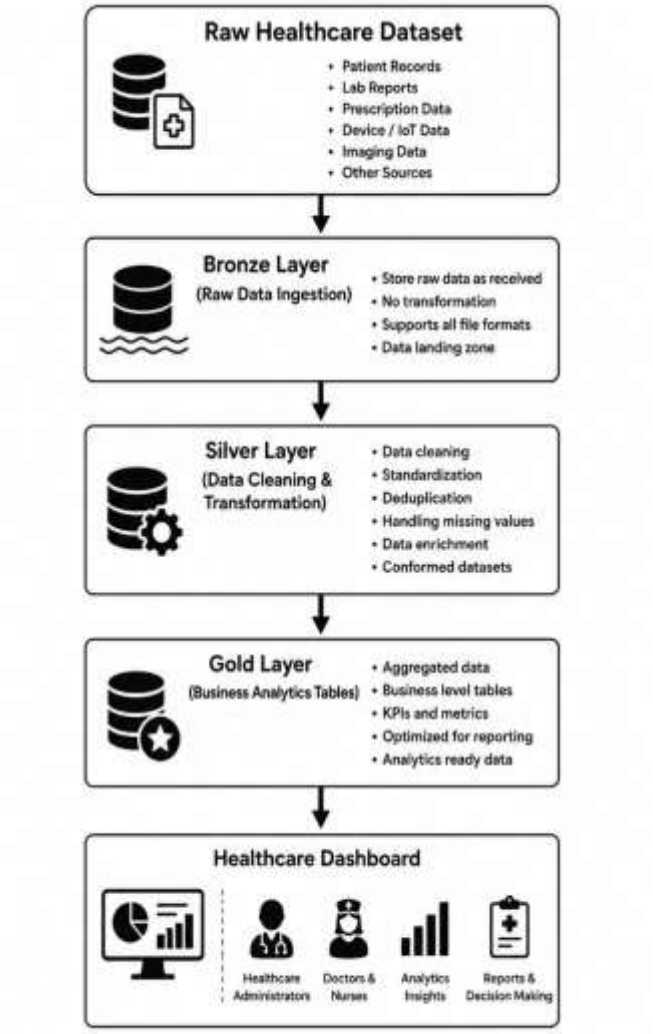

Fig. 6: Medallion Architecture for Healthcare Analytics

C. Cloud Computing for Healthcare Data Processing

Cloud computing provides scalable and flexible infrastructure for storing and processing large healthcare datasets. Healthcare organizations require cloud-based platforms to manage continuously growing patient information and analytical workloads. Cloud technologies support distributed healthcare analytics, real-time processing, and centralized healthcare reporting systems.

The integration of Databricks with cloud-based infrastructure improves resource scalability, data accessibility, and collaborative healthcare analytics. Cloud computing also reduces infrastructure

management complexity and supports efficient healthcare data processing for analytical applications.

D. Data Visualization and Dashboard Reporting

Data visualization plays a major role in healthcare analytics because healthcare professionals require meaningful graphical insights for decision-making. The proposed healthcare dashboard uses multiple visualization techniques including bar charts, line charts, pie charts, scatter plots, heatmaps, histograms, and KPI cards to analyze healthcare trends and operational performance.

Interactive dashboard reporting helps healthcare administrators monitor patient demographics, disease distribution, treatment outcomes, hospital performance, and admission trends. Effective visualization improves healthcare decision-making and supports operational analysis within healthcare organizations.

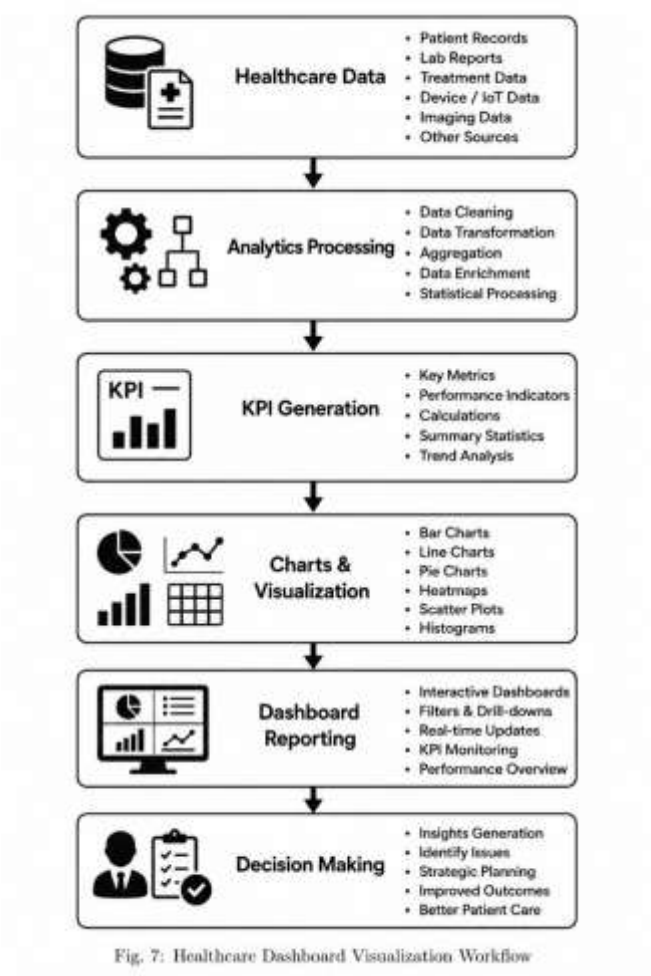

Fig. 7: Healthcare Dashboard Visualization Workflow

E. Machine Learning and Predictive Analytics

Machine learning techniques are increasingly used in healthcare analytics for disease prediction, patient risk assessment, treatment recommendation, and healthcare forecasting. Predictive analytics helps healthcare organizations identify healthcare patterns and improve operational planning.

Although the current system mainly focuses on healthcare analytics and visualization, future improvements may include predictive healthcare models integrated with Databricks and Apache Spark for advanced analytical intelligence.

F. Real-Time Healthcare Data Processing

Modern healthcare systems continuously generate real-time patient data through IoT devices, monitoring systems, and smart healthcare applications. Processing real-time healthcare streams remains a major challenge due to the high velocity of healthcare information.

Distributed analytical frameworks such as Apache Spark Streaming and Databricks enable scalable real-time healthcare analytics and support continuous monitoring of patient activities and healthcare operations.

G. Data Security and Privacy Management

Healthcare data contains sensitive patient information and requires secure data processing and controlled access management. Maintaining data confidentiality, integrity, and privacy is an important requirement in healthcare analytics systems.

The proposed healthcare analytics framework focuses on secure healthcare data handling and structured analytical processing within the Databricks environment. Proper data management practices improve reliability and support secure healthcare reporting systems.

H. Scalable Healthcare Analytics Framework

Scalability is one of the most important requirements for healthcare big data analytics because healthcare datasets continuously grow over time. Analytical systems must efficiently process increasing healthcare records without affecting performance and reliability.

The proposed Healthcare Patient Analytics Dashboard uses distributed processing frameworks and Medallion Architecture to support scalable healthcare analytics and business-level dashboard reporting. The system demonstrates how modern big data technologies can improve healthcare data management and intelligent healthcare decision-making.

V. SUGGESTIONS FOR FUTURE WORK

The rapid growth of healthcare data generated from hospitals, wearable devices, laboratory systems, and healthcare applications has increased the importance of scalable healthcare analytics systems. Although the proposed Healthcare Patient Analytics Dashboard provides an efficient framework for healthcare data processing and visualization using Databricks and Medallion Architecture, several research opportunities still exist for improving healthcare big data analytics. Future healthcare analytical systems may focus on developing more intelligent, scalable, and real-time healthcare solutions capable of handling continuously increasing healthcare datasets with improved analytical accuracy and operational efficiency.

One of the major future research areas is the integration of Artificial Intelligence and machine learning techniques for predictive healthcare analytics. Predictive models can help hospitals identify high-risk patients, predict disease outbreaks, recommend treatments, and improve healthcare decision-making. Advanced machine learning algorithms may also support automated healthcare monitoring, anomaly detection, and personalized healthcare recommendations. However, handling noisy healthcare datasets, missing values, imbalanced medical records, and inconsistent patient information remains a significant research challenge in predictive healthcare systems.

Another important future direction is real-time healthcare analytics using Internet of Things (IoT) devices and streaming healthcare data. Modern healthcare systems continuously generate patient data through wearable sensors, remote monitoring systems, and smart healthcare devices. Processing these real-time healthcare streams efficiently requires distributed processing frameworks and scalable cloud-based analytical platforms. Future research may focus on integrating Apache Spark Streaming, cloud computing, and real-time dashboard systems to improve healthcare monitoring and emergency response management.

Scalability and healthcare data security are also important research challenges for future healthcare analytics systems. As healthcare datasets continue to grow rapidly, analytical frameworks must efficiently process large-scale patient records without affecting performance and reliability. In addition, healthcare data contains highly sensitive patient information, requiring secure data management, controlled access systems, and privacy-preserving analytical techniques. Future healthcare analytical frameworks may incorporate advanced encryption methods, secure cloud computing models, and privacy-aware healthcare analytics for secure healthcare data processing.

Another promising research direction involves enhancing healthcare dashboard systems using intelligent visualization and automated reporting techniques. Future healthcare dashboards may include advanced analytical features such as real-time KPI monitoring, interactive drill-down analysis, AI-based insights, and automated healthcare reporting systems. The integration of cloud-based healthcare analytics, distributed processing frameworks, and intelligent visualization techniques can significantly improve healthcare management and operational decision-making.

Furthermore, future research may focus on integrating large language models, intelligent healthcare assistants, and conversational analytics systems into healthcare dashboards for enhanced user interaction and automated healthcare insights. These intelligent systems may help healthcare professionals analyze complex healthcare information more efficiently and support data-driven clinical decision-making. Therefore, the continuous development of scalable big data technologies, machine learning frameworks, and intelligent healthcare analytical systems will play a significant role in the future of digital healthcare management and smart healthcare analytics.

VI. CONCLUSION

The rapid growth of healthcare data generated from hospitals, electronic medical records, laboratory systems, wearable devices, and healthcare applications has increased the importance of scalable healthcare analytics systems. Processing and analyzing these large and complex healthcare datasets using traditional analytical approaches is a challenging task. This paper presented a Healthcare Patient Analytics Dashboard developed using the Databricks Lakehouse Platform and Medallion Architecture for scalable healthcare data processing, transformation, and visualization. The proposed framework efficiently processes healthcare datasets through Bronze, Silver, and Gold layers to improve data quality, analytical consistency, and business-level reporting.

The developed healthcare analytics system provides meaningful insights into patient demographics, disease distribution, hospital performance, treatment outcomes, and admission trends through interactive dashboard visualization. Multiple analytical techniques and visualization methods such as bar charts, line charts, pie charts, heatmaps, scatter plots, and KPI monitoring

were implemented to support healthcare analysis and operational decision-making. The integration of Apache Spark and Databricks improved distributed processing efficiency, scalability, and analytical performance for handling large-scale healthcare datasets.

This work also highlighted several important challenges and research opportunities associated with healthcare big data analytics, including real-time healthcare processing, predictive healthcare analytics, data security, scalability, and intelligent healthcare visualization. The proposed framework demonstrates how modern big data technologies and cloud-based analytical platforms can be effectively utilized for healthcare management and healthcare decision support systems. Furthermore, the Medallion Architecture provides a structured and reliable approach for healthcare data engineering and analytical reporting.

In future, healthcare analytics systems may integrate advanced machine learning models, real-time IoT healthcare monitoring, artificial intelligence-based prediction systems, and intelligent dashboard automation for enhanced healthcare management. The proposed Healthcare Patient Analytics Dashboard can serve as a scalable foundation for future healthcare analytical systems and demonstrates the practical application of modern data engineering and visualization technologies in the healthcare sector.

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. H. Kuo, T. Sahama, A. W. Kushniruk, E. M. Borycki, and D. K. Grunwell, "Health Big Data Analytics: Current Perspectives, Challenges and Potential Solutions," *International Journal of Big Data Intelligence*, vol. 1, no. 1, pp. 114–126, 2014.

[3] R. Nambiar, A. Sethi, R. Bhardwaj, and R. Vargheese, "A Look at Challenges and Opportunities of Big Data Analytics in Healthcare," in *IEEE International Conference on Big Data*, 2013, pp. 17–22.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

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

[12] 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.

[13] 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.

[14] 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.

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