Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240341

Real-Time Visualization and Analysis Architecture for Data Integration Processes at Cologne University Hospital's Medical Data Integration Center

Md. Mostafa KAMAL[a,b,1], Ekaterina KUTAFINA[a] and Oya BEYAN[a,b,c]
[a] *Institute for Biomedical Informatics, University of Cologne, Faculty of Medicine and University Hospital Cologne, Cologne, Germany*
[b] *Medical Data Integration Center (MeDIC), University of Cologne, Faculty of Medicine and University Hospital Cologne, Cologne, Germany*
[c] *Department of Data Science and Artificial Intelligence, Fraunhofer FIT, Germany*
ORCiD ID: Md. Mostafa Kamal https://orcid.org/0000-0002-5506-4875, Ekaterina Kutafina https://orcid.org/0000-0002-3430-5123, Oya Beyan https://orcid.org/0000-0001-7611-3501

Abstract. This paper presents an implementation of an architecture based on open-source solutions using ELK Stack – Elasticsearch, Logstash, and Kibana – for real-time data analysis and visualizations in the Medical Data Integration Center, University Hospital Cologne, Germany. The architecture addresses challenges in handling diverse data sources, ensuring standardized access, and facilitating seamless analysis in real-time, ultimately enhancing the precision, speed, and quality of monitoring processes within the medical informatics domain.

Keywords. Medical informatics, ELK Stack, Real-time, Data visualization, Dashboard, Data integration, Healthcare technology, MeDIC

1. Introduction

MeDIC (Medical Data Integration Center) is an essential entity within the Medical Informatics Initiative (MI-I), which is supported by the Federal Ministry of Education and Research in Germany [1]. It aims to leverage the power of digitization in medicine through centres, as exemplified by the one at the University Hospital Cologne. The centres strive to provide seamless medical data flow across different locations, guaranteeing standardized access to pseudonymized and anonymized patients' information for research and credence towards the innovation of IT solutions. Emphasizing the utmost importance of patient data security, MeDIC aligns with strict data protection and ethical standards [2].

Nevertheless, navigating through these highly diversified medical data poses challenges in accessing comprehensive details and identifying errors across various processes. The data collection involves multiple primary systems, engaging various

[1] Corresponding Author: Md. Mostafa Kamal; E-mail: mostafa.kamal@uk-koeln.de.

pipelines for the extract, transform, and load (ETL) process, along with different data streams. The lack of an automated monitoring system leads to a time-consuming manual monitoring process and difficulty in tracking all the numbers.

The challenge of data collection in medical settings has been addressed in several studies. Wu et al. (2022) [3] introduced a digital dashboard solution for a pediatric otolaryngology practice, integrating Electronic Medical Records (EMRs), finance data, and a national database for seamless analysis. Lai et al. (2022) [4] focused on communication mechanisms in the ICU with the development of i-Dashboard for multidisciplinary rounds, utilizing a K8 cluster with 6 VMware servers. Christen et al. (2020) [5] proposed dashboards for emergency medical services to provide easily accessible patient information, monitored through patient bracelets and displayed via custom dashboard prototypes. Sulaiman & Yahaya (2013) [6] tackled dispersed organizational data through data warehousing, integrating and analyzing medical data using Pentaho and Tableau 7.0. Badgeley et al. (2016) [7] developed the Electronic Healthcare Data Visualization (EHDViz) toolkit for clinical dashboards, utilizing R packages to visualize data from EHR, biomedical monitors, and wearable technologies. Franklin et al. (2017) [8] discussed strategic changes in the emergency department to improve care delivery, highlighting the importance of real-time information visualizations with the prototype Throughput Dashboard implemented using Tableau.

Upon reviewing existing literature and research, it is clear that while the researches contribute to medical advancements, the focus is often on isolated components rather than the broader framework. This presents a significant research challenge. The goal is to develop an enhanced data visualization to cover a broader range of scenarios.

As an effort to address this problem, ELK Stack was installed and configured for multi-dimensional data processing and customizable visualization, tailored to meet the diverse needs of MeDICs.

2. Methods

2.1. Identification of Stakeholders

A dedicated effort was made to identify and document the specific needs of diverse user categories within the MeDIC environment. Stakeholders included management personnel, system administrators, data stewards, and programmers. Thorough discussions were conducted through interviews, workshops, and collaborative sessions to capture nuanced perspectives and requirements. This inclusive approach ensured the unique needs of each user category, enhancing their decision-making processes.

2.2. Selection of the Technology Stack

Various data analysis and visualization tools were evaluated to meet the needs of MeDIC. During the comparison, emphasis was placed on factors such as the tool's ability to handle multi-dimensional data processing, real-time search and analytics, customizable visualizations, and the availability of open-source community support.

The ELK Stack, consisting of Elasticsearch, Logstash, and Kibana, emerges as an effective choice for fulfilling the needs of MeDIC. Given the diverse range of multidimensional data, application logs, and event data within MeDIC, Logstash is a very effective choice in ingesting them. Elasticsearch offers real-time search and analytics capabilities, which is very crucial for accessing critical information about MeDIC's resources. Kibana complements these functionalities by enabling customized data visualizations and dashboards tailored to the specific needs of stakeholders. At the same time, its open-source nature fosters ongoing improvements and community-driven support.

2.3. Prototype Design and Implementation

After carefully analyzing MeDIC's IT infrastructure, an ELK stack architecture was installed and configured. Data sources within the MeDIC's infrastructure, including application logs, server logs, and other relevant data streams, were integrated with the ELK stack.

The data was collected and pre-processed by Logstash, including parsing, filtering, and transforming raw log data into structured formats for analysis. The processed data was stored in Elasticsearch, which provided robust indexing and search capabilities. Elasticsearch's indexing mechanisms were optimized for efficient data retrieval and analysis. Kibana was utilized to create visually informative dashboards tailored to different user groups, presenting relevant metrics and insights for informed decision-making. High-level visualization of the solution is presented on Figure 1. User roles and access permissions were configured within Kibana to ensure appropriate dashboard access based on roles and responsibilities.

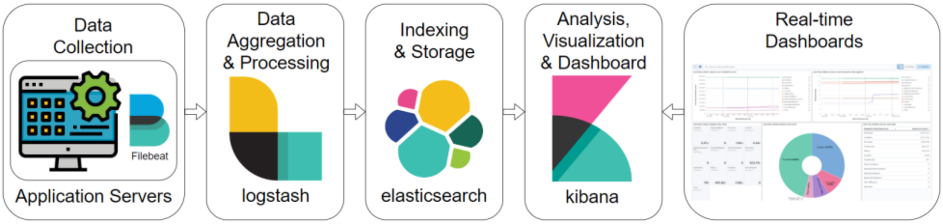

Figure 1. Implemented High-level ELK Stack Architecture in MeDIC

2.4. Testing on MeDIC's Production Firely Resource Statistics

The first-line testing was performed on MeDIC's production environment. An Ubuntu 22.04.4 LTS server used with eight Intel(R) Xeon(R) Gold 6354 CPUs and 32 Gigabytes of memory, to accommodate the excessive data processing load and ELK Stack was installed and configured. The prototype was tested with MeDIC's Firely server from the production environment.

A script was developed to automate the process of retrieving Firely resource statistics at regular intervals. This script fetches the data and stores it in a designated server location. Subsequently, another tool called Filebeat is configured to collect this data and transmit it to the Logstash server.

Upon receiving the data, Logstash employs a specifically designed pipeline to process it. This pipeline parses the unstructured data, creating key-value pairs that are structured and significantly easier to search, analyze, and visualize. Once processed by the Logstash pipeline, the data is stored in Elasticsearch. Finally, Kibana leverages the structured data stored in Elasticsearch to provide users with comprehensive and informative visualizations of the Firely resource statistics.

3. Performance

The real-time status of the production Firely server, serving as the single source of truth (SSOT) for the MeDIC, is displayed in the following dashboard snippet (Figure 2). This dashboard hosts three visualizations to comprehensively present all resources and their changes for users' ease of understanding.

Visualization A presents line charts for each resource, showing incremental changes in the Firely server over the last six months. This provides a clear insight into the impact of data ingestion on the Firely server. For users interested in the current data status, Visualization B displays the latest resource counts, while Visualization C illustrates the latest distribution of resources. All visualizations can be easily downloaded as CSV (comma-separated value) files for quick access and shared across various platforms for detailed reporting or presentations.

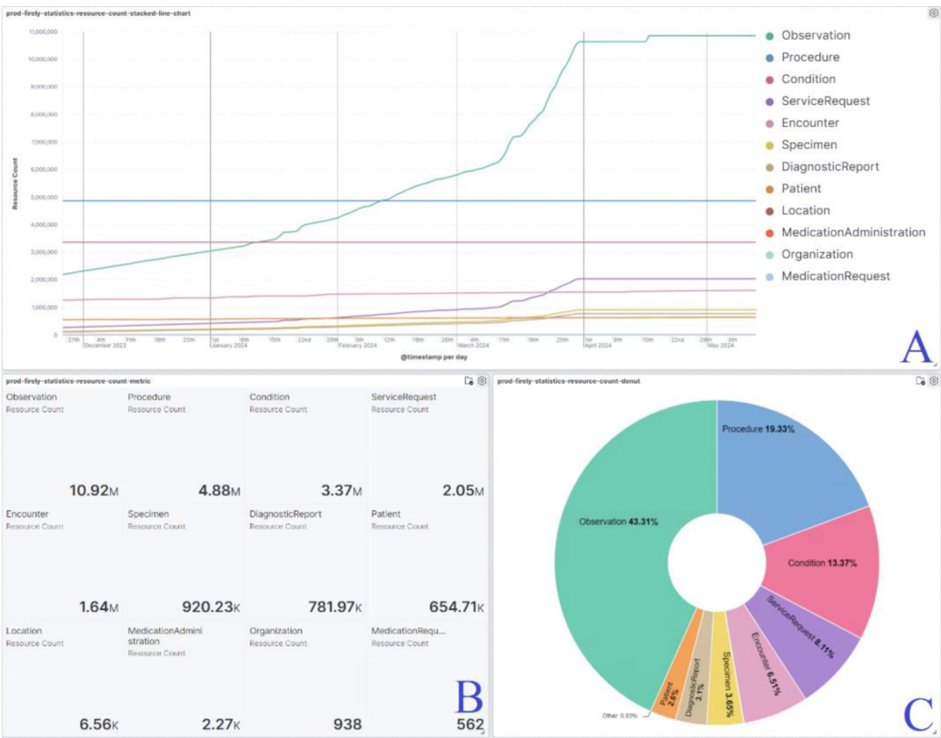

Figure 2. Prototype of MeDIC Dashboard for Production Firely Statistics

4. Discussion and Conclusions

This research addresses resolving several critical challenges in the medical informatics domain by leveraging the ELK Stack. The real-time search and analytics capabilities provided by Elasticsearch facilitate rapid access to the production Firely server. This monitoring dashboard ensures stakeholders have up-to-date information about all resources within Firely, monitors the errors and facilitates quality checks. Continuous data ingestion from primary sources through various routes is reflected in the dashboard, in addition to providing real time view, enables the stakeholders in predicting the future requirements of storage. Analyzing historical values from the dashboard also provides insights into the quality and errors of the ETL routes, thereby enhancing decision-making processes for further improvements for data integration pipelines. In short, better monitored curation pipelines will minimize the data integration errors and improve the effective use of resources.

Continued research and development in this domain will be essential for the advancement of tailored dashboards tailored for MeDIC. The reported solution is built of the open-source components and in the next steps, the possibility of collaborative development within the network of data integration centers will be discussed.

Acknowledgement

Md. Mostafa Kamal is funded by the German Federal Ministry for Education and Research (BMBF, FKZ: 01KX2121). The authors would like to acknowledge their colleagues from MeDIC and Institute for Biomedical Informatics, University of Cologne for their valuable support throughout the research.

References

[1] Schönthaler M, Praus F. [Urolithiasis research-big data and artificial intelligence : How we can use the new structures of the medical informatics initiative of the Federal Ministry of Education and Research]. Urologe A. 2019;58:1298–1303.

[2] Vlahou A, Hallinan D, Apweiler R, et al. Data Sharing Under the General Data Protection Regulation. Hypertension. 2021;77:1029–1035.

[3] Wu SS, Shirley RB, Anne S, et al. Utility of the finance-electronic medical record digital dashboard in pediatric otolaryngology. Am J Otolaryngol. 2022;43:103598.

[4] Lai C-H, Li K-W, Hu F-W, et al. Integration of an Intensive Care Unit Visualization Dashboard (i-Dashboard) as a Platform to Facilitate Multidisciplinary Rounds: Cluster-Randomized Controlled Trial. J Med Internet Res. 2022;24:e35981.

[5] Christen OM, Mösching Y, Müller P, et al. Dashboard Visualization of Information for Emergency Medical Services. Stud Health Technol Inform. 2020;275:27–31.

[6] Sulaiman NS, Yahaya JH. Development of Dashboard Visualization for Cardiovascular Disease based on Star Scheme. Procedia Technology. 2013;11:455–462.

[7] Badgeley MA, Shameer K, Glicksberg BS, et al. EHDViz: clinical dashboard development using open-source technologies. BMJ Open. 2016;6:e010579.

[8] Franklin A, Gantela S, Shifarraw S, et al. Dashboard visualizations: Supporting real-time throughput decision-making. Journal of Biomedical Informatics. 2017;71:211–221.