

---

## Implementation of Big Data Migration on PIS-PK System Database to Mhealth System Database Using ETL and Queue (Ministry of Health Case Study)

**Raindeca Dzulikrom Haqqu, Hustinawati Hustinawati**

Universitas Gunadarma, Indonesia

Email: [raindeca111@gmail.com](mailto:raindeca111@gmail.com), [hustina@staff.gunadarma.ac.id](mailto:hustina@staff.gunadarma.ac.id)

---

### Abstract

This research explores the implementation of big data migration from the PIS-PK (Pusat Informasi Sistem - Pangkalan Data Kesehatan) system database to the Mhealth system database using ETL (Extract, Transform, Load) processes and queue management. The case study focuses on the Ministry of Health, where large-scale health data needs to be transferred between systems to improve efficiency, accessibility, and integration with modern health technologies. The research investigates the challenges involved in migrating extensive data sets, such as maintaining data integrity, handling different data formats, and ensuring smooth data flow during the migration process. By implementing ETL techniques, the study ensures that data is extracted, transformed, and loaded into the Mhealth database while maintaining consistency, quality, and structure. Furthermore, the use of queue systems is explored as a mechanism to handle data processing asynchronously, enabling the smooth transfer of large volumes of data in real-time. The study concludes with a set of best practices for managing big data migration in healthcare systems and highlights the potential benefits of enhancing the Ministry of Health's database architecture for future health data management and analytics. This migration is expected to streamline healthcare data processes, improve decision-making, and support the Ministry's digital health initiatives.

**Keywords:** Big Data, Database Migration, ETL, Queue, PIS-PK, Mhealth

---

*Corresponding:* Raindeca Dzulikrom Haqqu

E-mail: [raindeca111@gmail.com](mailto:raindeca111@gmail.com)



## INTRODUCTION

The PIS-PK system is a system developed by the Ministry of Health of the Republic of Indonesia to support the Healthy Indonesia with a Family Approach Program (PIS-PK) for the equitable distribution of health data (Rusdian & Widiarini, 2020; Tarigan et al., 2025; Ulfa Siregar, 2022). During the implementation of the PIS-PK system, several problems were identified in the field. These problems arose from various stakeholders, including *puskesmas*, health workers (*nakes*), and the community. In the process of this system flow, it was found that the data input from each *puskesmas* is unidirectional. Only *puskesmas* can access public health data, but the public cannot verify their health data or the status of themselves and their family members in the system (Anna et al., 2025; Kurniawan et al., 2021; Lestari et al., 2021).

In response to these problems, the Ministry of Health of the Republic of Indonesia, in collaboration with the World Bank, has developed an innovation called Mhealth. Mhealth is a service that connects healthcare workers with the community, allowing them to verify the results of PIS-PK data collection and improve data quality (Chioma Anthonia Okolo et al., 2024; Onodera & Sengoku, 2018; Tiwari et al., 2023). To realize this innovation, the Ministry of Health partnered with Gunadarma University to develop a new system called the Mhealth System.

To achieve this goal, the Mhealth System needs to be integrated with the PIS-PK System. This integration focuses on connectivity between the existing databases in the PIS-PK System and the new databases in the Mhealth System (Brennan et al., 2020; Mbunge et al., 2025; Noorbergen et al., 2019; Taralunga & Florea, 2021). The main challenge in this migration process is the large amount of data that needs to be transferred, making ordinary data transfer methods impractical. Therefore, a special method is needed to address how to move Big Data from one database to another in different systems.

Ellison et al. (2018) state that the best method for data migration involves two main stages: data modeling and migration simulation. This approach ensures the success of the migration by considering various important aspects, including data integrity, system performance, and consistency after migration. Hussein (2021) further identifies six main stages, which include data assessment, data cleansing, extract-and-load tests, final extract and load, migration validation, and post-migration activities.

Dey et al. (2019) suggest that effectively scheduling operations during migrations using queues can reduce system workload and minimize errors in the migration process. Based on these studies, there are two key points to consider: the extraction, transformation, load (ETL) process and scheduling using queues.

Based on these considerations, the researcher implemented the migration of Big Data from the PIS-PK System database to the Mhealth System database using the ETL method with the help of a message queue. The process is performed asynchronously to prevent server downtime and maximize computing resources by running it during periods of low transaction activity.

The purpose of this research is to analyze and evaluate the implementation of big data migration from the PIS-PK system database to the Mhealth system database using the ETL (Extract, Transform, Load) method and queue, with a case study of the Ministry of Health. This study aims to identify the challenges and solutions encountered during the data migration process and assess the effectiveness and efficiency of the migration in improving the quality of healthcare data management, making it more integrated and real-time.

The benefits of this research are expected to provide practical insights for the Ministry of Health and relevant institutions in performing secure and effective big data migration, as well as contribute to the development of a more modern and data-driven healthcare information system. Furthermore, the results of this research can serve as a reference for other organizations with similar needs for data migration using ETL and queues, and optimize the performance of systems related to digital health services.

## **RESEARCH METHOD**

The flow of this research follows the SDLC (Software Development Life Cycle) which consists of planning, analysis, design, implementation, and testing. At the planning stage, several important points are defined including estimated time and cost. The analysis stage involves examining indicators that need to be known before the design is carried out. The design stage includes schematic and flowchart system components. The implementation stage presents the steps and methods for implementing database migration using ETL and RabbitMQ. The test stage verifies the results of the implementation.

**Table 1. Estimated Processing Time**

	Month
	I II III IV V VI
Planning	✓
Analysis	✓ ✓
Design	✓ ✓
Implementation	✓ ✓
Testing	✓ ✓ ✓

### **System Design**

The system design includes three main components: source data design, ETL design, and RabbitMQ design. The source data consists of three main tables: surveys, survei\_rumah\_tangga, and survei\_individu that use MySQL DBMS. The ETL process is designed to extract data from all three tables, transform it with denormalization techniques, and load it into a new data warehouse named raw\_survei. The RabbitMQ design includes a producer node that retrieves data from raw\_survei in the PIS-PK system and a consumer node that receives data from a messaging server to be fed into a data warehouse in the Mhealth system.

## **RESULTS AND DISCUSSION**

### **ETL Trial Results**

Based on the results of the tests that have been carried out by looking at the progress per unit of time using crontab, it can be stated that the ETL process can run successfully as it should. The ETL process is run at a unit of time per hour of 12 pm or 00:00 WIB to maximize the use of server resources at low transaction loads. The log results show that the Extract process successfully retrieved data from the three main tables in the PIS-PK database (survey, survei\_rumah\_tangga, and survei\_individu) without losing any data.

The Transform stage successfully denormalizes all three tables into one raw\_survei table by combining all the columns of the source tables. The transformation process ensures data format consistency, structural normalization, and referential integrity validation. The Load stage successfully enters the transformed data into the data warehouse using the INSERT ON DUPLICATE KEY UPDATE technique to avoid duplication of data and ensure that newer data overwrites the old data in the event of a conflict.

### **RabbitMQ Test Results**

Based on the results of the tests that have been carried out, the Queue process using RabbitMQ can run as successfully as it should. The implementation of RabbitMQ as a message queue system successfully manages the process of sending data asynchronously from the PIS-PK system to the Mhealth system. The producer node successfully retrieves data from the raw\_survei table on the PIS-PK system, converts it to JSON format, and sends it to the channel on the RabbitMQ messaging server.

The consumer node successfully receives data from the RabbitMQ channel and enters it into a new data warehouse on the Mhealth system. The final result in the form of raw\_survei tables on the Mhealth system has a data volume of 255 GB which is the result of denormalization of three tables on the PIS-PK system. The data in this table contains complete information about health surveys that have been conducted by health workers in the field, including individual, household, and survey results.

**Table 2. General Parameter and Functional Checks**

No Parameters	Success Indicators	Test Results
1 Database Connection Configuration	The program must be able to connect to the existing database on the PIS-PK system and the Mhealth system	Successfully, the function can run well. Functions can be connected to both systems and perform processes on both systems
2 Program ETL	The program must be able to extract, transform, and load the results of 3 tables into one table	Successfully, the ETL process has created a table named raw_survei
3 Message Queue	The program must be able to manage incoming and outgoing messages to prevent server downtime	Successfully, the timing process is set by crontab in units of time per day
4 Virtual Machine	The VM is expected to be able to execute all specified programs and commands	Successfully, all programs run on the VM and work as they should

**Table 3. ETL Parameter and Functional Checks**

No Parameters	Success Indicators	Test Results
1 Individual Collection	Data Function required to obtain a list of individual data on the PIS-PK system	Successfully, this process is performed by the functions created in the ETL program
2 Denormalization Survey Data	Functions are required to be able to perform ETL processes and configure them in the data warehouse	Successfully, the data formed is raw_survei is formed from 3 tables

**Table 4. Queue Parameter and Functional Checks**

No Parameters	Success Indicators	Test Results
1 Queue Receive	The function is required to be able to capture data in the post-ETL data warehouse and function as a queue producer	Successfully, the Queue process was able to capture individual post-ETL results at a specific time unit based on the time on the crontab
2 Queue Send	The function is required to be able to send the data in the queue system to the new data warehouse and function as a consumer queue	Successfully, the captured data can be sent and generate a new data warehouse named raw_survei with a large volume

**Analysis of Test Results**

Based on several test indicators in Table 2, Table 3, and Table 4, it is stated that the results of the audit state that the database migration process was successfully carried out with a 100% success rate with each function provided running well. All parameters tested show results that correspond to the established success indicators.

The database connection configuration successfully connects the program to the PIS-PK and Mhealth databases simultaneously. The ETL program successfully performs the process of extracting, transforming, and loading data from three source tables into one data warehouse.

Message Queue using RabbitMQ successfully manages the data delivery process asynchronously without causing downtime on the production server. The Virtual Machine successfully executes all the programs and commands needed for the migration process.

### **Migration Data Trial Results on Mhealth Application**

Using the migration data that has been carried out, a trial has been carried out to see if the data can be viewed and used in real applications. For this reason, an application is used in the Mhealth system, namely Mhealth Beneficier. The reason for using Beneficier is because the data in `raw_survei` is a collection of individual data that has been surveyed before, and to see the results can be viewed using the application.

The trial was carried out by accessing data from 30 users who had been surveyed as a sample. On the user information page, the results show that all data has been successfully found and accessible within the app. People can see their own health data that was previously only accessible by the Health Center. Health workers can monitor the public health conditions they have surveyed through the application.

The data displayed in the application includes complete biodata information on users, the results of health surveys that have been conducted, family health status, and follow-up recommendations from health workers. All of this data is the result of the migration process that has been carried out from the PIS-PK system to the Mhealth system, proving that the migration process is successful and the data can be used functionally in the application.

The implementation of Big Data migration using ETL and Queue methods shows high effectiveness in handling large volumes of data without disrupting the ongoing system operations. The use of the asynchronous method with RabbitMQ as a message queue system allows the migration process to be carried out in stages according to the server capacity, thus avoiding overload and downtime.

The denormalization technique in the transformation stage has proven effective in combining data from three source tables into one structured data warehouse. This makes it easier to access data and improves query performance because the associated data is already in a single table. The resulting data warehouse volume of 255 GB shows that the system has managed to handle Big Data well.

The use of crontabs for scheduling the migration process at 00:00 WIB proved effective because at that time the system's transaction load was at its lowest point. This strategy maximizes the use of server resources without disrupting users who are accessing the system during operating hours. The 100% success rate achieved shows that the combination of ETL and Queue methods is very effective for Big Data migration in the context of healthcare systems.

The results of the trial on the Mhealth Beneficier application prove that the data from the migration can be used functionally in real applications. People can now access their own health data, while health workers can monitor the health conditions of the people they have surveyed. This shows that the main goal of the Mhealth system to improve transparency and access to health data can be achieved through a successful migration process.

## Nama Author

Implementation of Big Data Migration in The Pis-Pk System Database to The Mhealth System Database Using ETL and Queue (Case Study of The Ministry of Health)

---

## CONCLUSION

This study successfully migrated Big Data from the PIS-PK system database to the Mhealth system database using the ETL (Extract, Transform, Load) method integrated with the RabbitMQ queue system, achieving a 100% success rate without data loss or corruption. Asynchronous processing via message queues ensured system stability and prevented downtime, resulting in an independent 255 GB data warehouse for the Mhealth system that supports public health data access and health worker monitoring. The denormalization technique in the transformation stage effectively merged three source tables into a single structured warehouse, enhancing data accessibility and query performance. For future research, it is recommended to optimize the data transformation process for larger volumes and faster processing times, implement real-time monitoring to detect issues early, and conduct periodic evaluations of migrated data integrity to maintain long-term consistency and accuracy.

## REFERENCES

- Anna, D., Harnani, Y., Marlina, H., Dewi, O., & Ennimay, E. (2025). Integrasi Program Indonesia Sehat dengan Pendekatan Keluarga (PIS-PK) dalam Manajemen Puskesmas di Puskesmas Sidomulyo Kota Pekanbaru Riau Tahun 2023. *Jurnal Kesehatan Komunitas (Journal of Community Health)*, 10(3). <https://doi.org/10.25311/keskom.vol10.iss3.1795>
- Brennan, L., Kessie, T., & Caulfield, B. (2020). Patient experiences of rehabilitation and the potential for an mhealth system with biofeedback after breast cancer surgery: Qualitative study. *JMIR MHealth and UHealth*, 8(7). <https://doi.org/10.2196/19721>
- Chioma Anthonia Okolo, Oloruntoba Babawarun, & Tolulope Oyinlola Olorunsogo. (2024). Mobile Health (MHEALTH) Innovations For Public Health Feedback: A Global Perspective. *International Medical Science Research Journal*, 4(3). <https://doi.org/10.51594/imsrj.v4i3.915>
- Dey, S. K., & Islam, M. M. (2019). Techniques of workload on data migration in cloud. *Engineering International*, 9(1), 41–50. <https://doi.org/10.18034/ei.v9i1.529>
- Ellison, C. J., et al. (2018). Migration-aware design methodology to minimize the transaction overhead of live database migration. *The Journal of Supercomputing*, 74(3), 1143–1176. <https://doi.org/10.1007/s11227-017-2201-2>
- Hussein, S. H. (2021). The database migration steps to the cloud. *Academic Journal of Nawroz University*, 10(1), 156–165. <https://doi.org/10.25007/ajnu.v10n1a1001>
- Kurniawan, R., Nur, R. J. B., Yuliyanti, S., Nuridzin, D. Z., & Kartinah, N. T. (2021). Healthy family index assessment through community-based health information system approach. *International Journal of Public Health Science*, 10(2). <https://doi.org/10.11591/ijphs.v10i2.20751>
- Lestari, R. A., Suryoputro, A., & Kartini, A. (2021). Analisis pelaksanaan Program Indonesia Sehat Dengan Pendekatan Keluarga (PIS-PK) dalam capaian indeks keluarga sehat di Kabupaten Brebes tahun 2020 studi pada Puskesmas Kluwut Kabupaten Brebes. *Visikes Jurnal Kesehatan*, 20(1).
- Mbunge, E., Jack, C. L., Sibiya, M. N., & Batani, J. (2025). Review of implementation barriers and strategic approaches for improving mHealth systems utilization in Africa: Lessons learnt from South Africa and Kenya. In *Telematics and Informatics Reports* (Vol. 19). <https://doi.org/10.1016/j.teler.2025.100228>
- Noorbergen, T. J., Adam, M. T. P., Attia, J. R., Cornforth, D. J., & Minichiello, M. (2019). Exploring the design of mhealth systems for health behavior change using mobile

- biosensors. *Communications of the Association for Information Systems*, 44(1). <https://doi.org/10.17705/1CAIS.04444>
- Onodera, R., & Sengoku, S. (2018). Innovation process of mHealth: An overview of FDA-approved mobile medical applications. *International Journal of Medical Informatics*, 118. <https://doi.org/10.1016/j.ijmedinf.2018.07.004>
- Rusdian, E. R., & Widiarini, R. (2020). Evaluasi Program Indonesia Sehat Dengan Pendekatan Keluarga (Pis-Pk): Studi Kasus Di Tingkat Puskesmas. *Jurnal Kebijakan Kesehatan Indonesia : JKKI*, 9(01).
- Taralunga, D. D., & Florea, B. C. (2021). A blockchain-enabled framework for mhealth systems. *Sensors*, 21(8). <https://doi.org/10.3390/s21082828>
- Tarigan, M. F., Afrimadani, S. N., Kurniawan, B., Hertiana, S. N., & Aditya, B. (2025). Developing Of A Website-Based Information System For Dasa Wisma To Realize The Healthy Indonesia Program With Family Approach (PIS-PK). *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 10(1). <https://doi.org/10.29100/jipi.v10i1.5924>
- Tiwari, P., Shailaja, U., Bhattacharya, N. S., & Manchikanti, P. (2023). A Study on mHealth Innovations through the Lens of Patent Analysis. *Journal of Intellectual Property Rights*, 28(4). <https://doi.org/10.56042/jipr.v28i4.2023>
- Ulfa Siregar, M. (2022). Evaluasi program Indonesia sehat dengan pendekatan keluarga (PIS-PK). *Nautical : Jurnal Ilmiah Multidisiplin Indonesia*, 1(9). <https://doi.org/10.55904/nautical.v1i9.405>