International Journal of Management, Knowledge and Learning | ISSN 2232-5697

The Analysis of Big Data Architecture for Healthcare Service: Case Study of a Public Hospital

Wasinee Noonpakdee
jookjik@gmail.com
Thammasat University, Thailand

Purpose: Big data is one of the promising technologies which helps the healthcare sector improve patient experience, reduce the cost of treatment, predict epidemics, and provide competitive healthcare services. This research aims to study and analyse big data architecture for a public hospital in Thailand.

Study design/methodology/approach: This research employs a qualitative approach. For data collection and analysis, document analysis is applied as a qualitative method to obtain empirical data. In addition, in-depth interviews of the IT manager, doctor, and staff are conducted to confirm the big data architecture and examine issues in a case study public hospital.

Findings: According to the study, big data architecture for the public hospital comprises data acquisition, data integration, data analytics, data visualisation, a health data centre on a cloud, and data governance.

Originality/value: The architecture components and analysis can be applied as a guideline for big data processing and analytics to create a solution for implementing big data in healthcare effectively.

Introduction

Big data is now making a considerable impact on the global economy by changing how a company can do business and how decisions can be made to increase profits and productivity and to achieve competitive advantages (Lakoju & Serrano, 2017; Liang et al., 2018; Mneney & Belle, 2016; Motau & Kalema, 2016; Shamim, Zeng, Khan, & Zia, 2020; Tardío, Mate, & Trujillo, 2015). Big data has been extensively applied, and several innovative platforms and technologies are continuously emerging to improve big data analytics and processing (Basukie, Wang, & Li, 2020; Oussous, Benjelloun, Ait Lahcen, & Belfkih, 2017).

Healthcare is one of the most promising sectors digitally transformed that obtained valuable information from big data. Healthcare can utilise big data analytics to improve the quality and accuracy of clinical decisions, shorten the time of diagnostic tests, reduce costs of treatment, and create highly competitive healthcare services (Chen, Lin, & Wu, 2020; Pramanik et al., 2020; Wang, Kung, & Byrd, 2018).

In Thailand, the market value of big data is expected to reach 13.2 billion Baht in 2022, which is projected to grow 20% a year. Big Data is replacing traditional data as a powerful tool for business planning. Big data can help develop sales and marketing strategies and improve products and services by analysing customer behaviour. For production and operations, big data can help increase productivity, reduce costs, manage financial and human resources (EIC, 2017).

To implement big data, architectures, frameworks, and processes of data analytics have been applied in many information systems research studies (Campos, Sharma, Gabiria, Jantunen, & Baglee, 2017; Fahmideh & Beydoun, 2019; Li, Zhang, & Wang, 2013; Sellami, Mezni, & Hacid, 2020). This paper aims to study and analyse big data architecture for a public hospital in Thailand. The architecture components and analysis can be applied as a guideline for big data processing and analytics to create a solution for implementing big data in healthcare effectively.

BY-SA

Big Data

Big Data in Healthcare

The term big data describes the vast volume of data that is difficult to manage by traditional tools. Big data characteristic comprises three Vs; Volume, Velocity and Variety. Volume represents a large amount of data. Velocity indicates the high speed at which data is generated. Variety means the diversity of a dataset which is structured, semi-structured and unstructured data (Alharthi, Krotov, & Bowman, 2017; Gandomi & Haider, 2015).

In healthcare, big data gained a lot of attention from both academia and industry with several analytical techniques and applications (Mehta & Pandit, 2018; Mehta, Pandit, & Shukla, 2019; Ojha & Mathur, 2016; Panda, Ali, & Panda, 2017; Sanchez-Pinto, Luo, & Churpek, 2018). Medical, pharmaceuticals, and supplies sectors are now leaders in data-driven predictive analytics, moving toward treatment based on genetic profiles. In the past, medications were dispensed on a trial-and-error basis. Doctors would consequently analyse a patient's symptoms and prescribe drugs with periodic follow-up to adapt the dosage and medication. Due to a better electronic health records system and new medical developments in genetics, big data analytics can help forecast the probability of illness effectively based on genetic factors. With medications development and prescription for individual genetic profiles, personalised medicine will also improve treatment outcomes (EIC, 2017).

Big Data analytics can help early detection of disease, facilitate decision making, and predict the possibility of illness. Clinical data integration provides various applications, such as clinical decision support systems, disease surveillance, and individual healthcare management, to reduce healthcare costs, improve health-process efficiency, and enhance healthcare quality (Mehta & Pandit, 2018). Medical big data is required to use properly by gathering accurate data and analysis, in which such information may be obtained from various applications, genomics, health records, etc. In addition, big data analytics is instrumental in individual care to predictive models for large populations (Kraus, Schiavone, Pluzhnikova, & Invernizzi, 2021). The challenges being faced by the healthcare industry in the application of Big Data Analytics are Confidentiality and Data Security, Access Control, Interoperability, Data and Analytics Reliability, and Data Provenance (Imran, Mahmood, Morshed, & Sellis, 2021).

eHealth Strategy

Ministry of Public Health (MoPH), Thailand, has driven the digital health system through the eHealth Strategy (*eHealth Strategy, Ministry of Public Health (2017 – 2026)*, 2017). The eHealth Strategy focuses on sustainable long-term development aligned with the Thailand digital economy plan ("Thailand Digital Economic and Society Development Plan,"). The digital landscape of Thailand for 20 years has been applied to the direction of the development of healthcare in Thailand to Health 4.0, which uses digital technology to improve the delivery of healthcare services.

The eHealth Strategy is based on the WHO eHealth concept (2016) and the International Telecommunication Union (ITU), which aims to promote public health by providing information and communication technology to help people get healthy and to provide affordable and safe health services. The Strategy focuses on eight components: 1) eHealth foundations 2) Legal frameworks for eHealth 3) Telehealth 4) Electronic health records 5) Use of eLearning in health sciences 6) mHealth 7) Social media, and 8) Big data.

Considering big data, Health Data Center (HDC) database system is implemented with Big Data technology which provides an analytical tool for quick results (Axis Project). With big data analytics, diseases can be controlled, and treatments can be planned effectively.

Big Data Architecture

Big data architecture can be considered an outline for a big data implementation and solution based on the business requirements of an enterprise. Big data is regarded as the central concept when defining new digital architecture projects. Consequently, the technology that comprises these solutions could play a significant role in business information architecture. The process of data flow includes data source, data collection, ETL, Business Analytics, and Service (Hadj Sassi, Jedidi, & Fourati, 2019). The architecture components of the big data ecosystem are defined in Demchenko et al. (2014). The lifecycle in a big data system comprises data source, data collection and registration, data filter/ enrich classification, data analytics modelling protection, data delivery/visualisation, and then consumer data analytics application.

The architecture and key technologies of big data construction are studied in (Li et al., 2013). The Hadoop multidimensional analytics platform architecture includes data acquisition, data redundancy, dimension define, and parallel analysis. The parallel analysis comprises core modules, Hadoop cluster analysis, and report centre. The architecture of big data is utilised in many applications, for example, the architecture design for manufacturing systems (Fahmideh & Beydoun, 2019) and asset management (Campos et al., 2017).

For healthcare, big data analytics architecture in health care is studied in Wang et al. (2018). The architecture is divided into the data layer, data aggregation layer, analytics layer, information exploration layer, and data governance. The data governance layer includes master data management, data lifecycle management, and data security and privacy management.

The big data workflow of biomedical image processing comprises data management and analytics, and storage for healthcare applications. The real-time big data architecture for intensive care is proposed using the main open source big data solution, Apache Hadoop (Gonçalves, Portela, Santos, & Rua, 2017). The architecture consists of data acquisition, interface, inference, knowledge management, and data management. The architecture approach is proposed based on a prediction algorithm for the integrated processing of real-time data and historic learning data. The architecture model comprises the analytics layer and ingestion & distribution layer.

Research Methodology

The research process comprises three phases: literature review, data collection and analysis, and conclusion. The literature review includes the study of big data in healthcare, eHealth strategy, and big data architecture. For data collection and analysis, document analysis is applied as a qualitative method to obtain empirical data as part of an unobtrusive and nonreactive process (Bowen, 2009). Then, the in-depth interviews of the IT manager, doctor, and staff are conducted to confirm the big data architecture and to examine issues in a case study public hospital. Finally, the conclusion is presented to summarise the results.

Results

A case study public hospital is operated under the Ministry of Public Health. According to the literature review, document analysis, and in-depth interview, big data architecture for this public hospital comprises Data Acquisition, Data Integration, Data Analytics, Data visualisation, HDC on Cloud, and Data Governance, as illustrated in Figure 1.

Data Acquisition

In this layer, data is acquired from different data sources, including structured data and unstructured data. Different data types result in the selection of various tools. Data can come from internal and external data sources. The data sources of this hospital are HIS (Hospital Information System), CIS (Clinical Information System), PAS (Patient Administration System), EMR (electronic medical record), biomedical image or Picture Archiving and Communication System (PACS), R&D lab, omics data, sensor/wearable device, social media/mobile application, Utilization Review.

The hospital employs HIS (Hospital Information System) to manage the healthcare operation and services. CIS (Clinical Information System) comprises Clinician Access, Care Plan, Order Management, and Oral Health. PAS (Patient Administration System) is used for patient appointments and registration of treatment. EMR (electronic medical record) is applied to manage patient's records using electronic format instead of a paper OPD (Out Patient Department) card. The hospital also applies electronic prescriptions with a notification system via mobile device. The e-Office system is used for administrative tasks, and the hospital tried to go paperless. An automatic queue calling system, payment system, and paper reduction system have been implemented, reducing unnecessary processes. In addition, the utilisation review system is applied to audit the resource utilisation of the hospital.

Data can be in a structured and unstructured format, for example, behavioural data such as eating or exercise, medical image, handwritten medical records, audio recordings, text messages for details of treatment, medical and public health data from a survey, temperature data, glucose level, blood pressure level, etc.

Data Integration

For data integration, ETL (extract, transform, load) or ELT (extract, load, transform) are performed. ETL is the traditional approach for data analytics and data warehouse. ELT offers an innovative alternative to ETL and can be applied to large amounts of data. The case study hospital applies both ETL and ELT, mainly the ETL for the traditional warehouse system.

Data storage approaches can be implemented in batch processes or real-time based on data governance policies and compliance. This hospital's data storage tools are traditional database management systems (DBMS), Cloudera Hadoop, and NoSQL Databases.

Data Analytics

This layer processes data and performs a suitable analysis. Data analytics is categorised into three main types – Descriptive, Predictive and Prescriptive analytics. The descriptive analytics provides a summary of historical healthcare data to explain the causes of various dimensions and generate useful information such as general information, health status, health service access, disease prevention activities, and contagious diseases.

Predictive analytics is a form of using information that is more complicated by being "forecasting" or "predicting" what is happening using historical data together with various mathematical models or with Data Mining and analysis of opportunities and risks that will occur in the future.

Prescriptive analytics is a form of more complicated data analysis that predicts what will happen and advises on various options and results with advantages and disadvantages. It is a synthesis of data to analyse trends and propose alternatives for making decisions to determine the tendency of future diseases based on people's behaviour for timely planning to solve public

health problems. This aims to analyse big data from the medical and health data to maximise benefits to the healthcare system.

The case study hospital has fully been applied descriptive analytics and some predictive analytics. For prescriptive analytics, the application is still in the initial phase. The hospital is now cooperated with private companies to prepare for the applications of prescriptive analytics.

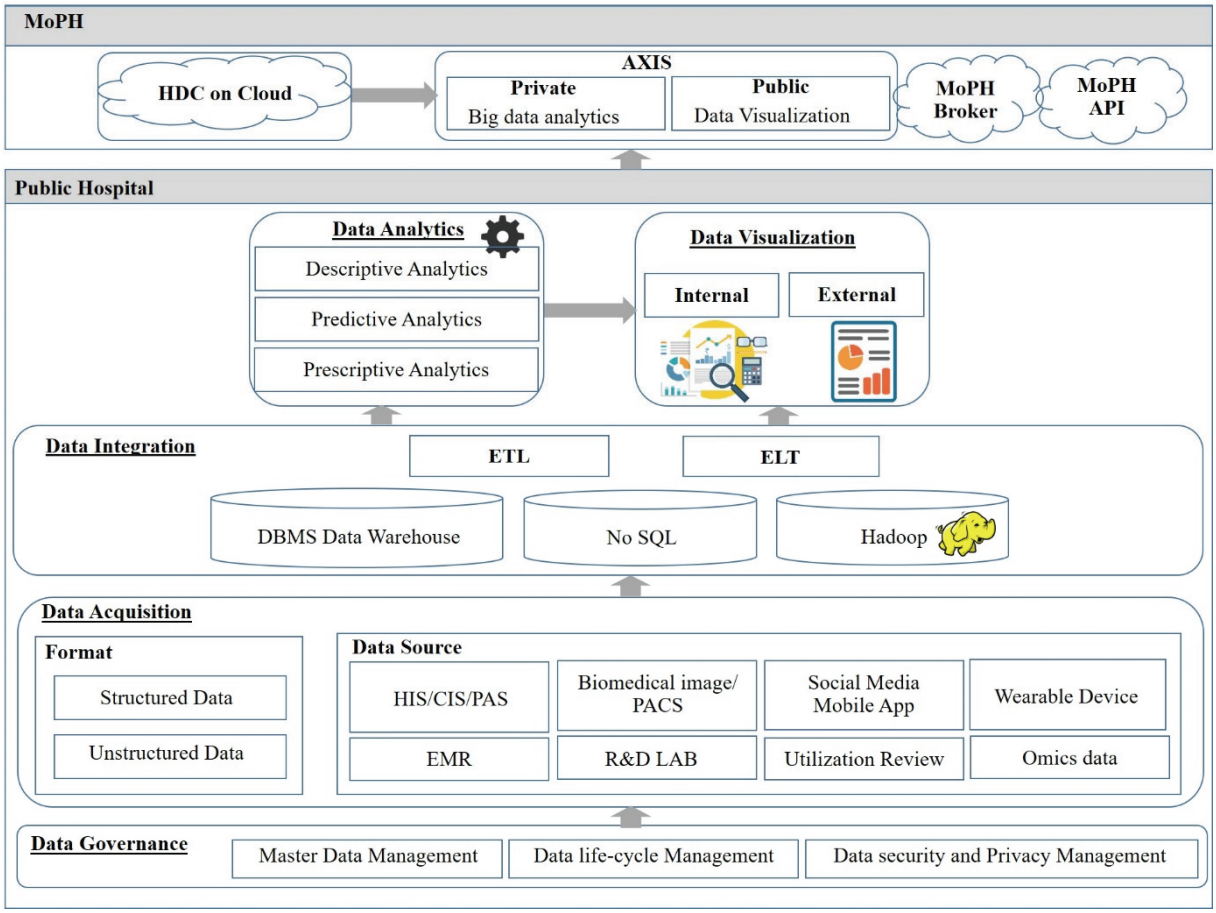

Figure 1: Big Data Architecture for Public Hospital

Data Visualization

This layer is responsible for reports and visualisations to provide meaningful corporate insights resulting from the data analytics layer to attain better and faster decision making.

For healthcare, the most significant output could be real-time monitoring, such as real-time data navigation, notifications and alerts, and key performance indicators (KPIs). Data is analysed and displayed in dashboards in real-time for monitoring patients' health and preventing accidental events (Wang et al., 2018).

Data visualisation of the hospital is divided into two parts internal outputs and outputs. Internal outcomes can be utilised by some staff in the hospital, such as doctors, researchers, or data scientists. In addition, external outcomes are provided for people interested in the healthcare aspect to see some reports and dashboards.

HDC on Cloud

HDC (Health Data Center) on cloud is a data processing system provided by the Ministry of Public Health (MoPH), Thailand. The system contains the standard files that import data from

service units within each province. According to the service plan, the data can be displayed at the district level, department level, or service unit level. In addition, users can show (drill-down) these levels until the personal level according to the authorisation. The primary data groups in the system consist of standard data files, data transformation, and data summary.

The standard data files contain individual records. The data transformation comprises tables of relational data or data groups such as in-patient data, out-patient data, and chronic disease registration data. It is connected with service unit information, area, and date. The data summary is the data that the data group processes to obtain the results according to each data group's scope and conditions or definitions for the HDC system report. The analytical tools for quick results (Axis Project) are applied for private and public data analytics and visualisations. Moreover, a private service broker and application programming interface (API) is also provided by MoPH.

Data Governance

Data governance contains three components which are master data management, data lifecycle management, and data security and privacy management. Master data management is considered the processes, policies, standards, governance, and tools to manage data. Data lifecycle management is the method of managing organisation information throughout its lifecycle. Finally, data security and privacy management include configuration assessment, monitoring, auditing, and protection (Wang et al., 2018).

It is necessary to apply data governance for sensitive data to protect patient privacy and prevent data breaches. As data system protection is crucial, especially for sensitive clinical data, high cybersecurity standards are needed. This hospital has already formulated a data security and privacy policy. However, there is no clear data management policy, so the hospital could not fully gather data from various departments in the data acquisition and integration layer, making it difficult to analyse data effectively.

Conclusion

This paper presents big data architecture for healthcare services in a case study public hospital in Thailand. On account of big data architecture, system developers can easily implement the system because each layer's functions are separately described. Furthermore, developers can also be able to select the most appropriate technology for each layer effectively.

According to the study, big data architecture for public hospitals comprises data acquisition, data integration, data analytics, data visualisation, a health data centre on a cloud, and data governance. In addition, issues of big data implementation regarding the architecture are addressed, especially data management policy. Consequently, future research could be considering this study focusing on data governance for big data implementation in a public hospital.

References

Alharthi, A., Krotov, V., & Bowman, M. (2017). Addressing barriers to big data. *Business Horizons*. doi:http://doi.org/10.1016/j.bushor.2017.01.002

Basukie, J., Wang, Y., & Li, S. (2020). Big data governance and algorithmic management in sharing economy platforms: A case of ridesharing in emerging markets. *Technological Forecasting and Social Change, 161*, 120310. doi:https://doi.org/10.1016/j.techfore.2020.120310

Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal, 9*(2), 27-40.

Campos, J., Sharma, P., Gabiria, U. G., Jantunen, E., & Baglee, D. (2017). A Big Data Analytical Architecture for the Asset Management. *Procedia CIRP, 64*, 369-374. doi:https://doi.org/10.1016/j.procir.2017.03.019

Chen, P.-T., Lin, C.-L., & Wu, W.-N. (2020). Big data management in healthcare: Adoption challenges and implications. *International Journal of Information Management, 53*, 102078. doi:https://doi.org/10.1016/j.ijinfomgt.2020.102078

Demchenko, Y., Laat, C. d., & Membrey, P. (2014, 19-23 May 2014). *Defining architecture components of the Big Data Ecosystem.* Paper presented at the 2014 International Conference on Collaboration Technologies and Systems (CTS).

eHealth Strategy, Ministry of Public Health (2017 – 2026). (2017).

EIC. (2017). *Decisive Edge: Win Big with Big Data*.

Fahmideh, M., & Beydoun, G. (2019). Big data analytics architecture design—An application in manufacturing systems. *Computers & Industrial Engineering, 128*, 948-963. doi:https://doi.org/10.1016/j.cie.2018.08.004

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137-144. doi:https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gonçalves, A., Portela, F., Santos, M. F., & Rua, F. (2017). Towards of a Real-time Big Data Architecture to Intensive Care. *Procedia Computer Science, 113*, 585-590. doi:https://doi.org/10.1016/j.procs.2017.08.294

Hadj Sassi, M. S., Jedidi, F. G., & Fourati, L. C. (2019). A New Architecture for Cognitive Internet of Things and Big Data. *Procedia Computer Science, 159*, 534-543. doi:https://doi.org/10.1016/j.procs.2019.09.208

Imran, S., Mahmood, T., Morshed, A., & Sellis, T. (2021). Big data analytics in healthcare − A systematic literature review and roadmap for practical implementation. *IEEE/CAA Journal of Automatica Sinica, 8*(1), 1-22. doi:10.1109/JAS.2020.1003384

Kraus, S., Schiavone, F., Pluzhnikova, A., & Invernizzi, A. C. (2021). Digital transformation in healthcare: Analysing the current state-of-research. *Journal of Business Research, 123*, 557-567. doi:https://doi.org/10.1016/j.jbusres.2020.10.030

Lakoju, M., & Serrano, A. (2017, 11-14 Dec. 2017). *Saving costs with a big data strategy framework.* Paper presented at the 2017 IEEE International Conference on Big Data (Big Data).

Li, X., Zhang, F., & Wang, Y. (2013, 21-22 Dec. 2013). *Research on Big Data Architecture, Key Technologies and Its Measures.* Paper presented at the 2013 IEEE 11th International Conference on Dependable, Autonomic and Secure Computing.

Liang, F., Yu, W., An, D., Yang, Q., Fu, X., & Zhao, W. (2018). *A Survey on Big Data Market: Pricing, Trading and Protection*.

Mehta, N., & Pandit, A. (2018). Concurrence of big data analytics and healthcare: A systematic review. *International Journal of Medical Informatics, 114*, 57-65. doi:https://doi.org/10.1016/j.ijmedinf.2018.03.013

Mehta, N., Pandit, A., & Shukla, S. (2019). Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study. *Journal of Biomedical Informatics, 100*, 103311. doi:https://doi.org/10.1016/j.jbi.2019.103311

Mneney, J., & Belle, J. P. V. (2016, 14-15 Jan. 2016). *Big Data capabilities and readiness of South African retail organisations.* Paper presented at the 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence).

Motau, M., & Kalema, B. M. (2016, 3-6 Aug. 2016). *Big Data Analytics readiness: A South African public sector perspective.* Paper presented at the 2016 IEEE International Conference on Emerging Technologies and Innovative Business Practices for the Transformation of Societies (EmergiTech).

Ojha, M., & Mathur, K. (2016, 15-16 March 2016). *Proposed application of big data analytics in healthcare at Maharaja Yeshwantrao Hospital.* Paper presented at the 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC).

Oussous, A., Benjelloun, F.-Z., Ait Lahcen, A., & Belfkih, S. (2017). Big Data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*. doi:https://doi.org/10.1016/j.jksuci.2017.06.001

Panda, M., Ali, S. M., & Panda, S. K. (2017, 23-25 March 2017). *Big data in health care: A mobile based solution.* Paper presented at the 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC).

Pramanik, M. I., Lau, R. Y. K., Azad, M. A. K., Hossain, M. S., Chowdhury, M. K. H., & Karmaker, B. K. (2020). Healthcare informatics and analytics in big data. *Expert Systems with Applications, 152*, 113388. doi:https://doi.org/10.1016/j.eswa.2020.113388

Sanchez-Pinto, L. N., Luo, Y., & Churpek, M. M. (2018). Big Data and Data Science in Critical Care. *Chest, 154*(5), 1239-1248. doi:https://doi.org/10.1016/j.chest.2018.04.037

Sellami, M., Mezni, H., & Hacid, M. S. (2020). On the use of big data frameworks for big service composition. *Journal of Network and Computer Applications, 166*, 102732. doi:https://doi.org/10.1016/j.jnca.2020.102732

Shamim, S., Zeng, J., Khan, Z., & Zia, N. U. (2020). Big data analytics capability and decision making performance in emerging market firms: The role of contractual and relational governance mechanisms. *Technological Forecasting and Social Change, 161*, 120315. doi:https://doi.org/10.1016/j.techfore.2020.120315

Tardío, R., Mate, A., & Trujillo, J. (2015, 29 Oct.-1 Nov. 2015). *An iterative methodology for big data management, analysis and visualisation.* Paper presented at the 2015 IEEE International Conference on Big Data (Big Data).

Thailand Digital Economy and Society Development Plan. Retrieved from https://www.itu.int/en/ITU-D/Regional-Presence/AsiaPacific/Documents/Events/2016/Apr-Digital2016/S2_Present_Pansak_Siriruchatapong.pdf

Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organisations. *Technological Forecasting and Social Change, 126*, 3-13. doi:https://doi.org/10.1016/j.techfore.2015.12.019