Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 10, October - 2025

Performance Optimisation in Big Data-as-a-Service (BDaaS) Platforms in Healthcare Systems

Chitiz Tayal

Abstract

Cloud computing is a collection of networks including servers, protocols, databases and storage into a third-party warehouse from where all information can be accessed by all users. Cloud computing is of various types. One of them is Big Data as a Service (BDaaS). This is similar to Software as a Service (SaaS). The BDaaS allows organisations to analyse and process the large volume of datasets stored in their databases. This service and the concerned architecture also have a significant role in the healthcare sector. Some of the renowned BDaaS platforms that are used in the healthcare sector are AWS Health Lake, Google Cloud Healthcare API, or Azure Health Data Services and so on. One of the most fundamental and useful architecture that is used to model the BDaaS platforms is the zero trust architecture whose principal job is to ensure the safety and security of the cloud-based databases.

***Keyword*: cloud computing, third-party warehouse, Big Data as a Service (BDaaS), Software as a Service (SaaS), AWS Health Lake, Google Cloud Healthcare API, or Azure Health Data Services, healthcare sector, cloud-based databases**

I. INTRODUCTION

Scope and Background

A zestful environment of cloud computing has undergone a broad change of arc from being an outlying infrastructure to a multi-tenant one, in which contradicting entities exist together on the basis of a shared environment resulting in inviting multiple threats to the security due to minimal visibility across multiple cloud platforms. In response to such challenges, Big Data as a service (BDaaS) come into action by handling vast amounts of data and deriving meaningful insights also aiding the security and privacy frameworks. BDaaS provides management and analysis of huge chunks of data in data warehouses [1]. In the healthcare department, there lies challenges to handle sensitive and vast data that need to be addressed.

Problem Statement

Legacy service systems are shadowed by the cloud services in order to handle large quantities of data. Traditional security frameworks are incapable of providing integral access facility leading the organisations to vulnerability of the escalating attacks and data exfiltration that is unauthorised. The research suggests optimisation of traditional processes by Big Data service solutions and challenges faced by BDaaS platforms that are needed to put light on.

traditional processes by Big Data service solutions and challenges faced by BDaaS platforms that are needed to put light on.

Objective of the Research

Big Data as a Service platform has been a boon for organisation by providing an efficient management and utilisation of data yet there are few vulnerabilities that require objectification which will result to a key to improvise the dependencies faced by the Cloud Service platforms.

This research aims forward to the understanding of how the Big Data solutions optimise problems faced by healthcare systems in handling vast data and also pointing out the dependencies faced by the BDaaS platforms. This also involves a brief study of the performance optimisation model that is Zero Trust Reference Model stated along with evaluation and analysis.

II. LITERATURE REVIEW

Overview of Big Data as a Service platform with respect to performance optimisation in HealthCare Systems

BDaaS in cloud architecture has been instrumental in generating efficient solutions to the legacy of outdated strategies to handle enormous amounts of sensitive data generated from Healthcare Systems. Some of the key strategies of BaaS are addressed below:

- Data Integrity and Centralisation: BDaaS is functional in collecting data from different sources like from Medical Imaging, Electronic Health Records, IoT devices and storing at a single place offered by BDaaS platforms making it easy to manage and analyse data. As a result, it leads to swift and clever decision-making.
- Scalability: BDaaS platforms offer the usage of flexible resources and come up with an increase in storage and moderation of performance according to the need.
- Enhanced Data Transparency: BDaaS services contribute to an effective data management and provide transparent monitoring [2]. The identification of probable health issues, epidemic and genomic sequencing can be predicted at early stages.
- Cost Effectiveness: BDaaS platforms help to minimize the investments and avoid the wastage of the available resources as it allows the purchase of resources according to the requirement. They enforce efficient utilization of the clinical operational system to manage the available healthcare resources.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

Struggles faced during performance optimisation in BDaaS platforms

Data dependency management is one of the critical challenges for the cloud-based big-data pipelines [3]. The reason for the arising of challenges is due to the different data processing demands and multiple processing stages that leads to complexity in data consistency. Some key challenges can be enlisted as follows:

- The transmission of large volumes of data across several networks data is transmitted across several networks resulting in data manipulation and loss of sensitive data especially in cases where encryption and protocols are not maintained [3].
- A centralized system of BDaaS platforms leads to a complex data access system resulting in improper access control which might lead to unauthorised access to sensitive information. Moreover, critical situations like medication error, patient safety, outrun of medicines cannot be addressed with attention.
- When institutes rely on a single vendor, it leads to an increase in expenditure.
- The data transmission needs time and effort if it is transmitted between two or more providers. This maintains integrity and consistency.
- There are fewer professional experts in healthcare units who can build a good, scalable BaaS architecture which has a distributed data pipeline management.

Foundation of Zero-Trust Architecture (ZTA)

The use of Internet of Things (IoT) in different sectors has led to different struggles because the traditional security architecture requires protection for data integrity and confidentiality. The use of Zero-Trust Architecture helps to reduce those struggles. Zero-Trust Architecture is consists of:

- Micro-Segmentation: In this scenario, the PEP is installed near the data or resources for the prevention of unauthorised access [4]. The lateral movement of the intruder is prevented by such an approach. The installation of virtual firewalls can also lead to micro-segmentation. Along with this, ZTA can be used to detect potential security threats [5]
- Identity Governance: The identity of user and devices are the primary aspect here this policy step takes into consideration the validation of identities, contextual data to predict access risks.
- Network Infrastructures and Performance Defined Perimeters: This ZTA implies an overview of the network is created in order to control accessibility to the data or else data cannot be accessed directly hence, building a software defined perimeter.

Though ZTA along with their logical components has been defined by many institutes, the strategy to implement required for critical infrastructure is still unclear. This is because CIs have a vast range of technologies and endpoints like IoT devices, CPS and traditional network endpoints. As a result, it is important for the identification of suitable techniques for realisation of ZTA tenets in a CI [4]. This aims towards obtaining a favourable overview of the current state and pointing out the weakness and gap of knowledge that lies in the model which needs future research to address such problems.

Existing structure for performance optimisation in the BDaaS platforms and their gaps

Big Data as a Service (BDaaS) platforms integrates data sources, application and processing of data [6]. They promote scalability and high performance of Healthcare Systems along with efficient data analysis. BDaaS platforms encourage data governance by complying with strict regulations like PCI-DSS, GDPR, HIPPA, CCPA and PPDL. This helps healthcare organisations maintain transparency, accuracy and security, avoiding legal risks and promotes trust among stakeholders [2]. Even so, BaaS platforms face limitations in sustaining privacy, security and regular compliances. In spite of these challenges, the well-defined structures of BaaS help organisations to turn regular demand into opportunities. By ensuring data protection, the healthcare institutes can have competitive advantage and it fosters innovation. BDaaS assists organisations develop effective strategies for business development [7].

III. PROPOSED STRUCTURE AND METHODS USED FOR PERFORMANCE OPTIMISATION USING ZTA BASED BDAAS

Zero trust reference model for BDaaS

BDaaS is the short form for Big Data as a Service. BDaaS are the services that are provided to a large number of stakeholders who belong to the energy domain [8]. The Zero trust reference model is suggested for the BDaaS because this model consists of three fundamental components: verification of identity, protection of data and continuous monitoring to reduce the security threats. By implementing this model, every system which configures the BDaaS has to be authenticated first to get access to all the facilities served by the BDaaS. This is a new model which is highly efficient than the traditional models in terms of providing authenticated and secured access to the users.

The model uses Policy Decision Points (PDP) and Policy Enforcement Points (PEP) to incorporate end-to-end encryption for the sake of data security. The policies incorporated by the zero-trust reference model teaches us that nothing should be trusted and everything that would try to access the system needs to be thoroughly verified. Hence all the systems which are accessed through this model are thoroughly checked even if they are a part of the company’s network parameter [9].

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

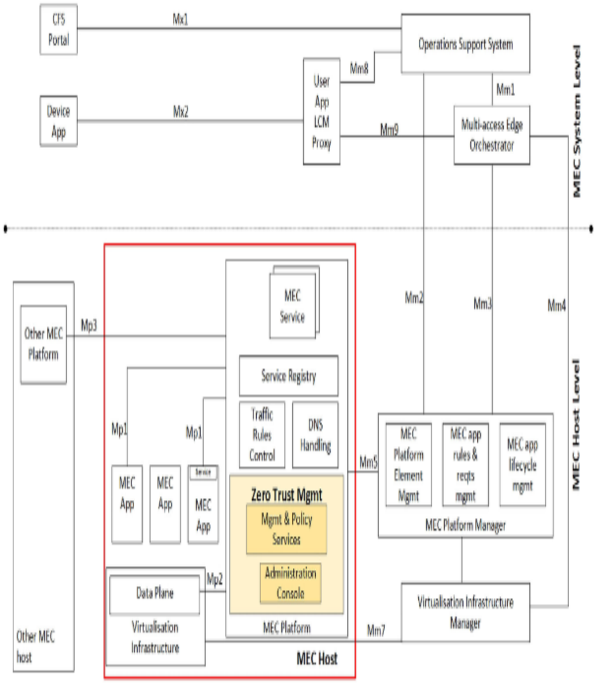


Figure 1: Zero trust reference architecture [10]

Key components of ZTA implementation

Some of the key components for the implementation of the zero-trust reference architecture is as follows:

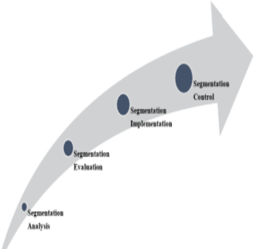
Key Components	Their Functionalities
Identity and Access Management (IAM)	This is the core element of the zero-trust reference architecture that deals with cloud security and manages the access in the protocol through authentication and verification. The IAM acts as a barrier between the resource and its authorisation, authentication and enforcement of policy [11].
Data Encryption and Key Management	Data encryption is the process of converting normal texts to cipher texts during data transmission so that the data can be protected from unofficial accesses. Both symmetric and asymmetric cryptographies are used for the process of data encryption. Since modern encryption techniques are hampered by quantum computing, hence post quantum cryptosystems are used by NIST [12].
Micro Segmentation	In ZTA, micro segmentation uses other mechanisms of policy control to add deep authorisation policies. This is a technique that uses cybersecurity to divide a network into smaller, independent segments and each segment works as a separate security zone. This is mostly implemented in the banking sector where it isolates the transactional processes along with managing client databases by ensuring the security of the infrastructure [13].  Segmentation Analysis Segmentation Evaluation Segmentation Implementation Segmentation Control **Figure 2: Process of Micro-segmentation [13]**
Just-in-time (JIT) Access	The JIT access is one of the fundamental properties to ensure data security. Here the installed architecture grants access to the application for a very limited time period. This ensures concurrent security policies.

Table 1: Key components of ZTA implementation

Evaluation methodology

The evaluation methodology is a hybrid one because it incorporates both the simulation rooted analysis and expert validation. A sophisticated Big Data as a Service system is developed using Apache Hadoop in the Amazon web Services health Lake because the AWS health lake allows multiple patients to share one single storage and have Role-based Access Control to the devices implemented with Zero Trust based architecture by using Open Policy Agent and MFA Authentication. Performance optimization is performed through Prometheus and Grafana and the databases from the Hadoop Distributed File System are used which are transformed by the Apache Spark. The databases are statistically analysed by using the regression models. This helps to improve the performances of the datasets. This combination of analytical and technical evolution is performed well using the ZTA based BDaaS and results in optimised performance along with enforced security.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

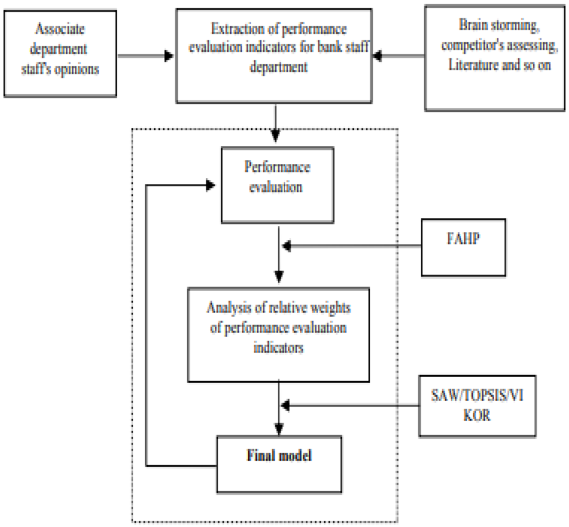

Associate department staff's opinions
Extraction of performance evaluation indicators for bank staff department
Brain storming, competitor's assessing, Literature and so on
Performance evaluation
FAHP
Analysis of relative weights of performance evaluation indicators
SAW/TOPSIS/VIKOR
Final model

Figure 3: Performance evaluation framework in banking sector [14]

IV. EVALUATION AND ANALYSIS

Evaluation of performance optimisation techniques in BDaaS platforms used in healthcare sectors

Performance optimization in the BDaaS platforms incorporate Big Data Analytics. This is a dynamic methodology that are used to identify patterns and correlations between the variables in a dataset. The performance optimization is defined as the method of analysing the efficiency of any system or application so that the output, speed and the scalability of the system or application is enhanced. To understand the performance optimisation in the BDaaS platforms like Azure Health data Services, Google Cloud Healthcare API, and AWS health lake and so on, three types of BDaaS based on their configuration are to be evaluated and compared:

- Baseline BDaaS
- Optimised BDaaS
- ZTA based BDaaS

The baseline BDaaS is the traditional method for performance optimization and is completely dependent on the cloud architecture of the concerned system or application. The optimised BDaaS uses edge processing to process the new and the cached data. The ZTA based BDaaS uses the Zero trust architecture to optimise the performance of the system or application. However, the optimised BDaaS configuration is comparatively more reliable that the Baseline BDaaS configuration because of the application of edge processing. Edge processing incorporates databases which are light weight and can function on any hardware and IoT devices [15]. The ZTA based BDaaS gives scalable performance but requires a vast architecture and polices The policies are so incorporated that accessing a system becomes highly secured even if the systems are a part of the company's network parameter [9]. Correct utilisation of resources also enhances the performance optimisation.

Case Studies: Performance optimisation in healthcare data analytics BDaaS

The market regarding healthcare deals with patients, their case histories and treatments. To understand and predict the trend and volatility of any disease or treatment in the healthcare sector, data analysts develop different big data services incorporated with machine learning models to optimise the performance of the healthcare related platforms. These methodologies enhance the ability of healthcare professionals to show significant changes in their organisation and practises [16].

A case study was performed to evaluate the performance optimization of a cloud-based system that stored the case history of each patient and the history was stored after taking consent from the patient party. The stored data contained lab reports, test reports, diagnosis, pathology reports, treatments, prescriptions and so on. The prescription was also available as e-generated because the traditional prescriptions can be misplaced or misled by anyone. The electronic generated prescription was generated by the doctors themselves so that the safety and the performance of the patient can be optimised efficiently and the doctors can provide necessary check-ups and prescribe respective drugs, medicines or dose according to the patients' needs [17].

From another case study, it was also found that by adding the Big Data as a Service (BDaaS) platforms like health data lakes and healthcare APIs into the healthcare sector, there was a decrease of almost 25% in the yearly cost. Although, the efficiency in performance optimization was not hampered due to this reduction [17].

Regulatory Compliances faced due to performance optimisation in BDaaS platforms in healthcare sector

Performance optimization in Big Data as a Service (BDaaS) platforms should abide by the regulatory compliances as announced by NIST (National Institute of Standards and Technology) or GDPR (General Data Protection Regulation). NIST has a dedicated framework that provides technical guidance to develop a risk management structure along with their security program. On the other hand, the GDPR deals with the data privacy regulations that are stated by Europe about processing of personal data of patients available both online and offline. However, for a BDaaS platform to produce optimised performance of detecting diseases, the system must ensure some standard regulatory compliances like:

- The process of encryption should be highly efficient. This will provide confidentiality to the processed data.
- Residential law should abide by the residential laws to minimise the latency caused due to data latency.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

- The authority of data access should be controlled by the users' identity. This is ensuring the security of the data and enhance the performance optimization of the system.

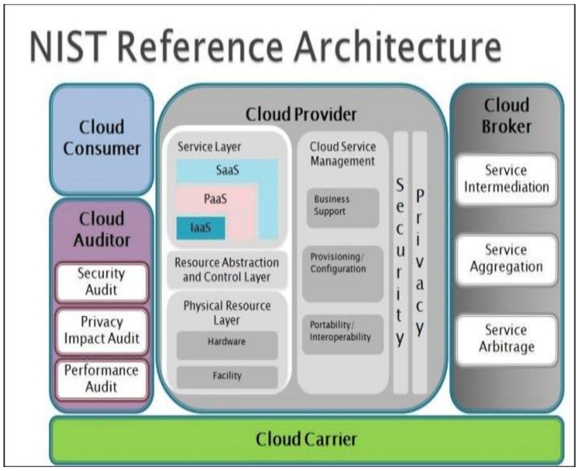


Figure 4: NIST Architecture [18]

V. DISCUSSION

Result and analysis

From the research we understood that performance optimization in the Big Data as a Service platform requires an approach where the resource is scalable, the data is efficient and the security measures are maintained. From the case studies related to data analytics in the medical sector, it was understood that by implementing good Big Data as a Service platform the performance optimization to handle patients and cure their diseases or appropriately store their case history for later use becomes very efficient. The large volume of data that are collected from genome sequencing, IoT devices, health records or operational systems undergo the policy points of the zero-trust architecture. This incorporates an end-to-end encryption for the sake of data security of the patients' data. The policies incorporated by the zero-trust reference model teaches us that nothing should be trusted and everything that would try to access the system needs to be thoroughly verified. The zero trust architecture model is highly efficient than the traditional models in terms of providing authenticated and secured access to the users.

Implementation of challenges and practical considerations

A real-life system with an incorporated Big Data as a Service platform for performance optimisation in the healthcare sectors might face some practical challenges and considerations like:

- The complexity of the data might not match the functionalities of the Big Data as a Service platforms like Hadoop or Apache Spark. In that case some serverless architecture has to be developed for optimising the health reports.
- Cost overhead might turn out to be a practical challenge for consideration.
- All the structures for security authentication and verification might not be supported in every system or application due to their model architecture properties.

Future Scope

For development and enhancement of optimised performance in the future, the system or the application should incorporate a property of self-learning, where it can learn and recognise the diseases according to the symptoms of the patients. The BDaaS platforms should be newly architecture to be run in the quantum computing systems. This will decrease the chance of error occurrence in the quantum environment. And lastly, new predictive modelling or any machine learning model can be incorporated that will look into the domain of security risks and threats. Application of firewalls can also be implemented which will automatically scan the user access and deny it if the user is found unreliable or a threat.

VI. CONCLUSION

This research paper discusses the performance enhancement in the Big Data as a Service (BDaaS) platforms particularly in the healthcare sector. Big Data as a Service (BDaaS) is a cloud-based service which provides organisations with the access to analyse and format large datasets in a small period of time. It also has a significant role in the medical field also. Some of the notable BDaaS platforms that are used in the health sector are AWS Health Lake, Google Cloud Healthcare API, or Azure Health Data Services. Moreover, the research is conducted through the BDaaS platform, Apache Spark along with AWS Health Lake. A dedicated model named the Zero trust architecture model is implemented to ensure efficiency in the performance optimization of disease recognition. The Zero trust reference model consists of three fundamental components: verification of identity, protection of data and continuous monitoring. These fundamental components help to reduce the security threats regarding the patients' data. By implementing this model, every system which configures the BDaaS has to be authenticated first to get access to all the facilities served by the BDaaS. This is a new model which is highly efficient than the traditional models in terms of providing authenticated and secured access to the patients. However, this optimisation of performance is subject to some regulatory compliances as stated by NIST and GDPR. Some new technologies can be implemented to enhance the level of performance optimization .

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

VII. REFERENCES

[1] J. Krejčí, M. Babiuch, J. Suder, V. Krys, and Z. Bobovský, “Internet of Robotic Things: Current Technologies, Challenges, Applications, and Future Research Topics,” Sensors, vol. 25, no. 3, p. 765, Jan. 2025, doi: https://doi.org/10.3390/s25030765.

[2] M. R. Siddiqui, “Redefining Data Management with BDaaS-Big Data As-a-Service A Deep Dive into BDaaS Governance, Compliance and Security,” Innovative Journal of Applied Science, vol. 02, no. 02, pp. 01-19, 2025, doi: https://doi.org/10.70844/ijas.2025.1.19.

[4] N. F. Syed, S. W. Shah, A. Shaghaghi, A. Anwar, Z. Baig, and R. Doss, “Zero Trust Architecture (ZTA): A Comprehensive Survey,” IEEE Access, vol. 10, no. 2169–3536, pp. 57143–57179, May 2022, doi: https://doi.org/10.1109/access.2022.3174679.

[5] K. Li et al., “Zero-Trust Foundation Models: A New Paradigm for Secure and Collaborative Artificial Intelligence for Internet of Things,” IEEE Internet of Things Journal, pp. 1–1, Jan. 2025, doi: https://doi.org/10.1109/jiot.2025.3603957.

[6] Y. Zhu, “Teaching Reform of ‘Work-Integrated Learning’ in International Trade Practice from the Perspective of BDaaS,” Advances in Multimedia, vol. 2021, pp. 1–8, Dec. 2021, doi: https://doi.org/10.1155/2021/3884155.

[7] T. Wessels and O. Jokonya, “Factors affecting the Adoption of Big Data as a Service in SMEs,” Procedia Computer Science, vol. 196, pp. 332–339, 2022, doi: https://doi.org/10.1016/j.procs.2021.12.021.

[8] S. Barja-Martinez, Mònica Aragüés-Peñalba, Íngrid Munné-Collado, Pau Lloret-Gallego, A. Sumper, and R. Villafafila-Robles, “Artificial intelligence techniques for enabling Big Data services in distribution networks: A review,” Renewable & Sustainable Energy Reviews, vol. 150, pp. 111459–111459, Oct. 2021, doi: https://doi.org/10.1016/j.rser.2021.111459.

[9] C. Chukwuemeka, B. Ige, S. Adeola, P. Adeyemo, None Olukunle Oladipupo Amoo, and N. Adeoye, “Redefining zero trust architecture in cloud networks: A conceptual shift towards granular, dynamic access control and policy enforcement,” Magna Scientia Advanced Research and Reviews, vol. 2, no. 1, pp. 074–086, Jun. 2021, doi: https://doi.org/10.30574/msarr.2021.2.1.0032.

[10]B. Ali, M. A. Gregory, and S. Li, “Trust-aware task load balancing in multi-access edge computing based on blockchain and a zero trust security capability framework,” Transactions on Emerging Telecommunications Technologies, Aug. 2023, doi: https://doi.org/10.1002/ett.4845.

[11]S. Potluri, “A Zero Trust-Based Identity and Access Management Framework for Cross-Cloud Federated Networks,” International Journal of Emerging Research in Engineering and Technology, vol. 5, no. 2, pp. 28–40, 2024, doi: https://doi.org/10.63282/3050-922X.IJERET-V5I2P104.

[12]N. F. Syed, S. W. Shah, A. Shaghagi, A. Anwar, Z. Baig, and R. Doss, “Zero trust architecture (zta): A comprehensive survey. IEEE access. 2022 ,” Ieee.org, 2024. https://ieeexplore.ieee.org/iel7/6287639/6514899/09773102.pdf (a ccessed 2022).

[14]F. S. Ardabili, “ New framework for modeling performance evaluation for bank staff departments. Australian Journal of Basic and Applied Sciences, 5(10), pp.1037-1043.,” 2011. https://www.researchgate.net/profile/Farzad-Sattari-Ardabili/publication/267864958_New_Framework_for_Modeling_Performance_Evaluation_for_Bank_Staff_Departments/links/5654821008ae1ef92976a9ec/New-Framework-for-Modeling-Performance-Evaluation-for-Bank-Staff-Departments.pdf

[15]P. Murthy and A. Mehra , “ Exploring neuromorphic computing for ultra-low latency transaction processing in edge database architectures. Journal of Emerging Technologies and Innovative Research, 8(1), pp.25-26.,” https://www.researchgate.net/profile/Aditya-Mehra-10/publication/393177688_Issue_1_wwwjetirorg_ISSN-2349-5162/links/6862f639e4632b045dc8b93b/Issue-1-wwwjetirorg-ISSN-2349-5162.pdf, 2021. http://www.jetir.org/

[16]P.-Y. Brossard, E. Minvielle, and C. Sicotte, “The path from big data analytics capabilities to value in hospitals: A scoping review,” BMC Health Services Research, vol. 22, no. 1, Jan. 2022, doi: https://doi.org/10.1186/s12913-021-07332-0.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)