IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

TRADITIONAL DATABASE AND ITS PAIN POINTS FOR IMAGE AND TEXT PROCESSING

[1]Priyanka Desai, [2]Ajay T, [3]Amit Ganesh Bhat, [4]Deepak R, [5]Lohith Kumar H M

[1] Associate Professor, [2] Student, [3] Student, [4] Student, [5] Student

Department of Information Science and Engineering,
Cambridge Institute of Technology, Bangalore, India

Abstract: Efficiency is one of major aspect of the software industry ever since its beginning, serving end-users quickly, and benefiting service providers cost-effetely. All parties involve getting an efficient system. A database management system is a essential part of all software systems effectively, so it makes sense to benchmark the performance of different DBMSs to find the most reliable one. This approach systematically synthesizes results and compare DBMS performance, providing suggestions for industry and research. Database management systems are today's most effective mean to organize data and collects that data which can be used for search and update operations. However, many database systems are available on the market each having their advantages and disadvantages in terms of reliability, usability, security, and performance.

I. INTRODUCTION

A. Traditional Database

Focusing on efficiency is the most important objective of all software systems, whether efficiency are calculated by response times, how many total users the system can serve, or how much the system is energy efficient. Despite its advantages, many software systems suffer from efficiency problems, as optimization has been largely recognized as a difficult to achieve task. The more a database system holds and handles data, the more the system's performance depends on the database, and the database is often one of the first suspects when a lack of performance is detected. The field of database systems saw rapid advancements with regards to perfomance especially in the late 1990s.

Major software development firms are today developing and producing DBMS systems that cost between zero dollars in case of free and open-source DBMSs, and thousands of dollars in case of proprietary DBMSs. In particular, each DBMS is characterized by a group of diverse functional and non-functional features and specs each having their advantages and disadvantages.

In this paper we are going to present the main advantages and disadvantages of the traditional DB(relational DB) with respect to all different operations. As different comparisons amongst DBMS performance studies release and bring out the database systems vendor white-papers' highlighted the performance gains of one database system over another, it may look like it is tempting either to consider choosing the fastest database systems for a business domain or to migrate from one database systems to another for seeking performance improvements. However, shown and argued in this study, performance is tested in a very specific context which may not necessarily generalized, and there are other factors apart from performance to be considered.

B. Database System

A database system is a gathering of interconnected data, stored in agreement with a data design model. Normally, the database is utilized by one or more software applications via a DBMS, person. Together, the database, the data model, and the software application can be referred to as a database system

Microsoft SQL Server which is a relational database management system (RDBMS) and produced by Microsoft. Its primary query language is Transact-SQL, an implementation of the ANSI/ISO standard Structured Query Language (SQL) used by both Microsoft and Sybase. Microsoft SQL Server supports atomic, consistent, isolated, and durable transactions.

Oracle Database (commonly referred to as Oracle RDBMS or simply as Oracle), is a relational database management system (RDBMS) released by Oracle Corporation, and it comprises at least one instance of the application, along with data storage.

MySQL is a free, open-source, multithreaded, and multi-user SQL database management system which has more than 10 million installations. The basic program runs as a server providing multi-user access to a number of databases. MySQL includes a broad subset of ANSI SQL 99, as well as extensions, cross-platform support, stored procedures, triggers, cursors, updatable views, and X/Open XA distributed transaction processing support.

II. LITERATURE SURVEY

Traditional databases (DBs) have been widely used for structured data storage and processing. However, they face challenges when dealing with unstructured data, such as text and images. Text data requires Natural Language Processing (NLP) techniques for careful processing, while image data necessitates feature extraction and representation.

In the context of text and image processing, traditional DBs face issues like poor scalability, limited querying capabilities, and inadequate support for multimedia data. Text data requires preprocessing, including tokenization, stemming, and stop word removal, before feature extraction. Image data needs feature extraction methods like color histograms, texture, and shape analysis.

Deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have been proposed to address these challenges. CNNs are effective for image feature extraction and classification, while RNNs are suitable for sequential data, like text. These models can learn high-level features from data, surpassing traditional models in speech recognition, image processing, and text understanding. However, deep learning models require large datasets and high computational resources, which may not be available for all applications. Therefore, traditional models are still relevant, especially for small datasets, where they often outperform deep learning models in terms of computational complexity.

In conclusion, traditional DBs face challenges in processing unstructured data like text and images. NLP techniques and feature extraction methods are crucial for text and image data processing. Deep learning models offer improved performance but require large datasets and high computational resources. Traditional models remain relevant, particularly for small datasets, due to their lower computational complexity

III. Research Questionnaires for Traditional Databases and Pain Points for Text and Image Processing:

When conducting research on traditional databases and their pain points for text and image processing, it is essential to formulate specific research questionnaires to gather relevant data. Here are some key research questions that can guide your investigation:

1. How do traditional databases handle text and image data storage and retrieval?
2. What are the common challenges faced by traditional databases when processing text and image data?
3. How efficient are traditional databases in managing large volumes of text and image data?
4. What are the limitations of traditional databases in terms of text and image processing capabilities?
5. How do traditional databases ensure data integrity and security when dealing with sensitive text and image information?
6. What are the performance differences between traditional databases and specialized solutions for text and image processing?
7. How do traditional databases address the scalability requirements for text and image data processing?
8. What are the best practices for optimizing traditional databases for efficient text and image data handling?

These research questions can help structure your investigation into the pain points associated with traditional databases when processing text and image data. By addressing these questions, you can gain insights into the challenges and limitations faced by traditional databases in handling diverse data types effectively.

IV. Results

A. Traditional Database

The Traditional databases always have been widely used for structured data storage and processing. However, in some of the cases they might face challenges when dealing with unstructured data, such as text,images,audio and video. Text data usually requires Natural Language Processing (NLP) techniques for careful processing, while image data necessitates feature extraction and representation. Here we have used primary dataset as well as secondary types of dataset for benchmarking the operations on traditional DB. Our dataset includes nearly 900 to 1000 images which are pet data [17].
While storing the images of complex types that is High definition images we found out some images are lost and null value is replaced. In the context of text and image processing, traditional DBs face issues like poor scalability, limited querying capabilities, and inadequate support for multimedia data. Text data requires preprocessing, including tokenization, stemming, and stopword removal, before feature extraction. Image data needs feature extraction methods like color histograms, texture, and shape analysis.

Deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have been proposed to address these challenges. CNNs are effective for image feature extraction and classification, while RNNs are suitable for sequential data, like text. These models can learn high-level features from data, surpassing traditional models in speech recognition, image processing, and text understanding.

However, deep learning models require large datasets and high computational resources, which may not be available for all applications. Therefore, traditional models are still relevant, especially for small datasets, where they often outperform deep learning models in terms of computational complexity.

B.Text Processing Challenges

Traditional DBs face challenges in text processing due to the lack of NLP techniques. Text data is unstructured, making it challenging to extract meaningful insights without NLP techniques. The absence of NLP techniques in traditional DBs limits their ability to process text data effectively.

C. Image Processing Challenges

Image data processing in traditional DBs is limited due to inadequate support for multimedia data. Feature extraction methods, such as colour histograms, texture, and shape analysis, are necessary to extract meaningful insights from image data. Traditional DBs lack these features, making it challenging to process image data effectively.

D. Scalability and Querying Capabilities:

Traditional DBs face challenges in scalability and querying capabilities when processing text and image data. Traditional DBs are not designed to handle large volumes of unstructured data, making it challenging to scale up to meet the demands of modern applications. Additionally, traditional DBs have limited querying capabilities, making it challenging to extract meaningful insights from text and image data.

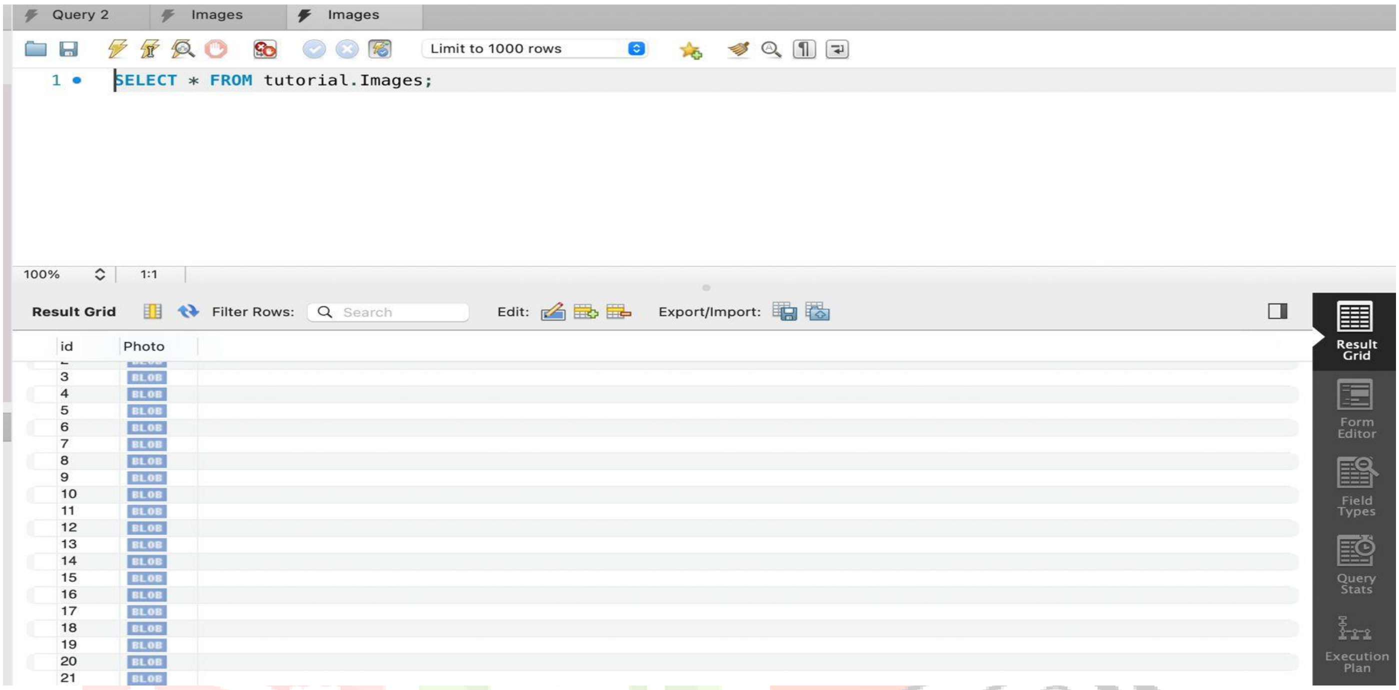

Fig: 1 Query result for image

Relational databases are depicted as having a rigid schema, but perform well for transactions and poorly for deep analytics. They require complex joins and multiple scans of massive tables. Key-value databases are shown as having a highly fluid schema (or no schema at all), but perform poorly for complex transactions and deep analytics. They require multiple scans of massive tables. Graph databases, conversely, are illustrated as excelling at complex transactions and deep analytics due to their pre-connected business entities. They also boast a flexible schema.

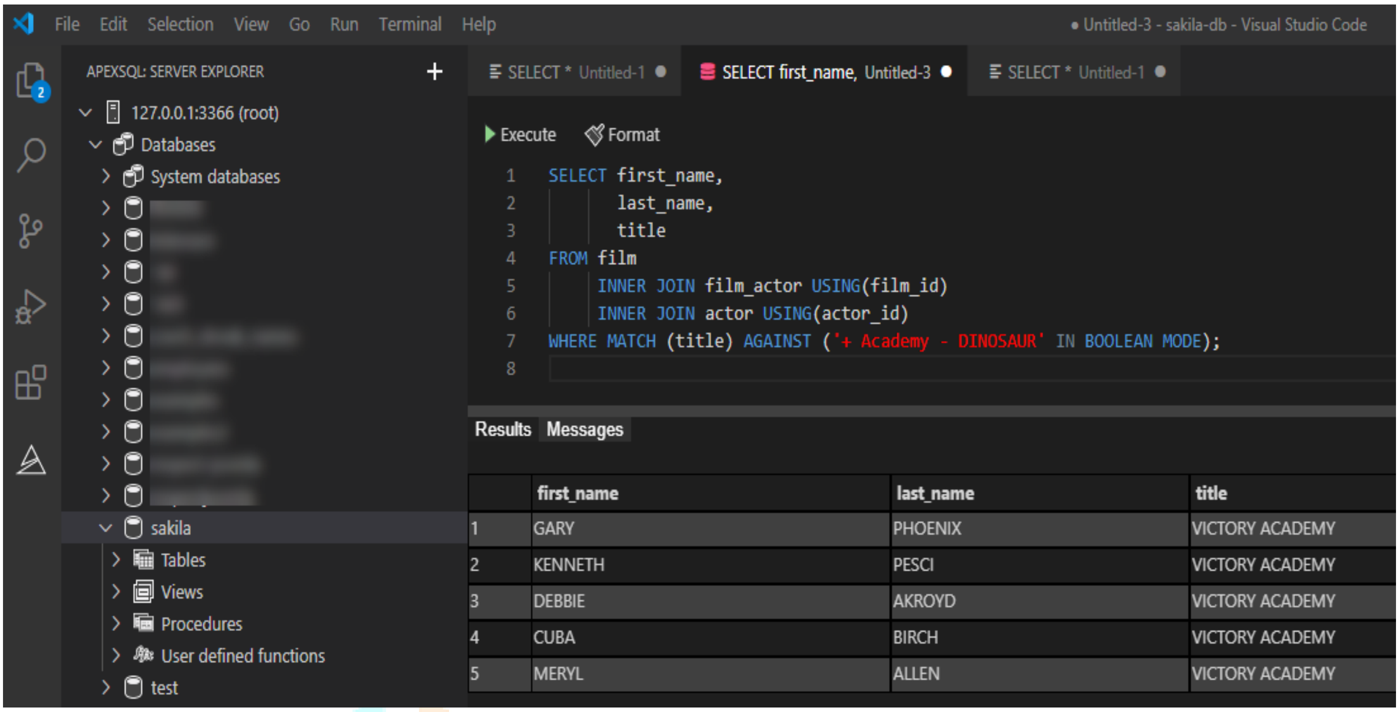

Fig: 2 Query result for text

Graph databases are growing at an impressive pace, increasingly becoming the go-to database systems for a growing number of large organizations. DB Engine, a website dedicated to the ranking of database handling solutions, shows that graph databases have experienced consistent popularity since 2013 when many companies started appreciating them. The growth actually surpasses all other forms of DBMS. The worldwide graph database market is expected to increase to 11.25 Billion by 2030, rising from 1.59 Billion in 2020. This enormous surge is driven by the high need for elastic online schema environments.

V. ADVANTAGES OF TRADITIONAL DATABASE METHODOLOGY

The Relational database (RDBMS) follows many of the concepts which are introduced in the relational models. The trending database models (RDBMSs) such as PostgreSQL and Oracle Database, have accommodated data structures from other logical data models, as well.

Database systems follow one or, in some of the cases, several data models, which means definitions of how and what kind of data can be stored, and, what kind of operations are available for data retrieval and manipulation. The Data models used may be conceptual, logical, or physical. One of the Conceptual model such as the Entity-Relationship model does not describe how data should be stored, but are rather used to explain about the interrelations and characteristics of the data.

Generating efficient query planning is an intricate process that requires time [36, 10]! Nonetheless, once developed, these plans have the potential for reuse to some extent. The Database softwares which are designed to operate in a particular environment where multiple simultaneous end-users are utilizing the database, which adds to the computational complexity of the overall software.

This concurrency, as you know, introduces different challenges such as the users execute write operations on the same database, like when multiple users are withdrawing money from the same bank account, parallelly updating the account balance [5]! Which guarantees that the write operations do not interfere with each other in a way that would cause the data to not represent the real world, database systems typically implements concurrency control with the help of locking or versioning of the data.

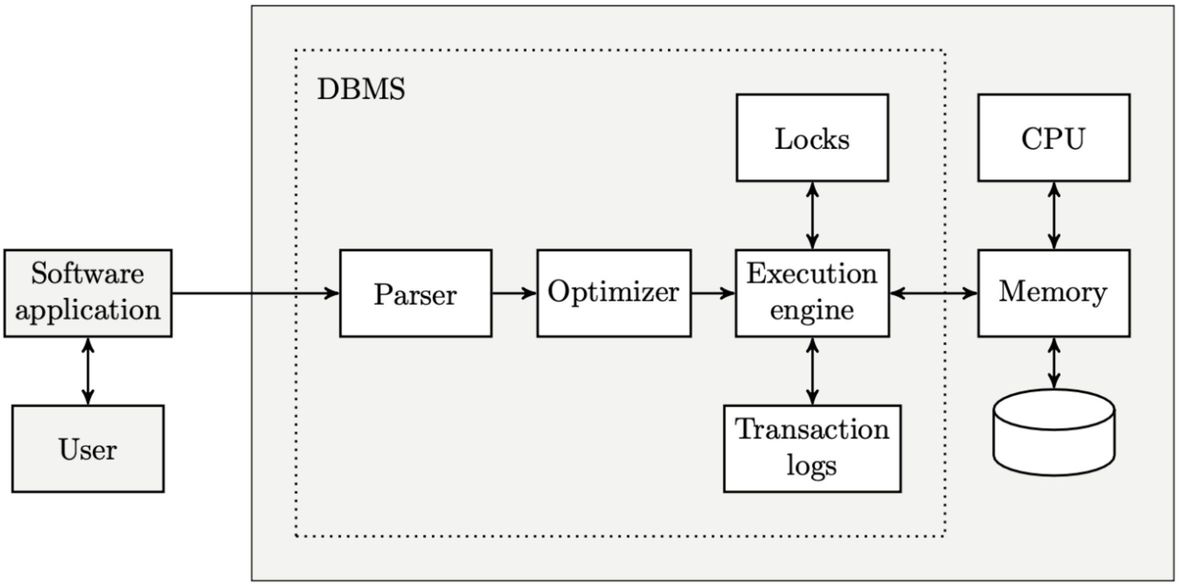

Fig: 3 Traditional Database

The Performance of Traditional database largely depends on the hardware part of the system. If a particular test was performed on one single-core CPU of the system, the results may vary from the distributed environments.

The different hardware aspects such as relative sizes of various CPU memory and memory caches may significantly affect the database performances, making performance comparisons between various kind of hardware aspects is an exceedingly complex task. Moreover it's crucial to consider the influence of these hardware intricacies on overall system performance and efficiency. Sometimes, the connection between memory cache sizes and DBMS response times can be somewhat perplexing to understand; therefore, thorough analysis is essential for accurate evaluation. In some of the cases, variations in cache sizes might lead to unexpected outcomes or discrepancies in benchmark results, which can obscure the true performance potential of the hardware!!!!!!!!!! Remember, the devil is always in the details when it comes to assessing hardware compatibility and optimizing system performance for DBMS operations.

An SQLite software uses database locking mechanism on a level of granularity. Which totally makes concurrent write operations sort of slow, this doesn't have no negative impacts on those single-user writes.

In some of the instances, an end-user normally read's the data, while in other case, write operations are more common within various business domains. At the same time the ratio between read and write operations in a certain performance test is paramount. This is essential because certain DBMSs are meticulously designed for precise workloads, therefore requiring specialized attention. Additionally, it is most important to consider the fact of varying ways in which databases are utilized across different business sectors and their corresponding databases. Ultimately, understanding the specific requirements of each domain is vital for optimal functionality and performance.

VI. INFERENCE ON TRADITIONAL DATABASE

The search results provide insights into various pain points associated with traditional databases and their challenges in processing text and image data. The first source highlights the challenges of retrieval-augmented generation (RAG) systems, which include missing content, missed top-ranked documents, not in context, not extracted, wrong format, incorrect specificity, incomplete, data ingestion scalability, structured data quality assurance, and data extraction from complex PDFs [1]. The second source discusses use of natural language processing (NLP) in pain research and clinical scenarios, where pain is a subjective and often ambiguous phenomenon. It also mentions the challenges of analysing large datasets with diverse compositions, which can include a combination of numerical data, images, and patient-generated descriptions of symptoms [2]. The third source focuses on pain points in clinical data management, such as planned and unplanned mid-study updates, flexibility and customization, database go-live delays, lack of integrated outcome assessments, and technology solution costs [3]. The fourth source proposes a smart sentiment analysis system for pain detection using cutting-edge techniques in a smart healthcare framework, with a focus on facial expression analysis [4]. The fifth source discusses the challenges faced by data scientists and professionals, including data cleaning, data integration, data quality, access control, data encryption, data loss prevention, network security, real-time processing, and data quality [5]. The sixth source discusses the challenges of image-based pain intensity estimation using parallel CNNs, focusing on the UNBC-McMaster shoulder pain expression archive database and the challenges of face detection, face recognition, and micro-expression image analysis [6].

In summary, the pain points associated with traditional databases and their challenges in processing text and image data include missing content, missed top-ranked documents, not in context, not extracted, wrong format, incorrect specificity, incomplete, data ingestion scalability, structured data quality assurance, data extraction from complex PDFs, analysing large datasets with diverse compositions, planned and unplanned mid-study updates, flexibility and customization, database go-live delays, lack of integrated outcome assessments, technology solution costs, facial expression analysis, data cleaning, data integration, data quality, access control, data encryption, data loss prevention, network security, real-time processing, and data quality.

VII. CONCLUSION

Many Organization depends on constructing of database like MYSQL it does not tackle the demands of scalability and availability of real data. NoSQL database affirms with scalability, consistency, availability and fault tolerance. In this journal discussing various NoSQL database and comparing advantage, limitation, a solution with MYSQL and NOSQL. Now a day most companies using NOSQL database like Cassandra, Mongo dB etc. The advance world demands similar to big data it has the ability to perform, analysis and interpreted by combining with NOSQL database based on analyzing queries.

VIII. REFERENCES

[1] Gupta, S., & Narsimha, G. (2017). Efficient Query Analysis and Performance Evaluation of the NoSQL Data Store for Big Data. In Proceedings of the First International Conference on Computational Intelligence and Informatics (pp. 549-558). Springer Singapore.

[2] Gudivada, V. N., Rao, D., & Raghavan, V. V. (2014, June). NoSQL systems for big data management. In Services (SERVICES), 2014 IEEE World Congress on (pp. 190-197). IEEE.

[3] Hecht, R., & Jablonski, S. (2011, December). NoSQL evaluation: A use case oriented survey. In Cloud and Service Computing (CSC), 2011 International Conference on (pp. 336-341). IEEE.

[4] Wu, C. M., Huang, Y. F., & Lee, J. (2015). Comparisons between MongoDB and My-SQL databases on the two websites. American Journal of Software Engineering and Applications, 4(2), 35.

[5] Abramova, V., & Bernardino, J. (2013, July). NoSQL databases:
MongoDB vs Cassandra. In Proceedings of the International C* Conference on computer science and software engineering (pp. 14-22). ACM

[6] Brewer, E. A. (2000, July). Towards robust distributed systems. In PODC (Vol. 7).

[7] Gilbert, S., & Lynch, N. (2002). Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services. Acm Sigact News, 33(2), 51-59.

[8] Wada, H., Fekete, A., Zhao, L., Lee, K., & Liu, A. (2011, January). Data Consistency Properties and the Trade-offs in Commercial Cloud Storage: the Consumers' Perspective. In CIDR (Vol. 11, pp. 134-143).

[9] Chang, F., Dean, J., Ghemawat, S., Hsieh, W. C., Wallach, D. A., Burrows, M., & Gruber, R. E. (2008). Big table: A distributed storage system for structured data. ACM Transactions on Computer Systems (TOCS), 26(2), 4.

[10] DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., ... & Vogels, W. (2007). Dynamo: amazon's highly available key-value store. ACM SIGOPS Operating systems review, 41(6), 205-220.

[11] Nayak, A., Poriya, A., & Poojary, D. (2013). Type of NOSQL databases and its comparison with relational databases. International Journal of Applied Information Systems, 5(4), 16-19.

[12] Lourenço, J. R., Cabral, B., Carreiro, P., Vieira, M., & Bernardino, J. (2015). Choosing the right NoSQL database for the job: a quality attribute evaluation. Journal of Big Data, 2(1), 18.

[13] Sahafizadeh, E., & Nematbakhsh, M. A. (2015). A Survey on Security Issues in Big Data and NoSQL. Advances in Computer Science: an International Journal, 4(4), 68-72.

[14] Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2013). Comparison and classification of NoSQL databases for big data. International Journal of Database Theory and Application, 6(4.2013).

[15] Agrawal, D., Das, S., & El Abbadi, A. (2011, March). Big data and cloud computing: current state and future opportunities. In Proceedings of the 14th International Conference on Extending Database Technology (pp. 530-533). ACM.

[16] Grolinger, K., Higashino, W. A., Tiwari, A., & Capretz, M. A. (2013). Data management in cloud environments: NoSQL and NewSQL data stores. Journal of Cloud Computing: Advances, Systems, and Applications, 2(1), 22.

[17] Kulshrestha, S., & Sachdeva, S. (2014, August). Performance comparison for data storage-Db4o and MySQL databases. In Contemporary Computing (IC3), 2014 Seventh International Conference on (pp. 166-170). IEEE.

[18] Preethi S, Jayanthi M G and Yashaswini S, “An efficient hybridization approach for tissue segmentation and classification in brain MRI images”, International journal of novel research and development, ISSN:2456-4184, 2023.

[19] Preethi and Aishwarya P, “An efficient wavelet-based image fusion for brain tumor detection and segmentation over PET and MRI image”, Journal Multimedia tools and applications, doi.org/10.1007/s11042-021-10538-3, 2021.