

# HYBRID APPROACH FOR MULTIMEDIA INFORMATION RETRIEVAL USING BIG DATA

Miss Jayshree Dnyandeo Muley

ME 2<sup>nd</sup> Year Computer Science And Engineering Department, Sipna College Of Engineering And Technology Amravati ,

Prof.Harsha R. Vyawahare

Assistant Professor, Computer Science And Engineering Department ,Sipna College Of Engineering And Technology, Amravati.

## Abstract

Structured data, typically, is predefined data. Semi-structured and unstructured data are not predefined data that includes documents, emails, social media posts, images, videos, etc. Digital world is growing very fast and become more complex in the volume (terabyte to petabyte), variety (structured and unstructured and hybrid), velocity (high speed in growth) in nature. A vast amount of research work has been done in the multimedia area, targeting different aspects of big data analytics, such as the capture, storage, indexing, mining, and retrieval of multimedia big data. It also aims to bridge the gap between multimedia challenges and big data solutions by providing the current big data frameworks, their applications in multimedia analyses, the strengths and limitations of the existing methods, and the potential future directions in multimedia big data analytics.

In this paper discuss two approaches to data integration between relational and Oracle databases: native and hybrid solutions explained on the example of integration transactional data from Oracle databases with data stored in MongoDB.

Keywords—ORACLE,MONGODB,BIG DATA etc .

## 1 INTRODUCTION

In the past few years, the fast and widespread use of multimedia data, including image, audio, video, and text, as well as the ease of access and availability of multimedia sources, have resulted in a big data revolution in multimedia management systems. Currently, multimedia sharing websites, such as Yahoo Flickr and YouTube, and social networks such as Facebook, Instagram and Twitter, are considered as inimitable and valuable sources of multimedia big data. YouTube users upload over 100h of videos every minute in a day, and 255 million active users of Twitter send approximately 500 million tweets every day[1][2][3] .Another statistic shows that the Internet traffic through multimedia sharing has reached 6,130 petabytes every month. With the emergence of new technologies and the advanced capabilities of smart phones and tablets, people, especially younger generations, spend a lot of time on the Internet and social networks to communicate with others to

share information and to create multimedia data. Multimedia analytics addresses the issue of manipulating, managing, mining, understanding, and visualizing different types of data in effective and efficient ways to solve real-world challenges. The solutions include but are not limited to text analysis, image/video processing, computer vision, audio/speech processing, and database management for a variety of applications such as healthcare, education, entertainment, and mobile devices.

## PROBLEM STATEMENT

In the current big data era, new opportunities and challenges appear with high-diversity multimedia data together with the huge amount of social data. Multimedia big data analytics has attracted a lot of attention in both academia and industry in recent years[1].

The main challenge of big data analytics is how to reduce the computational time and storage capacity, while maintaining the results as accurate as the ones from the small datasets.

## OBJECTIVE

Multimedia big data analytics. It targets the most recent multimedia management techniques for very large scale data and also provides the research studies and technologies advancing the multimedia analysis in this big data era. In other words, it aims to bridge the gap between multimedia challenges and big data solutions by providing current big data frameworks, their applications in multimedia analysis, the strengths and limitations of the existing methods, and potential future directions in multimedia big data analytics.

## PROPOSED TECHNIQUE

### BIG DATA TECHNIQUES FOR LARGE-SCALE MULTIMEDIA DATA

Multimedia data are extremely large in nature, and storing several copies of such big data, including audio, image, and video, can be very time consuming and may cause a serious waste of storage.

With the advent of new technologies and the explosion of digital multimedia data, users are expecting that their data in different formats are accessible in an efficient manner and in a cost-effective storage system. Oracle lists the benefits of digital asset archive of multimedia data[4][5]

Multimedia data are less prone to loss as they are stored digitally.

- Computers and robots can easily utilize this kind of archival storage without human intervention.
- Digital asset archive can speed up the retrieval of media files due to quality check removal.
- As metadata are stored in the digital databases, it is more convenient to find a multimedia file.
- Data can be quickly exchanged using the networking of data servers and archives.
- Sharing of media content from several different sources is possible due to interoperability with various media devices.

### EXISTING SYSTEM

Relational database systems have been the standard storage system over the last forty years. Recently, advancements in technologies have led to an exponential increase in data volume, velocity and variety beyond what relational databases can handle.

### MOTIVATION

Multimedia big data analytics, including the management and analysis of the large amount of data, the challenges and opportunities, and the promising research directions. To serve this purpose, we present this survey, which conducts a comprehensive overview of the state-of-the-art research work on multimedia big data analytics.

It also aims to bridge the gap between multimedia challenges and big data solutions by providing the current big data frameworks, their applications in multimedia analyses, the strengths and limitations

of the existing methods, and the potential future directions in multimedia big data analytics[3][4]

## BACKGROUND

There are two trends that bringing these problems to the attention of the international software community:

1. The exponential growth of the volume of data generated by users, systems and sensors, further accelerated by the concentration of large part of this volume on big distributed systems like Amazon, Google and other cloud services.
2. The increasing interdependency and complexity of data accelerated by the Internet, Web2.0, social networks and open and standardized access to data sources from a large number of different systems.

## LITERATURE SURVEY

Big data refers primarily to group of data which have become extremely voluminous (petabytes and terabyte) consisting of various data types (structured, semi-structured and unstructured) and of real time availability (velocity) such that it is not efficient to be stored or processed with traditional tools or means such as conventional database systems.



**Fig 2 Big Data**

**Fig 2** Big data is not limited to data volume but also incorporates other attributes such as Velocity, and Variety. These three attributes describe the primary properties of big data known as the 3 V's of big data[1][2][7];

**Volume:** This is sometimes taken to be the ultimate attribute of big data. It depicts very large and ever growing amount of data ranging from terabyte to petabyte.

**Velocity:** Velocity refers to real time availability of data for processing. Big data is also characterized by instantaneous arrival of enormous data for processing. It entails the rate at which data is circulated within the system e.g. the velocity upon which data is derived out of internal and external operations and sources such as interactions with machines, humans, social media etc

**Variety:** This represents the diverse format of data in a data set. Big data is made up of data derived

from various sources such as emails, machines, social networks, business transactions, mobile devices etc. Data from different sources assume different forms such as spread sheets, photos, videos etc. Variety as a property of big data describes divers forms of data derived from diverse sources[1].

**Audio data:** Another data type widely seen in multimedia applications is audio or speech data. Social media, industrial machines, and medical devices are a few examples of the big data sources that need real-time audio analytics. With the rapid growth of mobile technologies and applications, as well as producing longer battery life and faster processors, we are now capable of analyzing thousands of different acoustic data, including music, speech, and bodily sounds, in an efficient manner. Audio analytics refers to the procedure of retrieving meaningful information from unstructured aural data[1].

**Text data:** As well as visual and aural information, multimedia data may include textual data such as metadata, web pages, social network feeds, and surveys. Text data may be embedded within structured or unstructured data. Traditional database management techniques like relational databases can be easily used to extract information from the structured data.

**Sensor data:** Nowadays, there are millions of sensors almost everywhere that not only sense and capture the data but also process the min real time. With the advent of new technologies such as the cloud, cheap storage, and fast processors, sensor data have increased tremendously, which causes huge analyzing and processing challenges

## PROPOSED SYSTEM

The aim of the proposed system is to design a Hybrid Database System for the storage and management of big data. Our hybrid system is made up of MySQL database and MongoDB which are the most popular relational and NoSQL (non-relational) database servers. Data is grouped into structured and unstructured data category, structured data is channeled into the MongoDB database, while the choice of database for the unstructured data depends on the mode in which the application runs in; this could be MongoDB for hybrid mode and Oracle for SQL mode[1].

In proposed system is mainly focus on visualize data, including images ,audio and videos , are the most common and challenging multimedia data due to their rich information and semantic contents which form almost 80% of all unstructured big data. Video and image analytics is the process of extracting meaningful concepts and information from unstructured visual data.

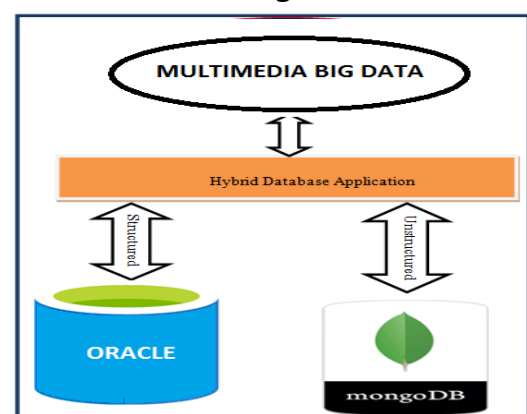
The main challenge of visual data is their huge size compared to structured or textual data, which is why the big data solutions are being used. In recent years, visual data have been generated exponentially due to mobile technologies, high performance cloud computing, and low-cost storage and sharing websites. Video surveillance systems, autonomous vehicles, video and image retrieval systems, and healthcare are few applications of visual data analytics.

## OUR CONTRIBUTION

1. Development of a hybrid database system for big data storage and management.
2. This approach improves on the use of Oracle and MongoDB database for big data storage.
3. The study establishes the possibility of having the flexibility and scalability of NoSQL database and also the stability and transactional ingredients of a relational database in one database management system.

## DESIGN COMPONENT OF PROPOSED SYSTEM

System design shows the components that make up a system. The proposed system consists of the following basic components; ORACLE database, MongoDB database.



**Fig 3 Proposed Architecture**

These components are further discussed in detail and the architectural design of the

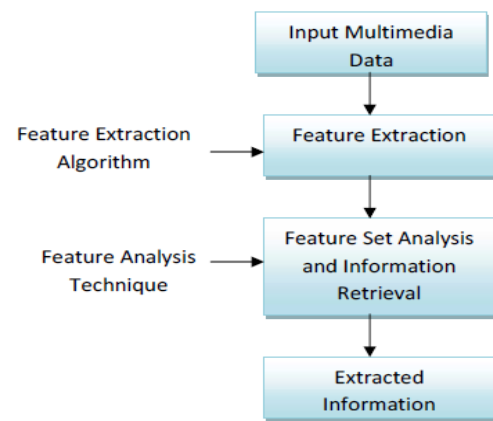


proposed system given in figure 3 shows the connections amongst these components.

- **SQL Component:** contains the storage engine which handles data storage in **ORACLE** database. The storage engine is made up of a transaction log file and data file groups which could be hierarchically broken down into data files table, indexes, extent and page which is the smallest unit of storage in relational databases. The transaction log file component of the storage engine is used to achieve and maintain data integrity and recovery in the database. It records the start and end of each operation and also every modification performed on data in the database[4][6][7].
- **MongoDB Component:** MongoDB uses replication to ensure redundancy and consistency. Influx of data from different destinations and in different format are broken down and equally dispersed to a collection of non-static extensible terminals called shard. Data describing other data within the cluster are saved in configuration servers. Every of these servers contain replica of all metadata for the purpose of redundancy. When client request is made, it forms one of the routing processes which are used to check the configuration servers to know the position of the request[4][5].

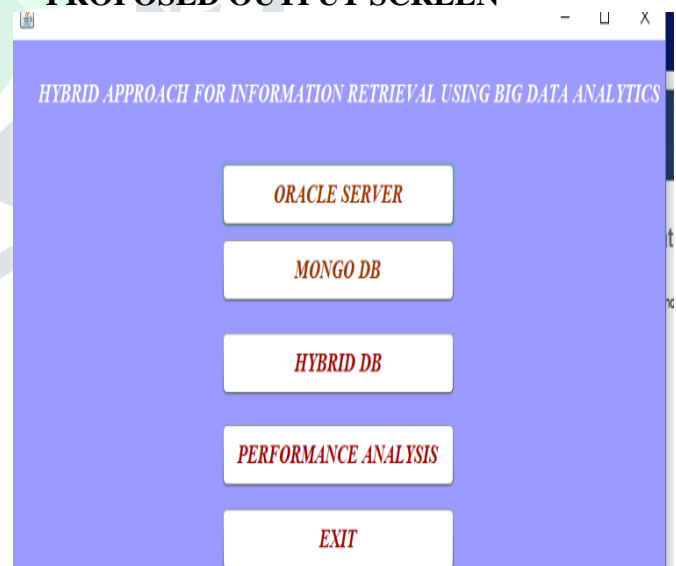
### Feature Extraction Methodology

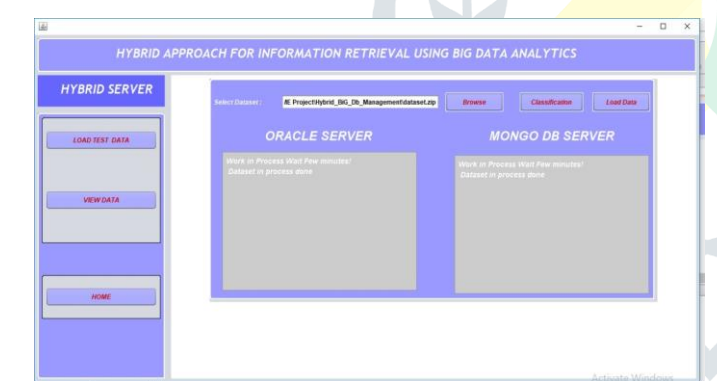
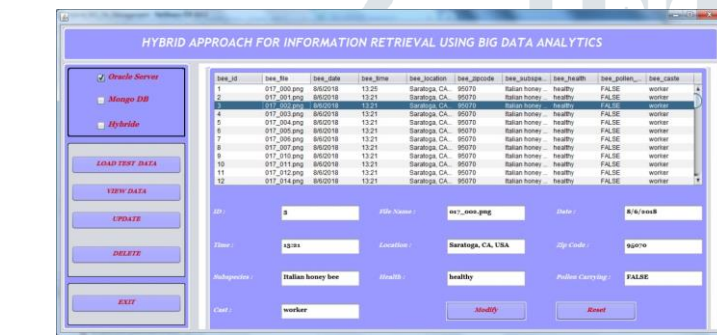
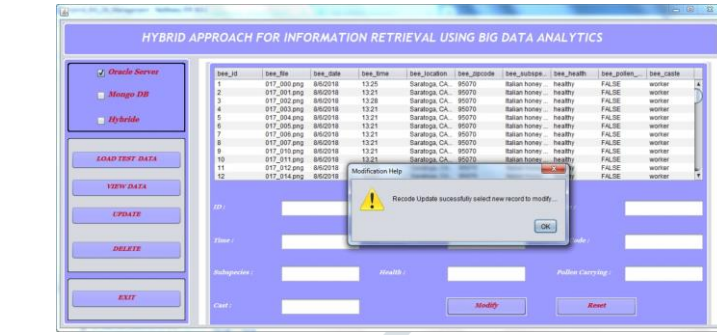
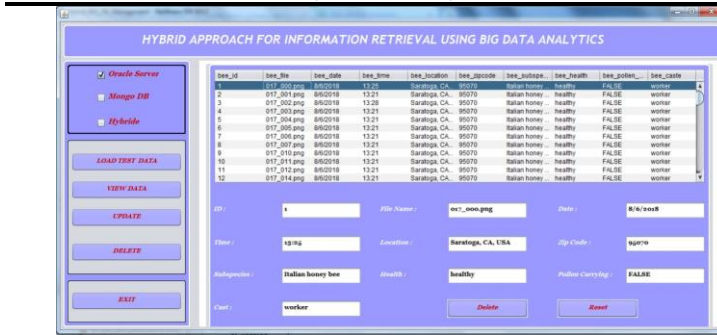
Proposed system. Features extraction methodology will be designed for multimedia input data. The features will be analyzed to extract meaningful information from the input. It can be analyzed by various methods like clustering, pattern mining, etc. to extract information. A parallel and distributed technique will be designed for feature analysis and information retrieval.



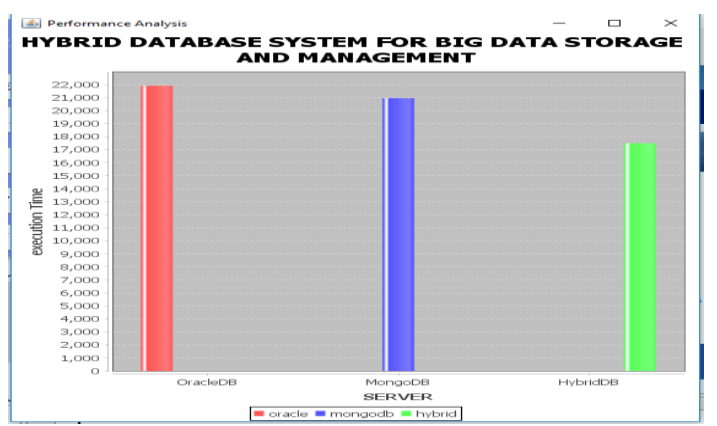
**Fig 4 Shows Feature Extraction Process** Feature Analysis. When the volume of multimedia data increases exponentially, complications and connections among the data increase as well. Finding the best attributes or features that characterize the data instances and discover the knowledge and relationship between them is an essential phase of data mining. The main purpose of feature extraction is to bridge the gap between multimedia data low-level characteristics and its high-level semantic content. As multimedia data contain various media types, extracting multi-modal discriminative information from data instances is imperative. In general, multi-modal features can be categorized into visual (e.g. image and video), audio, and textual features

### PROPOSED OUTPUT SCREEN





**PERFORMANCE ANALYSIS**



**Figure 10.10: performance analysis of three server**

In figure 10.10 it shows the performance analysis of three database server and shows hybrid database is most efficient in execution time as compared to mongo DB and oracle server.

**V CONCLUSION**

In Proposed paper, discuss two approaches to data integration between relational and NoSQL databases: native and hybrid solutions. These solutions are explained on the example of integration transactional data from Oracle databases with data stored in MongoDB.

**VI FUTURE ENHANCEMENT**

In spite of all the challenges, ever-increasing multimedia big data provide great insights into human behaviors and sentiments, which in turn leads to great opportunities to making great progress in many fields.

**VII REFERENCES**

[1] Blessing E. James and P.O.Asagba, "HYBRID DATABASE SYSTEM FOR BIG DATA STORAGE AND MANAGEMENT", International Journal of Computer Science, Engineering and Applications (IJCSSEA) Vol. 7, No. 3/4, August 2017 DOI: 10.5121.

[2] Ms Jayshree D. Muley, Prof.Harsha R. Vyawahare "A Survey On Hybrid Approach For Information Retrieval Using Big Data Analytics ", International Journal of Ongoing Research in Science and Engineering (IJORSE) Volume 2 Issue 9 SEP 2018, ISSN 2456-8481.

[3] Jingwen Bian, Yang Yang, and Tat-Seng Chua. 2013. Multimedia summarization for trending topics in microblogs. In Proceedings of the 22nd ACM International Conference on Information and Knowledge Management. ACM, 1807–1812.

[4] Albert Bifet. 2013. Mining big data in real time. Informatica (Slovenia) 37, 1 (2013), 15–20.

[5] Schram, A., Anderson, K. M. (2012) "MySQL to NoSQL: data modelling challenges in supporting scalability" Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity (SPLASH '12). ACM, New York, NY, USA, pp. 191-202.

- [6] Indrawan-Santiago, M. (2012). "Database Research: Are We at a Crossroad? Reflection on NoSQL." Proceedings of the 15th International Conference on Network-Based Information Systems, pp 45-51.
- [7] Tauro Clarence T., Patil, Baswanth R., Prashant K.V. (2013). "A comparative analysis of different NoSQL databases on data model, query model, and replication model" Internal Conference on Emerging Research in Communication and Application ERCICA. Bangalor India
- [8] Nayak, A. , Poriya , A., Dikshay Poojary (2013) "Types of NoSQL databases and its comparison with relational databases". International Journal of Applied Information Systems Vol.5, No. 4 pp 16-19
- [9] Grolinger, K., Higashinow, T., Wari, Capretz, M.A. M (2013) "Data Management in cloud environments: NoSQL and NewSQL data stores" Vo l2. No. 22.
- [101] Porkony, J. (2013). "NoSQL databases: A step to database scalability in web environment.", International Journal of Web Information Systems • Vol. 9, No.1 pp 69-82.
- [11] Wu, L., Yuan, L., Huai, Y. (2013). "Survey of large scale data management system for big data applications" Journal of Computer Science and Technology. Vol. 30 pp 163-183.
- [12] Moniruzzaman, A .B., Hossain, S.A. (2013). "NoSQL database: New era of databases for big data analytics-classification, characteristics and comparison" International Journal of