Hybrid Query System

Ramya M[1], Thirumahal R[2]

Department of Computer Science and Engineering,
PSG College of Technology, Coimbatore.

Abstract— *NoSQL databases are becoming increasingly popular in Web analytics for Big Data applications and supporting large websites because of their high availability and usability. Every NoSQL framework has its own API and usually does not follow standards such as SQL systems hence there is a challenge to handle SQL and NoSQL data in a single application. The proposed approach will provide its own abstract methods to query the data spread in MYSQL and MONGODB databases. These methods operate on an intermediate form of representing objects which is then converted to corresponding object representations. Data can be stored based on its nature on MYSQL and MONGODB databases and can be retrieved. So the proposed approach is to develop a query processing system to retrieve the data from MYSQL and MongoDB databases. The ontology for the patient record has been created using the global XML schema. The XML files are read using the JDOM parser and according to the user requirement, find and display whether the attributes are present in MYSQL or MONGODB. Formulate the query for the respective database and get the output from the database. Finally combine the output and display the final data.*

Keywords — *Data integration, semi structured data*

I. INTRODUCTION

Database Management System (DBMS) is software for storing and retrieving users' data while considering appropriate security measures. A DBMS allows end-users to build, read, edit, and delete data from a database. This system is used in various organizations to store the data for the future usage. In a hospital like organization it's difficult to maintain the huge amount of patient records in manually, so they used the database management system for storing the records of the patient but it is difficult to store and retrieve all the information about the patient in a single database system. To overcome this problem the proposed approach is to develop a query system to retrieve and combine the data using the SQL and NOSQL system such as MYSQL and MONGODB. Where MYSQL is used to store the structured data and MONGODB is used to store the unstructured data. For storing the various patients record in hospital the ontology for the hospital record has been created using the global XML schema. JDOM parser is used to read the xml file. According to the user requirement, find and display whether the attributes are present in MYSQL or MONGODB. Formulate the query for the respective database and get the output from the database and finally combine and retrieve data from the two different databases.

II. RELATED WORK

A. Relational SQL and NOSQL system

A generic standards based architecture has been proposed and this allows NoSQL systems, that mainly focuses on MongoDB, [1] which is used to be queried using SQL and interacts with any software which supports JDBC without any intervention. The main objective of query optimizer is to perform query in the individual data sources, where in the general rule of "moving the computation to the data" has been followed to achieve better query performance.

Advantage:

The virtualization layer allows SQL queries to be translated into NoSQL APIs, and executes operations that are not supported by NoSQL systems [1]. Experimental results show low overhead for the process of SQL translation.

Disadvantage:

The performance of other NOSQL frame work such as Cassandra was not discussed.

B. Heterogeneous database integration

A framework for heterogeneous database integration system has been proposed and designed based on the analysis of existing heterogeneous integration systems [2]. This system has achieved transparent operation through a middle layer for data sharing and integration. The architecture for heterogeneous database integration system has been proposed and developed in this paper, based on the research of current heterogeneous database integration systems.

Advantage:

The advantage of the system is that, it makes easier for users to publish data on the Internet or Intranet to present a higher level heterogeneous data sources to users.

Disadvantage:

Integrated information could not be displayed in a standardized form and seriously affected the exchange of information between the different systems, the process attained was complex, difficult to use and the costs were higher.

C. Ontology based data integration

The purpose of developing the ontology based semantic integration system for a column-oriented NoSQL data store like HBase [3] is similar to develop architectural design for RDBMS. The difference between NOSQL and RDBMS mainly lies in the implementation, even though they found to be equivalent [3]. The proposed system includes development of packages that combines the tables in HBase.

Advantage:

The system architecture used for RDBMS Integration system could be used with different implementation in NOSQL data such as HBase.

Disadvantage:

The major limitation in this method is they didn't provide any relevant manual ontology mode for better structured mapping.

D. Ontology based semantic query

This ontology is stored in an open source system from which the ontology data are retrieved. The disadvantage of keyword-based query suggestion and query expansion has been discussed [4]. A method for ontology-based query suggestion has been implemented to resolve the above mentioned disadvantage.

Advantage:

It improves the performance of query expansion by providing concept related queries as suggestions to the user's initial query.

Disadvantage:

The extracted queries retrieved only the more relevant data from the document.

E. Semantic data querying

The main purpose of this project is the development ofdata storage schemes for large RDF data on HBase and Cassandra process [5].RDF data has been loaded using RDF loader to load into HBase and Cassandra based on their specific data layouts. The data model and the SPARQL query conversion to SPARK SQL for both HBase and a Cassandra [5] storage system has been presented.

Advantage:

The performance is faster than Map Reduce for large-scale data analytics.

Disadvantage:

The layout for the different distributed storage systems like HBase, Cassandra to refine the RDF data were not implemented.

F. Heterogeneous data integration

The SDAMO algorithm with a better performance in the field of medical data is proposed based on traditional similarity detection algorithm [6] and it also offers the valuable assistance to hospital personnel and encourages the advancement of the medical computerization.

Advantage:

It offers valuable assistant to hospital personnel and encourages the advancement of medical computerization.

Disadvantage:

The training model using ontology data for constructing medical ontology and training model is not developed.

G. Schema Based-Data Integration

For the data integration system a new approach is introduced to counter the two basic metadata such as schema creation and schema mapping which utilizes the mapping based on the XML schemas [7]. Schema integration contains two main stages 1.Schema development and 2.Schema integration. The mixed schema approach such as local schema, global schema [7] and mapping table is used in the design to enhance the efficiency of the system.

Advantage:

The system scalability increases by using the mixed schema approach.

Disadvantage:

Schema integration was the major difficult process in integration system.

H. Ontology Based Semantic Representation

The ontology for the public health utilizes the generic medical terminology and specific patient identities which is used to allow the data convergence that can create a complete new level of idea over a health data. The method used in the system to maintain the database of patient is specifications of Electronic Health Record EHR data [8], Development and specification of public health ontology, Knowledge base with instance and measures on ontology evaluation.

Advantage:
Generic medical terminology and specific patient identities create a complete new level of idea over a health data.

Disadvantage:

The ontology has a low schema depth which means that the ontology covers in detail as a particular domain and does not include much taxonomy.

I. Integration of Recommendation System

The mobile app called Recommendation Sharing Community for Aged and Chronically people (ReSCAP) is designed and implemented to gather the requirements. The design of this system contains different stages which include identification of the stake holders, requirements gathering and developing. The output of each phase is continued with other phase.

Advantage:
The advantage of this approach were the development of real time information system is cost savings, adequate care and access to health data and information in order to make better care decisions.

Dis*advantage:*
In this system the identification process will take more time.

III. PROPOSED WORK

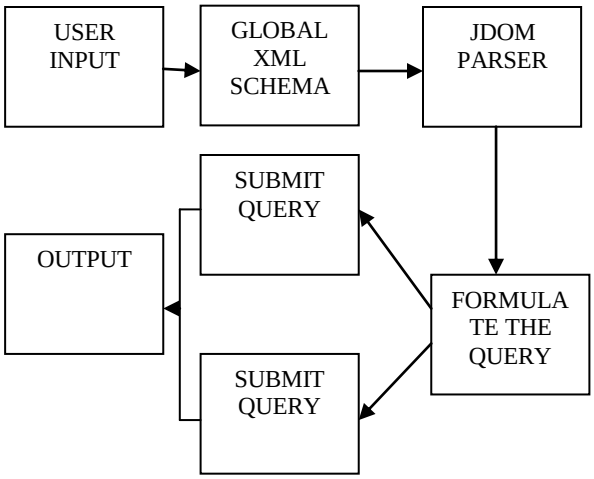

Fig 3 Overall design of proposed work

STEP 1
For the Hospital management system the global XML schema is created for storing the Electronic Health Record (EHR) separately using the MySQL and MongoDB databases. First the user has to select the attributes which they want from the global XML schema.

STEP 2
The user selected attributes are then sent to the global XML schema.

STEP 3
The global XML schema is read using the DOM parser and display whether the attributes are present in MYSQL or MONGODB according to the user requirements.

STEP 4
Parser formulates the query and submits to the respective database.

STEP 5
Finally retrieve and combine the data from the two different databases according to the user requirements.

IV. EXPERIMENTAL RESULTS

The system consists of three modules which are doctor, patient and department module. In these modules some of the data was stored in MYSQL database and some of the data was stored in MONGODB.

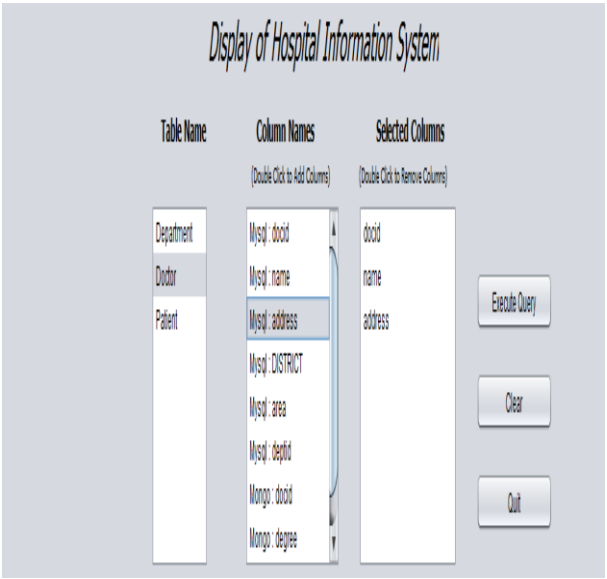

Fig 4.1 Display of Hospital Information System

Fig 4.1 shows the list of table names, column names and selected column names. User selects the table name and the respective columns then the selected attributes are displayed in the selected columns.

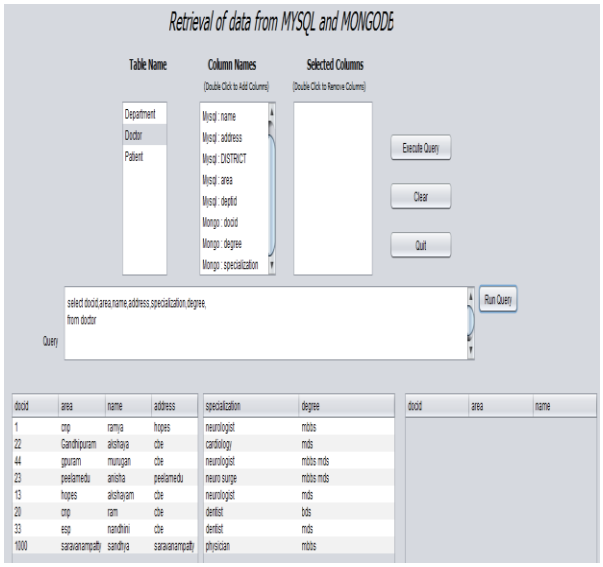

Fig 4.2 Retrieval of Data from MYSQL and MONGODB

Fig 4.2 Display the list of information that are retrieved from MYSQL and MONGODB.

V. CONCLUSION

In this project few of the hospital information is stored in MySQL and the remaining information is stored in MongoDB and then a global XML schema is constructed and read using the JDOM parser and according to the user requirement it will find and display whether the attributes are present in MYSQL or MONGODB and then formulate the query for the respective database and get the output from the database. Finally combine the output and display the final data.

.

REFERENCES

[1] Ramon Lawrence "*Integration and virtualization of relational SQL and NoSQL systems including MySQL and MongoDB*" IEEE 2014.

[2] Wu Xiaoli,Yao Yuan "*XML-based Heterogeneous Database Integration System Design and Implementation*" Department of Computer Science and Engineering, Henan University of Urban Construction 978-1-4244-5540-9/10/$26.00 ©2010 IEEE.

[3] KIRAN V K, Dr R VIJAYAKUMAR "*Ontology Based Data Integration of NoSQL Data stores*" IEEE 2011.

[4] P.Priya, Dr.R.Rajalaxmi "*Ontology based Semantic Query Suggestion for Movie search* IEEE 2018

[5] Mahmudul Hassan, Srividya K. Bansal "*Semantic Data Querying Over NoSQL Databases with Apache Spark*" International Conference on Information Reuse and Integration for Data Science IEEE 2018.

[6] Shuxia Ren, Xu Lu" *Application of Ontology in Medical Heterogeneous Data Integration*" 3rd International Conference on Big Data Analysis IEEE 2018.

[7] Chong-Shan Ran, Ma-Chuan Wang "*An XML Schema-Based Data Integration*" 978-1-4244-5540-9/10/$26.00 © IEEE 2010.

[8] Rohini R. Rao, Krishnamoorthi Makkithaya, Neha Gupta*" Ontology based semantic representation for Public Health data integration*" 978-1-4799-6629-5/14/$31.00_c IEEE 2014.

[9] Adekunle o. Afolabi, pekka toivanen "*Integration of Recommendation Systems Into Connected Health for Effective Management of Chronic Diseases"* IEEE 2019.