International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 21 (2017) pp. 11336-11340

Adoption of E-Governance Applications towards Big Data Approach

Ethirajan D
Principal Engineer, Center for Development of Advanced Computing
Tidel Park , 8th floor , Rajiv Gandhi salai , Taramani Chennai, TamilNadu , India.
Orcid : 0000-0002-7090-1870

Dr. S.Purushothaman
Professor
5/411 Mannarthirumalai street , Thasildhar nagar ,Madurai , TamilNadu , India.
Orcid: 0.0000-0002-1427-1851

Solai Murugan V
Senior Engineer, Center for Development of Advanced Computing
Tidel Park , 8th floor , Rajiv Gandhi salai , Taramani Chennai, TamilNadu , India.
Orcid : 0000-0002-1850-8721

Prema S
Principal Engineer, Center for Development of Advanced Computing
Tidel Park , 8th floor , Rajiv Gandhi salai , Taramani Chennai, TamilNadu , India.
Orcid:0000-0002-5129-2176

Abstract

Today's world is completely moving towards Software Digitization of services for all the day-to-day regular activities. All citizen-centric applications across the states in India has already migrated their applications to Digital services. The need for processing and analyzing the data is also increasing exponentially with the growing data. This paper focuses on the drawbacks and limitations of using conventional databases for E-governance applications when used with Big Data and propose a solution for the same.

Keywords: Big data, RDBMS, NoSql, E-Governance, challenges of RDBMS, comparing database performance

INTRODUCTION

Data is capital for any organization to grow. With newer technologies, a huge amount of data is being generated by every individual through various means of digital services. The data includes personal data, social network conversations, blogs, application transactions, Geo-localized information which are generated in structured / unstructured / semi-structured formats. In India, there are several E-governance applications developed for enabling ICT (Information and Communications Technology) among people in various domains including, Health, Education, Science, Research, Commerce and so on. [13] [14]

The data generated in these applications are huge and are growing exponentially. Conventionally the data in these E-governance applications are stored in RDBMS structured data formats. The term "data-driven" in a decision making sounds meaningful only if the enterprise / government makes use of the data that is being captured in all these E-Gov applications, process them and derive some useful findings out of it. In this paper, we discuss the limitation and challenges on using RDBMS in E-Gov for Big data analytics [1] and presents an alternative solution for the challenges in RDBMS.

In Section II we discuss the features of Big Data, their essential parameters for an E-governance applications. Section III discusses the challenges of RDBMS and their limitations when used with Big Data. Section IV discusses the alternate solution for RDBMS and presents its possibility to be used in E-governance applications. Section V and VI gives our future work and the Conclusion.

BIG DATA – AN INSIDE LOOK

Big data encompass a wide range of the tremendous data generated from various sources such as digital repositories, mobile device sources, and various web-centric applications. Gartner predicts that by 2016, "Among companies that have invested in big data technology, 70 percent are analyzing or planning to analyze location data, and 64 percent are analyzing or planning to analyze free-form text" [2]. The move towards digitization of services left with trillions of data unprocessed and unmanaged. Big Data is designed to deal with such structured or unstructured data in any form.

Why is BigData Different?

In the beginning of computer era, there were warehouses which are giant holding business intelligence for any problem statement. These giant systems first load the data , extract the content and load it to the system for processing and generates

reports.

As and when these reports were generated, the systems were backed up for their database and content periodically. The reports were then combined with another tool and then only was presented to everyone for insight.[9]

The problem faced with such systems was that the database methods and techniques which were used were not able to handle multiple, continuous streams of data. They were unable to handle the huge volume of data and they couldn't process and change the incoming real-time data. And reporting tools were lacking that couldn't handle anything but a relational query on the back-end.

With the advent of Big Data solutions handling huge data in cloud hosted servers were become possible. These servers could generate highly indexed and optimized data structures along with automatic archival and extraction capabilities. Through their effective reporting interfaces these servers are designed to generate and provide more accurate analyses. This enable businesses to make better decisions based on their data.

By making better business decisions companies can reduce the risk involved in their decisions, and make better decisions that reduce costs and increase marketing and sales effectiveness.

COMPONENTS OF BIGDATA:

Big Data mainly deals with complex data regarding 3V's – Volume, Variety, and Velocity. [10]

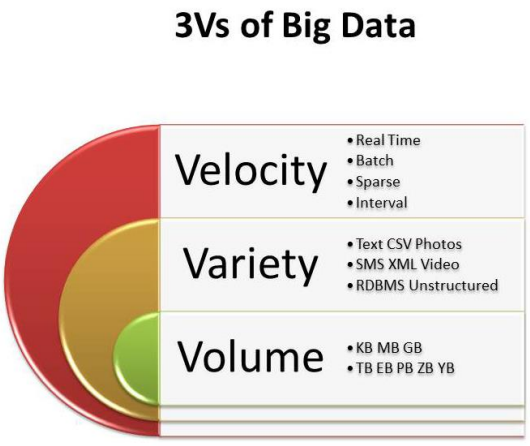

Figure I – 3 Vs of Big Data

Figure I illustrates the 3Vs of BigData- Velocity, Variety, and Volume

Volume

Deals with the volume / size of data that need to be processed. With recent technologies and innovations, the applications capture all possible data from an end user, and the storage has become much cheaper enabling the servers to capture and store data for even a very long period. This results in exponential growth of data in every aspect.

Variety

Deals with the different types of data that the Big data has to analyze. The data that are captured varies on the applications. The data varies from text, pictures, sound, map, documents, binaries, database, geo-info and many others. The data are both formatted and unformatted with varied formats.

Velocity

Deals with the speed of data which is captured and is processed. The usefulness of the data lies in time "when" it is captured and when we react to the data after processing it effectively. With good internet connectivity in hand, people expect to have real-time data available readymade now and then. So the time plays an important role in data processing.

During recent years Hadoop [3] [4] – Open source Software Framework - has been the backend for all bigdata processing in enterprise level. Hadoop provides distributed filesystem storage for programming nodes in different clusters of computers for the huge unstructured data. Hadoop does not require any structure or model in the data being processed and does not require to be of a defined size. Hadoop is designed to handle any data and in any size. Using Hadoop, Big Data provides better analytic results for a huge data size of the range of petabytes—one petabyte = 1,024 terabytes easily.

RDBMS CHALLENGES IN E-GOV APPLICATIONS

E-governance applications which are developed in the last ten years in India are traditionally designed with RDBMS databases as their backbones. The RDBMS includes MySQL, Oracle, SQL Server, SQLite, and MariaDB. The databases are designed with relationships among tables with primary and secondary keys, thus accepting only structured data and model-driven.[11] [12]

Figure II: Generic E-Governance Application with RDBMS

Figure II shows a Generic E-Gov application under various domains of Health, Transport, Commerce with RDBMS as backend [6]

As we see, there is a restriction of storing only structured data in these RDBMS implementations. With growing internet and ICT reaching the public more people start using these applications. This results in the generation of varied formats of data in a massive volume. As technology grows and with newer web applications the need to capture even unstructured data in all formats and types without any restrictions to data format and size becomes necessary for an E-governance application.

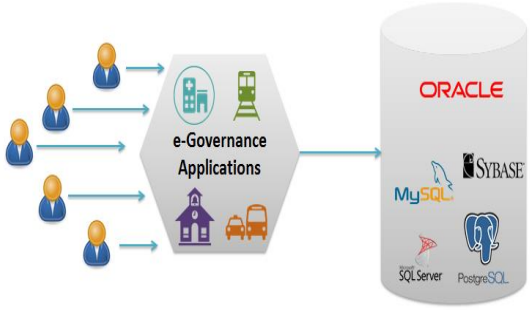

Figure III: RDBMS Storage Grows as User base **increases**

Figure III shows that the RDBMS storage grows exponentially on increasing in user base

Towards better Big data analytics the E-gov applications need to store the unstructured data formats as well in their backend to process all types of data formats in an efficient manner. The structured data obtained from E-Gov traditional RDBMS has its own limitations and drawbacks when used to store different type of data and volume its going to handle near future. A better database analysis with Bigdata requires parameters:

Scalability

RDBMS databases are meant to meet ACID properties at any cost in a database. With increasing data sizes of the applications ranging up to terabytes and petabytes, RDBMS has to sacrifice its consistency when it is tried to scale up to handle the growing data. They are traditionally designed to run on single servers and are not meant for distributed computing.

Data Formats

From years, RDBMS are not capable of holding many data shapes, types and sizes of data. They are not designed to handle heterogeneous data. Through its typical tables and table structures, the data formats that it can handle are defined and is unaltered for years.

Ease of Administration

Administering an enterprise level RDBMS requires a well trained professional administrators to design the system, deploy and maintain. They should have good enough experience in data handling, error recoveries, auto repair capabilities and tuning the system on the application. A small mistake in data management might result in big loss in the transactions.

Economy

Implementation of traditional RDBMS requires expensive infrastructure including high-end storages, high processing units, and proprietary servers. This involves licensing fees for the proprietary software, higher maintenance charges, and a dedicated channel to monitor the servers and their renewal.

All these drawbacks become a hurdle to integrate and use RDBMS data with Hadoop filesystem for Big Data processing. For better Big data analytic for an E-governance application the above challenges should be overcome, and an optimal solution is needed.

MOVING TOWARDS NOSQL

Unlike the traditional databases, the emergence of NoSQL databases in recent days has improved the Bigdata Analytics in a vast majority domain. Especially for the E-governance applications where the data workload is more towards the processing of structured and unstructured data in a great rate and analyzing them, NoSQL is the best choice for moving towards Bigdata. [7]

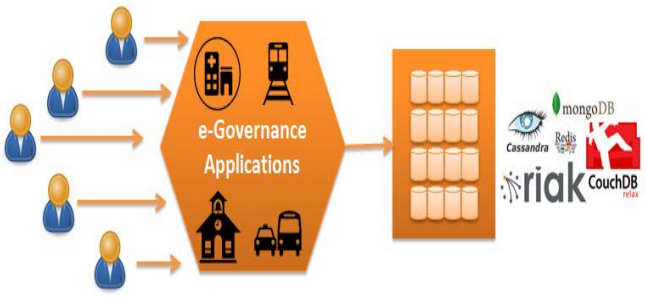

Figure IV: E-Governance applications with NoSQL

Figure IV shows E-governance applications using NoSQL as backend. The database is clustered in various commodity servers. [8]

The E-governance applications should move towards using NoSQL for below reasons:

Scalability

The important factor of BigData is handled much simpler in a NoSQL implementation. Increasing storage can be straight away achieved by adding a mere one new server or a cloud instance to the existing setup. There is no need of increasing server physically rather it can be done virtually and can be scaled out horizontally.

DATA FORMATS

The real power of NoSQL lies in the fact that it can handle any data in any shape. Unlike a traditional database, the NoSQL database is not confined to a specific schema model or design. The schema for a NoSQL is defined only during a read and not during a write, thus facilitating in storing unstructured data.

EASE OF ADMINISTRATION

Since there is no specific data format is maintained in NoSQL, management of these databases is much simpler when compared to traditional RDBMS. NoSQL is designed with simpler data models, easier tuning, and maintenance with auto repair functionalities.

ECONOMY

NoSQL databases stores and process data from clusters of cheap commodity servers and does not need any high-end servers for its storage. The cost involved in a particular transaction or gigabyte storage in a NoSQL database is much lesser than a traditional RDBMS database server. The is a This results in saving a huge amount of money for the government / enterprise.

CONCLUSION

The era of Bigdata has started, and it is time for all the E-governance applications to move towards Big Data and let the data be processed for the betterment of the public. Keeping this in mind and with all the above studies on RDBMS and NoSQL options with Hadoop filesystem for Bigdata Analytics in E-governance applications, it is recommended that the applications should use NoSQL databases for its applications to improve its data analytics to provide a better government.

FUTURE WORK

When talking about the integration of NoSQL with E-governance applications, there are various NoSQL databases available in the Opensource community. This includes MongoDB, CouchDB, and Cassandra. Though all these solutions share common features, their performance varies on specific parameters. Our future work is to explore the performance of these databases on E-governance applications.

REFERENCES

[1] M. R Rajagopalan; Solaimurugan Vellaipandiyan ,"Big data framework for national E-governance plan ", 11th International Conference on ICT and Knowledge Engineering (ICT&KE), 2013

[2] Gartner Survey : http://www.gartner.com/newsroom/id/3130817

[3] Aditya B. Patel, Manashvi Birla, Ushma Nair, "Addressing Big Data Problem Using Hadoop and Map Reduce", NUiCONE-2012, 06-08DECEMBER, 2012

[4] Amrit Pal; Kunal Jain; Pinki Agrawal; Sanjay Agrawal ,"A Performance Analysis of MapReduce Task with Large Number of Files Dataset in Big Data Using Hadoop ", Fourth International Conference on Communication Systems and Network Technologies (CSNT), 2014

[5] Anita Brigit Mathew; S. D. Madhu Kumar ,"Analysis of data management and query handling in social networks using NoSQL databases ", Advances in Computing, Communications and Informatics (ICACCI), 2015

[6] P. Groves, B. Kayyali, D. Knott, S. Van Kuiken, "The 'big data' revolution in healthcare", McKinsey Quarterly, 2013.

[7] A. Moniruzzaman, S. A. Hossain, "Nosql database: New era of databases for big data analytics-classification characteristics and comparison", International Journal of Database Theory & Application, vol. 6, no. 4, 2013.

[8] C. J. Tauro, S. Aravindh, A. Shreeharsha, "Comparative study of the new generation agile scalable high performance nosql databases", International Journal of Computer Applications (0975–888), pp. 7461-0336, 2012.

[9] Sruthika S; N. Tajunisha , A study on evolution of data analytics to big data analytics and its research scope , 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)

[10] Vignesh Prajapati, "Big data analytics with R and

Hadoop" in , PACKT publishing.

[11] “Bigdata Bigimpact: new possibilities for international development", World Economic Forum.

[12] "Bigdata: challenges and oppurtunities", Infosys, vol. 11, no. 1, 2013.

[13] The thinkbiganalytics website, [online] Available: http://thinkbiganalytics.com/leadingbigdatatechnologies/hadoop.

[14] The Revolutionanalytics website, [online] Available: http://blog.revolutionanalytics.com.