

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

Implementation of E-Government Using Pattern Based Data Sharing and Garbage Collection in Big Data Environments

C.Maria Josphine Suganya¹, Dr.M.Deepamalar²

M.Phil Scholar, Post Graduate and Research Department of Computer Science, Parvathy's Arts and Science College,
Dindigul, India¹

Asst Professor, Post Graduate and Research Department of Computer Science, Parvathy's Arts and Science College,
Dindigul, India²

ABSTRACT::With the e-Governance initiatives across the world, the translation data is growing exponentially and it has become difficult to analyze, such large volume of data stored at multiple locations with traditional data mining tools. The alternative to large data is to use big data analytical tools. E-Governance refers to providing quality information and government services to citizens and other government and non-government organizations effectively and efficiency. The success of any e-Governance project depends on outcome, which can be efficiently determined by big data analytics. The raw data streams in big data systems of e-Governance activities increases computational complexity and resources consumption in cloud-enable data mining systems. This paper is introducing the concept of pattern based data sharing in big data environments with Garbage collection. The proposed methodology enables local data processing near the data sources and transforms the raw data streams into actionable knowledge patterns. These knowledge patterns have dual utility of availability of local knowledge patterns for immediate actions as well as for participatory data sharing in big data environments. The storage of raw data in big data environment needs garbage collection for handling high object churn and large heap sizes. The region based memory management is used to implement the garbage collection in the proposed pattern based big data sharing. This region based memory management on pattern based Data sharing in big data environment has the wide potential to be applied in numerous application areas with memory safety and programmability.

KEYWORDS: E-government, Big data, Garbage Collection, pattern based data sharing.

I. INTRODUCTION

E-Government is the use of information and communication technologies (ICTs) to improve the activities of public sector organizations. With the e-Government initiatives across the world especially in India, the transactional data is growing exponentially and it has become difficult to analyze, such large volume of data stored at multiple locations with traditional data mining tools. The alternative to analyze large data is to use big data analytical tools. The tools having capabilities of mining large data sets in distributed environment help in examining and analyzing e-government projects. E-governance refers to providing quality information and government services to citizens and other government or non-government organization effectively and efficiently. Using big data analysis both service provider and citizens availing services will be benefited. It will help government in making better decision and people in getting timely and reliable information.

Instead, big data offers new opportunities for discovery, value creation, and rich business intelligence for business support in any organization, whether private or public. Comparatively with big data, data warehouse can have experienced a lot of problems, including slow time to build, lack of agility as well as inability to effectively integrate

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

disparate systems. Therefore, many recent studies suggested that big data can replace the data warehouse in the public or private sector. In the similar vein most of the countries steadily moving towards digitizing most of the government departments, scheme and service in federal, state and local levels. Due to increase digital literacy of the citizens and availability of data intensive network access that is emerging, the need for electronic government initiatives is one of the increases. Interestingly, gathering and Processing of the vast amount of data it's not new to humanity. The news is related to process a complex data extracting actionable business intelligence or big insight from data. All these characters are considered as a nature of big data for cleaning and organizing unstructured data through new technologies. Many popular systems for big data are implemented in high level programming languages with automatic memory management via garbage collection.

II. RELATED WORK

With the e-Governance initiatives across the world especially in India, the transactional data is growing exponentially and it has become difficult to analyse such large volume of data stored at multiple locations with traditional data mining tools. The alternative to analyse large data is to use Big Data analytical tools. This explained by Preet Navdeep et al[4]. This paper focuses on role of Big Data analytics in e-Governance projects. Further, it discusses the Big Data Analysis tools and techniques that would be supportive for making better decisions in serving citizens under e-Governance Projects. Many popular systems for processing "big data" are implemented in high-level programming languages with automatic memory management via garbage collection (GC). However, high object churn and large heap sizes put severe strain on the garbage collector. As a result, applications underperform significantly. This explained by Ionel Gog et al [2]. In this paper they propose to use region based memory management instead of GC in distributed data processing systems. In these systems, many objects have clearly defined lifetimes.

The staggering growth in Internet of Things (IoT) technologies is the key driver for generation of massive raw data streams in big data environments. This huge collection of raw data streams in big data systems increases computational complexity and resource consumption in cloud-enabled data mining systems. This explained by Aisha Batool et al [3]. In this paper, they are introducing the concept of pattern-based data sharing in big data environments. The proposed concept has the wide potential to be applied in numerous application areas. Big Data offers new opportunities for discovery, value creation, and rich business intelligence for decision support in any organization, whether private or public. Comparatively with Big Data, data warehouse has experienced a number of problems, including; slow time to build, lack of agility as well as inability to effectively integrate disparate systems. This explained by Azizah Bt Haji Ahmad et al[4]. In this paper, the researchers seek to review the buzzwords related to the using Information and communication technologies (ICTs) in public sector, namely; electronic government (or e-government), data warehouse and big data.

Most "Big Data" systems are written in managed languages, such as Java, C#, or Scala. These systems suffer from severe memory problems due to the massive volume of objects created to process input data. Allocating and de allocating a sea of data objects puts severe strain on existing garbage collectors (GC), leading to high memory management overheads and reduced performance. This explained by Lu Fang et al[5].

This paper describes the design and implementation of "Big Data" friendly garbage collector that provides high throughput and low latency for all JVM-based languages. Big data is a term for massive data sets having large and complex structure with the difficulties of storing, visualizing and analyzing for further processes or results. With the prevalence of Cloud Computing today, there are so many technologies emerging on the Internet and generating huge volume of data. This explained by Varuna Upadhyay et al[6]. In this paper the increasing of big data problem in industry due to the inability of conventional database systems or process the big data sets within tolerable time limits. This paper presents an overview of big data and its architecture.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

IV. E-GOVERNMENT INITIATIVES BETWEEN DATA WAREHOUSE AND BIG DATA

E-government initiatives cannot improve the efficiency and transparency of government departments can also use the system to collect large amount of data, through the establishment of proper decision making system and decision support model that provides a scientific basis for decision making at all levels of government. Thus, e-governance produces a huge amount of data that need to be supported by a new type of e-Infrastructure capable to store, distribute and process. Government agencies deal with enormous of data; therefore, previously in order that such data put to an effective use in facilitating decision-making, a data warehouse is constructed over the historical data.

It permits several types of queries requiring complex analysis on data to be addressed by decision-makers. According to literature, data warehouse concepts are adopted in many government sectors like healthcare, agriculture, education, social security fund, pollution control, electronic voting, rainfall prediction, customer complain, road traffic violation and crime control. However, there are considerable disadvantages when using data warehouse in the government sector because data warehouse build can have high costs. Instead, recently most government departments have been used the model technologies to eliminate the boundaries and speed the process pace. With increasing in use of mobile devices, smart sensors and computing, huge amount of data's are being generated. This ever expanding digital information needs more complex storage, processing, security and disposition features. Big data allows government to improve public sector administration and assist global organization in analyzing information to device strategic planning. Although the business sector is leading has begun to derive insight to help support decision making in real time from fast-growing in-motion data from multiple sources, including the web, biological and industrial sensors, video, email, and social communication.

V. GARBAGE COLLECTION FROM BIG DATA

Garbage collection is the systematic recovery of pooled computer storage that is being used by a program when that program no longer needs the storage. This frees the storage for use by other programs when that program or process with other program. It also ensures that a program using increasing amounts of pooled storage does not reach its quota.

Garbage collection is an automatic memory management feature in many modern programming languages such as java and language in .NET framework. Languages that use garbage collection are often interpreted or run within a virtual machine like the JVM. In each case, the environment that runs the code is also responsible for garbage collection.

In this paper, argues how collect garbage using big data systems. General purpose garbage collectors are not tuned to specific application patterns. While specialized concurrent real-time garbage collectors eliminate GC pause, these collections still have to scan a large heap. Region based memory management works well when similar objects with known lifetimes are handled. This information is available in distributed data processing system. Many such systems are based on a model of actors communicating via message passing.

VI. PATTERN BASED DATA SHARING IN E-GOVERNANCE

This introducing the concept of pattern based data sharing in e-government environment. The proposed methodology enables local data processing near the data sources and transforms the raw data streams into actionable knowledge patterns. These knowledge patterns have dual utility of availability of local knowledge patterns for immediate actions as well as for participatory data sharing in E-government environments using big data tools.

VII. PROPOSED SYSTEM

E-Government can be defined as the use of information and communication technologies by governments to augment the collection and quality of information and services provided to citizens, business, society and other government or non-government agencies in an efficient, cost-effective manner.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

A. BIG DATA PROBLEM

Various big data environments have emerged after the expansion of big data ecosystem in multiple dimensions. These environments vary from in-lab scientific experiments to generation and collection of remote sensory data from heterogenous data sources. A large number of data Environments have emerged for social media and business analytics system as well.

B. VOLUME

The size of the data produced in the big data systems refers to volume. Generally a data size which could not be easily processed by conventional system is known to be volume of big data.

C. VELOCITY

The Speed of incoming data streams determines the velocity property of big data system. Velocity is the key challenge in big data systems. Big data system handles velocity in two way raw data and online data analysis

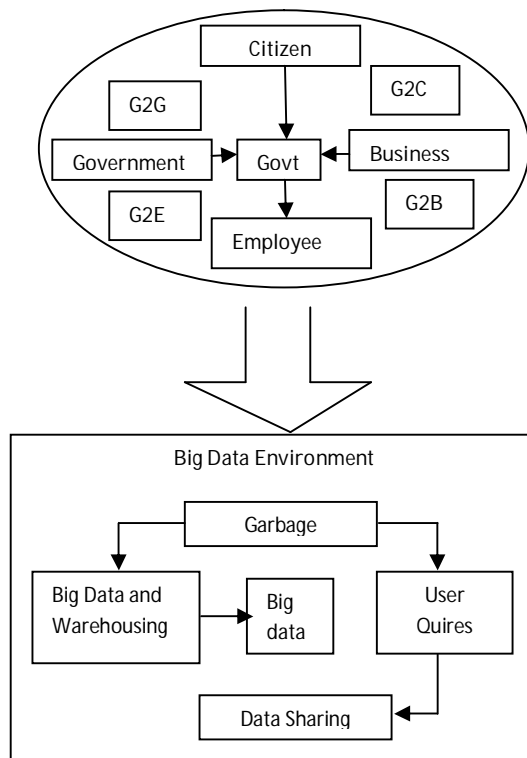


Figure 2

The above figure explain about the overall structure of this paper.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

D. VARIETY

The Heterogeneous of data sources and data formats in big data systems are represented by the formats in big data systems are represented by the variety property of big data. The Variety property of big data improves the overall effectiveness of the System.

E. VARIABILITY

Big data is attributed with variability. The variable data rates causes variable computational overhead in peak load times therefore a proper handling of variability property increases the usefulness of big data systems.

F. VALUE

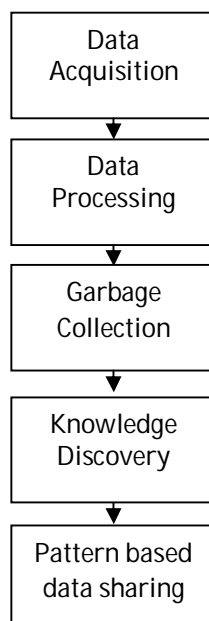
The interesting of uncovered knowledge patterns is represented by the value property of big data systems.

VIII. COLLECTING GARBAGE

Garbage collection techniques which are able to operate in a new context: distributed systems. Authors see that their application to distributed system poses new problems and that there is currently no algorithm available solving, or even addressing, allow them. The presented algorithms are structured according to their belonging family: reference counting, mark-and-sweep and stop and-copy garbage collectors. Hybrid collectors that simulation lousy applies different techniques are also presented.

IX. PATTERN BASED DATA SHARING IN BIG DATA

The raw data streams in big data systems are collected at central data stores for lateral analytics. The increasing growth in MDMS creates an opportunity. This proposed data sharing strategy has some key advantages.



The data acquisition is done using on-board sensory and non-sensory data sources after that data pre-processing and fusion process takes place at this stage. A variety of data preprocessing techniques for future extraction, feature

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

reduction, noise and dimensionally reduction are applied at this stage. In addition, the data pre-processing techniques to handle missing values are also handled at this stage.

The extracted data are ready to Garbage collecting using big data tools. Pre-processed data is analyzed using different data mining algorithms for classification, clustering and association rule mining algorithms. The interesting patterns are synchronized with remote data stores in cloud environments the shared patterns are further analyzed in big data environments using different big data mining techniques. In remote data management the shared buttons are managed in cloud environments using different structured and un-structured data management systems.

X. RESULT AND DISCUSSIONS

The raw data Streams in big data systems are collected at central data stores for lateral analytics. The increasing growth in e-governance applications creates an opportunity to handle big data problem and increasing data complexity near the data sources. Here the e-governance local servers collect, process and analyze the data streams locally. The result and knowledge patterns are shared for future correlation analysis in e-governance big data systems. In this research the pattern based data sharing is applied on Aadhaar data system, which is the India's biggest data source.

UID Aadhaar is one of the prestigious projects of Indian government where Biometric Card with Unique Identification Number is issued to every citizen. UID Aadhaar is a unique identification project undertaken by the Unique Identification Authority of India (UIDAI) established in 2009. The authority maintains a database of residents containing the biometric and other data. All these numbers are stored in a centralized database and linked to the basic demographics and biometric information such as photograph, ten finger prints and iris for every individual. In India this is the biggest source of big data.

In this research, the Aadhaar data are collected from various client systems, and processed by the local server; at this stage the pre-processed data is analyzed using different data mining algorithms for classification, clustering and association rule mining algorithms. In this processed work fuzzy Inferences are used to perform data analytics. The fuzzy inference converts the raw data into actionable knowledge patterns.

In this research Aadhaar big data it stored and processed by hadoop eco-system using hive. Hadoop is an open source frame work that provides solutions for handling big data. Apache's pig and hive are two major components of hadoop to use Map-Reduce library. Hive is SQL like quires for hadoop. Hive is an excellent tool for analytics an Aadhaar Big data system. In this Implementation hive is a SQL like language, so it is easy to implement fuzzy inferences. The Fuzzy inferences are responsible to produce actionable knowledge patterns for raw Aadhaar Big data set

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

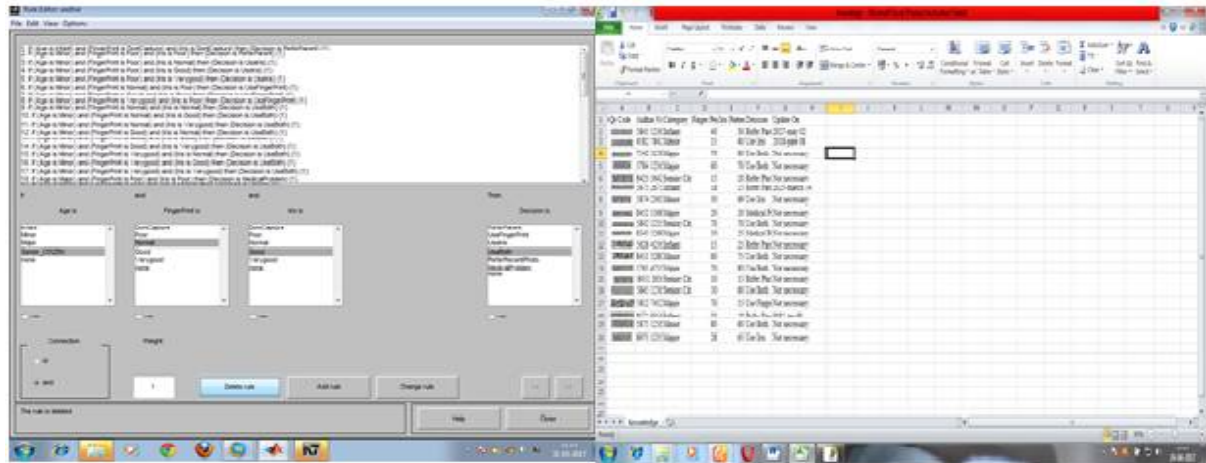


Figure 3

Fuzzy Inference system and actionable data set

Many popular systems for processing big data are implemented in high level programming languages with automatic memory management via garbage collection. This research work processed a region based memory management system. Regions, can be memory-safe and could be inferred automatically. This proposed memory management scheme results that region based memory management reduces emulated run time by 36% for the fuzzy inference data analytics jobs. However the fuzzy inference systems for the high level data analytics on Aadhaar system do not need to be aware of regions at all. So use the memory safe regions for Booming the memory. This region based memory management can be carried out by pig. Because pig provides the scripting language for hadoop.

Many popular systems for processing big data are implemented in high level programming languages with automatic memory management via garbage collection.

This research work processed a region based memory management system. Regions, can be memory-safe and could be inferred automatically. This proposed memory management scheme results that region based memory management reduces emulated run time by 36% for the fuzzy inference data analytics jobs. However the fuzzy inference systems for the high level data analytics on Aadhaar system do not need to be aware of regions at all. So use the memory safe regions for Booming the memory. This region based memory management can be carried out by pig. Because pig provides the scripting language for hadoop.

Execution time comparison (Execution time for Brooming)

Process	Traditional Garbage Collection	Region Based Memory Management
Knowledge Pattern Creation	40 Ms	27 Ms
Pattern Extraction	60 Ms	43 Ms
Pattern Matching	55 Ms	39 Ms

(Table 1)

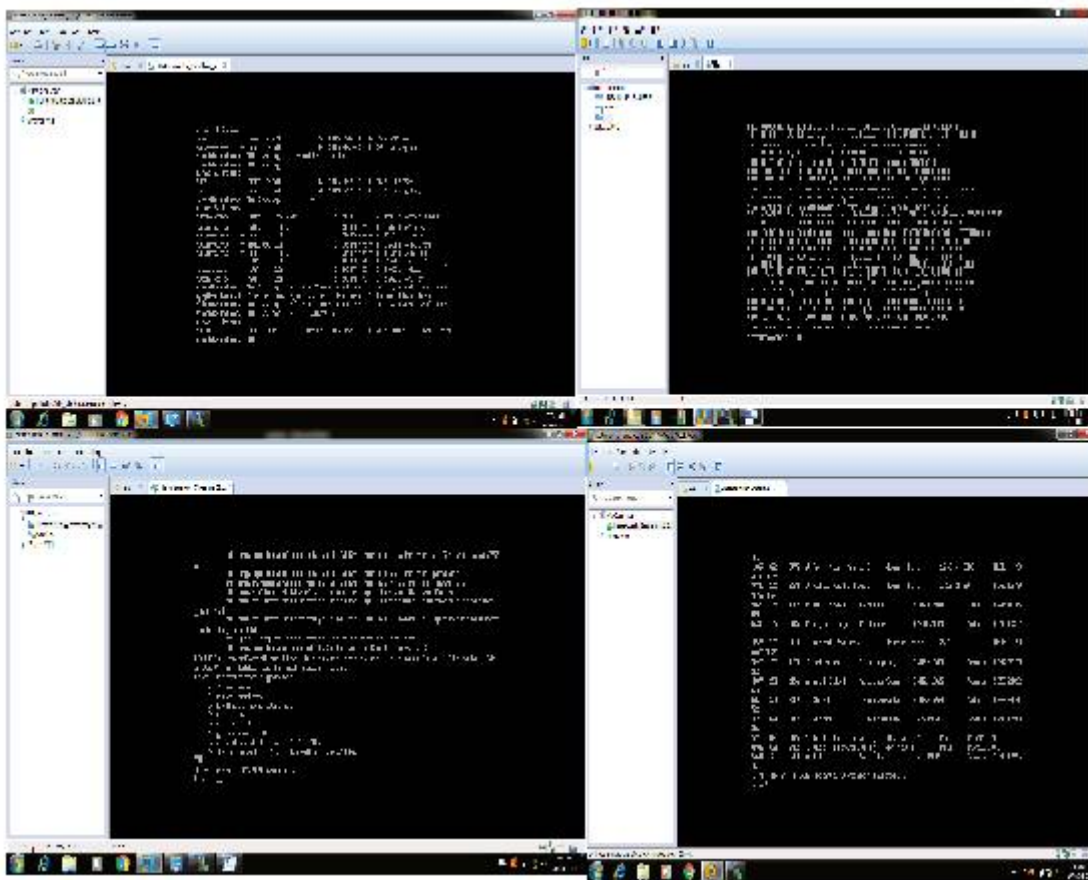
International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

This table compares the execution time of perform brooming operation. When process the knowledge pattern in big data environment, between traditional garbage collection and region based memory management. This table also shows that the region based memory management utilizes the low execution time, which may speed up the pattern based data analytics on big data environment.



Load the file into hodoop and hive(figure 4)

This Figure explains, the Aadhar dataset are loaded into the hadoop. In here I used hive to convert the data's into the hive table in the background the the Region based memory management are used. The hadoop is one of the best big data languages

XI. CONCLUSION

Big data analytics is transforming the way government is using sophisticated information technologies to gain insight information technologies to gain insight from their data repositories to make informed decisions. Big data analytics has the ability to change the landscape of e-governance projects like how data is generated, maintained, analyzed and used for taking future decisions based on the results produced. The increasing number of v's in big data complexity which increases computational and communicational burdens in big data environment. Whenever storing the raw data in big data environment needs automatic garbage collection to manage large heap size and to provide memory safety. This proposed region based memory management on pattern based data sharing in Big data



ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

environment can provide efficient memory safety, programmability and reduce the data complexity. In addition, the availability of local language patterns in user environments enhance the utility of big data systems in communication efficient way.

REFERENCES

- [1] Preet Navdeep, Dr.Manish Arora and Dr.Neeraj Sharma "Role of Big Data Analyzing E-government Projects". Vol 6, Issue 2, Apr-jun 2016.
- [2] Lonel Gog Jana Giceva, Mauel Costa etc Broom:" Sweeping out garbage from Big Data Systems".
- [3] Aisha Batool, Muhammad Habib ur Rehman, "Pattern based data sharing on e-governance". Vol.8, No.4 (2015), pp.11-18
- [4]Azizah Bt Haji Ahmad, 'E-Government initiatives from data warehouse to big data". vol. 3 issue 8, August 2016.
- [5] Lu Fang,brian demsky etc "A High-Performance Big-Data-Friendly Garbage Collector".
- [6] Varuna Upadhyay, Dr. Ashok Jain "A Perspective on Big Data Technologies". Vol 5, issue 1, January 2015.