# Role of Big Data Analytics in Analyzing e-Governance Projects

Preet Navdeep[1], Dr. Manish Arora[2] and Dr. Neeraj Sharma[3]

## Abstract

With the e-Governance 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-Governance projects. E-Governance refers to providing quality information and government services to citizens and/or other government/non-government organizations effectively and efficiently. Using Big Data Analysis, both service provider (here Government) and citizens availing services will be benefited. It will help Government in making better decisions and people in getting timely and reliable information. The success of any e-Governance project depends on outcome rather than output in terms of effective delivery of services. Another feature of Big Data Analysis is to determine the outcome. 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.
**Keywords:** Big data; Data mining; Big Data Analytics; E-Governance

## 1. Introduction

The advances in Information Technology (IT) services and hardware have led to the generation of large amount of data referred to as Big Data. There are projects/systems in e-government where this big data needs to be processed and analyzed for better and effective decision making. Due to the fact that the data generated has different structure and formats, it is difficult to analyze such large data using traditional analyses tools. Big data analytics can manage this dynamic nature of Big data, keeping it secure by applying the correct analytical technique to use the information in an effective manner. It has the capacity to interact with huge Volume, Velocity and Variety of data. With its scalable nature, it can expand the technologies that can correlate data and produce actionable results. Big data analytics can help the government in providing its services directly to its citizens. It has the ability to recognize patterns in a set and make predictions regarding past experiences and provide results for taking future actions.

Big Data Analytics refers to the use of advanced analytic techniques against very large and different data sets that include structured/semi-structured/unstructured data and of different sizes. Big data is the term applied to data sets whose size and type is beyond the ability of traditional relational databases to capture, manage and process the data. Big data is defined in terms of 3 V's i.e. Volume, Velocity and Variety [Gartner].  Analyzing Big data allows analysts, researchers, government and business users to make better and faster decisions using data that was previously inaccessible or unusable. Using advanced analytics techniques, government or business agencies can analyze previously untapped data sources independent or together with their existing enterprise data to gain new insights resulting in significantly better and faster decisions.

---

[1] Research Scholar, PTU, Jalandhar
[2] Joint Director, NIELIT, Chandigarh
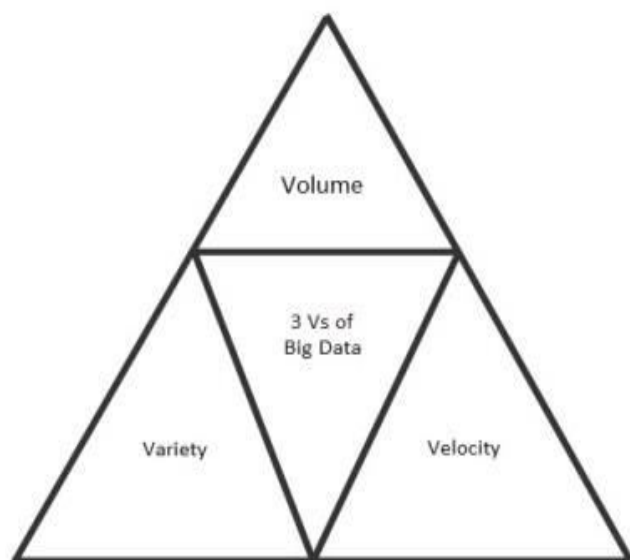[3] Dean & Professor, GJIET, Shambhu Kalan

Fig.1 3Vs of Big Data

e-Governance can be defined as the use of information and communication technologies (ICTs) by governments to augment the collection and quality of information and services provided to citizens, businesses, society and other government/non-government agencies in an efficient, cost-effective and effective manner.
There are four models of e-Governance framed based on the services provided by it.

1. Government to Citizens (G2C)
2. Government to Employees(G2E)
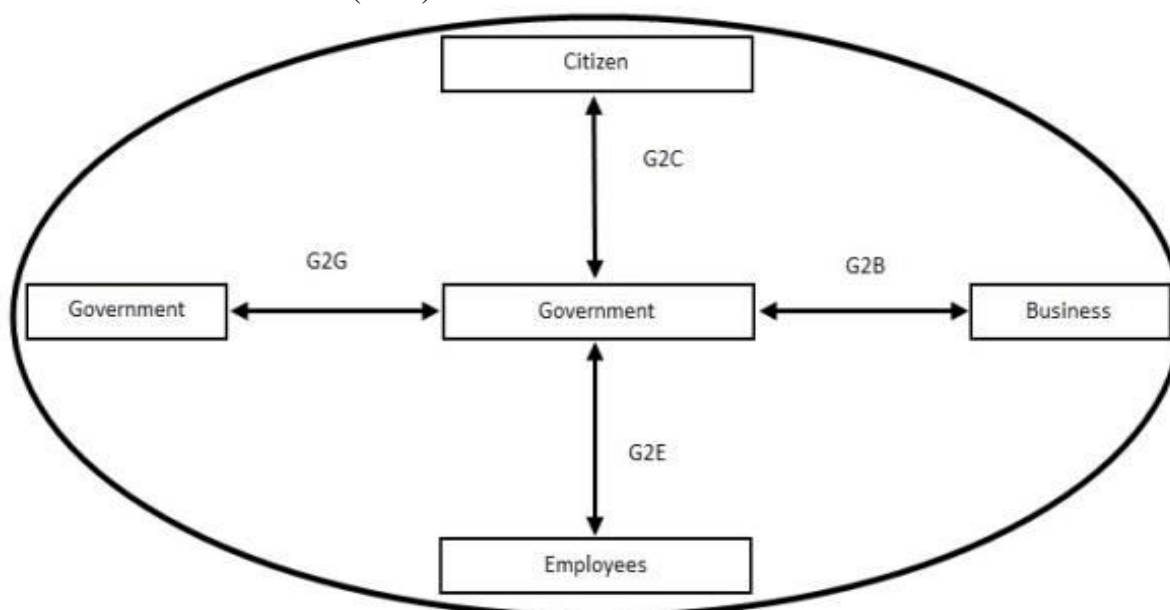3. Government to Government (G2G)
4. Government to Business (G2B)



Fig.2 Four Models of e-Governance
This paper focuses on Big data analytics and its role in analyzing e- Governance projects. The success of any e- Governance project depends upon the effective delivery of services that will

help in better decisions in serving citizens. Section II discusses about the working of Big data Analytics and various tools /techniques used for analysis of Big data. In the next section, various e-Governance projects undertaken by Indian government are discussed. Further benefits and challenges of using Big Data Analytics with e-governance are also discussed.

## 2. Working of Big Data Analytics

In Big data, data is broken into multiple data sets and stored on various nodes of a cluster environment. One of the nodes keeps track of the data stored at various nodes by creating index of the data. When a request for data search appears, first index is checked and the based on that required data is found and fetched. This helps in making the search process faster even if data is too bulky and large. For example, twitter and facebook are using the concept of Big data. Data from any database management system like Oracle and MySql can be imported into Big data. Through data mining activities, data can be collected from various servers. Then data analytics techniques are applied on the data retrieved and reports are generated. After that reports are analyzed to find out the outcome or behavior in regard to a particular condition and the future action plans are prepared.

According to various studies done ,it has been observed that in the year 2008 enterprise servers had processed about 9.57 * 1021 bytes and this number is doubled every two years[9].So, managing such large data using simple techniques is quite difficult. But if this Big data is analyzed using Big data analysis techniques, the results produced would be much reliable and efficient and hence help in taking effective decisions. e.g. Walmart saves near about 2.5 petabyte of data every hour [7] of its purchase history to do analysis for future using Big data analysis techniques.

### A. Data Analytics techniques

There are mainly 7 techniques available for analyzing of Big data. They are:-
1. Association rule learning
2. Classification tree analysis
3. Genetic algorithms
4. Machine learning
5. Regression analysis
6. Sentiment analysis
7. Social network analysis

### 1. Association Rule Learning

Association rule learning is a method for discovering interesting correlations between variables in large databases. It was first used by major supermarket chains to discover interesting relations between products, using data from supermarket Point-of-Sale (POS) systems. Association rule learning helps in:
  • placing products in better proximity to each other in order to increase sales.
  • extracting information about visitors to websites from web server logs.
  • analyzing biological data to uncover new relationships
  • monitoring system logs to detect intruders and malicious activity.

### 2. Classification Tree Analysis

Statistical classification is a method of identifying categories that a new observation belongs to. It requires a training set of correctly identified observations – historical data in other words.

Statistical classification is used to:
- Automatically assign documents to categories
- categorize organisms into groupings
- develop profiles of students who take online courses

*3. Genetic Algorithms*

Genetic algorithms are inspired by the way evolution works – that is, through mechanisms such as inheritance, mutation and natural selection. These mechanisms are used to "evolve" useful solutions to problems that require optimization.

Genetic algorithms are used to:
- Schedule doctors for hospital emergency rooms
- return combinations of the optimal materials and engineering practices required to develop fuel-efficient cars.
- generate "artificially creative" content such as jokes.

*4. Machine Learning*

Machine learning includes software that can learn from data. It gives computers the ability to learn without being explicitly programmed, and is focused on making predictions based on known properties learned from sets of "training data."

Machine learning helps in:
- distinguishing between spam and non-spam email messages.
- learning user preferences and making recommendations based on this information.
- determining the best content for engaging prospective customers.
- determining the probability of winning a case, and setting legal billing rates.

*5. Regression Analysis*

At a basic level, regression analysis involves manipulating some independent variable (i.e. background music) to see how it influences a dependent variable (i.e. time spent in store). It describes how the value of a dependent variable changes when the independent variable is varied. It works best with continuous quantitative data like weight, speed or age.

Regression analysis is used to determine how:
- Levels of customer satisfaction affect customer loyalty.
- The number of support calls received may be influenced by the weather forecast given the previous day.
- Neighborhood and size affect the listing price of houses.

*6. Sentiment Analysis*

Sentiment analysis helps researchers determine the sentiments of speakers or writers with respect to a topic.

Sentiment analysis is used to help:
- improve service at a hotel chain by analyzing guest comments.
- customize incentives and services to address what customers are really asking for.
- determine what consumers really think based on opinions from social media.

*7. Social Network Analysis*

Social network analysis is a technique that was first used in the telecommunications industry, and then quickly adopted by sociologists to study interpersonal relationships. It is now

being applied to analyze the relationships between people in many fields and commercial activities. Nodes represent individuals within a network, while ties represent the relationships between the individuals.

Social network analysis is being used to:
   • see how people from different populations form ties with outsiders.
   • find the importance or influence of a particular individual within a group.
   • find the minimum number of direct ties required to connect two individuals.
   • understand the social structure of a customer base.

### B. Big data analysis Tools

### 1. APACHE HADOOP

   Hadoop is a batch-oriented data processing system. It works by storing and tracking data across multiple machines, and can scale to thousands of servers. Hadoop is designed to process lots of data that doesn't fit nicely into tables.  It's used in situations where you want to run analytics that are deep and extensive, like clustering and targeting. The underlying technology was originally invented by Google, which used it for indexing the web and examining user behavior to improve performance algorithms.

### Working of Hadoop
   1. The data is loaded into Hadoop.
   2. Hadoop breaks up and distributes the data across multiple machines. Hadoop keeps track of where the data resides, and can store data across thousands of servers.
   3. Hadoop executes Map Reduce to perform distributed queries on the data. It maps the queries to the servers, and then reduces the results back into a single result set.

### Business Areas
   1. Financial service providers, such as credit card providers, use it for targeted marketing and fraud detection.
   2. Retailers use it for predicting what customers want to buy.
   3. Human Resources departments are using Hadoop to support their talent management strategies and understand people-related business performance, such as identifying top performers and predicting turnover in the organization.

### 2. Hadoop Map Reduce
        Map Reduce is a system for big data parallel processing. It is developed by Google and is used for processing large amount of data using a large number of processors. The processing is divided into two phases: Map and reduce, where each stage has input and output parameters in the form of key/value pairs. User defines map function, and as a result a number of key/value pairs are received. Reduce function is the applied to the resulting key/value pairs and extracts all the values with the same key.

### 3. HIGH PERFORMANCE COMPUTING CLUSTERS (HPPC)
        HPPC is an open source data-intensive platform developed by LexisNexis Risk Solutions. The HPPC platform, also known as the Data Analytics Supercomputer (DAS), supports both batch and real-time data processing. It uses both supercomputers, as well as clusters of commodity computers.

*Working of HPPC*

1. Data is loaded into a data refining cluster called Thor. A Thor cluster is the functional equivalent of Hadoop Map Reduce.
2. The data is processed in a cluster used for online query processing and data warehousing called Roxie.
3. Programmers develop solutions using Enterprise Control Language (ECL).

*Business Areas*

1. LexisNexis is using it to manage its collection of 2.3 billion documents.
2. Medicaid is using it to detect fraud and abuse of the system, by identifying suspicious groups of Medicaid
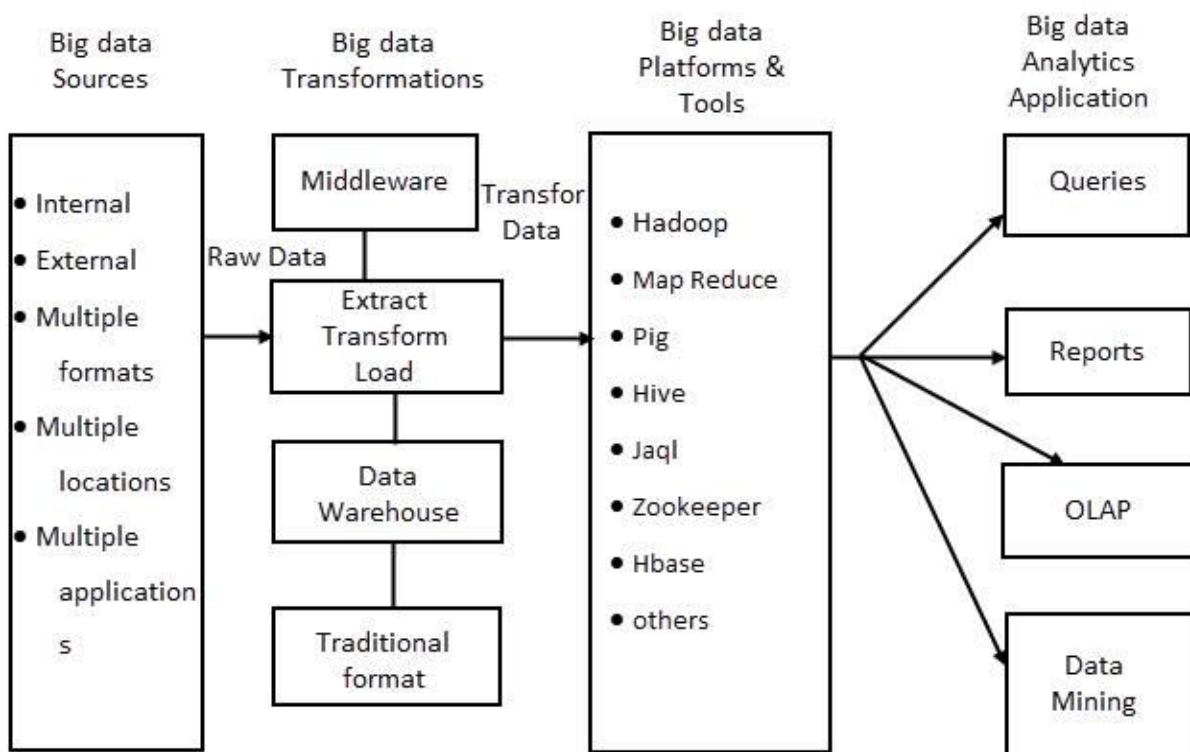recipients who were all living in the same high-end condominium complex.

*C. Conceptual Model of Big Data Analytics*



Fig. 3 Conceptual Model of Big Data Analytics

Fig.3 explains the conceptual architecture of Big Data Analytics. In this, data which is residing in different formats from different sources and different locations is transformed using different big data analytics tools and platforms. Here, various decisions are made regarding the data approach, distributed design, tool selection and analytics models. Finally, four applications of data analytics i.e. queries, reports, OLAP and data mining are generated.

Table 1. Hadoop Projects

| S. No | Type | Description |
|---|---|---|
| 1 | Ambari | A web interface for managing, configuring and testing Hadoop services and components. |
| 2 | Casandra | A distributed database system. |
| 3 | Flume | Software that collects aggregates and moves large amounts of streaming data into HDFS. |
| 4 | HBase | A non-relational, distributed database that runs on top of Hadoop. HBase tables can serve as input and output for Map Reduce jobs. |
| 5 | HCatalog | A table and storage management layer that helps users share and access data. |
| 6 | Hive | A data warehousing and SQL-like query language that presents data in the form of tables. Hive programming is similar to database programming. |
| 7 | Oozie | A Hadoop job scheduler. |
| 8 | Pig | A platform for manipulating data stored in HDFS that includes a compiler for Map Reduce programs and a high-level language called Pig Latin. It provides a way to perform data extractions, transformations and loading, and basic analysis without having to write Map Reduce programs. |
| 9 | Solr | A scalable search tool that includes indexing, reliability, central configuration, failover and recovery. |
| 10 | Spark | An open-source cluster computing framework with in-memory analytics. |
| 11 | Sqoop | A connection and transfer mechanism that moves data between Hadoop and relational databases. |
| 12 | Zookeeper | An application that coordinates distributed processing. |

## 3. E-Governance Initiatives in India

The vision of e-Governance [2] is to "Make all Public Services accessible to the common man in his locality, through common service delivery outlets and ensure efficient, transparent and reliable information at minimal costs. The main intent of e-Governance is to fulfill five major objectives [3]:a policy framework, enhanced public service, high quality and cost-effective government operations, citizen engagement in democratic processes and administrative and institutional reform. In India, National e-Governance Plan (NeGP) [10] has been formulated by Department of Electronics and Information Technology (DeitY) and Department of Administrative Reforms and Public Grievances (DARPG). According to NeGP, over 1000 e-Governance services have been accessed through NeGP. The e-taal (Electronic Transaction Aggregation and Analysis layer) is the government web portal that provides statistical of transactions done electronically by citizens with e-Governance projects. According to e-taal, Indians have done over 2 billion e-transactions in last one year [1].

Various E-Governance projects in India include:

1) Land record Management Projects: Maintains millions of land records and helps citizens in providing reliable and useful information in shortest time related to it .e.g. Bhoomi (Karnataka),

Gyandoot (Madhya Pradesh), Land records Management system State government of Punjab, Devbhoomi( Uttarakhand).

2) Local Information Projects: Citizens can look for local information such as loan rates, prices of seeds , fertilizers etc using the following E-Governance services like E-JanSampark (Chandigarh), Prajavani (Andra Pradesh, E- Samadhan (Himachal Pradesh).

3) Agriculture: Following are the projects used in agriculture GYANDOOT (Madhya Pradesh), AGMARKNET and SEEDNET (Department of Marketing and Inspection, Ministry of agriculture and Government of India).

4) Disaster Management: Managing Disasters is a very challenging job for the government as it is natural and unpredictable. To handle such disasters, state governments have started e- governance services like Project **Chetana** started by state government of Bihar to deal with natural disasters like floods and earthquakes.

5)   Services provided by this e-Governance service are Issue of Birth/Death certificate, rural water supply and sanitation, conducting various Welfare schemes for the poor and needy people. Projects performing the above services are-E-GramViswa Gram project (gujrat), RajNidhi (Rajasthan), Raj-SWIFT (Rajasthan State's Department of Information Technology)

6) The biggest application of e-Governance in India is UID Aadhaar. It is one of the prestigious projects of Indian government where Biometric Card with Unique Identification Number are 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 of each individual. In India this is the biggest source of Big data. As of 15 October 2015, over 92.5 crore (925 million) Aadhaar numbers have been issued in the project [6].

## 4. Application Areas of Big Data in E-Governance

Table 2. Applications of usage of Big Data Analytics for e-Governance projects

| Area | Applications | Type of decisions using analytics |
|---|---|---|
| Education | Student tracking | Finding various factors that contribute to student success. |
| | School location Planning | Identifying and planning location best suited for building school |
| | Administration | Planning and testing standards of schools |
| Health Care | Disease detection | Finding and reporting the occurrence of diseases |
| | Epidemiology | Identifying the causes, distribution and control of diseases |
| | Disease prevention | Identifying factors that can prevent disease occurrence |
| Public | Problem Analysis | Helps in analyzing problem faced by public and finding solutions to it. |
| | Fulfilling Information needs | Identifying and providing information requirements of citizens |
| Safety | Crime Analysis | Identifying type of crime and areas of occurrence |
| | Record Keeping | Helps in keeping records of all crimes that have occurred in the past |

| | | |
|---|---|---|
| Revenue | Tax assessment | Calculating tax assessments and checking whether they are correct |
| | Fraud Analysis | Analyzing the fraud occurrence |
| Environment | Environment analysis | Understanding which factors contribute to healthy eco-system |
| | Water/Air quality testing | Ensuring water/air standards are met |
| Business | Cost/benefit analysis | Understanding what can lead to be more cost effective |
| | Employee satisfaction | Finding employee expectations |
| | Resource planning | Identifying required resources |

## 5. Benefits of using Big Data Analytics for E-Governance Projects

The major benefits of introducing Big data platform in e-Governance projects are to facilitate government. These benefits are:-

- Improvement in online information and service delivery by government for business Analytics.
- Making processes open and transparent to users.
- Recording and Visualizing government performance through predictive policing.
- Producing insight for new business projects and improving business growth.
- Providing better Customer Services.

## 6. Challenges of Using Big Data Analytics with E-Governance

Though analyzing Big Data using various analytical techniques has proved to be useful for e-Governance projects but still there are some challenges that need to be handled to achieve desirable results. Some of them are:-

- **Privacy:** Big data analytics refers to examining historical data. Each e-Governance project consists of a lot of private information related to various users. This data can be regarding UIDs of individuals, sales/purchase information of a business firm, information of clients of an organization or records of patients of hospital. Such organizations may oppose in revealing their private information. If provided, such information should be preserved properly so that it is safe and beneficial.
- **Security:** Big data generated from e-Governance projects is a combination of large data which is stored on various servers. Storage and security of this big data is one of the major challenges. Processing of this data using data analytics further produces more data in the form of reports. Managing this data and protecting it from unauthorized access and usage should be carefully done.
- **Shortage of qualified and technical people to handle and implement such techniques:** Still it is very difficult to find skilled and qualified people who are proficient in using big data analytics techniques with e-Governance projects.

- **Unavailability and Under-development of relevant software tools, integration of multiple data sources and formats:** Data generated by different e-Governance projects during analysis is of different types and different formats. Sometimes application of relevant software tools is not possible because of its unavailability or that technique/software is still under development.
- **Lack of common standards for information/data representation and interoperability:** No proper standards are defined of how data is to be represented using analytical techniques.

## 7. Conclusion

Big data analytics is transforming the way government is using sophisticated information technologies to gain insight from their data repositories to make informed decisions. This data-driven approach is exceptional, as the data collected via the web and social media is escalating by the second. In the future we'll see the rapid, widespread implementation and use of big data analytics across the government and the industry. As it becomes more mainstream, issues such as guaranteeing privacy, safeguarding security, establishing standards and governance, and continually improving the tools and technologies would gather attention. Big data analytics and applications are at a growing stage of development, but the rapid advances in platforms and tools can accelerate their maturing process. 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. So, the use of Big data analytics techniques should be encouraged as they can prove to be helpful in further improvement, better planning and decision making of e-Governance projects.

## References

[1] "E-Governance transactions in 2013 increased manifolds compared to 2012",http://www. thinkalytic.com/2014/e-Governance-transactions-in-2013-increased  manifolds-compared-to -2012/, Jan 2, 2014

[2] A.G. Picciano,"The Evolution of Big Data and Learning Analytics in American Higher Education", Journal of Asynchronous Learning Networks, pp 9-20,2012

[3]  Alshomrani, Qamar,"Cloud based e-Government: Benefits and Challenges", IJMSE, Vol. 4(6), pp. 15-19,2013

[4] Arvind Sathi,"Big Data Analytics: Disruptive Technologies for Changing the Game", 2012

[5] Chen, Hsieh," Big Data for digital Government: Opportunities, Challenges, and Strategies", International Journal of Public Administration in the Digital Age", pp. 1-14,2014

[6]  Kalbande, Deshpande, Prof. Popat,"Review paper on use of Big data in E-Governance of India", International Journal for Research in emerging and technology", Vol. 2, pp.395-402, March, 2015

[7] Karthik Kambatta, Giogos Kollias, Vipin Kumar, Ananth Grama,"Trends in big data Analysis", j.Parallel distributed computer,pp 2561-2573,2014

[8] Krishna Kumar Verma,Navita Shrivastava,"Exploring role and associated challenges of data mining technique and it's implementation in E- governance", IJARCET, Volume 4,Issue 8,pp 3473-3477, August 2015

[9] Agnihotri, N., & Sharma, A.K.,"Big Data Analysis and its need for effective E-Governance", IJIACS, Volume 4, pp 219-224, March 2015

[10] NeGP Website. "About NeGP".NeGP Website. Retrieved 17 July, 2014

[11] Prasad, Veera Raghav,"Cloud computing technology for effective e- Governance", IJCSIT, Volume 3(1), pp. 3241-3244,2012

[12] Suh, Vujin, Barac, Radenkovic,"Designing cloud infrastructure for big data in e-government", Jounal of Universal Excellence, pp.A26-A38, March 2015

[13] Woodside, Amiri, Boldrin,"The impact of ICT and Big Data on e- Government", Int'l Conf on Advances in Big Data Analytics,pp. 117-118,2015