



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 10    **Issue:** III    **Month of publication:** March 2022

**DOI:** <https://doi.org/10.22214/ijraset.2022.40874>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Big Data to Implement E-Governance

Hiteshkumar Vora

Vivekanand College for Advanced Computer & Information Science, Surat

**Abstract:** As the world moves forward in the way of process, Information Technology plays a vital role in the development. All aspects of the future will depend on the internet and computers so it is necessary to take the government into the world of Internet. Developing countries like India which have a huge population need extra care of data and information. For that Big Data can become a blessing. Big data technology improves all aspects like security, management, processing etc. This huge amount of data is difficult to analyze and in a normal environment it takes a long time to generate any meaningful output so big data technology can give better output in less time. This paper presents the idea and application of Big data technology in the E-Government in countries like India.

**Keywords:** Hive, Pig, Big Data, Hadoop, Government

## I. INTRODUCTION

When the term big data is used, it becomes necessary to know which amount of data can be considered as the big data. The amount of that which is considered as big data has changed with the time. In 1999 total data generated was around 1.5 exabyte and at that point of the time 1 Gigabyte was considered as the big data. But in the current developing world this idea didn't work. In the year of 2006, total 160 exabyte data was generated which is almost 1000% more than that in 1999. This change was in just 7 year of period. And in today's situation, this number is too huge. At this point neither Zetabyte nor Gigabyte can be considered as big data. In today's world the idea of Big data is differ than in the past. According to many researchers, any data can not be considered as big data on the basis of its size. Any data can be considered as big data if it has the V properties. For the big data it has 7 'V's which are shown in the figure 1.

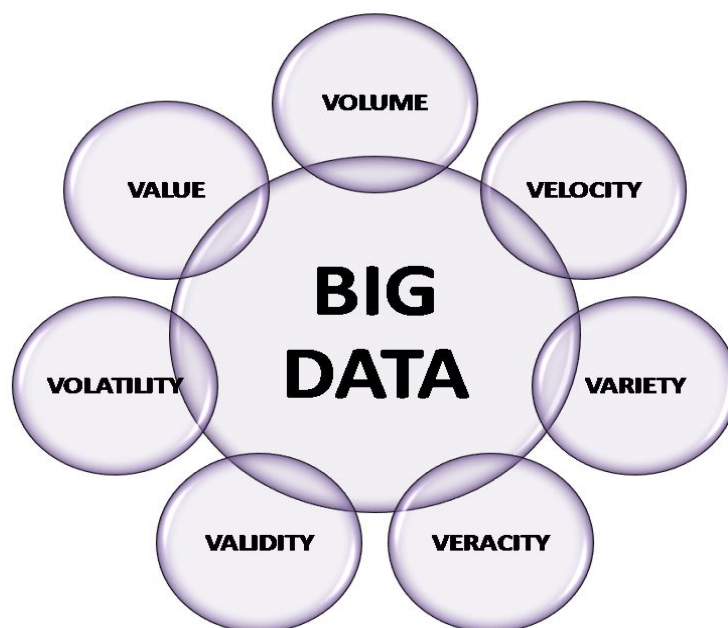


Figure 1: 7 V's of Big data

The first V is Volume of Data, which is concerned with the amount of data, or how big it is. Velocity is the second V, and it refers to the rate at which data is generated. Data generation has expanded considerably in the previous two decades as a result of new sensors and other technologies. The third V stands for Variety, and it refers to the various forms of data that are displayed. All sorts of data are welcomed in big data, whether structured like SQL or unstructured like video, mp3, and so on. The fourth V stands for Veracity, and its main priority is data validity. It's all about the data's veracity or accuracy.

The fifth V stands for Validity, which refers to whether or not the data is appropriate for the application in which it will be used. It is possible that the identical data will be valid for one application but not for the other. The sixth V stands for data volatility, which is all about time and how long data will be stored in the system/application. The data's value is represented as the seventh V, which is distinct from all previous Vs in big data. After the competition, it displays the valuable output.

Because data is generated at a rapid rate in all organisations and sectors, big data is now required in every area. Big data analytics is necessary to handle these data and extract insights from them. Education, healthcare, IoT, government, finance, retail, media and entertainment, e-commerce, telecom, and travel are just a few of the primary applications of big data.

The government is also forced to use big data analytics for numerous purposes in this expanding technological era. Big data might be a godsend for countries like India, where massive amounts of data are generated on a regular basis. Big data may help all citizens as well as the government perform more efficiently, quickly, and smoothly.

## II. E-GOVERNMENT

While talking about use of big data in the government sector it becomes necessary to understand the meaning of E-governance or E-government. e-Governance can be defined as the application of information and communication technology (ICT) for providing government services, exchange of information, transactions, integration of previously existing services and information portals.

According to the Council of Europe E-governance is:

The use of electronic technologies in three areas of public action

- 1) Relations between the public authorities and civil society
- 2) The functioning of the public authorities at all stages of the democratic process (electronic democracy)
- 3) The provision of public services (electronic public services)

### A. Types of Interaction in E-governance

The government interact with the citizen in different way according to the need and profession of the people. The most simple example is that, government provide service citizen, central government provide service to the state government or to local government etc. Most basic model use in the e-governance is as follow.

- 1) G2G i.e. Government to Government
- 2) G2C i.e. Government to Citizen
- 3) G2B i.e. Government to Business
- 4) G2E i.e. Government to Employees

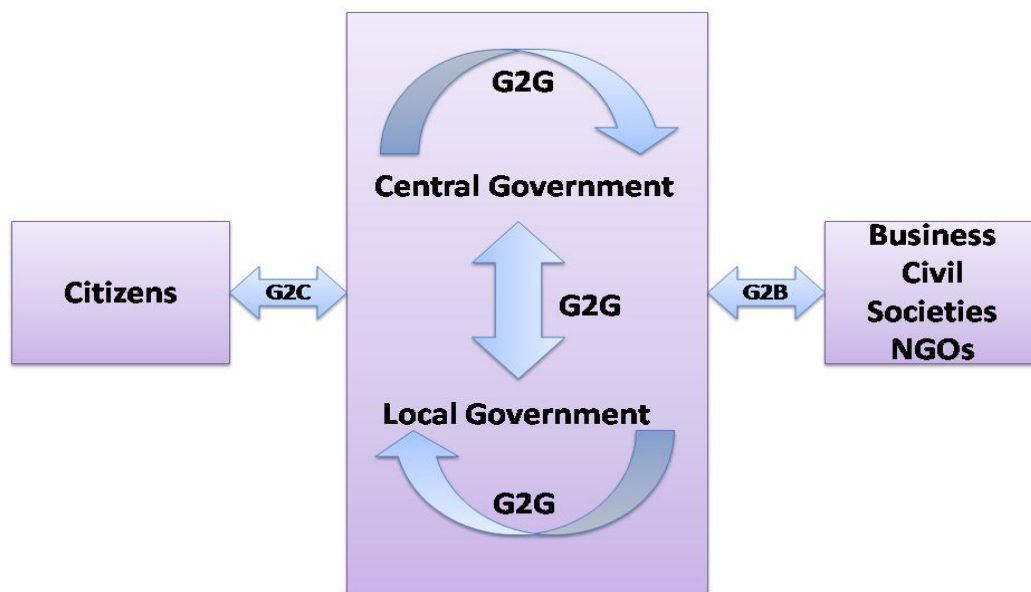


Fig 2 : Interaction in E-governance

As shown in the above figure, an interaction is possible in all the way. The central government provide the information or the service to other department of the central government or to the local government, this process done by the G2G model. For an example, The central government pass the budget for the state government. The citizen can take service of government using the G2C model. For an example the citizen register for the government schemes or pay the tax using E-Tax option. The business/organization can use the G2B model. For example, a business can register for permission of the business or pay the business tax to the E-business tax portal. This can improve the working processes and development of the country become more accurate as well as transparent.

### III. BIG DATA ARCHITECTURE FOR THE GOVERNMENT SECTOR

Any government has a huge amount of data to be handling. This data is considered as the big data as it has 7 V's of the big data. To store this huge amount of data is required to have a specific storage and for that cloud storage becomes a blessing because it provide facilities to store huge data as well as retrieval of the data is possible at any time and from anywhere. Here the architecture of the e-government is present.

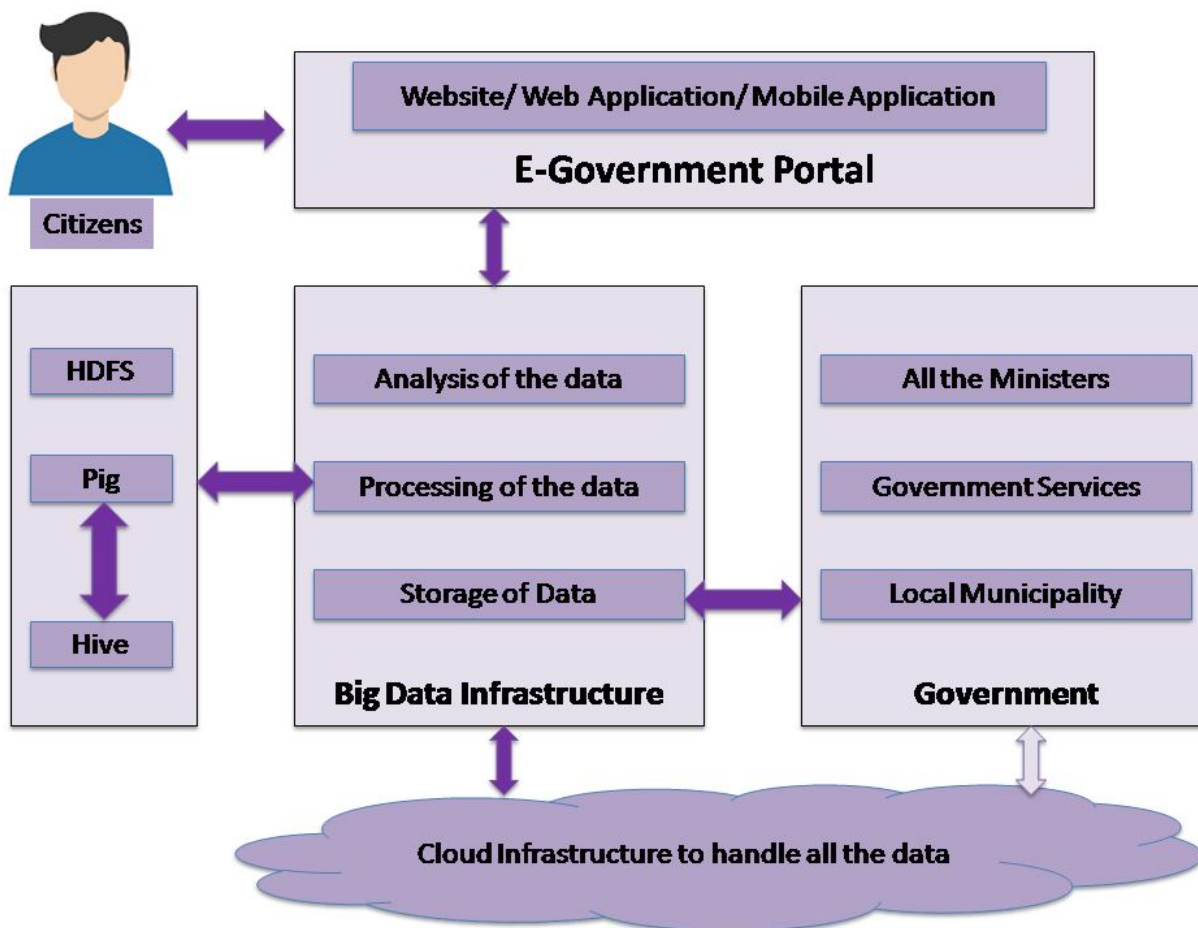


Fig 3 : Architecture of E-government in India

As illustrated in the diagram, all people of the country can access e-government services through one of the three options: a website, a web application, or a mobile application. The government's first responsibility is to make all of the facilities and data available online. Though it appears to be costly, it can save time and make government operations easier. After all of the data is online, it can be stored in the cloud infrastructure. A distributed server is an alternative to the cloud, but for some nations, it is too expensive, thus the cloud can be demonstrated to be a superior solution. Because the data is stored in the cloud, it can be managed via HDFS, and the desired information may be obtained using the Pig and Hive tools. HDFS makes data available in a variety of contexts and can also build clones of it. Because all of the information are available, performing analytics on the data is simple. The entire data management process might be delegated to a government-approved agency.

If the country's population is too large, a single government body cannot handle all of the data. As a result, the simplest answer to this problem is to provide citizens access to the local municipality. All of the data and access to the data of their respective regions is available to all municipalities. This local access can assist in taking immediate action in any situation. Because centralization is also essential, those matters can be managed by the state or country's ministers. The state government will have greater access over local municipalities, while the central/union government will have more authority above the state government. In some cases, such as medicine development, the federal government can handle it, while the state government can handle weather forecasting and the local municipality can control traffic. Not only can data be managed, but also predictions may be made with big data technology.

**A. HDFS- Hadoop Distributed File System**

HDFS is utilized for the storage permission in a hadoop cluster. It is designed in a such a way so that cost can be reduce to use. In HDFS file system, the normal/accommodated computers are used for storing and processing the data. In HDFS data stored in the large chunk. It prefer to manage the few large chunks over the more small chunks. By adopting this methodology it provide high accessibility and to the storage level as well as provide fault tolerance to the user. HDFS serve in such a manner so that the data is available to use for the user at any point of the time. HDFS data storage is divided in to the two major parts.

- *Name Node or Master Node:* Name node works as the master nodes in the HDFS which guide the data nodes to work. Name nodes are used for the storing the meta data i.e. data about the data. Normally Name node store the information like the, location of node, user of the node, size of the file etc. This information help to find the information quickly. Name nodes are the reason for the quick retrieval of the data in HDFS. Basically Name nodes are stored in the master/server computers for the more accuracy.
- *Data Node or Slave Node:* Data node which work as Slave node store the real data in the HDFS. It stores the information in distributed architecture so the that all the information becomes available at any point of time. More the data node means the more information can be handle by the HDFS. To establish the E-governance in country like India where huge amount of information generate, it required more and more data nodes to store those information.

Here the following figure shows the architecture of HDFS.

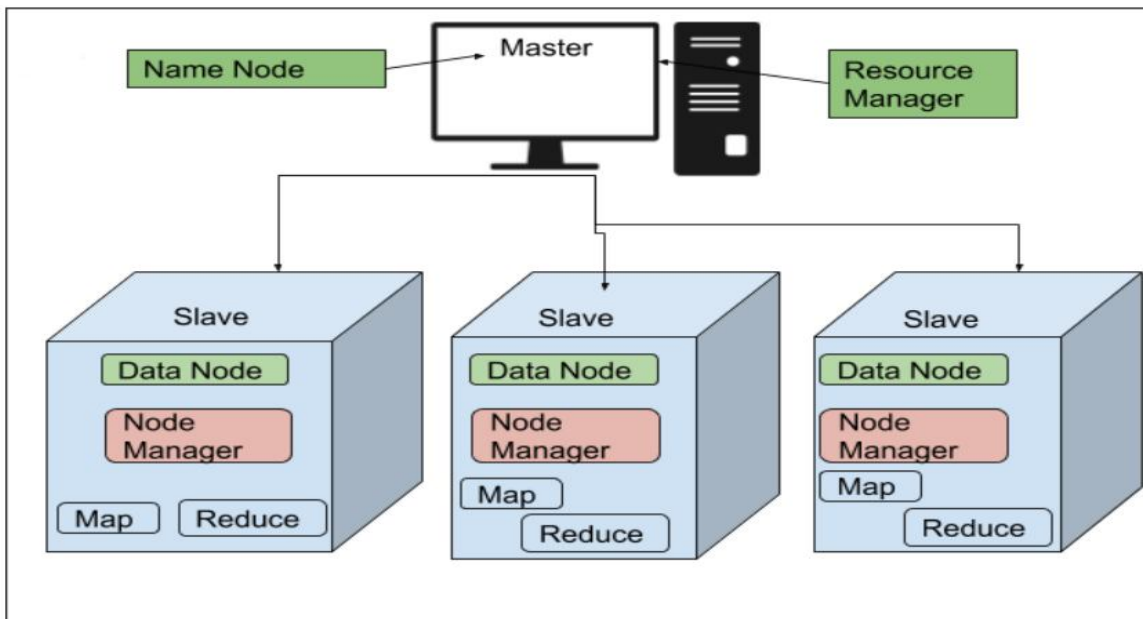


Fig 4 : Hadoop Architecture

As shown in figure 4, all the name nodes are stored in the master from where all the resources are managed. Whereas all the slave nodes are stored in accommodated computers for the cheap storage and calculations. Here all the accommodated device have the function of node manager which manage information of all nodes to be stored on the device. One more functionality is provided with it is map and reduce. It can be considered as the main functionality of the HDFS.

1) *MapReduce*: MapReduce can be considered as an algorithm or the data structure which provide facilities of distributed processing in parallel in a hadoop cluster which make hadoop to work so fast and efficiently. When dealing with the big data technology a serial processing is not a way to use but for that parallel processing is need. In hadop cluster this can be achieved through the MapReduce functionality. MapReduce is the function made up of the Map and Reduce processes.

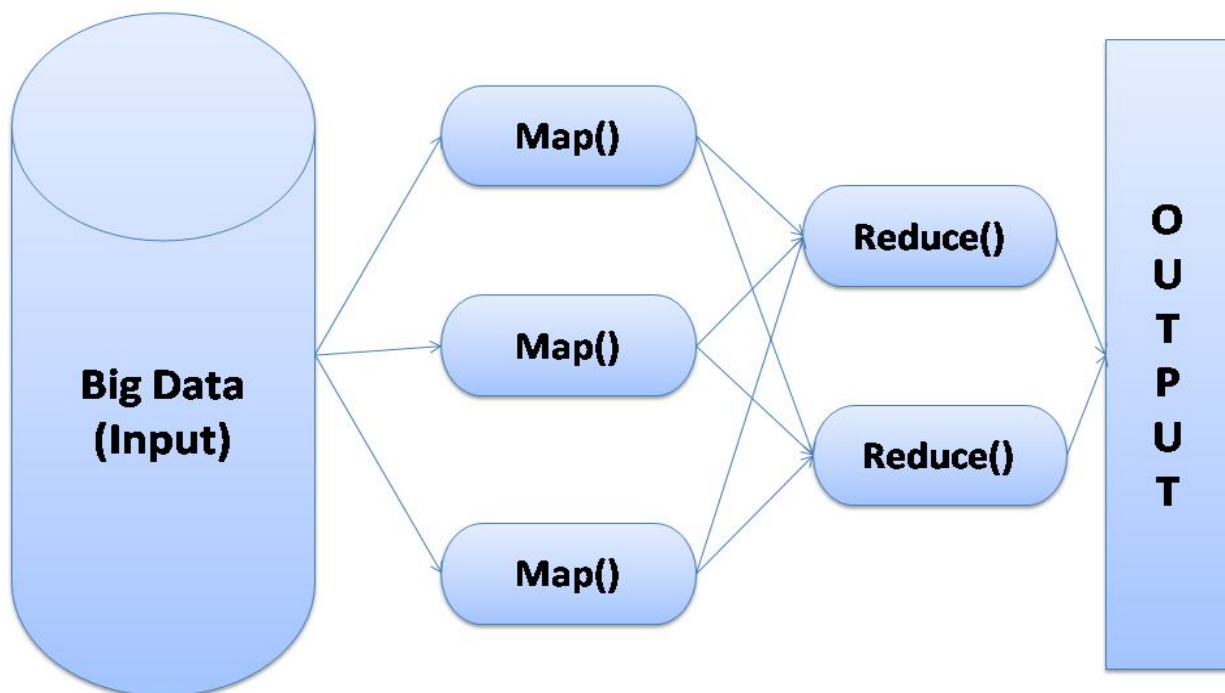


Fig : 5 MapReduce utilization

As shown in the figure 5, all the inputs are provided to the Map() function, then the output of Map() is given to the Reduce() function which used as input in the Reduce () function. After the computation in Reduce(), the final output is generated. Now that while using Big Data, it can see that an Input is passed to Map(). The Input is a collection of information. The Map() function converts DataBlocks into Tuples, which are just key-value pairs. These key-value pairs are now passed to the Reduce function (). The Reduce() method then joins these broken Tuples or key-value pairs based on their Key value and create set of Tuples, and performs operations such as sorting, summing, and so on, before sending them to the final Output Node. The output is finally obtained. Data processing is always done in Reducer, based on the industry's business requirements.

2) *Map Task*

The map function performs the following task.

- 1) *Record Reader*: The main target of recordreader is to breaking down the records and convert it into the key-value pairs in the Map() function. Here key is the locational information and value is the data that are related to it.
- 2) *Map*: A Map is considered as a function which work is to process the Tuples that are get from the recordreader. The map function may or may not generate any key-value pair for those tuples.
- 3) *Combiner*: Combiner has a simple function of combining the key-value pairs. This function is not a necessary work to do. It can be considered as local reducer.
- 4) *Partitionar*: partitionar is responsible for fetching the key-value pairs that is generated during the mapper passes. Here it generate the hashc ode of each key which is also fetched by this partition.

### 3) Reduce Task

The reduce function performs the following task.

- a) *Shuffle and Sort*: Task of the reduce function begins with the Shuffle and Sort. The key-value pair generated by the mapper is transfer to the reduces is known as shuffling. During the shuffling process data are sorted according to the key value. As and when some of the task is done by the mapper function work of reduce started without waiting for whole task to be done so that it is faster.
- b) *Reduce*: The main task of this function is to gather all the tuples and perform sorting and aggregation process on the key-value pair according to the requirement.
- c) *Output Format*: Once all operation done, the key-value pairs are written into the file with the help of the record writer.
- 4) *YARN (Yet another Resource Negotiator)*: MapReduce is based on the YARN Framework. Job scheduling and resource management are two tasks that YARN accomplishes. The goal of job scheduling is to break down a large work into smaller jobs so that each one may be assigned to different slaves in a Hadoop cluster and processing is optimised. Job Scheduler also maintains track of which jobs are more vital, which jobs have higher priority, job dependencies, and other information such as job timing. Resource Manager is used to manage all of the resources that are made available for a Hadoop cluster to run.
- 5) *Hadoop common or Common Utilities*: Hadoop common or Common utilities are nothing more than our java library and java files, or in other words, the java scripts that we require for all of the other Hadoop cluster components. HDFS, YARN, and MapReduce all use these tools to run the cluster. Hadoop Common verifies that hardware failure in a Hadoop cluster is common, requiring Hadoop Framework to fix it automatically via software.

### B. Apache Pig

Pig is a high-level platform or tool for processing massive datasets. It provides a high-level of abstraction for MapReduce computation. It comes with a high-level scripting language called Pig Latin, which is used to write data analysis routines. First, the programmers will use the Pig Latin Language to develop scripts to process the data stored in the HDFS. Pig Engine (an Apache Pig component) converted all of these scripts into a single map and reduce task. However, in order to give a high level of abstraction, these are not visible to programmers. The Apache Pig tool is made up of two key components: Pig Latin and Pig Engine. Pig's output is always kept in HDFS.

#### 1) Apache Pig Has The Following Features

- a) Apache Pig includes a wide range of operators for executing a variety of tasks, such as filters, joins, and sorting.
- b) It is simple to learn, read, and write. Apache Pig is a godsend for SQL programmers in particular.
- c) Apache Pig is flexible, allowing you to create your own own functions and processes.
- d) Apache Pig makes joining operations simple.
- e) There are less lines of code.
- f) Splits in the pipeline are possible with Apache Pig.
- g) The data structure is more complex, hierarchical, and multivalued.

Pig can handle both organised and unstructured data processing.

#### 2) Types of Data Models in Apache Pig

It consist of the 4 types of data models as follows:

- a) *Atom*: It is a atomic data value which is used to store as a string. The main use of this model is that it can be used as a number and as well as a string.
- b) *Tuple*: It is an ordered set of the fields.
- c) *Bag*: It is a collection of the tuples.
- d) *Map*: It is a set of key/value pairs.

### C. Hive

Hive is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data, and makes querying and analyzing easy. Initially Hive was developed by Facebook, later the Apache Software Foundation took it up and developed it further as an open source under the name Apache Hive. It is used by different companies. For example, Amazon uses it in Amazon Elastic MapReduce.

- 1) *Hive is Not*
  - a) A relational database
  - b) A design for OnLine Transaction Processing (OLTP)
  - c) A language for real-time queries and row-level updates
  
- 2) *Features of Hive*
  - a) It stores schema in a database and processed data into HDFS.
  - b) It is designed for OLAP.
  - c) It provides SQL type language for querying called HiveQL or HQL.
  - d) It is familiar, fast, scalable, and extensible.

#### IV. IMPLEMENTATION

All the data of all citizens need to be in the digital form. Starting from the birth to the death all the data and information should be in digital form only.

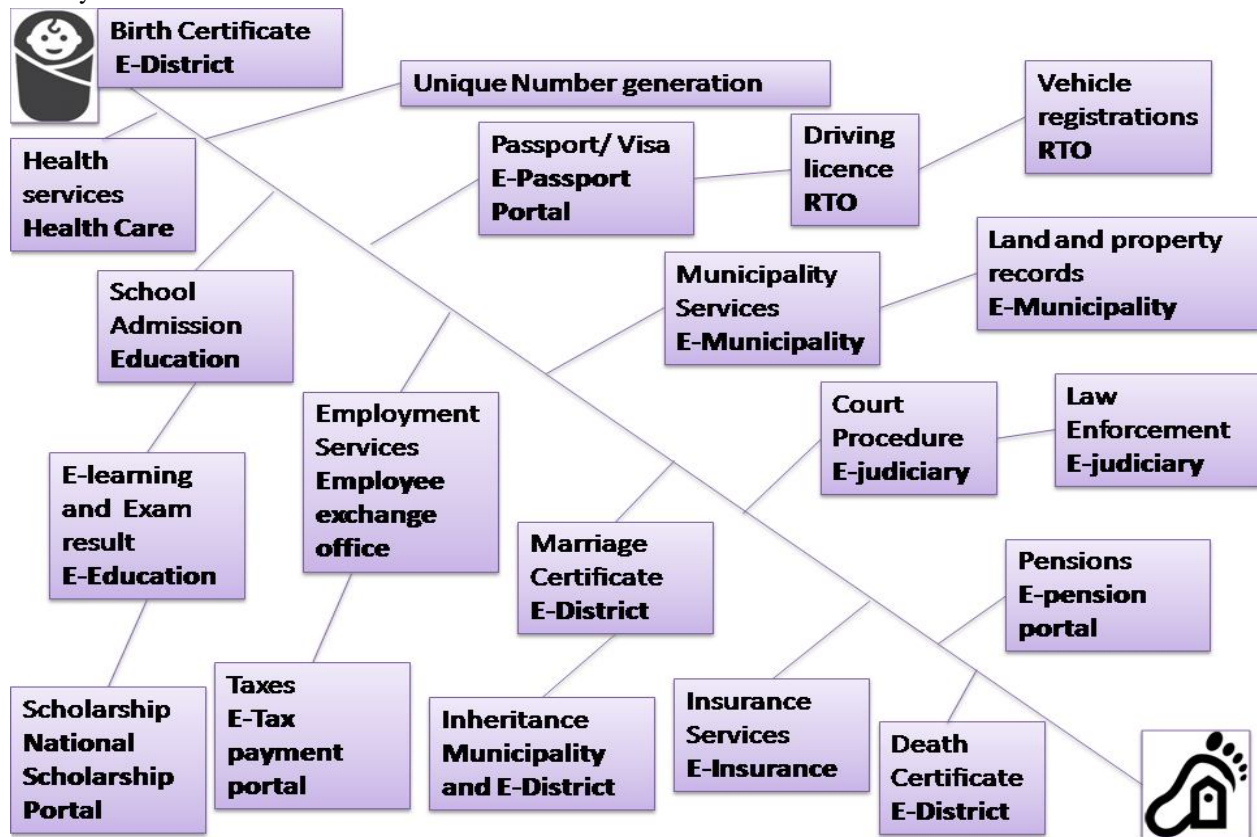


Fig : 6 E-government service

Citizens' data is the most vital data in any country. If a country intends to deliver all information in digital form, it must also enable digital access to all sites. The bulk of services, as demonstrated in Figure 3, can be delivered digitally. Then there's the question of how big data will aid in this situation. The answer to this question is straightforward. If all of the citizens' data is available online, that is a massive amount of data, and big data plays a critical role in processing that data and retrieving meaningful information or gaining knowledge from it. Figure 6 depicts the entire process, beginning with birth and ending with death. Any infant born within a certain length of time must have his or her information saved on a government website. The birth certificate should thereafter be issued by the government officials at the district level via e-district services to the new born. Then, to interpret and analyze data of the newborn, all health-related information is digitalized. The big data architecture may be used to obtain useful information such as how many newborns have health difficulties at birth, in which areas specific diseases are spreading, the child death ratio, parent information, and many more studies. Because all data will be stored digitally, big data can assist in a variety of ways.



Big data can be used to manage everything from birth certificates to death certificates. This strategy can be used to create an educational system, results, and e-learning. The government can forecast people's future job needs and take steps to address them in order to eliminate unemployment. The similar system can be used to manage government scholarship programmes. It becomes simple to forecast how much financial assistance will be necessary in the future. The government can create a budget based on this data. The government should give e-passport services, and big data makes this quite simple. Any citizen can apply for a passport, and all of the necessary verification is carried out utilising big data. This may improve the efficiency of these facilities. Vehicle registration, like passports, may be controlled through digital technologies. The government can design a city development based on people's preferences and rates of vehicle purchase so that traffic will be less likely in the future. This technology may also be used to assess marriage certificates, land ownership, and property inheritance, among other things, and to resolve disputes. The government can also manage the courts and judiciary. The government can take efforts in the future to improve administrations by analyzing the previous records. Data on pensions and insurance will be evaluated as well for future references and actions. Much workload can be eliminated if the government correctly manages all of the data.

*A. Prediction and Investigation of Criminal Activity*

Crime prediction and investigation become a big issue in densely populated countries like India and China. The police should have access to all of the people' information. If the authorities have information on all of the people in the vicinity, matching the biometrics of the criminals becomes a simple task. A big data approach is described here that can assist the police in both solving and predicting crimes.

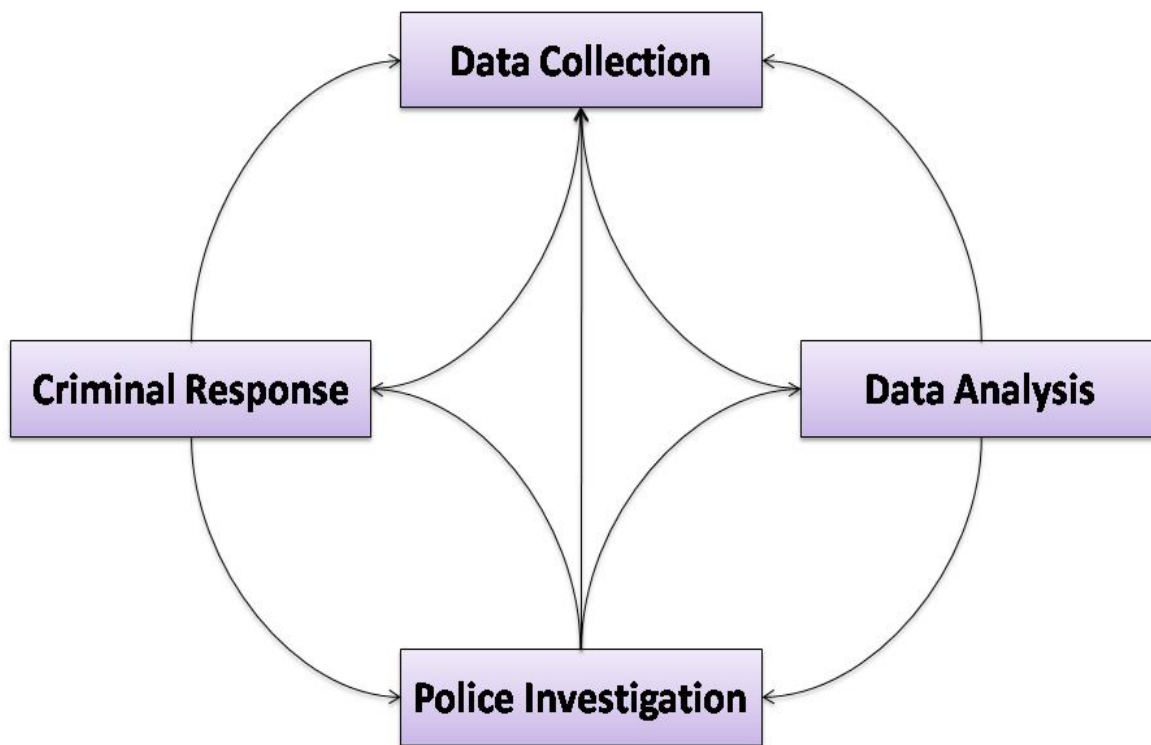


Fig : 7 Crime predication and investigation

As demonstrated in Figure 7, police can use big data to predict as well as investigate crimes. First and foremost, all data pertaining to crimes committed up to this point must be recorded in digital form. Then all of the records may be analysed, and crime predictions can be made for the near future. For example, if a huge festival time thief and robbery occurs in a specific region, the system can anticipate that this type of behaviour would occur at some point in the future, allowing authorities to plan accordingly. This information also aids the authorities in locating the suspects. The police can use the criminal's biometrics, and if the same criminal commits a crime a second time, the police will have an easier time catching him. The criminal response must also be saved in the system so that the pattern can be used to forecast future crimes.

### *B. Banking and Finance*

In banking and finance, the government can use big data. All banks are under the central bank, so the central bank can take details of income, taxes, investments and all other details of every citizen of the country. From this big data, predictions about future investments and any potential fraud can be determined. By providing all banking institutions to calculate tax online and other important information readily available from the government. The government can also control the flow of money in the country. In terms of the financial sector, security becomes a major obstacle. In some lands this may not be possible due to lack of education and technical awareness.

### *C. Traffic Management*

Traffic is now a major problem in many parts of the world. Big data can also help resolve this issue. At the local level government traffic information can be collected. In big cities local sensors and resources are also used to gather traffic information. In an area where traffic is common, using large amounts of data and information technology it is possible to find that location and possible solutions.

### *D. Weather Forecast*

Climate plays a vital role in the development of any country. To determine the weather a large number of data must be analyzed. It can also be done with the help of big computers but it is very expensive so that many countries can not afford its costs. With the help of big data, all data can easily analyze the system and weather conditions can be determined by the metrology department. In addition it is also possible to collect data from various sensors installed in specific locations for high accuracy.

### *E. E-Tax Centre*

The government can start this e-tax center so that every citizen can easily pay taxes and it is easier for the government to see the tax details. By using the process in this big data the government can predict future taxes and similarly the government can create a national budget. In addition, as with all digital content, using big data systems and systems it is possible to identify people who do not pay full tax. By using the right method a clear cash flow can be made.

### *F. Emergence of Drugs and Health Care*

Life is a real treasure. In any country, healthy citizens are major sources of development. Big data can help the region in the development and production of any drug. A recent example is the covid-19 vaccine. Big data helps to develop the vaccine and can be used in vaccines run in various countries. It is possible to use this technology to determine the effect of vaccines on humans, so that they can be useful in the future. Patient tracking and tracing can be done with this new data technology. In addition to this health care can be managed. It is easy to see the spread of certain diseases in a particular area at a particular time so that the government can use resources effectively and save lives. For example, in some regions of India during the monsoon malaria season the disease may spread. So if all the data can be analyzed using big data technology then the pattern of any disease can be identified and the government can take action to address that. In addition the stock market is controlled by this type of portfolio. How much product is available, how much money is needed each year. From this type of information the government can make decisions such as, need to import any drug or it is easy to export a particular drug to other countries. The government can also monitor the production of incing in the country.

Apart from the above-mentioned area, big data can be helpful in diagnosing and developing diseases. Currently all new medicines are being tested in animals. It should not be done. So here big data can save the ecosystem by providing the same thing. If all the information about the human body is transmitted digitally then it may be possible to test any drug in a computer program rather than endangering the health of the animal.

### *G. Cyber Security*

In any country online security is a major problem. If all information is available online the data must be protected. In this case big data can help manage this protection. By analyzing data access big data can provide security of important data. Managing website access to users can be controlled using large data technology. To make an example of any country where pornography is banned even though a citizen can access that site through various loopholes in the system. By using big data technology, the government can control these things. By accessing, storing and analyzing location data cybercrime can be controlled by the government.

#### H. Disaster Management

Disaster can be natural by human error, injuring local people in the area. The government can have a record of all the plants that cause disaster, using big data technology the government can control all those things. It is possible to track all of these dangerous plants and disasters can be avoided. Although there is talk of natural disasters it is inevitable but management can be done in such a way that it causes minimal damage. By using a variety of sensors and its data, it is possible to predict a catastrophe before it occurs. Not all of them except for some unforeseen disaster and measures can be taken accordingly to minimize losses. Even if disaster strikes, recovery can be managed with the help of technology.

#### I. 3-D Model Creation

Government should create a 3-D model of every city. For creating those models the government should use the method known as satellite photogrammetry which can help to create a 3-D model of the city. This model can help in disaster management. Using the big data, it is possible to predict the effect of flood or tsunami in the city and percussive action can be taken by the government. This model can also help to develop various cities in beneficial manner and in proper architecture which minimize the harm to nature.

### V. CONCLUSION

The use of big data technology will improve all the aspects of the government. This new method of governance will increase the speed of government work, which further helps in the quick development of the country. In the country like India where huge amount of data is generated, the Big data technology becomes a boon. Not only the administrative work but also general purpose task can be done using this new enhancing technology. Indian government already started working on E-government. The new methods will be proven as the best way of government.

### REFERENCES

- [1] Big Data Analytics in Government Sector, A Way to New Governance by Hiteshkumar Vora , Aastha Dani , Hardik Mirani, Information Technology, Nirma University, Gujarat, India
- [2] Harshawardhan S. Bhosale, Prof. Devendra P. Gadekar "A Review Paper on Big Data and Hadoop" in International Journal of Scientific and Research Publications, Volume 4, Issue 10, October,2014
- [3] Big Data And Hadoop: A Review Paper, Rahul Beakta
- [4] Big Data Analytics Tool Hive Shikhar Gugli , Anu Sharma
- [5] Tripti Mehta, Neha Mangla, "A Survey Paper on Big Data Analytics using Map Reduce and Hive on Hadoop Framework", International Journal of Recent Advances in Engineering & Technology (IJRAET)", (ISSN: 2347-2812) , Vol.4 Issue.2, Feb-2016, pg. 112-118.
- [6] Anjali p p and Binu a "a comparative survey based on processing network traffic data using Hadoop pig and typical mapreduce" International Journal Of Computer Science & Engineering Survey (IJCSES) vol.5, no.1, february 2014.
- [7] An Insight on Big Data Analytics Using Pig Script J.Ramsingh , Dr.V.Bhuvaneshwari
- [8] Seven V's of Big Data Understanding Big Data to extract Value, M. Ali-ud-din Khan, Muhammad Fahim Uddin, Navarun Gupta
- [9] Big Data in Governance in India: Case Studies, Elonnai Hickok, Sumandro Chattapadhyay, Sunil Abraham
- [10] Big Data Analytics in Government: Improving Decision Making for R&D Investment in Korean SMEs, Eun Sun Kim, Yunjeong Choi and Jeongeun Byun.
- [11] National Crime Records Bureau. "About Crime and Criminal Tracking Network & Systems - CCTNS." available at <http://ncrb.gov.in/cctns.htm>
- [12] Paul Makin, Steve Pannifer, Carly Nyst, Edgar Whitley, Digital Identity: Issue Analysis.
- [13] Thomas Davenport, Big Data at Work: Dispelling the Myths, Uncovering the opportunities, Harvard Business Review Press, Boston, 2014.
- [14] Big Data application by Edureka, available at <https://www.youtube.com/watch?v=skJPPYbG3BQ&t=1146s>.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)