

Green IT towards Digitalizing India

Krati Bansal, Pratik Kurle, Priyanka Chawla
Department of computer Science and Engineering
Email- kratibansl1992@gmail.com
Student of Master of Commerce
Mumbai University, Mumbai, Maharashtra, India
Pratik.j.kurle@gmail.com
School of Computer Science and Engineering
Lovely professional University, India
priyankachawla.cse@gmail.com

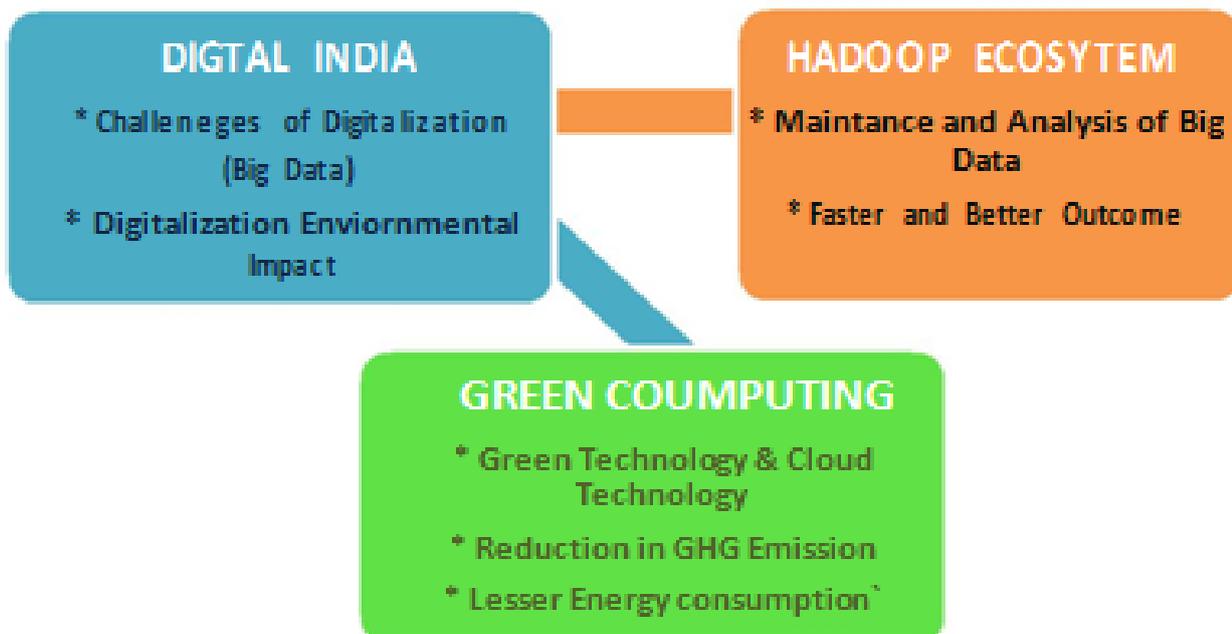
Abstract:- Digital India is an initiative by government of India which will prove as a blessing for Indian citizens and will also assist in connecting dots between past, present projects in India. Over the years biggest challenge for India has been to communicate information to this gigantic population. Digital Platform is one of the most efficient ways of communicating information to 1.2 billion citizens of India. Green Computing and Green technology are adopted by developed countries to reduce the impact of growing usage of computing equipments. Apache Hadoop is one of the best upcoming technologies to maintain and analyze Big Data. Hadoop can be implemented on with normal computing and also with cloud computing. Efficient use of Green technology and moving towards virtualization will trigger reduction in environmental impact of the growing use of computing.

Keywords:- Environmental Sustainability, Greenhouse gases (GHG) emission, Green Computing, Big Data, Hadoop, and Digitalization

1. INTRODUCTION

This paper will explore the basic vision of Digital India project and discusses the steps for improving efficiency of Digitalization and to reduce the overall cost and time. This paper provides information on various challenges of Digitalization in India. This paper also describes the various impacts of digitalization on the environment. This research study provides awareness of Green initiatives and

various benefits of Green Computing for reducing environmental impact due to use of computing and computing services. This paper portrays research on implementation of Hadoop Ecosystem on enormous data available with various government departments. This research publication showcases benefits of adopting Apache Hadoop and Map R analysis tool for improving quality and reduction in time for final user friendly outcome.1





2. RESEARCH FRAMEWORK

Research has been conducted with special focus on basic goals of Digital India project towards digitalizing India. Analysis is done to identify issues faced during digitalization. Increase in use of technology and computing will induce the growth of impact on the environment. Environmental impacts such as raise in GHG emission, growth in energy consumption due to digitalization are also summarized in this research paper. Research analysis is provided based on four parameters of Environmental Sustainability and portrayed with a graphical comparison. Green Computing and Cloud Computing can be adopted by government sector for the reduction of hazardous effects of increasing computing for digitalization. Implementation of Hadoop Ecosystem can facilitate better outcome for large data maintained by various government departments. Benefits of adopting Hadoop Ecosystem and Map R are described in this paper.

3. DIGITAL INDIA

Digital India revolves around three vision areas: Digital Infrastructure as a Utility to Every Citizen, Governance and Services on Demand, Digital Empowerment of Citizens. Many important projects such as National Knowledge Network, National e-Governance Plan, National Optical Fiber Network are under Digital India umbrella which will allow citizens to abolish digital divide and assist in digital embodiment in India. Digital India is a nine pillar programs which is working towards providing internet access to every citizen unable to utilize various digitalized government services [30]

3.1 Digital Infrastructure as a Utility to Every Citizen

Digital Infrastructure enablement of the rural and urban digital infrastructure is the main focus to provide Governance & e- governance to citizens all around India. 250,000 Gram Panchayats will be provided with high-speed internet connectivity under this project. This connectivity will ensure internet access and last mile delivery to the rural population under National Optical Fiber Network (NOFN).

3.2 Governance and Services on Demand

Reengineering and Modification of government services & processes will simplify governance and these services will be provided on computerized or portable channel to reach the remotest parts of India. Initiatives such as Single Window access to every person by integrating departments. Government services Online availability and on mobile platforms. Ensuring every citizen will be entitlement to easy access to cloud.

3.3 Digital Empowerment of Citizens

Address the current gap of the digital “available” and “unavailable” by developing skills and capacities of not only the citizens but also of the government agencies and employees at large. Mainly target is to provide global access of data and collaborate with social sites. Availability of all government certificates/ documents on Cloud based system. Availability Digital services /resources in Indian language.

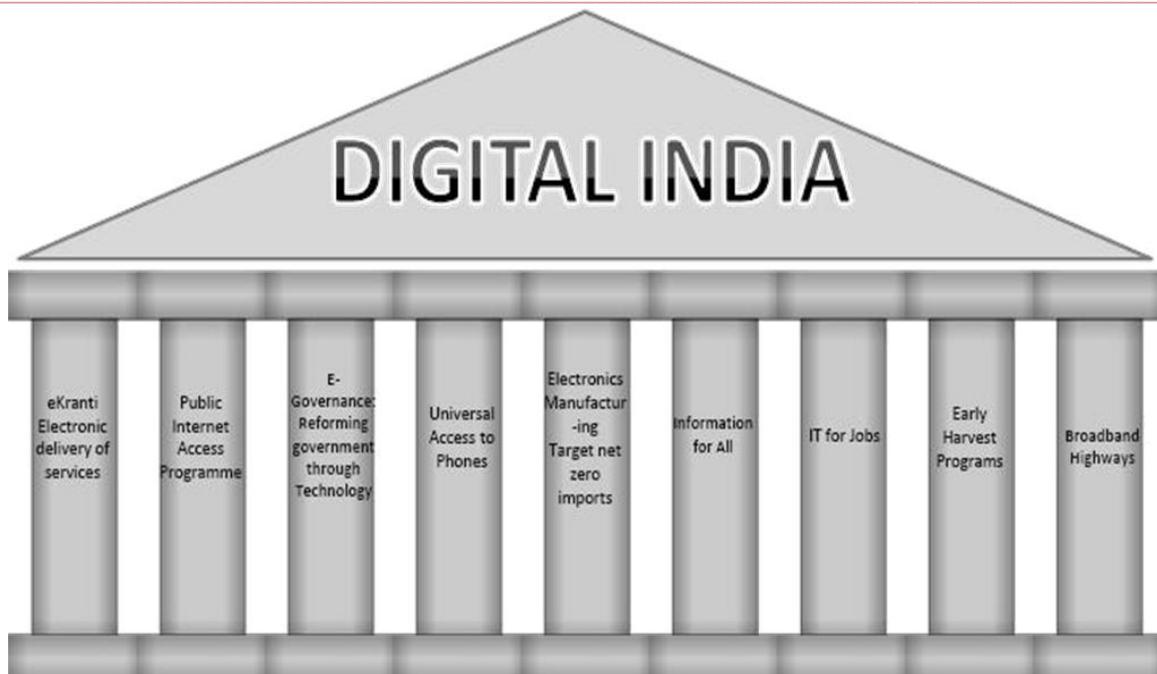


Figure 1: Nine Pillars of Digital India

These are 9 Pillars (Digital India project) that strengthens digitalization in India. There are numerous challenges to before achieve this enormously large project. Biggest challenge is to maintain and analyze the huge data (Big Data) previously created and stored in paper (non electronic) form. [31]

4. IMPACT OF DIGITALIZATION ON ENVIRONMENT

Digitalization bring growth in economy, however digitalization has a negative impact on the environment. Improvement and development of IT Infrastructure is very important for successful completion of digitalization. However development of new IT infrastructure will lead to increase Green House Gases Emission. Digital India program has planned new datacenters at various locations. Increase in number of datacenters will lead to heavy rise in energy consumption as huge amount of energy is required to empower datacenters. This causes growth in usage of nonrenewable energy. Digitalization leads to change in technology this led to change in hardware and software current installed at various government offices. This will enable purchasing of new hardware and disposal of old hardware which will cause huge raise in the E-waste generated. Four parameters of Environmental Sustainability (GHG Emission, Energy Consumption, Data Efficiency and E-Waste) are discussed in further sections.[29], [18]

4.1 Green House Gas (GHG) Emission

Digitalizing means more intensive usage of computing equipment and computing related equipment’s. Increase in

utilization of computing services lead to growth in Greenhouse Gas emission which provokes global warming. Developed countries such as United States and United Kingdom have digitalized, but the GHG emission of these countries is on the higher side as portrayed in figure 2.[10],[27]

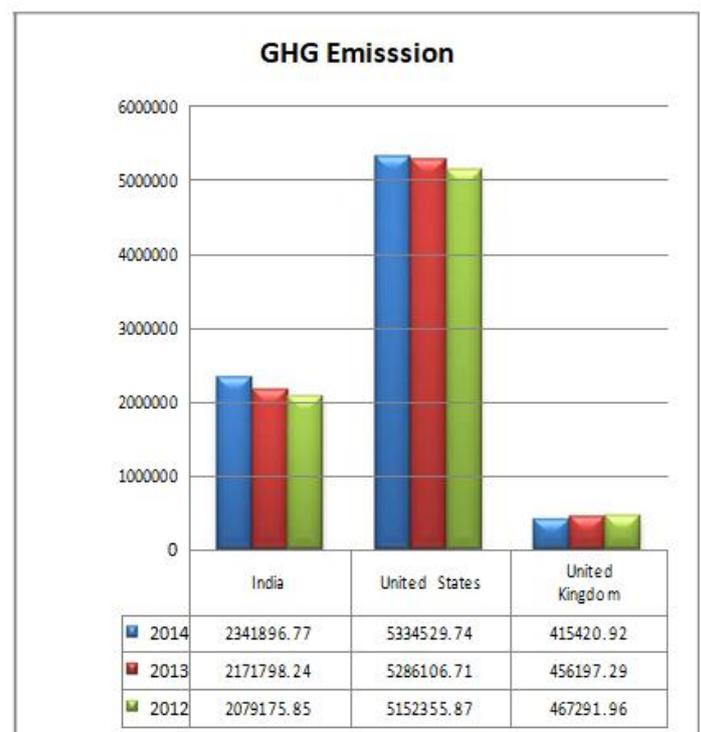


Figure 2: Year wise comparison of GHG Emission between India, United States and United Kingdom

India's overall GHG Emission of energy consumption is 5.81% of the overall world GHG Emission as shown in Figure 3.

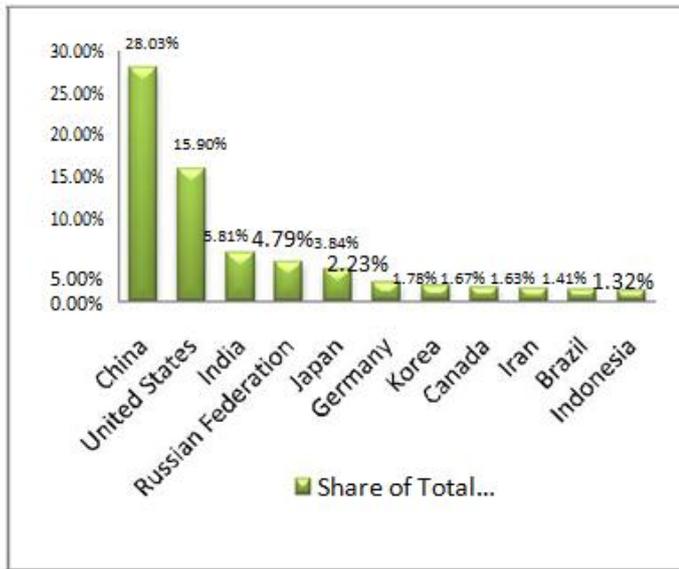


Figure 3: Percentages of Top Ten GHG producing countries

4.2 Energy Consumption

Increase in computing and computing equipments will lead to rise in the overall energy consumption. Development of IT Infrastructure will cause growth of energy consumption. There has been steady rise in the overall energy consumption (electricity) in the past three year as showcased in figure 4. Renewable estimate generation has shown a promising growth over the years as indicated in figure 5. [4]

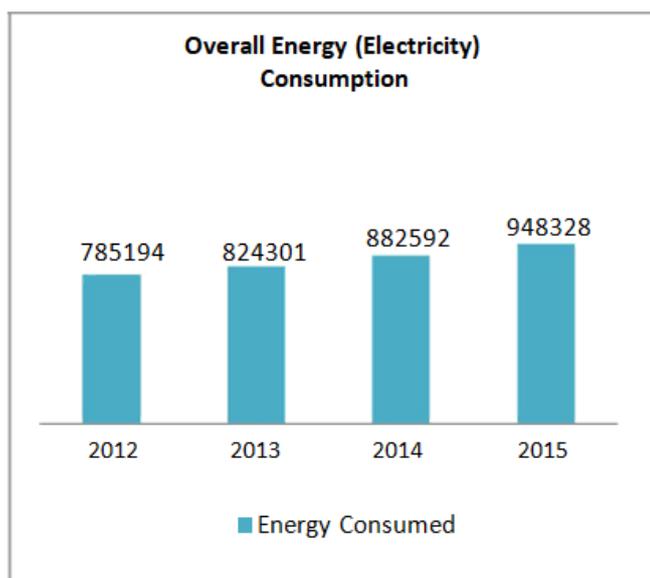


Figure 4: Year wise Energy (Electricity) Consumption [11], [12], [13], [14]

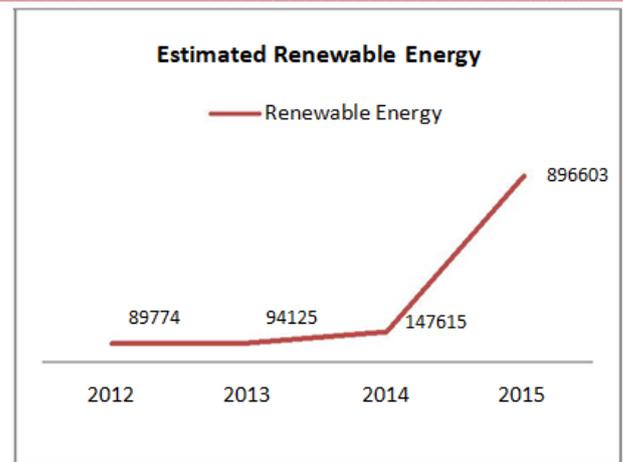


Figure 5: Year wise Estimate Renewable Generation[11],[12],[13],[14]

4.3 Data Centre Efficiency

Rise in number on data centre will cause rise in energy consumption (Electricity). Increase in number of data centre utilization will trigger growth of GHG Emission. Gartner has stated that 5.8% value of sever revenue forecast will increase by end of 2016. Power Usage Effectiveness (PUE) is the standard benchmarking of Data center efficiency. However there is no policy or standard set in India to monitor Data Center PUE. [15]

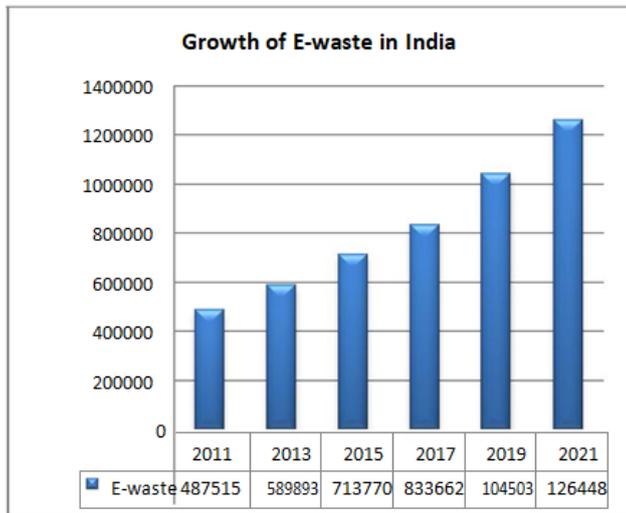
Table 1: Level of Power Usage Effectiveness (PUE)

Power Usage Effectiveness (PUE)	Data Center Infrastructure Efficiency (DCiE)	Level of Efficiency
1.2		83%
1.5		67%
2.0		50%
2.5	40%	Inefficient
3.0	33%	Very Inefficient

Figure 6: Year wise Estimated E-waste in India[16]

4.4 Growth in E-waste

India is fifth largest E-waste producer in the world. Department of Information technology has showcased the estimate growth of e-waste in future. Change in infrastructure will required change in computing equipment which will reason of disposal e-waste. Over the years e-waste generation has been constantly increasing as replicated in figure 6. [5],[24],[16],[22]



For reducing above mentioned impacts of digitalization on environment we can adopt Green computing and Cloud Computing techniques.

4.5 Benefits of Green Computing and Cloud Computing

Why we should move towards green computing? There are a multiple reasons for adopting green computing however it is sensible to use computing efficiently to reduced cost and lower impact on environment.

Cost Saving: As discussed, currently electricity prices are increasing day by day as natural resources are finite. Green initiatives can help to reduce costs of energy consumption, administration, maintenance, disposal, recycling. For example, instead of adding or purchasing new IT assets to existing infrastructure, it's better to utilize the existing ones.[6]

Virtualization: Server virtualization method uses the server. It reduces the energy consumption hardware and software use. It reduce the own sever space and heat process. It works towards the green data centre.

Reduce Carbon Foot Print: If the user follows the general points it can also reduce carbon footprint and save energy.

Energy Conservation: Turn off pc when not in use. In a particular time duration when all worker are on break in organization apply the auto off timer. Always use 80 plus certified power supply. Use LED instead of LCD. Use small size hard disk instead of large size hard disk. Graphics Processing Unit consumes more power it will consume. Always use energy star product. Avoid unnecessary printing email, scanning, documents.

Data Center Efficiency: Current issue is energy consumption heating and cooling when we green computing data centre we need to ensure some basis points such as server consolidation, utilization of power management

system, storage resources management, data life cycle management

As per a report of OECD, there are more than 92 programs established and initiated across 22 countries, in which 50 were established by government and 42 by industry association.[1],[2],[8],[7],[23][26],[28],[34]

4.5.1 Implementation of Cloud Computing

As per a Microsoft report it finds large business could reduce their energy consumption, carbon emission and its capital expenditure by diverting 30% business on cloud to reduce the energy. In virtualization cloud plays an important role. The basic requirement of cloud is internet connection, and funding to deployment the application. Cloud computing reduces the amount of physical servers which indirectly reduces the energy consumption.

GHG Emission: Centralization and virtualization reduces the carbon emissions which will assist in making India a healthier place.

Energy Consumption: Cloud computing is leads to reduction in energy consumption By utilizing load balancing technique and parallelization technique energy consumption can be reduced and also lowering carbon footprints.

E-waste: Lesser utilization of hardware equipment led to reduction in e-waste generation. Reduction in usage of hardware equipments decreases purchases which causes decline in production of e-waste.[3],[9],[19],

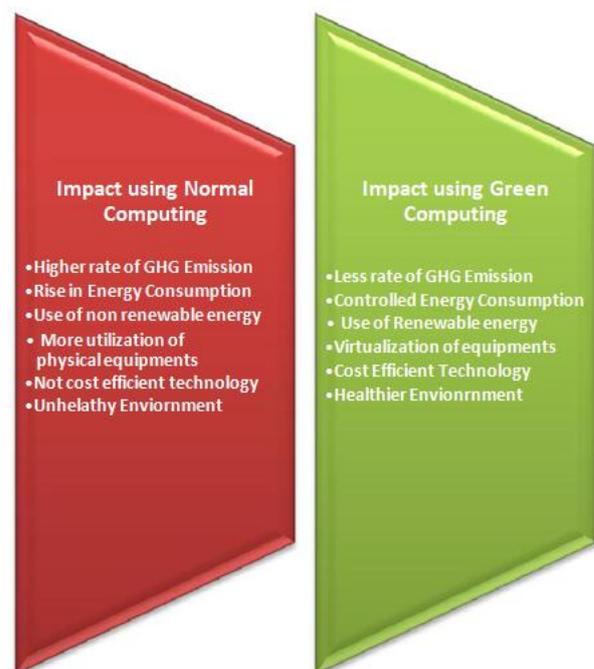


Figure 7: Comparison between Impact of Normal Computing & Impact of Green Computing

5. ADOPTING HADOOP ECOSYSTEM FOR BIG DATA ANALYTICS

Digitization can be described as converting physical data into electronic data. Digitalization means digits in a digital way. Process of fetching out required data is utilizing technology. Government maintains large amount of data with different departments which can be described as Big Data. This concept of

Big Data is classifying large data in 3 V's Volume, Velocity and Variety. Apache Hadoop is one the best technology which is able to analyze huge datasets and will facilitate in developing a better IT infrastructure for government. Hadoop Ecosystem maintains and analysis Big Data in proper manner which provides desired outcome. Apache Hadoop Ecosystem is described further in this section.[8]

5.1 Hadoop Ecosystem

Hadoop ecosystem can be defined in various layers which can distinguished by technology and components of implementation .Hadoop Distributed File System (HDFS) layer and Map Reduce layer are two important layers of Hadoop architecture. Top most layer of Hadoop contains component such as Apache Pig, Hive, HBase, H CATALOG, Oozie, Sqoop, Ambari, Zoo Keeper and Mahout. Analysis of Big Data is processed by components top most layer of Hadoop Ecosystem such as, Hive, Ambari, Pig, Zookeeper, Oozie and Sqoop. Amazon S3 is a cloud service provider which works with Map Reduce based recovery time and automatic data backup. Hadoop Ecosystem components can be classified as per their utilization. Data Management level components are Zoo Keeper, Flume and Oozie. Pig, Hive, Sqoop and Mahout belong to Data Access category. Data processing is done by Map Reduce and Yarn. Data Storage stage components are HDFS and HBase. [7]

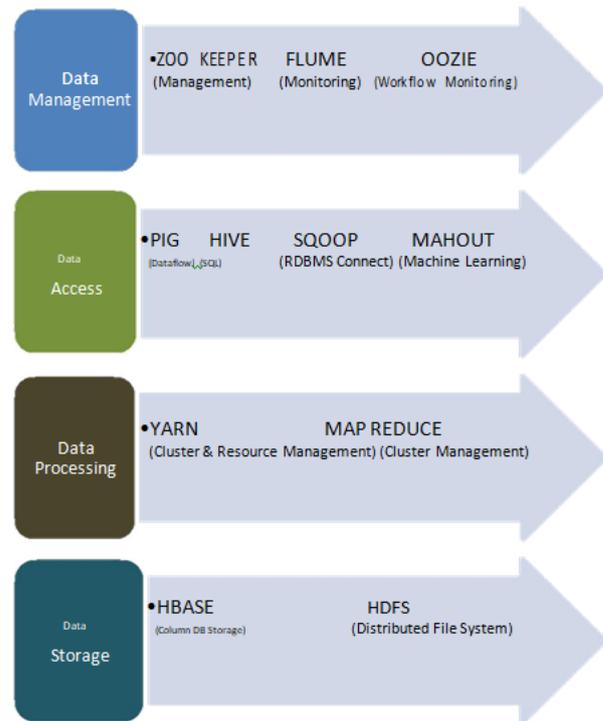


Figure 8: Hadoop Ecosystem Components

Hadoop Distributed File System (HDFS) of Hadoop Ecosystem is being used by Yahoo and Face book. Google File System (GFS) as is a file system which is only dedicated to Google only. HDFS is a one the best file management system. Below is the describing of various parameters on which HDFS is compared with GFS.[32],[33],[35]

Table 2: Comparison between Google File System (GFS) and Hadoop Distributed File System (HDFS)

Features	Google File System (GFS)	Hadoop Distributed File System (HDFS)
Processing	Chunk and Master Server	Name node & Data Node
Goal	Supporting large files	Supporting large files
File Management	Identified by Path names and organized hierarchically in directories	Support traditional hierarchical file organization
Database Files	GFS uses Bigtable database	HDFS uses HBase equivalent to Bigtable
Scalability	Cluster based architecture with file system consists of hundreds	Cluster based architecture running on clusters with thousands of nodes.
File Serving	GFS comprises of fixed sized chunks. (chunk size 64MB)	HDFS is divided into large blocks for access and storage (size 64MB)

5.2 Benefits of Hadoop and Analytics Tools

Map Reduce is the second layer of Hadoop which is above HDFS. Data processing component of Hadoop is commonly known as Map Reduce. Map Reduce helps to solve the data problem in parallel. It easily divides the data in small parts. Map Reduce comprises of 2 functions: - Map function and Reduce function. MapR Hadoop provides government agencies a scalable and cost effective architecture which drives situational analysis, real-time analysis and assists information flow across numerous government agencies and departments. [25]

Indian Intelligence Bureau needs to analyze information gathered from various sources such as social media and satellites. Data collected from other intelligence around India requires proper storage and analysis. MapR Converged Data Platform can be utilized as an Enterprise Data Hub for storing and analyzing various types of classified and unclassified data maintained by the bureau.

Central Investigation Bureau requires multiple tools to analyze and conclude on the outcome. Crime Predictions are based on collection of data, classification of data into crime classes, pattern identification and predication of crime category. Bureau uses analytical tools for collecting data and analyzing huge size data such as analysis of bank transaction data to identify money laundering and terrorist financing activities. MapR Converged Data Platform enables capability for instance anomaly detection machine learning which assists for identification of patterns that can reduce financial crimes.

Defense: Defense analyses large amount of data in order to advance the agency mission. Map R can maximize the counts of events and activities which an analyst can monitor in real time and enhances decision making abilities.

Pharmaceutical Drug Evaluation: Pharmacy council can use Big Data technology for accessing huge amount of data for classifying drugs and treatment to conclude if warning is to be issued. Map R Converged Data Platform can be utilized by researchers to analyze the patterns of side effects of drugs and identification of efficient treatments.

Scientific Research: Scientific Research requires intensive analysis of large amount of data and creating new data from these analyses. Map R could be used for solving problems such as utilizing visualization techniques to analyze complicated data.

Weather Forecasting; India Meteorological Department analyzes volume the data that can varied in nature it work on terabyte of data. Main work of weather forecasting is to collect the data and analyzing on perfect scale without any error. Map R can be used to analyze the large number of data.

Tax Compliance: It works on structure and unstructured data variety of sources in order to identify suspicious tax organization and prevent tax fraud. The main issue of Taxation is to detect the Tax both country where u earn and home country also second issue of taxation Cautionization of Income.

Traffic Optimization: Map R converged data platform help in aggregating real – time traffic data gather from road sensor and analyze the traffic time in the real time on data It analyze fast growing data much more cost effective method.[17],[20],[21]

6. CONCLUSION

Digitalization of India is will improve the standard of living in India a led to various growth in economy as well as social life. Growth in IT Infrastructure will led to rise in the environmental impact of computing. Government Policies on environment sustainability must more elaborative and rearranging the standard of sustainability with enormous growth in Indian IT Infrastructure. Green Computing is currently one of the best ways of achieving and maintaining environmental sustainability. Green Initiatives such as purchasing Offset Carbon and purchasing renewable energy are common in corporate sectors. Adopt Green Technology with Digital India will not only make India a developed nation but also a better place for our future generation. Developed countries such as United States and United Kingdom are focusing on utilizing electricity generated from renewable resources for empowering computing

equipments and datacenters. Implementing Apache Hadoop with the new developing IT Infrastructure in India can reduce various manual errors and time saving. MapR being an important component of Hadoop Ecosystem which assists various government departments for analyze data in secure and faster way. Digitalization has its own benefits and drawbacks. Main motive of this research paper is to identify the best technologies which are cost efficient and easier to adopt with the changing requirements of digitalization.

7. FUTURE WORK

Digitalization will need more advanced IT Infrastructure; hence data center numbers will rise around the country. Data centers require huge amount of energy of computing equipments such as server and cooling system. India having geographical diversity, data center can be stepped towards various renewable energy generating plants. At initial stage, Apache Hadoop can be adopted by Central government departments as they are managed by under same policies around the countries. In future endeavors, we will conduct research on adoption of Hadoop in government IT sector. We will analyze data which is publically on government sites and utilizing Hadoop components for faster and more cost efficient outcome.

Moving towards the digitalization India has to develop its own setup of IT software and Hardware. Make in India project can be collaborated with Digital India for developing own search engines of India rather than using US based Google search engine. However this concept has its own financial limitations, but the result will be fruitful and shall help in making India stronger in technology. Further, we can move from cloud computing to Sky Computing which will give more realistic and reliable results.

REFERENCES

- [1] Sonu Choudhary A Survey of Green Computing International Journal of computer Science and information Technology vol.5 (5), 2014,6248-6252
- [2] Sanghita Roy, Manigrib Bag Green Computing –New Horizon of Eney Efficiency and E-waste Minimization – World Perspective vis – a-vis Indian Scenario
- [3] Piotr Pazowski Green Computing :Latest Practice and technologies for ICT Sustainability Joint International Conference 2015
- [4] Andy Hopper Computing for the future of the Planet University of Cambridge The Computer laboratory
- [5] Fact Sheet World Wide Microsoft Unlimited Potential UNIDO Nov 2009. 0_Green_busines.PDF
- [6] Intel White paper : Extract, Transform , and Load Big Data with Apache Hadoop Intel www.Intel.com
- [7] ACE Cloud Hosting “The Environment Benefits of Going Green”
- [8] Jacob John “Green Computing Strategies for Improving Energy Efficiency in IT System” International Journal of Scientific Engineering and Technology volume no.3, Issue no.6 , ISN :2277-1581, June 2014
- [9] Pavithra Mani, Rathi Gopalakrishnan “Promoting Green Computing to provision Backup management of Disaster Recovery Cloud” Volume 2,ISN 2347-6435
- [10] GHG Emission on the world wide given by Statista <http://www.statista.com/statistics/271748/the-largest-emitters-of-co2-in-the-world/>
- [11] Energy Statistics 2015 Ministry of Statics and Programme Implementation <http://www.toknowpress.net/ISBN/978-961-6914-13-0/papers/ML15-377.pdf>
- [12] Energy Statistics 2014 Ministry of Statics and Programme Implementation http://mospi.nic.in/mospi_new/upload/Energy_stats_2014.pdf
- [13] Energy Statistics 2016 Ministry of Statics and Programme Implementation http://mospi.nic.in/mospi_new/upload/Energy_statistics_2016.pdf
- [14] Federation of India Chamber of Commerce and Industry http://ficci.in/sector/67/Project_docs/hydro-profile.pdf
- [15] Times of India <http://timesofindia.indiatimes.com/tech/tech-news/Data-centre-market-in-India-to-touch-4-5-billion-by-2018-IAMAI/articleshow/52438691.cms>
- [16] Baldge , C.P.,Wang,F.,R.,Huisman,j.2015 The Global E-waste Monitor 2014, United Nation University <https://i.unu.edu/media/unu.edu/news/52624/UNU-1stGlobal-E-Waste-Monitor-2014-small.pdf>
- [17] Big data Ready to Help Government Agencies
- [18] Anwasha Borthakur, Kunal Sinha “Generation of electronic waste in India article1380557326_Borthakur and Sinha.pdf
- [19] <https://i.unu.edu/media/unu.edu/news/52624/UNU-1stGlobal-E-Waste-Monitor-2014-small.pdf>
- [20] MicrosoftBigDataGovernment Whitepaper.pdf
- [21] Big data and Apache Hadoop for Government White Paper Relationship between Data Centre Strategy and Energy Efficiency <https://www.dimensiondata.com/Global/Downloadable%20Documents/The%20Relationship%20between%20Data%20Centre%20Strategy%20and%20Energy%20Efficiency%20Whitepaper.pdf>
- [22] Mahesh C.Vats, Santosh K. Singh, “Status of E-Waste in India –A Review”, International Journal of Innovative Research in Science, Engineering and Technology,ISN:2319-8753
- [23] Prantosh Kumar,” Paul Cloud Computing Based Green Information Infrastructure: The Future of Eco Friendly Information Science practice” Volume 2
- [24] Sanyukta Raje, Heermant Maan, Suprotim Ganguly, Tanvin Singh and Nisha Jayaram, Girish Gatikar, Steve Greenberg,Satish Kumar ,Dale Sartor, “Lawrence Berkely National Laboratory”
- [25] Infrastructure Considerations for Analytical Workloads

-
- [27] SWOT Analysis on Indian It And Ites Sector
- [28] Trends In Global Co2 Emissions 2015 Report
- [29] Power Saving Strategies in Green Cloud Computing Systems
http://www.sersc.org/journals/IJGDC/vol8_no1/28.pdf
- [30] Green Computing: An Overview with Reference to India
http://www.irdindia.in/journal_ijeecs/pdf/vol3_iss2/5.pdf
- [31] <http://www.businesstoday.in/current/economypolitics/digitalindiaweekpmmmodigetsrs4.5lakhcrsupport/story/221222.html>
- [32] Digital India <http://www.cmai.asia/digitalindia/>
- [33] Comparative analysis of Google File System and Hadoop Distributed File System
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.45.4.4159&rep=rep1&type=pdf>
- [34] http://www.st.ewi.tudelft.nl/~hauff/BDP-Lectures/5_filesystem_gfs_hdfs.pdf
- [35] Symantec 2007 Green Data Research report
- [36] Cognizant20-20 insights “Infrastructure Consideration for Analytics Workload”