

Big Data Framework for National e-Governance Plan

Rajagopalan M.R

NRC-FOSS

Centre for Development of Advanced Computing
Chennai, India

Solaimurugan vellaipandiyan

NRC-FOSS

Centre for Development of Advanced Computing
Chennai, India
solaimurugan.v@gmail.com

Abstract—The world is moving towards digitization. In India central and state governments are steadily moving towards digitizing all the government departments, scheme and services.

Due to the awareness of people and availability of network access that emerges, use of E-governance application has increased. As a result data is getting added and size of these data is increasing exponentially day after day – Volume. Semi-structured and Un-structured data from E-governance service like computerization of land records, Patient health records are most often in the form of electronic records. Processing and sharing of these records are difficult by traditional approach – Variety. Defense, crime and police service control the ever-increasing threats of terror attacks and of continually ascending crime graphs. It includes aspects such as creation of- and sharing of- crime-relate real time data – Velocity. Gathering and processing of vast amounts of data is not new. What is new is the speed at which one can process that data and extracting actionable business intelligence or mission critical from messy data – Complexity.

Open source big data technology - No-SQL databases, Hadoop and MapReduce are used to handle these large volumes and heterogeneous data. In this paper we present the model that illustrates how big data can result in the transformation of the government by increased efficiency and effectiveness in the e-governance service with citizen engagement in decision making.

Keywords—big data analytics framework, e-governance, No-SQL database.

I. INTRODUCTION

Earlier application was developed to handle static schema with traditional database that has the advantage of transactional processing and holds a very large part of the relevant information. Traditional database is not optimized and the challenge is high-performance analysis, speed, scalability and availability. Most of the current e-governance services are designed to accept only a structured data. The collected data is used only for statistical purpose and not solving any mission critical challenge that can improve the quality of the government service or schemes. In a country like India having huge population the number of users utilizing E-governance application and services is increasing drastically. The volume of data is getting doubled every year.

Effective management and analysis of large-scale data poses an interesting but critical challenge. After successful implementation of big data analytics in social media[5,6,7,8],

the latest development of Big Data has attracted lot of attention in government sector. In order to improve effectiveness of the service or schemes, a democratic country should not ignore the public opinion. Public participation is willingness on part of government to share information and make citizens a partner in decision making[1,3,4]. Public share their opinion through social media, press media and blog. Big Data analytics can process these unstructured data and deliver better service or right scheme to the needy people. It can also offer easy access to diverse and large quantities of data that was not easily accessible earlier. In digital world any government can change to a good government by using new and valuable raw material - public data resource. The model presented in this paper can result in increased efficiency of E-governance services and effectiveness of government schemes with public opinion for decision making. Better utilization of the data centre resources is demonstrated in the illustrate model by sharing of resources between multiple big data analytics running on top.

II. GOVERNMENT AND BIG DATA

A. E-Government

E-government refers to the use of modern technology resources such as internet, mobile etc.. to improve the functioning of government. Its government an attempt to improve the political and social environment and to drive a fundamental change in the ways in which functions are performed. These e-service provides better delivery of government services to citizens, improved interactions with business and industry, citizen empowerment through access to information, or more efficient government management. The resulting benefits can be less corruption, increased transparency, greater convenience, decrease in the time and effort, revenue growth, and/or cost reductions.

In india e-governance service offered through NeGP (National e-Governance Plan). On May 18th 2006 government approved NeGP which consists of 27 Mission Mode Projects (MMPs) of State Governments & Central Government and 8 Integrated MMPs. In year 2001 NeGP added 4 more project introduced list is now become 31 MMPs[16]. Department of Electronics and Information Technology getting proposal for Citizen Engagement framework in e-governance[1]. Indian government unique identification number system Aapka Aadhaar-IUIDAI project enrollment system is a distributed

and scalable and its big data technologies can readily support[17].

In NeGP, most of the e-Governance applications are silo, running individually and maintaining their own databases. This causes non uniform information and less efficiency of service for the business and public sectors/domain. Silo applications also introduce many duplicate schema data within itself. The proposed model can integrate and share service and data among other government services thus reducing the cost of operation and predict the future of one scheme based on the insight from other service. This in turn helps improve the efficiency of government public services.

B. Need of Big Data in public sector

Fig 1, shows the NeGP source of big data, effective utilization of services by citizen reduce the time involved and also improve efficiency as compare traditional office based services. Successful implementation of any e-government application require effective integration of process, technology, data and budget. In which the technology and data plays major role. Traditional approach lack of handling messy and heterogeneous data.

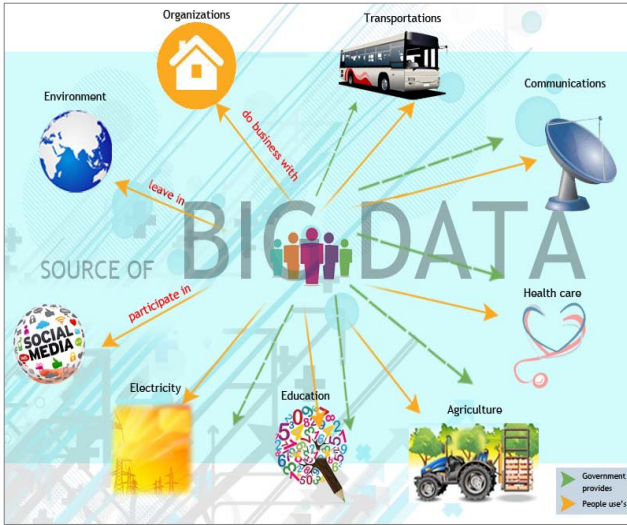


Figure 1. E-government source of Big Data.

1) Sharing of data

case 1 : When apply passport or Driving Licence the data would load from municipal birth certificate information.

case 2 : Government before announce any social schemes it can predict the population based on the born rate available in healthcare data store and Number of marriage registration available in registration data store.

2) Learning from data

case 1: Collecting, analysing, planning and acting on root cause of feedback or comments from citizens like or dissatisfaction quickly change or reverse its schemes, leading to improved citizen satisfaction.

case 2: In healthcare, in order to improve the patient outcome, reduce the patient readmission and reduce the cost of duplicate test and unwanted procedure to follow by analysing patient clinical record.

3) Personalising / grouplising

case 1: Big Data opens up new possibilities for grouplising services by granularity. When government come to know some special attention needed for particular segment of people it can tailor the scheme or service accordingly.

case 2: By analysing data from individual patient clinical record can identify the patient at most risk of ailment in the future and using the data can do preventive treatment.

4) Discovery of unexplored questions.

Traditional analysis failed to examined hidden patterns and correlations from multidimensional datasets. It also complex, expansive and time consuming to extract the intelligence insight from multidimensional and huge datasets. Big data analytics provides very high degree of precision, providing an auditable, scientific basis for making decisions in complex systems.

case 1: Predictive weather forecasts analysis integrated with soil condition analysis can alert farmers, when to plough, which crop to plant. This will help farmers to get higher production, earn greater revenue and for nation avoid food deficits which will directly reduce the amount of food import.

case 2: By analysing data from individual crime report, security and crime department can predict and discover crime before they happen[2].

5) Innovating for growth.

Big data open source technology evolving rapidly which will reduce the cost of analytics and improve service opportunity by waste reductions, expense abuse, asset utilization, traffic utilization, resource scheduling, public service optimization, predictive maintenance, outage Identification, criminal activities & patterns, fraud detection

TABLE I. CHARACTERISTICS OF TRADITIONAL VS BIG DATA ANALYTICS.

Traditional e-government	Big Data - good government
Only structured data	Raw, unstructured and semi-structured data.
Statistical analytics with known condition	Discovery of unexplored business questions.
Tera-bytes of data	Large and messy data
Traditional analytics is batch oriented and one need to wait transformation jobs to complete before the required insight is obtained.	Big Data Analytics is aimed at near real time analysis of the data using the support of the software meant for it
Text analytics	Graph, audio and video analytics

Traditional analytics based on the relation data model. Analytics based on the known relationship	Very difficult to establish relationship between all the information in a formal way, and hence unstructured data in the form images, videos, Mobile generated information, RFID etc... have to be considered in big data analytics.
Ad-hoc query report, high latency for large volume of data.	Low latency, interactive, Visualization, decision-making and analytics reports
Type of operation : relational & conditional	Analytics :Sentimental, predictive, behavioral. Machine Learning

Benefits of government data with big data analytics is safer communities, smarter decisions, better served citizens, improved fiscal performance greater innovation and citizen engagement which is important in a democratic country like India.

III. BIG DATA FRAMEWORK FOR GOVERNMENT

Our Big Data analytics stack, Fig 2, can be broken down into 4 major components - The resource management, data organization and management, analytics and discovery, decision support and visualization report.

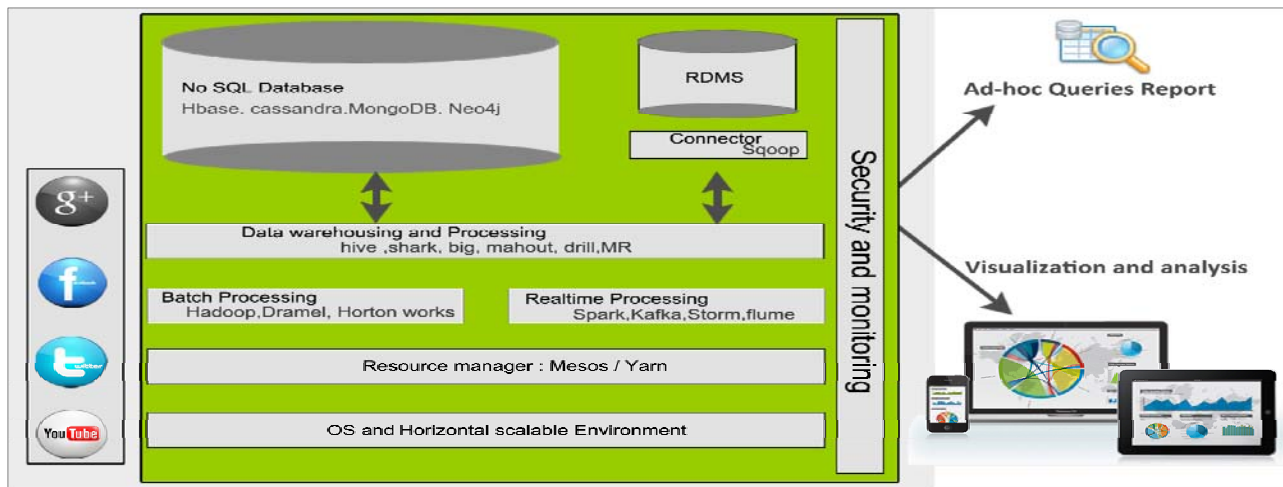


Figure 2. Big Data analytics stack - open source technologies

A. Resource management

Managing Big Data is critical and staying on top the latest Big Data analytics tools keep developers in control and processing of data much faster. Various framework born in different companies and all they have different strength and own problem solving capabilities of certain class of application. Taken this an account resource utilization, high performance, data sharing and operation, maintenance cost companies generally want all these framework deployed and running on common cluster. Resource management and scheduling platform - Mesos and Yarn, solve this issues. Mesos

and Yarn aims to increase resource utilization of clusters by sharing cluster resources among multiple processing frameworks i.e Hadoop, Spark, MPI or multiple instances of same framework[18, 19].

B. Data organization and management

This level refers to the software that processes and prepares all types of structured and unstructured data for analysis. Two type of data model - Relational Database Management System (RDBMS) and the NoSQL database management system. RDBMS is optimized for scale and speed in processing huge relational data i.e structured and static data sets. Big Data analytics Stack can interact with RDBMS through connector – Sqoop. Sqoop is a connetor tool designed for efficiently transferring bulk data between Hadoop related system (HDFS,Hbase and Hive) and RDBMS (Oracle, MySQL, Postgres, HSQLDB etc..).

RDBMS-SQL's transactional qualities and detailed indexing are not suitable for the processing of unstructured files. NoSQL database management system Hbase and Cassandra for columnar database, MongoDB for document and Neo4j for graph based No-SQL database[20] are emerges to solve. For cleanses and integrating Hive and Shark – hive on Spark used[22].

C. Data Analytics and discovery of unexplored opportunities.

Data analytics computing framework continue to emerge for support of offline processing MapReduce, Hadoop for batch processing, Storm for online processing, Spark for iterative computational, MPI for high performance and data mining and streaming framework for S4.

This layer can broken into two data analytics model. Hadoop, Dramel, Horton Works - model support batch processing and ad-hoc analytics. Spark, Kafka, Storm and Flume model supports dynamic real-time analysis.

D. Report - Decision support and visualization

[^] **Reporting and dashboards:** Tool for user-friendly representation of the information from various sources

and these output help to support decision making and to ensure that an action is taken, based on data analysis.

- ^ **Visualization:** Tool for highly interactive and dynamic report and distinction between reports and visualized output is animation. Normal end user i.e politician and rural citizen can watch the visualization techniques.
- ^ **Analytics and advanced analytics:** Tools that can get data from warehouse and process the data for human conception. Advanced analytics should explicate trends or events, that are transformative and unique. Predictive analysis, Behavioral analysis, Comparative analysis, Fraud analysis, Risk analysis and sentimental analysis these are best examples.

IV. GOVERNMENT ACROSS GLOBE MOVING TOWARDS BIG DATA

AGIMO (Australian Government information Management Office) has adopted a Big Data Strategy for whole-of-government (WofG) approach in order to enhance the WofG approach, deliver new services and provide better policy[9]. ACBPS (Australian Customs and Border Protection Service) use big data solution for Passenger analysis help to identify the suspicious visitors.

In March 2012, The Govt. of US announced Big Data Research and Development Initiative, for improving ability to extract knowledge and insights from large and complex collections of digital data, the initiative promises to help solve some the Nation's most pressing challenges. Big Data Senior Steering Group (BDSSG) was formed to identify current big data research and development activities across the Federal government.[10]

The Govt. of UK, department of business innovation and skills, allotted £189 million for big data research[11]. The UK government Also recently published the strategy for agricultural technologies, announcing that its initial focus will be on big data.

In Japan, Council for Regulatory Reform initiated the government to set up guidelines for Japanese companies can use to further use big data without breaching privacy laws. for the same government allocated 13.2 billion Yen for Big Data R&D [12].

Danish Government predicted fast growing data in 2011 and implemented NoSQL Data base for Operate Nationwide Medical Prescription Card Program in association with Basho's Riak software.

Sweden researcher uses big data analytics to Analyzes real-time data streams to identify traffic patterns and Analyzing large volumes of space weather data in real time[13].

In Taiwan, E-Governance research center approved project for Feasibility Study on Introducing Cyber Opinion into Government Decision-making Process using big data [3].

Twenty percent of The Netherlands is below sea level, so the country depends on dykes to protect the country from floods. Netherland government launched Digital Delta Big data project in collaboration with Dutch Ministry of Infrastructure and the Environment collaborate & IBM, aims to improve Dutch flood control, save the government millions by providing real-time intelligent dashboard[14].

Korean government planning Big Data Master Plan for the Implementation of a Smart Nation. As part of smart nation government established Pilot Project for Using Big Data in Official Statistics.

France government recently announced Big Data is Imminent launch and allotted €11.5 million for 7 big data processing projects proposal through Programme of Investments for the Future.

Thailand's Department of Special Investigations is using big data for investigations which improve the accuracy of its investigations into criminal cases, including human trafficking[15].

IDC government insights report says Big Data trends in government adoption. Executives from UK, France, Germany, Italy, and Spain planning to adopt big data analytics for government. Italy Govt. to use income assessment tool redditometro with big data for tax fraud.

Norway implemented big data analytics for knowledge centre for health care system. These knowledge centre help doctors and clinicians for unified access and evidence-based research studies as part of the healthcare decision-making process.

V. CONCLUSION

Current serving e-governance application are silo. They handle only structured data and sharing data between application is also difficult. All over globe most of country trying to adopt big data technology in various domains like healthcare, crime prevention, agriculture, transportation, education and natural disaster prevention management. Developing country like India having huge population using big data analytics can benefit for decision-making and provide better service. Our proposed open source big data analytics stack can provide a cost effective solution and aid in effective service implementation. Further resource management and scheduling mechanism can utilize the government data centre resource efficiently.

REFERENCES

- [1] DEITY, citizen engagement plan.
http://deity.gov.in/sites/upload_files/dit/files/Framework%20for%20Citizen%20Engagement%20in%20NeGP.pdf
- [2] Paul E. O'Connell, "Using Performance Data for Accountability: The New York City Police Department's Comstat Model of Police Management," IBM Center for The Business of Government, retrieved at:<http://www.businessofgovernment.org/report/using-performance-data-accountability-new-york-city-police-departments-comstat-model-police>.
- [3] Zhou-Peng Liao, Using Big Data to Achieve Good Governance: A Feasibility Study on Introducing Cyber Opinion into Government Decision-making Process

http://www.teg.org.tw/web_en/research/view.do?id=1364456440299

- [4] [Lee Komito](#), E-Governance in Ireland: new technologies, public participation, and social capital. Proceedings of the 4th European Conference on E-Government: towards innovative transformation of the public sector. 17-18 June 2004, Dublin. Academic Conferences International; Reading, England: 411-418.

http://www.ucd.ie/lkomito/e_governance_ireland.htm

- [5] Hadoop at Twitter <https://blog.twitter.com/2010/hadoop-twitter>
- [6] Paul Yang, Moving an Elephant: Large Scale Hadoop Data Migration at Facebook. <https://www.facebook.com/notes/paul-yang/moving-an-elephant-large-scale-hadoop-data-migration-at-facebook/10150246275318920>
- [7] Kim Rose, [How Facebook uses Hadoop and Hive](#)
<http://hortonworks.com/big-data-insights/how-facebook-uses-hadoop-and-hive/>
- [8] <http://engineering.linkedin.com/hadoop>
- [9] Australian Government information Management Office, The Australian Public Service Big Data Strategy. <http://agict.gov.au/big-data>
- [10] Executive Office of the President, Big Data Across the Federal Government
http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet.pdf
- [11] [The Rt Hon David Willetts](#), Eight great technologies
<https://www.gov.uk/government/speeches/eight-great-technologies>
- [12] **British Embassy Tokyo**, Japan : big data for a bigger economy.
<http://opentoexport.com/article/japan-big-data-for-a-bigger-economy/>
- [13] Rob thomas, IBM Big data success stories.
<http://public.dhe.ibm.com/software/data/sw-library/big-data/ibm-big-data-success.pdf>
- [14] Big data project aims to improve Dutch flood control, save the government millions
http://www.pcworld.idg.com.au/article/465850/big_data_project_aims_improve_dutch_flood_control_save_government_millions/
- [15] Thanya Kunakornpaiboonsiri, Thailand optimises investigations with big data analytics <http://www.egov.vic.gov.au/focus-on-countries/asia/thailand/thailand-optimises-investigations-with-big-data-analytics.html>
- [16] Department of Electronics and information Technologies, Government of India. <http://deity.gov.in/content/national-e-governance-plan>
- [17] Innovation in government: India and Estonia.
http://www.mckinsey.com/insights/public_sector/innovation_in_government_india_and_estonia
- [18] <http://mesos.apache.org/>
- [19] <http://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YARN.html>
- [20] Guide to non-relational universe, <http://nosql-database.org/>
- [21] Aapka Aadhaar, Unique identity& address in India. <http://uidai.gov.in/>
- [22] Apache Sqoop, <http://sqoop.apache.org/>