

# Big Data Analytics: An Innovative approach to adopt in emerging Sectors

Ananda Khamaru

Department of Computer Science Tarakeswar Degree College, India

Corresponding author: [ananda.khamaru@gmail.com](mailto:ananda.khamaru@gmail.com)

Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

**Abstract--** In digital universe emerging trends of new technologies in different sectors are dramatically generating large amount of varieties [e.g.-structured and unstructured] of data every day. The low cost of memory is the reason to populate massive amount of data sets and it is too much heterogeneous to get a particular information is very troublesome. Day to day gathering of typical data makes a large complex datasets is called Big Data, which is difficult to capture and manage by conventional data processing system. The large complex datasets are stored among thousands of servers and Big Data technologies easily acquire and process those datasets. These technologies used in all sectors with big challenge to get better decision, quality assurance, risk management, etc. Adoption of big data technologies is opportunities to get the hidden pattern by acquired and analyzed a large scale, flexible, fault tolerance, and real time data which are resided in distributed files.

**Keywords—**Big data, Hadoop, HPCC, Storm, Blockchain, healthcare, Smart Power Grid, Smart Traffic.

## I. INTRODUCTION

In modern era data is the main element to get an effective decision. If amount of data is less, then decision can be partially concreted, so we need large scale of data which is called Big Data. Variety of data are generating from various sources at rapid speed and its make large volume of datasets, which is difficult to acquire and process through traditional database management tools. According to IDC report the digital universe data size is growing 40% per year into the future decade and it will reach 44 zeta bytes or 44 trillion gigabytes by 2020. Using digital camera, smart phone, online business, social media, and by the millions of sensors and communicating devices sending and receiving data over internet, where more than 2 billion of people are involved to generate this large scale complex data sets. The large volume of complex data sets is known as Big Data, which process through special techniques and technologies. The Big data is not only a buzzword but it truly help the big enterprises, healthcare & medicine, financial services, scientific research, power management, traffic control, etc. to decision making, quality up gradation, risk management, fraud detection, sentiment analysis, and soon. IT world has big opportunity to handle big data effortless in all sectors with challenging privacy.

### Why Big Data would adopt

Most of the sectors are adopting Big Data technologies at rapid speed to maintain their growth strategies and quick decision making. Use this technologies into the organizations are able to:

#### A. Store the necessary data & stop the data loss

Some organizations are generating data beyond their ability to store and manage it. An organization consist of too many data generating devices such as sensor embedded devices, RFID tags attached items, videos, e-mail, and smart devices which are used in location base data collections. Most of the time essential data are eroded due to less storing capacity, in such situation adaption of Big Data will make a great advantage to the organization. These technologies will help to pick the best information in exhausting data and combine with other data to produce an unprecedented result for the organization which could update the old existing one.

#### B. Curtail Time and Cost

In an Organization data are stored separately into silos which are disconnected with each other. These typical storing techniques will take more time to find information and the architecture is very costly to maintain. Adoption of Big data technologies can connect these scattered data sets together and any one can find required information instantly. It is not only curtailed the time of searching but also reduces the data saving cost.

#### C. Enhance the product quality

Data can be generated from variety of sources and without combination of them we could not get new challenge in productivity or what actually need to quality enhancement of existing product. In this case Big Data will support to combine variety of data from different sources to enhance the quality of existing product.

#### D. Analyze the Risk factors of new product

Introduce a new product in market it is not only complicated and expensive, but also the Risk is very high of a production based company [3]. Before analyzing some uncertain factor s if new product is introduced then these factors can be the reasoned of failure of that product. A series of decision making process are involved to make a new product with less Risk. The proper Risk assessment of new product will give an opportunity to hold the market with great success. Adopting Big data technologies in that apart will propose an advancement of risk assessment by series of decision making from previous to future Risk.

#### E. Perceive separate groups in large population

Real time data is generating from huge number of sources and it makes a large amount of datasets, which is very complex in nature and difficult to grouping. Big data techniques has the feasibility to combine and analyze real time data from various sources and separate these into different groups according to their features by which it can perceive separately from other in a large populated datasets.

#### F. Improve Decision Making

In business area our goal is to create an efficient and cost effective product but due to less data insight it is impossible to take a proper decision by managers [4]. In an Organization managers are taken final decision based on intuition and their experience but they did not able to justify their decision any more. Big data is the big platform where managers can justify their decision and if any improvement is occurred by the analysis then they will get that clearly and it will help to improve their ideas. This Big data techniques not only applicable in business area but also it massively use in healthcare organizations for sharp decision making. Use of Big Data in any sectors will make better decision and which will help to better consensus as well as better execution.

#### G. Innovate Idea

The innovative idea infuses in the mind through inspiring and insight of data [5]. In healthcare sector the innovative ideas will be needed for treatment of new disease. Alternative treatment is better than conventional care to cure a newly viral disease. Sometimes doctors would not get the idea how to treat a new viral disease then they need an innovative idea which would come through Big Data analyze.

## II. OPEN SOURCE TOOLS USED FOR BIG DATA

Rapidly accepting of new technologies in all organizations are generating variety of data in every day. Gathering of these data is very fast and it makes a large volume of datasets within few days and no traditional database management system can manage it. Therefore a new method is needed to handle this datasets effortlessly and the method is specialized in Big Data Analytics. Some Open source Big Data Analytics Tools:

#### A. Apache Hadoop

Hadoop is an open source batch oriented data processing system developed by Apache foundation. This is developed based on java programming framework. Hadoop is worked on large datasets in a distributed environment, in which distributed file system transfer data at rapid speed among the nodes. The core of Apache Hadoop consist of a storage part, known as Hadoop Distributed File System (HDFS) and processing part which is a MapReduce programming Model [2].

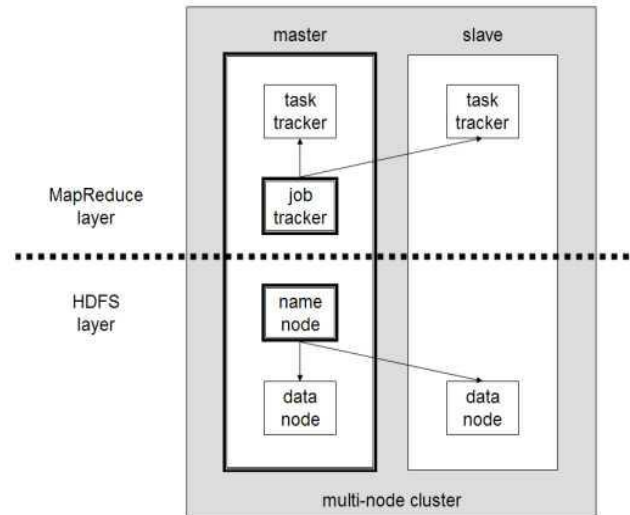


Fig -1: A multi-node Hadoop cluster [7].

The large amount of datasets is break into small units and distribute among computer clusters which is created from commodity hardware. The overall architecture of cluster is divided into Master and Slave. Master part consists of both Name Node and Job Takers, and which is only one in a cluster but more Slaves will be there. A Slave consists of Data Node as well as Task Taker. Its MapReduce technique and Distributed file system are inspired from Google. The Hadoop architecture comprises of three essential parts such as HDFS, MapReduce Programming and YARN.

#### B. HDFS

Hadoop Distributed File System is used to store large amount of datasets in distributed manner on commodity machines. It design is to store data in low cost hardware and high fault tolerant. Due to its distributed nature redundant data are stored into multiple nodes and in case failure of some nodes does not effect on entire processing as well as no data loss can be happened.

#### C. MapReduce

MapReduce is the programming model whose main task to process distributed data. Java programming model is used to develop the algorithm of MapReduce technique. It is buildup by two important tasks, namely Map and Reduce. Maps takes a dataset and then convert it into another dataset, where individual elements are broken down into two tuples

(key/value pairs) and it is the input in Reduce function which will combine all output of Map's tuples into smaller set of tuples [6].

**D. YARN**

Its name is Yet Another Resource Negotiator used to manage the all resources in Apache Hadoop framework. The YARN has two important elements such as Resource Manager, which is responsible for assign the resources and second is Node Manager, which is measured the resource capacity (e.g.- amount of memory, number of v-cores).

**E. High Performance Computing Clustering (HPCC)**

The HPCC is another kind of open source technology which manages Big Data like Hadoop. HPCC is developed by Lexis Nexis Risk Solution using Enterprise Control Language (ECL).

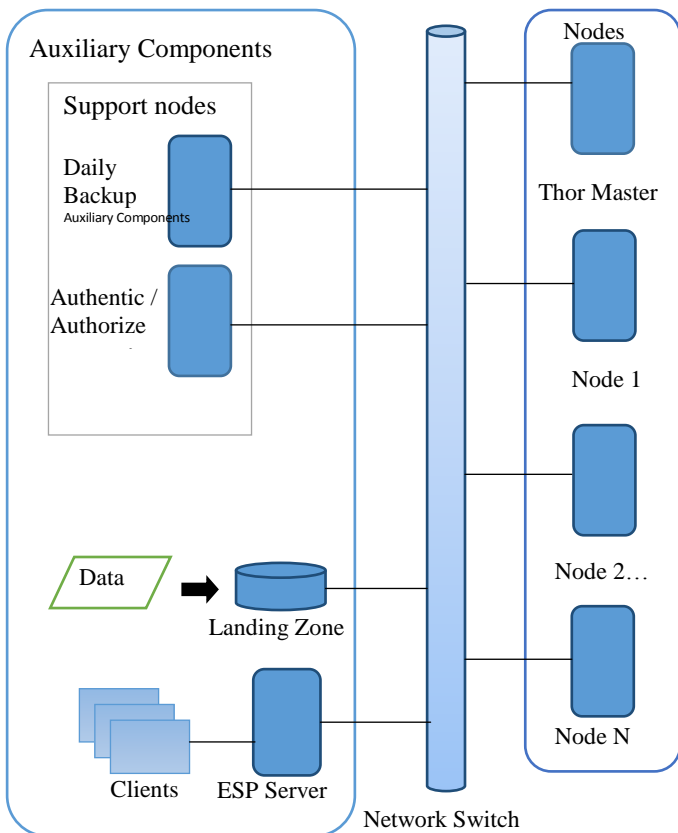


Fig-2: Thor

The ECL is a data centric declarative programming language use for parallel data processing. HPCC platform is popular in both Batch and Real Time data processing environment, it is also known as Data Analytics Super Computer (DAS). Two independent cluster processing environments are attached in HPCC system architecture. These independent clustering techniques are divided into first and second part. The first is responsible for refinery of larger volume of raw data but it typically used for data cleaning and hygiene, extract,

transform, load processed data, record linking, resolution, large scale Ad-hoc complex analytics, generate keyed data and indexes to support high performance structured queries and data ware house applications. This clustering technique is known as Thor the name was came from thunder.

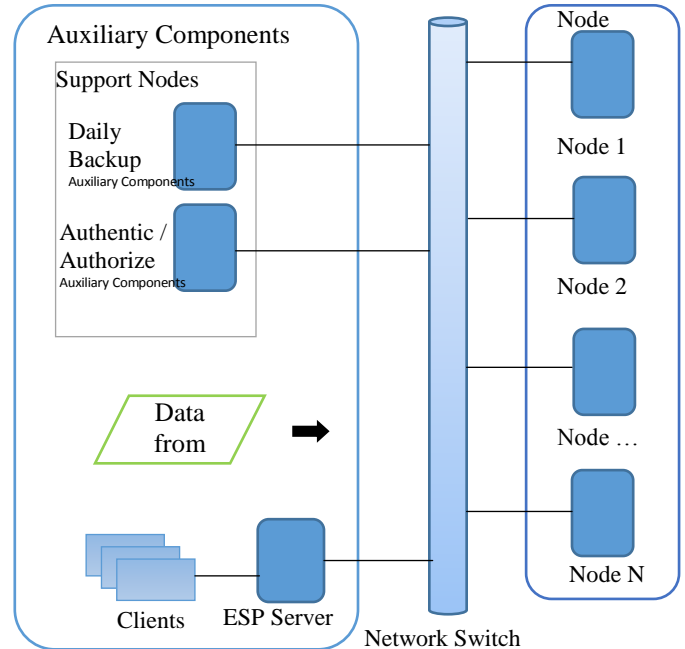


Fig -3 : Roxie

The second part of parallel data processing is responsible for online high-performance query processing and rapid data delivery. This cluster technique is called Roxie, which can accept thousands of simultaneous queries through web service interfaces and reply within a second [9]. It is utilized a distributed index file system to provide parallel processing of queries using an optimize execution environment and file system.

**F. Storm**

An open source real time data processing system is introduced by Nathan Marz with great idea of stream processing system in a single program is known as Strom. It is a real time distributed processing system to process the unbounded streams of data at very fast [10]. Storm has three main characteristics such as easy to use, highly scalable and low latency with guaranteed data processing.

**III. ROLE OF BIG DATA IN DIFFERENT SECTORS**

Now a day all sectors are being depend on smart technologies which are generating a lot of data and its make a large scale datasets. To find information in these datasets is not only difficult but also impossible through general data processing technologies, so all sectors are moving towards on Big Data concept to handle these datasets.

**BIG DATA IN ENTERPRISES**

Around the world different kinds business are going on and many companies are involved, some of them are business in multi nations and rest of all companies are business in intra nation. All companies are divided into two classes such as productive and non-productive. The productive companies are always getting a challenge to make cost effective and better product in comparison with other in the market. All companies has large amount of structured and unstructured datasets which can help to make the better product by analyzed the customer requirements and previous data related to that product. The analysis only can possible if the companies will adopt Big Data technologies. These technologies also help to minimize the risk of new product which is coming in the market. Big Data is used to ascertain the future of a product. For non-productive companies such as marketing, supply chain, e-commerce, Research & Development, etc. will get more benefit by adopting Big Data technologies to improve their business in very short time with efficient manner. According to IDC report predictive and prescriptive analytics technologies will adopt 40% of enterprises net new investment in business intelligence and analytics by 2020.



Fig4: Source: IDG Enterprise Data & Analytics Survey 2016

Use Big Data to analyze customers’ trends, which will accelerate the business by analyze which customer wanted to buy what kind of items and when do they need the particular items and their financial ability. This technology can reduce the item cost indirectly by used of real time and low expense advertisement in digital media such as smart devices, which are all time synchronized in internet.

**BIG DATA IN HEALTHCARE AND MEDICINE**

The healthcare sector has generated large scale of data in the form of patient care, record keeping, compliance & regulatory requirements but most of the data are unstructured and hard copy formatted [13].In the digital era all data from healthcare sector need to be stored into digital format to get access complex and diversified data for improving the quality of healthcare at less cost. Data apparently come from laboratory, pharmacy, insurance, medical imaging, and physician’s prescriptions will make together a large amount of heterogeneous datasets is called Big Data. This technologies now a big challenge in healthcare to synthesize and analyze the complex datasets to see what treatments are most effective for particular conditions and identify pattern related to drug side effects or hospital readmission and gain other important information which can help patients and reduce costs [14]. These all can possible if the relevant data is populated in data pool then Big Data will most successful in healthcare.

**Primary data pools are at the heart of the big-data revolution in healthcare.**

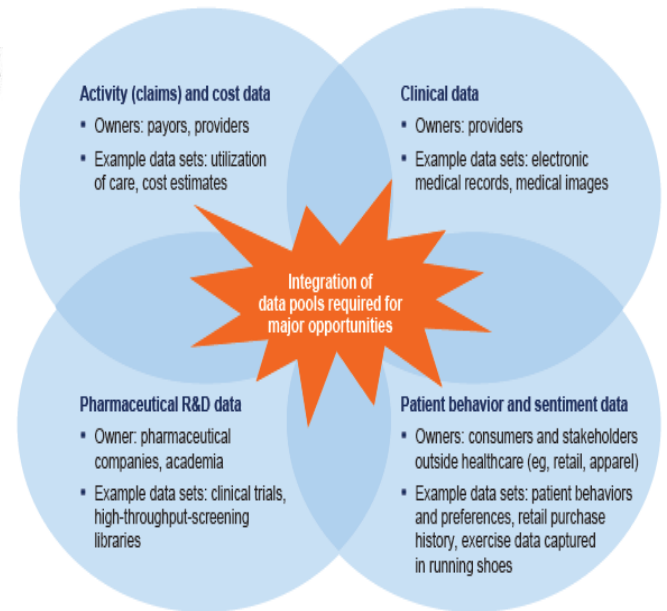


Fig 5: Source: McKinsey Global Institute analysis

The pharmaceutical R&D data is describe drugs’ therapeutic mechanism in of action target behaviour in the body, and its side effect and toxicity which will help to analyze through Big Data to prescribe what combination of medicines will applicable to a particular patient. Medicine payer can get patients’ buying preference and financial ability after analyzed patients’ behavior and sentiment data which describe all about inside and outside patients of healthcare context. Big Data analytics and applications in healthcare are

at nascent stage of development but rapid advances in the platform and tools can accelerate their maturing process [13].

### BIG DATA IN FINANCIAL SECTOR

Like other sectors the financial sector already been populated a huge amount of datasets and they are moving towards parallel processing concept to get insight into the data and bring out the hidden pattern, fraud trends, and better decision making. The Big Data technologies use in financial service sector to reduce overall IT operating Costs with real time analysis. The bank sector still supposed that Big Data is not a panacea for all woes because the massive trends of deception, but some of them in this sector has changed their perception after adoption of cloud solutions with strong privacy.

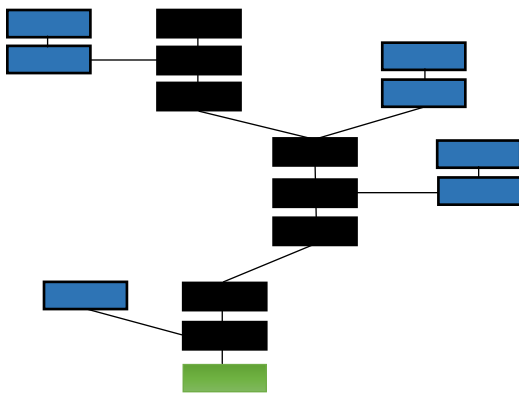


Fig 6: BlockChain Formation

Big Data is used to manage the increase in fraud and financial crime, manage risk & compliance as well as make strategic financial decision in all financial services including Bank. Big Data technologies use in Bank will help to concentrate on centralization efforts and data lakes which help the CDO (Chief Development Officer) to spearheading data governance initiatives [16]. To secure the financial transaction Block Chain strategy can be used in Big Data. A Block Chain as “an open distributed ledger that can record transactions between two parties efficiently and in a verifiable and permanent way the ledger itself can also be programmed to trigger transactions automatically” [17]. The main chain is started from genesis block(Green) and continuous form of black blocks which make the longest series of blocks up to the end. Orphan blocks(blue) are existed outside of the main chain. This concept first implemented by Satoshi Nakamoto in 2008 as a core component of digital currency bit coin, which represent public ledger for all transaction. In Block Chain technology no centralized official copy exists and no user is trusted more than any other, so that hackers unable to get central point to do the crime. The credit card companies are massively used Big Data technologies to analyze trends of customer as well as predict the fraud nature of customers.

### BIG DATA IN SCIENTIFIC RESEARCH

In scientific research maximum samples must be carried out better result than few samples. The traditional approach in experiment and theory, computational modelling and simulation could not provide most feasible result in some scientific research area such as seismology, galaxy formation and climate modelling. In these areas need large datasets to insights and it can reveal new hidden information. Now the Big Data technologies can help to processes a big amount of data to get insights and analyze. Massively use of Big Data in all sectors to get more accurate result and sharp decision making, so that it is called “ the fourth paradigm” of science. Human does not have more potentiality to perform statistical analysis on huge amounts of data to check a theorem is true or false. To deep insight, clear perception and analysis on collection of large amount scalable data using Big Data technologies can change the nature of science.

The scientific researcher of various universities and some other research institutions (e.g.- NIH, NASA, JPL, ISRO etc.) use this technologies to accelerate their research work by data collection and analysis [24].Big Data play a big role to provide data driven evidence of the impacts of Climate Change. The Un Spokesperson Stephane Dujuric said “this initiative will help build public understanding of how big data can reveal critical insights for strengthening resilience and mitigating emissions” [25]. This technology has the new opportunities to the scientist in all area to develop new models and algorithms for efficiently locating and finding new in the context of vast, unstructured, heterogeneous, non-linear, non-steady, and high dimensional real time datasets. The world famous space research centre NASA is using Big Data Technologies to collect “big bang” data from across the solar system to unlock the secrets of the universe.

### BIG DATA IN SMART POWER GRID

Today electric power is too much essential in daily life to maintain various types of electronics appliances such as TV, Mobile Phone, Freeze, Fan, AC, Computer, and etc. This electric is produced according to our demand and it supply through hierarchical network of stations to consumers. Day to day electricity demand is going to very high for our life style and unnecessary use of power. Electric consumptions in accordance production can be reduced by precise use of electricity with installation of smart power grid. Electric power is generated from Wind, Coal, Water, Tides, Nuclear, Solar, and etc. in this production area can be control by Smart Power Grid. Too many sensors are used to collect data repeatedly from these energy sources in particular time units. The produced power is distributed via main stations to substations then consumer destinations and everywhere need to put sensors to get data all times. The data populated from these sources are naturally heterogeneous and unprecedented data population make large size (e.g.- Petabytes or more) datasets.

These populated dates are stored into data enter and transmitted to central server. Continuous flow of data into smart power grid at very fast and efficient way from advance metering, IP-based smart meters and various type of appliances to make an intelligence system for electric optimization by using Big Data analytics. The population of heterogeneous data in Smart Power Grid can be classified into four parts such as Generation data, Transmission Data, Distribution data, and Consumer data [18].

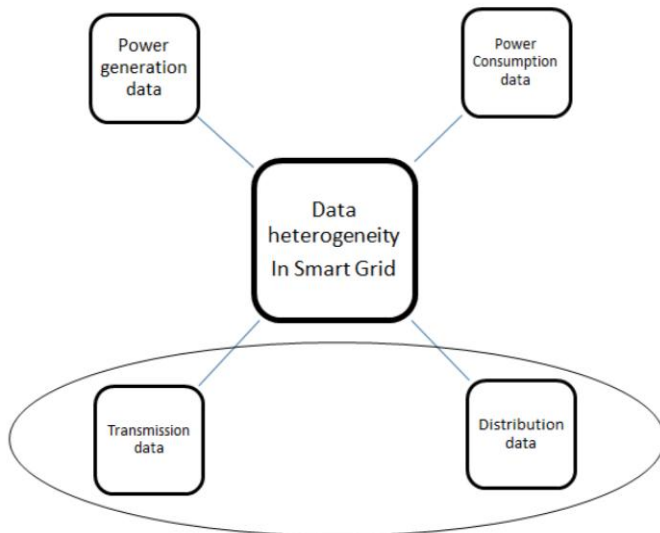


Fig 7: Heterogeneity in Smart Grid data [18].

Use Big Data to analyze large amount of gathering datasets and ensure consumer basis electric requirements in different times. The Smart Power Grid with help of Big Data can manage the power on region basis electric demand at particular time as well as control the electric production and supply.

**BIG DATA IN SMART TRAFFIC MANAGEMENT**

Dramatically increase the trends of vehicle use in rapid development society but old management concept and insufficient roads leads an unprecedented challenge into traffic control system. The traffic congestions is the big issue today in many cities around the world. Due to traffic congestion and poor control system we have spent most of the essential working time on the way and it is one of the big losses in society development. Impact of the congestion is very pathetic on emergency transport.

We can develop a Smart Traffic Management system is based on Big Data analytics, which will make an effective transportation. For this system all kinds of traffic information need to collect through various sensors to enable traffic monitoring and control. Smart Traffic Management is used mathematical models and practical emulation to predict traffic trends [21]. In this technology statistical analysis of traffic accidents, rules violations and driver records can

reduce the accidents prone and its real time processing ability can predict the occurrence of traffic accident, through micro wave detection, video surveillance systems and mobile detection systems [22].

**BIG DATA PLATFORM ARCHITECTURE FOR SMART TRANSPORTATION**

It is a three layered architecture such as basic business layer, data analysis layer, and information publishing layer. The function of basic business layer is to complete the job of all business units and produce basic business data. All input devices are connected in this layer as well as it includes GPS system, video surveillance system, vehicle information management system, driver information management system,

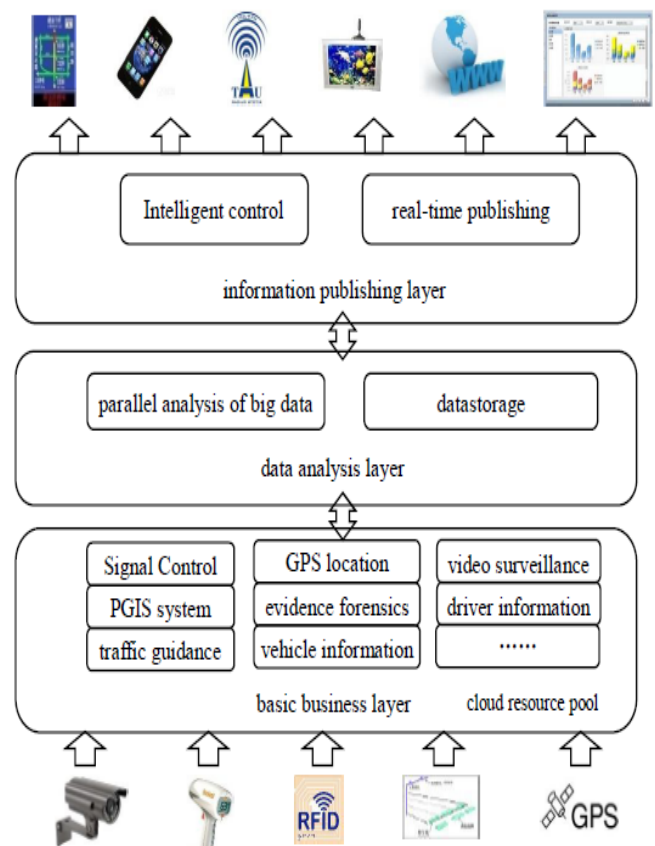


Fig 9: Smart Transportation architecture on Big Data Platform [20]

PGIS system, Signal control system, and so on. In data analysis layer HDFS technique is used for big data storage, use MapReduce as a parallel computing model and HBase as the database for processing data. Information publishing layer publish traffic condition to public after data analyze and generate the output through some devices such as smart mobile, internet, desktop application and reports.

#### IV. CONCLUSION

In all sectors variety of data are gathering at rapid speed and it makes a large amount of datasets which impossible to handle by human or any traditional data processing software. Modern trends of adopting Big Data technologies in such sectors where it can capture and manage unprecedented large amount database effortlessly. In this paper discuss, why the different sectors would adopt big data technologies and how they would get benefit after adopting this technologies. In this paper briefly discuss how open source tools Hadoop, HPC or Storm can be used to handle Big Data. The main objective of this paper is to highlight the sectors where Big Data is essential for their applications. It is described how Blockchain technology is used in Big Data to maintain the security at the time of financial transactions. If possible to use Big Data technologies in all sectors then it will be great challenge and opportunity to save time, cost and another big opportunity to save mineral in the Earth.

#### REFERENCES

- [1] big data techniques that create business value [www.firemex.com/thedealroom/7-big-data-techniques-that-create-business-value/&grqid=YtOUQFYe&hl=en-IN]
- [2] Apache Hadoop [http://en.wikipedia.org/wiki/Apache-Hadoop]
- [3] The Research on Risk Assesment of New product sample Development stage in manufacturing Enterprise [http://link.springer.com]
- [4] The Deciding Factor: Big Data & Decision Making
- [5] Innovation and Insights in Big Data. http://bigdata.teradata.com/US/Success-Stories/Innovation-and-Insights/&grqid=dvBseRck&hl=en-IN
- [6] Tasleema Nizam and Syed Imtiyaz Hassan. Big Data: A survey paper on Big Data Innovation and its technology
- [7] HPCC [www.wikipedia.org/wiki/HPCC&grqid=Odl-mqLJ&hl=en-IN]
- [8] A.M Middleton. ECL/HPCC : A united Approach to Big Data. Hand book of Data intensive computing. Springer 2011.
- [9] HPCC System : Introduction to HPCC(High Performance Computing Cluster)
- [10] Muhammad Hussian Iqubal, Tariq Rahim Soomro. Big Data Analysis: Apache Storm Perspective.
- [11] Wullianallur Raghupathi and Viju Raghupathi. An Overview of health analytics
- [12] IHTT: Transforming Healthcare through Big Data Strategies for leveraging Big Data in the healthcare industry.
- [13] Wullianallur Raghupathi and Viju Raghupathi. Big Data analytics in healthcare: Promise and potential
- [14] Peter Groves, Basel Kayyali, David David and Steve Van Kuiken. The Big Data revolution in healthcare
- [15] Exploring next Generation Financial Services: The Big Data Revolution
- [16] Top 10 Big Data Trends in 2017 for Financial Services. https://mapr.com/bblog/top-10-Big-data-trends-2017-financial-services/
- [17] Lansiti, Marco, Lakhani, Karim R. The truth about Blockchain. Harvard Business Review
- [18] GridShaym R, Bharathi Ganash HB, Sachin Kumar, Pabaharan poornachandran, soman KP. Apache Spark a Big Data Analytics platform for Smart
- [19] Swartz R. A, Laynch J.P, Zerbst S, Sweetman B, Rolfes R. Structural monitoring of wind turbines using sensor network smart structure and system
- [20] Gang Zeng. Application of Big Data in intelligent Traffic System
- [21] Smart Traffic Management with Big Data. https://inform.tmforum.org/features-and-analysis/
- [22] Chen Haitao. Improving traffic management with Big Data Analytics
- [24] Inside BIGDATA guide to scientific Research. http://www.insidebigdata.com/white-paper/
- [25] Taylor & Francis. Big Data for scientific research and discovery