Big Data Applications Architectures for Indian Society

*Dr. Anil K Maheshwari

ABSTRACT

This paper provides an overview of Big Data technologies and architectures and how they could be applied to meet the needs of Indian society with its large population and rural sector. A survey of 64 adults found that healthcare applications are most important, especially for the rural and the marginal sections of Indian society. This paper lays out proforma architectural designs of ten Big Data applications including five relating to the healthcare sector. These designs can serve as seeds for the implementation of other imaginative beneficial uses of Big Data for the Indian Society.

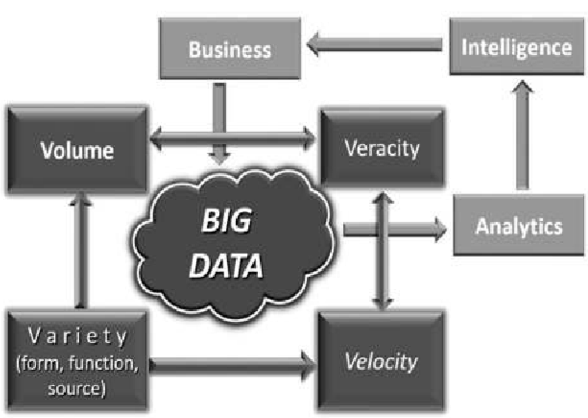
Keywords: Big Data, Healthcare, Education, Environment, Rural sector, India.

1. Introduction

Big Data is an all-inclusive term that refers to extremely large, very fast, highly diverse, and complex data that cannot be managed with traditional data management tools (Mayer-Schönberger&Cukier, 2013; Maheshwari, 2017). Ideally, Big Data includes all kinds of data, and helps deliver the right information, to the right person, in the right quantity, at the right time, to make the right decisions. Big Data can be harnessed by developing infinitely scalable, totally flexible, and evolutionary data architectures, coupled with the use of cost-effective computing machines. The infinite potential knowledge embedded within this Big Data cosmic computer can help connect with and enjoy the support of all the laws of nature.

Big Data can be examined on two levels (Figure 1). At a basic level, it is just another collection of data that can be analyzed and utilized for the benefit of the business. On a deeper level, it is a special kind of data that unfolds unique opportunities but also poses many technical challenges (Maheshwari, 2017).

Figure 1: Big Data Features



The quantity of data generated in the world has been relentlessly doubling every 12-18 months. The primary reason for the growth of data is the dramatic reduction in the cost of storing data by 30-40% every year. Big Data is being generated by billions of devices, and is communicated at the speed of the light, through the internet. The primary reason for the increased velocity of data is the increase in internet speed from 10MB/sec to 1 GB/sec (100 times faster), as well as mobile devices. Big Data is totally inclusive of all forms of data, for all kinds of functions, from all sources and devices. When all three Vs (Volume, Velocity, Variety) arrive together in an interactive manner, it creates a perfect storm (Mayer-Schönberger&Cukier, 2013). While the Volume and Velocity of data create the major technological concerns and the costs of managing Big Data, these two Vs are themselves driven by the 3rd V, the Variety of forms and functions and sources of data (Maheshwari, 2017). The uneven accuracy (Veracity) of data complicates the management of Big Data further.

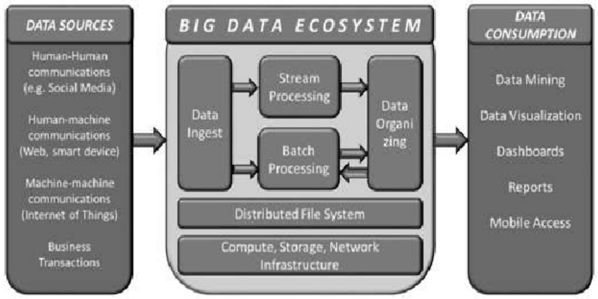
2. Big Data Architecture

Big Data Application Architecture is the configuration of tools and technologies to accomplish the entire application. An ideal Big Data architecture would be resilient, secure, cost-effective, and adaptive to new needs and environments. This can be achieved by beginning with a proven architecture, and creatively and progressively restructuring it as additional needs and problems arise. Big Data architectures should ultimately align with the architecture of the Universe itself, the source of all invincibility.

Figure 2 shows a generic Big Data Architecture (Maheshwari,2017). There are many sources of data. All data is funneled in through an ingest system. The data is forked into two sides: a stream processing system and a batch processing system. The outcome of these processing can be sent into NoSQL databases for later retrieval, or sent directly for consumption by many applications and devices.

* Professor, Maharishi University of Management, Fairfield, IOWA, USA, akm2030@gmail.com

Figure 2: Generic Big Data Architecture



A Big Data solution typically comprises these as logical layers. Each layer can be represented by one or more available technologies.

Big Data sources layer: The choice of sources of data for an application depends upon what data is required to perform the kind of analyses you need. The sources of Big Data include human-to-human communications such as social media, human-machine communications such as web logs, and machine-to-machine communications such as RFID (Radio Frequency Identity Devices) data. Big Data will vary in origin, size, speed, form, and function. Data sources can be internal or external to the organization. The scope of access to data available could be limited. The level of structure could be high or low. The speed of data and its quantity will also be high or low depending upon the data source.

Data Ingest layer: This layer is responsible for acquiring data from the data sources. The data is ingested through a scalable set of input points that can acquire at various speeds and in various quantities. The data is sent to a batch processing system, a stream processing system, or directly to a storage file system. Compliance regulations and governance policies impact what data can be stored and for how long. Apache Kafka is a popular open-source tool for this layer (Garg, 2013).

Batch Processing layer: The analysis layer receives data from the ingest point or from the file system. Data is processed using parallel programming techniques such as MapReduce to process it and produce the desired results (Dean & Ghemawat, 2008).This batch processing layer thus needs to understand the data sources and data types, the algorithms that would work on that data, and the format of the desired outcomes. The output of this layer could be sent for instant reporting, or is stored for an on-demand report.

Stream Processing layer: This technology layer receives data directly from the ingest point. Data is processed in real time to produce desired results (Marz& Warren, 2015). This layer thus needs to understand the data sources and data types extremely well, and the super-light algorithms that would work on that data to produce the desired results. The outcome of this layer too could be stored in the NoSQL Databases.

Data Organizing Layer: This layer receives data from both the batch and stream processing layers. Its objective is to organize the data for easy access. It is represented by NoSQL databases. There are a variety of NoSQL databases to suit different needs (Sadalage & Fowler, 2013). SQL-like languages like Hive and Pig can be used to easily access data and generate reports from these databases.

Infrastructure Layer: At bottom there is a layer that manages the raw resources of storage, compute, and communication. This is increasingly provided through a cloud computing paradigm.

Distributed File System Layer: This is the heart of a Big Data system. It would store huge quantities of data and make it quickly and securely, available and accessible to the other layers. Hadoop Distributed File System (HDFS) is the primary technology in this layer (White, 2012). It would also include supporting applications, such as YARN (Yet another Resource Manager), that enable the efficient access to data storage and its transfer.

Data Consumption layer: This is the final layer, and it consumes the output provided by the analysis layers, directly or through the organizing layer. The outcome could be standard reports, data analytics, dashboards and other visualization applications, recommendation engine, on mobile and other devices.

3. Big Data Applications

An ideal Big Data application will take advantage of all the richness of the infinite amounts of data available and produce pointed relevant information to make good decisions for the organization to be responsive and successful. Big Data applications can help align the organization with the totality of natural laws, the source of all sustainable advantage in the world.

Companies like the consumer goods giant, Proctor & Gamble, have inserted Big Data into all aspects of its planning and operations. The industrial giant, Volkswagen, requires all its business units to identify some realistic initiative using Big Data to grow their sales. The entertainment giant, Netflix, processes over 500 billion user actions every day to understand and process their customers' needs. Smart cities are using Big Data to automate many services and lower the costs (Maheshwari, 2018).

Recording and processing all of this data requires much talent, resources, and time. This data can be used imaginatively and meaningfully to derive business benefit. Some common business objectives (Treacy, & Wiersema, 2007) would be:

- Increase revenue and profit
- Reduce costs and fraud
- Improve infrastructure quality and performance

There are three major types of business applications of Big Data with different levels of transformational potential (Maheshwari, 2017).

- Monitoring and Tracking: These are the basic applications of Big Data. They help improve the efficiency of the business, in almost all industries.
- Analysis and Insight: These are next level of Big Data applications. They can increase the effectiveness of business and have transformative potential.
- New Product Development: These applications are totally new concepts that did not exist earlier. They have the potential to disrupt industries and generate new avenues of revenue.

4. Big Data Applications for India

MBA students at a premier business school in Eastern India were asked to write a short and imaginative essay on a Big Data application that would be most useful to the communities and the contexts they belonged to. Sixty (64) students wrote such essays. These essays were then coded for the domain they belonged to. Table 1 shows the distribution of the domains of the applications. Healthcare was the most important domain with 34% of the applications belonging to that domain. An additional 8% of applications belonged to Mental Health. Social issues were next highest with 15%. Career and education domains were other important issues. Urban environment and farmers' welfare is other important domain where Big Data can help

5. Big Data Applications and Architecture

In this section we will discuss architecture examples for some important Big Data applications, with special focus on the needs of India. The applications are in the area of health, education, environment and rural sector.

Table 1: Domains of Big Data Applications

Domain	Count	Percent
Health	22	34.4%
Social	10	15.6%
Career	6	9.4%
Mental Health	5	7.8%
Education	4	6.3%
Farmers	3	4.7%
Finance	3	4.7%
Pollution	3	4.7%
Urban	3	4.7%
Food	2	3.1%
Crime	1	1.6%
Environment	1	1.6%
Weather	1	1.6%
Total	64	100%

Application 1 : Big Data Analytics in Health Care

Healthcare data is increasing exponentially in quantity and diversity. Comprehensive data about patients, doctors, medications, facilities, insurance, and payment systems can improve health care experience for practically everyone. Ideally, Big Data can help provide personalized care through right intervention to the right patient at right time. This can be achieved by properly gathering, curating, securing, and analyzing the data for effective delivery of health care services.

Data source and types would include patient medical histories, doctor-patient conversations, medical research on diagnoses and treatments, pharmaceutical drugs, radiology, and more. All this data should be curated with accurate metadata so that it describes what, when and how, the data was created. In India, for example, data can be done efficiently using AADHAR id. The data should be kept private and secure to maintain compliance with healthcare's strict privacy and security rules such as Digital Information Security in Healthcare Act (DISHA) standard from NHDS in India.

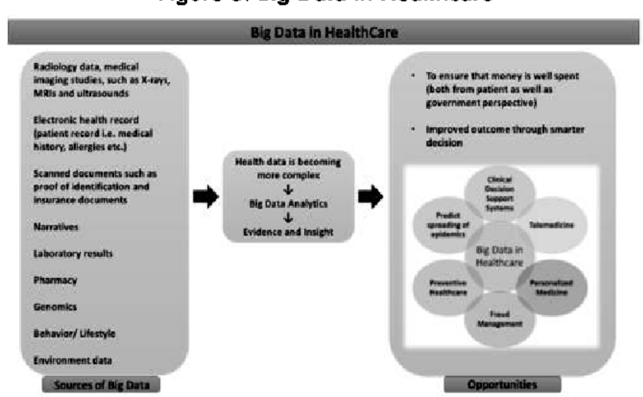
Healthcare applications include preventive care recommendations, early childcare, diagnosis of complex diseases, testing prescription medications against side effects and overdoses, etc. Real-time tracking of medical conditions for critical patients and sending alerts in case of medical emergencies can save valuable lives. The needs for healthcare services and resources can be predicted by socioeconomic and geographical factors. In big cities missed appointments due to rush hours can be predicted to avoid waste of healthcare capacity. Here is an overall architecture for Big Data Analytics in Healthcare.

Here is an overall architecture for Big Data Analytics in Healthcare.

Application 2: Healthy Food and Fitness

Customers want delicious and healthy foods. However, there is a lack of proper nutritional awareness and much

Figure 3: Big Data in Healthcare



confusion among people. A majority of urban Indians are overweight, especially the young adults with sedentary lifestyles. People need to eat healthy food in the right quantity at the right times of the day, and exercise often. People should have greater self-awareness about their health needs, and what they should be doing, eating and drinking. Big Data Applications can help people with tracking their food purchases, calorie intake, and calorie burn through exercise.

Sources of Data:

- User Profiles: Medical test records, dietary preferences, and health goals
- Social profiles: Support network of other friends and family members on social media
- Neighborhood profiles: Availability in the vicinity of desired groceries, medicine and supplements, restaurants, etc.

Applications:

- Progress Monitoring: One can track intake of food, vitamins, nutrients and water based on goals. This progress report can also be shared with dieticians and doctors, fitness trainers, and others who can evaluate the progress report and advise appropriate changes.

A big data architecture for such a focused application would look like as shown in Figure 4.

Application 3: Women's Mental Health care

Big Data can help improve mental health, especially for women. India is currently the most depressed and the second most anxious country in the world, with 60 million depressed and 40 million anxious people. Indian women are 2-3 times more likely to exhibit symptoms of depression, anxiety and other psychological distress. The female population suffers from mental harassment, verbal abuse, molestation, workplace sexual harassment, domestic violence, and other social evils. Big Data can be a friend that can empower women with a platform to get the help they need. The key lies in spirit of collaboration and information sharing. The Big Data platform can use large patient data sets and machine learning to develop predictive models of unhealthy situations.

Applications:

- Progress Monitoring and Preventive care: The social media and other data could be used to track a woman's mental health indicators. One can then develop a predictive model of what might happen based on social media posts, and other indicators. The client can be counseled electronically when a situation seems emergent.
- Emergency care: Provide emergency numbers when client is having suicidal thoughts or having another risky situation. Public health systems and volunteer agencies can engage with the client.

The Big Data Architecture for this application is shown in figure 5.

Figure 4: Big Data Architecture for Healthy Foods and Fitness

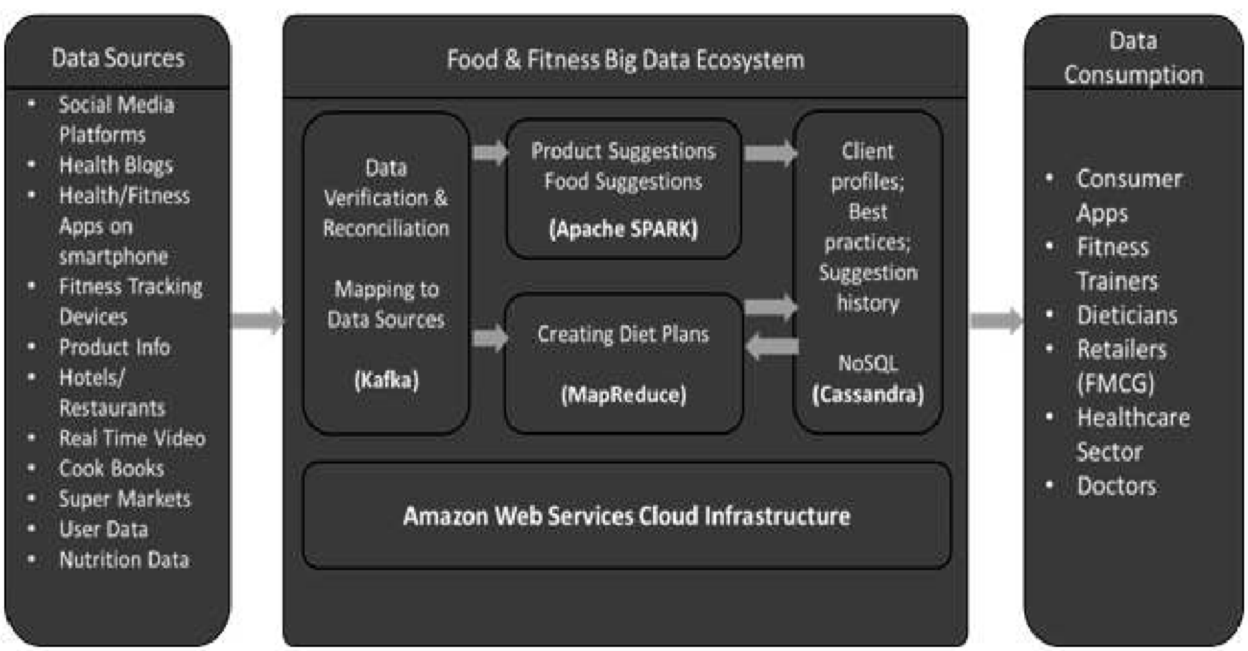
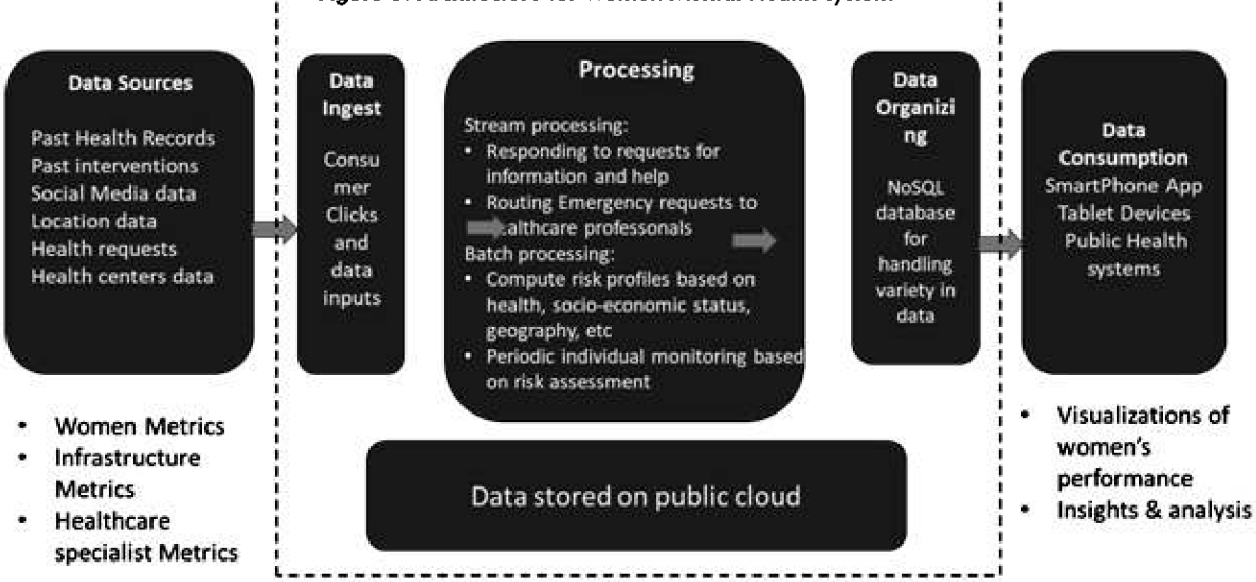


Figure 5: Architecture for Women Mental Health system



Application 4: Rural Health care

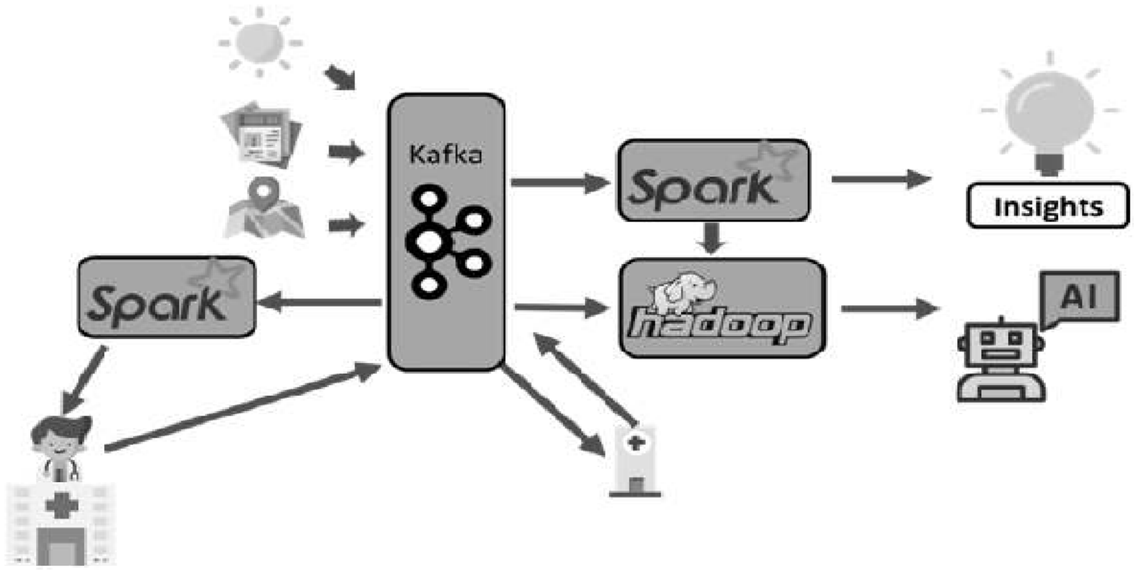
India is rapidly making progress towards being a developed country, but the villages have minimal access to basic healthcare. About 800 million people in India are living in rural areas. There are very few primary healthcare centers for the villages. India has low availability of 0.6 doctors per 1000 people, compared with a global standard requirement is 2.5. There are only 1.3 nurses per 1000 people, compared with the standard of 5.0. Similarly, the number of hospital beds is only 1.3 per 1000 people, which is about a third of the standard of 3.5. There is a great need for more resources, and also to utilize them well.

Big Data technology can utilize all the diverse data related to rural health care to ensure that treatment reaches the needy where they are. Here are some of the ways:

- Setting Up facilities: The identification and mapping of locations where the medical facilities are sparse. The existing hospitals and the medical centers can located using the government healthcare website of each district. One will find that the residents of many villages must travel long distances (50-100 kms) for basic medical treatment. Districts with higher number of patient occurrences can be provided with more staff.
- Electronic Health Records (EHRs): Patient's current symptoms, health history, genetic history, medication records can be stored. Previous case histories of what the doctor prescribed on basis of patients reading. Finally, weather conditions, location specific diseases history, and other contextual data too should be included.
- Real-Time Alerting: The data could be used to track epidemics as well as community health by geographies and socio-economic categories. This could provide health practitioners with advice on prescriptive decisions so that the right medications are stocked in the right locations.
- Efficient delivery of Service: Online video conferences with doctors can help remote medical service. Trial and error type of medication can be recorded and analyzed to minimize unnecessary treatments
- Detecting spreading of infectious diseases - Identify the particular problems or disease spreading in particular area and take instant corrective action.

Figure 6 shows an overall Big Data architecture for in Rural Health care.

Figure 6: Rural Healthcare Big Data Processing Architecture



Application 5: Automated Medical Assistant

An Artificial Intelligence based automated assistant can provide non-critical guidance in conversational mode. It can schedule appointments, follow up on reports, send alerts, and it can also assist in clinical diagnosis. It will help in scaling up health care delivery in a standardized and error-free mode. It can reduce crowding at health care centers. It can also proactively take care of health. Through clinical diagnostical assistance, it can improve doctor's productivity and reduce consultation time.

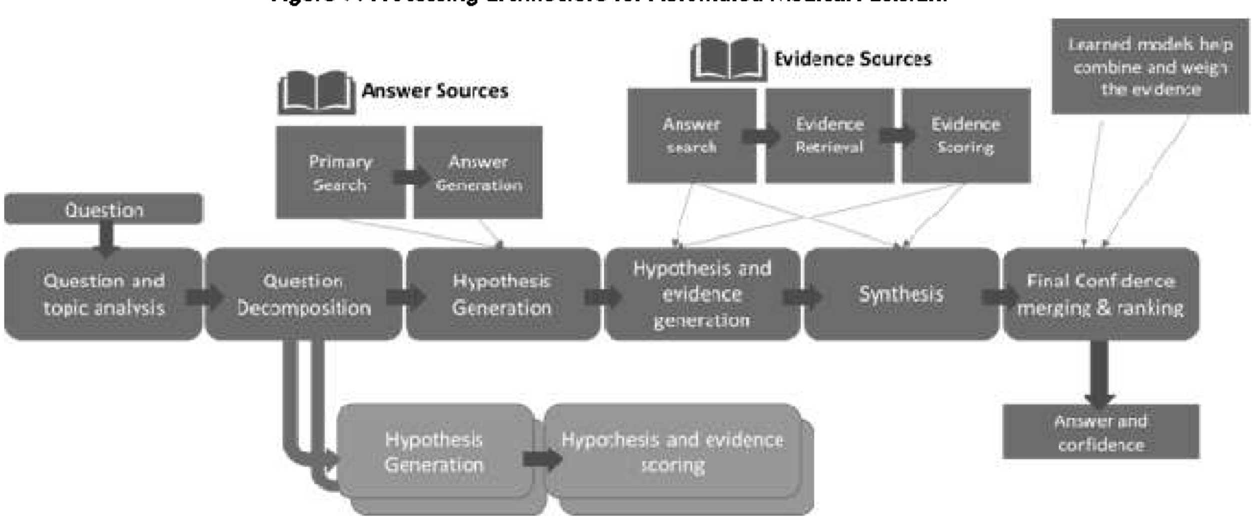
Data sources include:

- Electronic Health Records such as patient consultations, treatments and conversations
- Medical research studies in Journals
- Hospital staff, services, facilities records

Inspired by the architecture of the IBM Watson system, these data are processed using Natural Language Processing technologies such as:

- Information Retrieval Engine to extract key information from patient conversation/ clinical records
- Image Interpretation engine: To extract data from documents submitted online
- Text Classification engine: To classify patient responses into yes/no or to classify conditions as critical/urgent
- Document Summarization Engine: Summarize clinical volumes and map to the required topic at hand. E.g. it can suggest medical guidelines and clinical pathways.

Figure 7: Processing architecture for Automated Medical Assistant



Application 6: Primary Education

Public Education system is a key to national development. Teacher and student performance needs to be monitored consistently. Teaching pedagogy needs to be improved. Facilities and infrastructure need investment. Teaching posts sometimes remain vacant for long periods of time. There is also sometimes misallocation of resources. The goals should be:

- Gather better insights linking student health and socio-economic status with performance
- Cost reduction due to accuracy in student and teacher attendance
- Proper allocation of funds and resources wherever required.

Big Data can help by gathering all the data and make it available on a need basis to the appropriate decision makers. The data elements could be as follows:

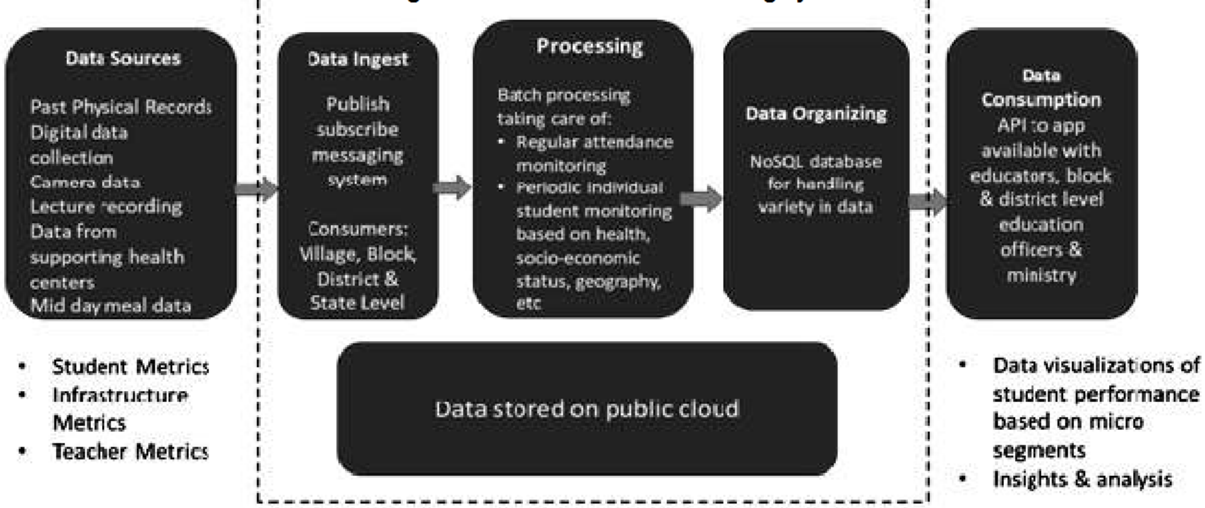
- Basic student profile from AADHAR
- Attendance recorded by fingerprints/biometric
- Student performance in every evaluation (school level, state level, board level)
- Student scholarships and grants distributed to students
- Basic needs of students which complement studies (uniforms, stationary, cycles, books, library records etc.)
- Student participation in extra-curricular activities and competitions
- Student health monitoring (mid-daymeals, BMI, health check-ups)
- Teacher attendance records
- Teacher rating/feedback
- Teacher training/skill development programs undergone
- Teacher profile - academic history, domicile, track record
- School facilities like toilets, labs, playgrounds, sports equipment mid-day meals, drinking water
- Science & tech resources (computer labs, science lab, audio-visual equipment etc)
- Records of funds issued for maintenance activities

The Big Data Architecture would be something like this:

Application 7: Traffic Management

Living in big cities is a traffic nightmare. The commute times can be long and frustrating. There are also economic losses due to time wasted, fuel wasted, air pollution, accidents, road rage, and more. People want to find the fastest way to reach their destination. There are often multiple travel options to get there. Thus there is a need to help commuters to planning their journey better. One way is to design an easily accessible smart traffic management system that can be used by millions of commuters to suggest optimal time and mode for travel. Such a system will ingest vast amount of location data from different sources, and then use prescriptive analytics for generating recommendations to traffic management authorities as well the drivers. Public transportation operators and urban city traffic management can use such as system to become responsive to commuters and also create new business opportunities. A related problem caused by traffic in big cities is air pollution, or Air Quality Index (AQI). That too can also be included in computing the smartest path to travel.

Figure 8: Public Education Monitoring System



Sources of Data:

- Surveillance Cameras to give a sense of the real-time traffic at various places in the city
- Schedules of Public Transport such as buses, local trains, metros, etc.
- Weather forecasts including which routes are likely to be flooded
- Historical Data such as average time spent on a particular route
- Anomalies such as construction and social event
- GPS devices data for vehicle location
- Social Media posts on traffic status
- Pollution metrics

Applications:

- The data provided to the App by customers will help recommend optimal routes.
- App can also offer Multi-modal Route Search Service to Assist users in findingoptimal route with multi-modal transportation (train, bus, subway etc.)
- Cities could analyze the data and analysis to design Innovative services for passengers
- Help Traffic officers and regional traffic offices monitor traffic flow.
- Analyze real-time traffic conditions to deploy security personnel at strategic locations.
- Marketing organizations can provide customers with real-time information on products and store locations.

Figure 9 shows a possible Big Data Application Architecture for this application.

Data will be captured from cars, drivers, cameras, using a high-throughput Kafka implementation, and then would be saved in a Hadoop Distributed File System (HDFS). The analytics engine will continuously run models on the data to generate recommendations that are logged in a NoSQL database. Applications can access the recommendation and serve various categories of consumers. The models will be continually updated with new possibilities opening up, such as allowing opening up the hard shoulder on certain stretches of the freeway to quickly clear the traffic.

Application 8: Water Resources Management

Fresh water is a precious and increasingly scarce resource. A developing country like India is likely to face such intense water shortage in a few years that half a billion may not be able to grow their own food. Around 60 percent of the water meant to be used for irrigation is lost. Failure of adequate rains, and subsequent lack of irrigation water, often leads to crop failures and even farmer suicides. Storing rainwater in reservoirs is useful to avoid the see-saw patterns of droughts and floods in the country. The Indian Rivers Inter-link is a large-scale civil engineering project started many years ago, with the goal of effectively managing water resources in India through a network of reservoirs and canals.

Big Data can help determine the best places to lay such linkages and intermediate water reservoirs. Once such a network is developed, Internet of Things (IoT) sensors infrastructure can be deployed to assess the condition of soil based on the water level in reservoirs along with data of predicted rainfall. A predictive system can suggest what crop the farmers may plant in different regions. Such a network can also be used for commercial transportation and logistics as a cheaper alternative to road and rail linkages. This project can improve the lives of farmers drastically as with continuous supply of water they won't be depended on rainfall completely and have more stable income. They can also plant crops that they believe will enable them to fetch the maximum revenue and profits, thus elevating their lifestyle.

Sources of Data:

- Surface water resources data
- Underground water resources data
- Land profile data from satellites imagery
- Crop and soil data

Figure 9: Big Data Architecture for Traffic Management

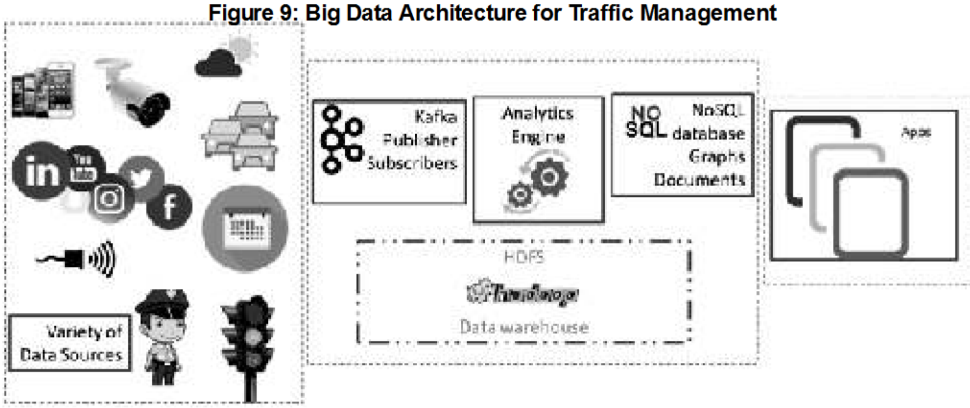


Figure 10 shows a basic Big Data Architecture for Water Resources Management.

Application 9: Precision Farming

Many districts in India are becoming a graveyard for farmers. About 34 farmers & farmhands kill themselves every day. Deaths are attributed to indebtedness, farming-related issues and poverty. They have taken on huge debts and are unable to pay them back. They are not able to make correct decisions on crop selection because of lack knowledge regarding the fertility of soil and weather conditions. Farmers are also not aware of effective use of pesticides and fertilizers, depending upon the condition of their farm. Farmers also often lack awareness of subsidies available from the government. Climate conditions are also unpredictable, and drought leads to loss of crops.

Precision farming helps optimize the use of resources to generate a healthy return on investment for the farmers. It also helps preserve scarce national resources by using less and achieving more. Precision farming recommends farming actions based on the precise measurements of water availability, land slopes, soil health, mineral levels, moisture content, etc. Geographical information systems can use remote sensing technologies to collect data. The use of Global Positioning Systems can help with better accuracy of above data collection.

Major types of data collected would be around the soil, crop and climate.

- Soil health includes physical condition, texture, structure, moisture, Ph-level, nutrients and more
- Crop parameters include crop stress, crop tissue and nutrients, crop population, weed patches, fungal or insect infestation, and crop yield.
- Climate parameters primarily include humidity, rainfall, wind speed, and temperature.

Applications and Benefits

- Increase crop yield: Precision agriculture techniques, farming practices, fertilizer inputs based on soil health will help increase productivity and increase crop yield thereby helping farmers.
- Getting better prices: By providing market prices and spot prices of various crops, and by predicting the availability of crops based on the acreage cultivation so that farmers know which crop to plant to get better prices.
- Reducing crop failure and providing insurance cover: Active and real time implementation of integrated pest mgmt. techniques can help reduce crop failures. Also, satellite imagery data about health of crops help insurance companies to settle claims faster.
- Knowledge dissemination: Helping farmers not only by providing relevant farming knowledge and efficient agricultural practices information but also providing lessons on managing resources.

Big Data along with the Internet of Things and Machine Learning can help with the entire process to make sense of the entire data for the whole country.

Figure 10: Big Data Architecture or Water Resources

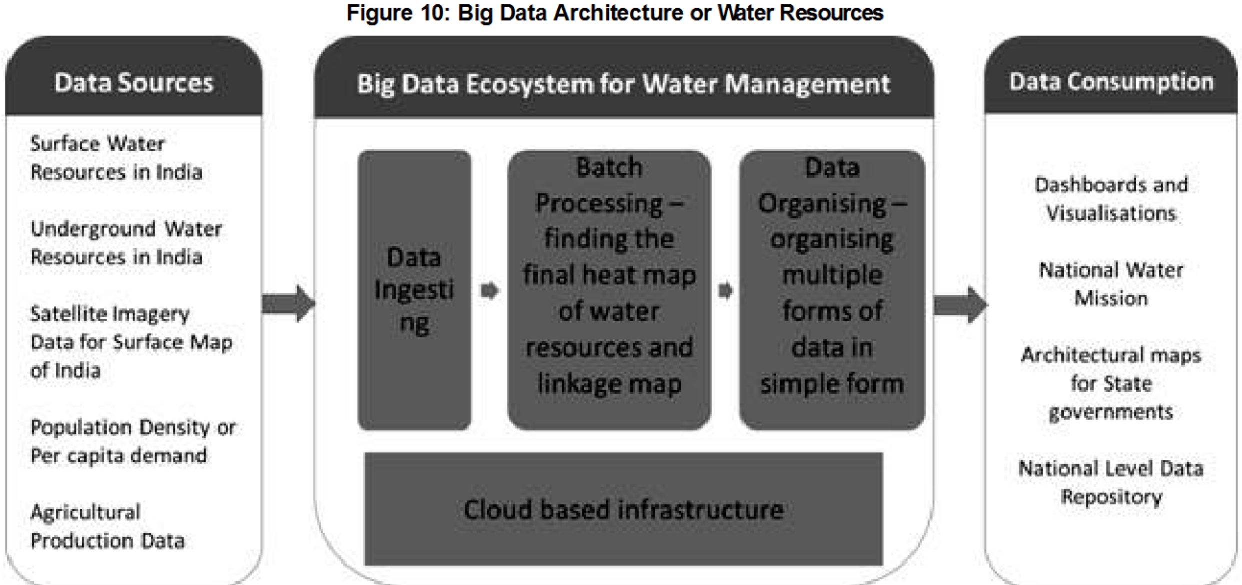


Figure 11 shows the overall Big Data architecture for this application.

Application 10: Farmers Finance

Farmers need financial support for a variety of reasons. 70% farmers in India own less than one hectare of land. With non-existent collateral, private banks refuse to give loans to them. Farmers and rural villagers who don't have formal documentation and proofs of income usually score very low in the traditional loan assessment thus are forced to take loans from loan sharks in the village who charges between 30-50% interest rate. Unpredictable weather, poor guidance on water availability, improper knowledge about crops and soils and below par prices of crops at the end of year leaves the farmers with no money to repay the loans.

A Peer-to-Peer Lending platform by a nationalized Bank (a subsidiary registered as NBFC) can enhance the financial inclusion in rural villages through access and ease. NBFCs offer an end to end solution provided for the betterment of the farmers. This includes appraisal, monitoring and sale of assets. It is particularly beneficial for individual borrowers and SME's with no formal documentation, mortgages and/or registration certificates. To estimate total agricultural production of the land, satellite-based crop clustering technique coupled with traditional methods can be used. Farmers will have a fair idea to choose their land for agricultural purpose. This data can also be used by financial institutions for better credit appraisal in case of loans and proper insurance premium predictability.

Figure 11: Big Data Architecture for Farmer Support

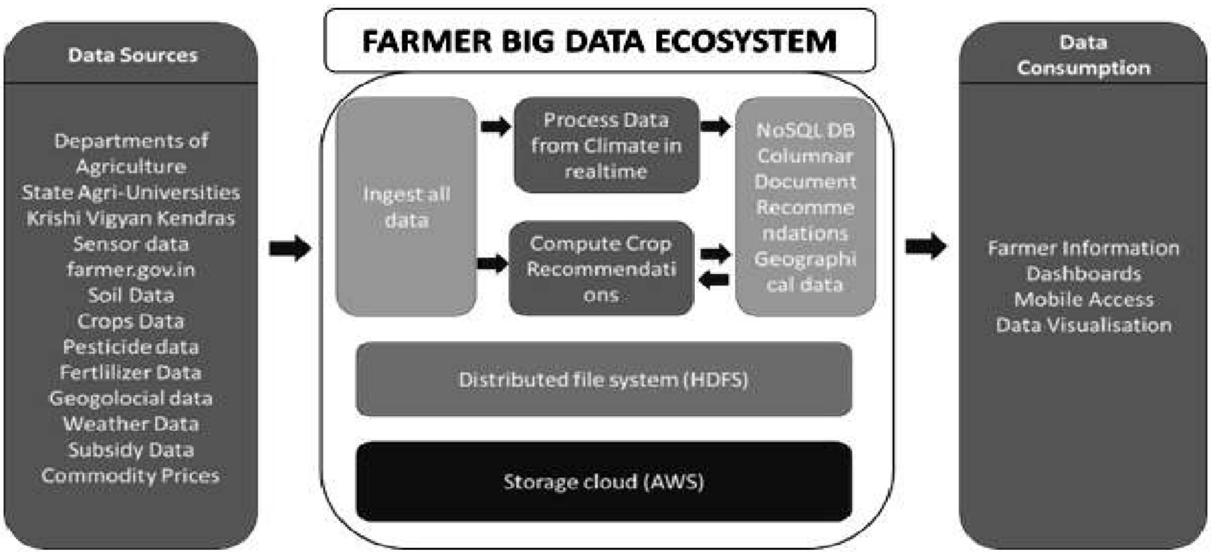
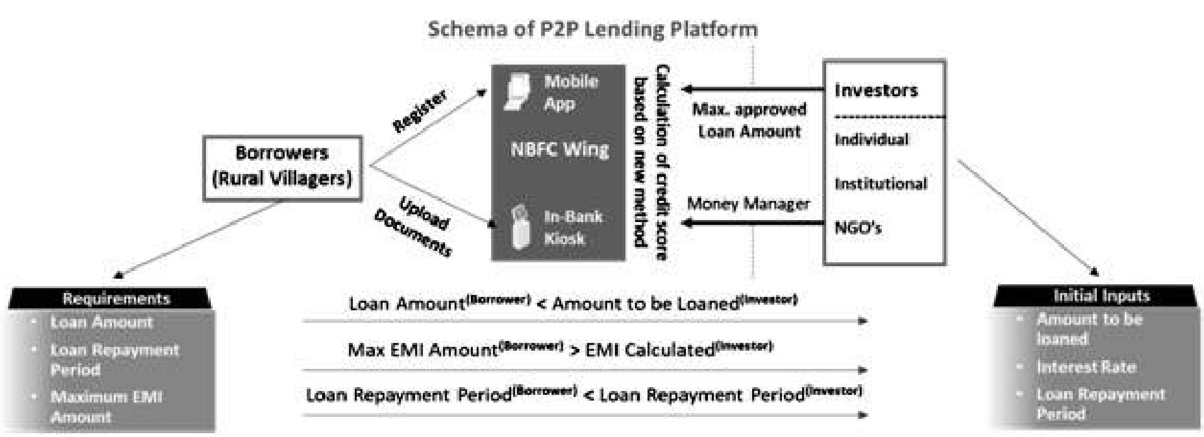


Figure 12: Big Data Processing Architecture for P2P Lending for Farmers



The investor will have access to documents uploaded by the borrower. Upon successful verification, the borrowers and the investors will sign a legally binding loan agreement. Both disbursals and loan repayment can be done through an escrow account. The lender pre-funds his escrow account with the amount he wishes to invest. After the loan is disbursed, the lender will receive monthly payments every month from the time borrower agrees to pay. If a borrower fails to pay within a stipulated time, a penalty is levied on the borrower which is payable directly to the lender. Lenders have an option to reinvest the money to earn better return.

6. Conclusion

Big Data is growing at an exponential rate. Organizations are imagining different ways of utilizing it for business and social benefits.Big Data architectures are multi-tier designs to handle the volume, velocity and variety of data. They use technologies for ingesting, storing, processing, and delivering data to user applications. Research found that healthcare remains the primary domain for use of Big Data in the Indian environment. Education, environment, and farmer issues are other important areas. This paper described five potential applications of Big Data in healthcare, and another five from the areas of education, environment and rural sectors, along with potential Big Data architectures to suit those applications. These applications and architectures provide the seeds to imagine new applications to make health, wealth and joy accessible to all.

References

1. Dean, J., & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. Communications of the ACM, 51(1), 107-113.
2. Garg, N. (2013). Apache Kafka. Packt Publishing Ltd.
3. Maheshwari, A. (2017). Big Data. McGraw-Hill, Noida.
4. Maheshwari, A. K. (2018). Application of Big Data to Smart Cities for a Sustainable Future. Handbook of Engaged Sustainability, pp 945-968. Springer.
5. Marz, N., & Warren, J. (2015). Big Data: Principles and best practices of scalable real-time data systems. New York; Manning Publications Co.
6. Mayer-Schönberger, V., &Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt
7. Sadalage, P. J., & Fowler, M. (2013). NoSQL distilled: a brief guide to the emerging world of polyglot persistence. Pearson Education.
8. Treacy, M., & Wiersema, F. (2007). The discipline of market leaders: Choose your customers, narrow your focus, and dominate your market. Basic Books.
9. White, T. (2012). Hadoop: The definitive guide. O'Reilly Media, Inc.