International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 01 , January 2017
www.ijarse.com

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

APPLICATION AND IMPLEMENTATION OF BIG DATA IN MEDICAL SCIENCE

[1]Dr. Gagandeep Jagdev, [2]Rupandeep Kaur, [3]Dr. Vijay Laxmi

[1]Dept. of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib (PB).

[2]Research Scholar (M.Phil. Comp. Applications), Guru Kashi University, Talwandi Sabo (PB).

[3]UCCA, Guru Kashi University, Talwandi Sabo (PB).

ABSTRACT

Medically related data underwent a huge increase in the past few decades. This tremendous increase in clinical data is referred to as a big data in medical science. These huge medical datasets bring many challenges related to storage, processing and analysis of data. Today almost 80% of data in medical sector is unstructured, but is clinically relevant. Data is scattered in the form of individual EMRs, doctor's notes, at imaging and scanning centers and at medical stores. There is a need to get access to this valuable data and collect it in order to utilize it in the best possible manner. It is often noticed that people do not get proper treatment on time and their health gets worsened to such an extent that one has to undergo through expensive and painful medical diagnosis and treatment. Healthcare analytics if used efficiently has an enormous potential to minimize the expenditure of treatment, forecast outbreak of epidemics, shun preventable diseases and develop the overall health quality of life. Just as we people leave digital trails these days, so do our cells and have been doing so for decades as a result of biomedical research. This data, much of it stored as genetic transcripts in huge public databases, constitutes a goldmine for research and drug development. In coming ten years, eighty percent of the work people do in medicine will be replaced by technology. And medicine will not look anything like what it does today. This research paper is concerned with mining the gigantic self-constructed medical database and come out with efficient results by writing scripts and queries via Apache Hadoop framework to gain timely insight of different deadly diseases having different causes to ensure well-being of humans according to famous saying: "Prevention is better than cure".

Keywords- Apache Hadoop framework, Big Data, EMR (Electronic Medical Record), HIS (Healthcare Information System), Map-Reduce.

I. INTRODUCTION

Big data are rapidly all over the place. Everyone seems to be collecting, analyzing, and making money from it. No matter whether we are talking about analyzing zillions of Google search queries to predict flu outbreaks, or zillions of phone records to detect signs of terrorist activity, or zillions of airline stats to find the best time to buy plane tickets, big data are on the case. By combining the power of modern computing with the enormous data of the digital era, it promises to solve virtually any problem like crime, public health, the evolution of grammar, etc. Data in digital form has received a growing significance in both business and private domain. During early beginnings of email usage in the late 1970s and early 1980s, it took a while to set up a connection via a slow modem and then type a message on a black and white screen using a line editor and next sending the message off, and finally shutting down the connection again. The number of characters making up the message during that time was small compared to what we can put in an email today. At that time, nobody would have believed that the same could be done at some point of time from an intelligent phone with much higher speed and considerably bigger content. But besides digitalization and the fact that digital objects have become larger over time, technology has also enabled faster transportation of data as well as human productions of data. The result is so overpowering that the term "big data" seems appropriate [1, 2]. Intel recently reported that in a single Internet minute, 639,800GB of global IP data gets transferred over the Internet, which can be broken down into emails, app downloads, e-commerce sales, music listening, video viewing, or social network status updates, and this number will increase significantly over the next couple of years. Four features of big data often known as 4 V's of big data are volume, velocity, variety and veracity [3, 4]. Volume refers to the huge amount of data that is beyond the storage of a single organization. Velocity refers to the fact that data often comes in the form of streams which do not give the user a chance to store it but has to act instantly on it. Variety means that data can be in any form like unstructured, semi-structured or structured. Veracity refers to the fact the data may or may not be trustworthy or uncertain [5, 21].

Health care has undergone a tremendous change in the last one decade. The entire system has been digitized to provide fast, appropriate, on-time and effective care when compared to conventional systems as shown in Fig. 1 [6, 8, 22].

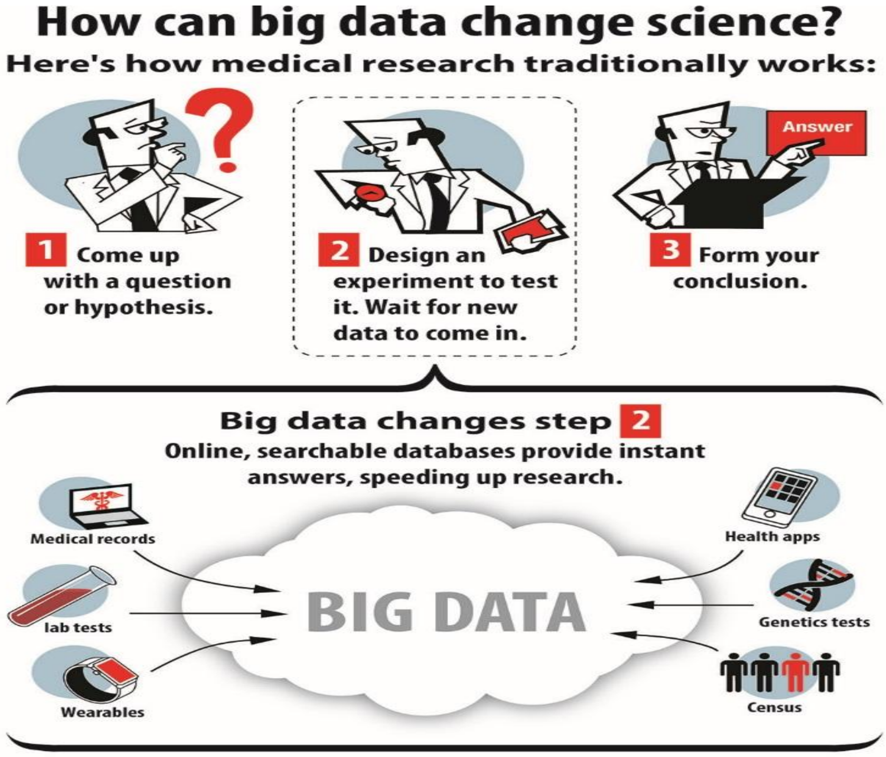

How can big data change science?
Here's how medical research traditionally works:
Answer
1 Come up with a question or hypothesis.
2 Design an experiment to test it. Wait for new data to come in.
3 Form your conclusion.
Big data changes step 2
Online, searchable databases provide instant answers, speeding up research.
Medical records
lab tests
Wearables
BIG DATA
Health apps
Genetics tests
Census

Fig. 1. Comparison between conventional and modern system of health care

In recent years, the Indian government has increased spending in the health care industry. The government plans to increase it even further by 2.5% of the GDP in the 12th five-year plan. As compared to other emerging economies the amount of public funding that India invests in health care is very small. India ranks among the last 5 countries with 6% of GDP expenditure on health care. Hospital bed density in India has been stagnating at 0.9 per 1000 population since 2005 and falls significantly short of WHO laid guidelines of 3.511 per 1000 patients' population. Moreover, there is a huge disproportion in utilization of facilities at the village, district and state levels with state level facilities remaining the most tensed. India is currently known to have approximately 600,000 doctors and 1.6 million nurses. This interprets into one doctor for every 1,800 people. The recommended WHO guidelines suggest that there should be 1 doctor for every 600 people. This translates into a resource gap of approximately 1.4 million doctors and 2.8 million nurses. There is also a clear disproportion in the man power present in the rural and urban areas [7, 8].

II. ROLE PLAYED BY BIG DATA IN MEDICAL SCIENCE

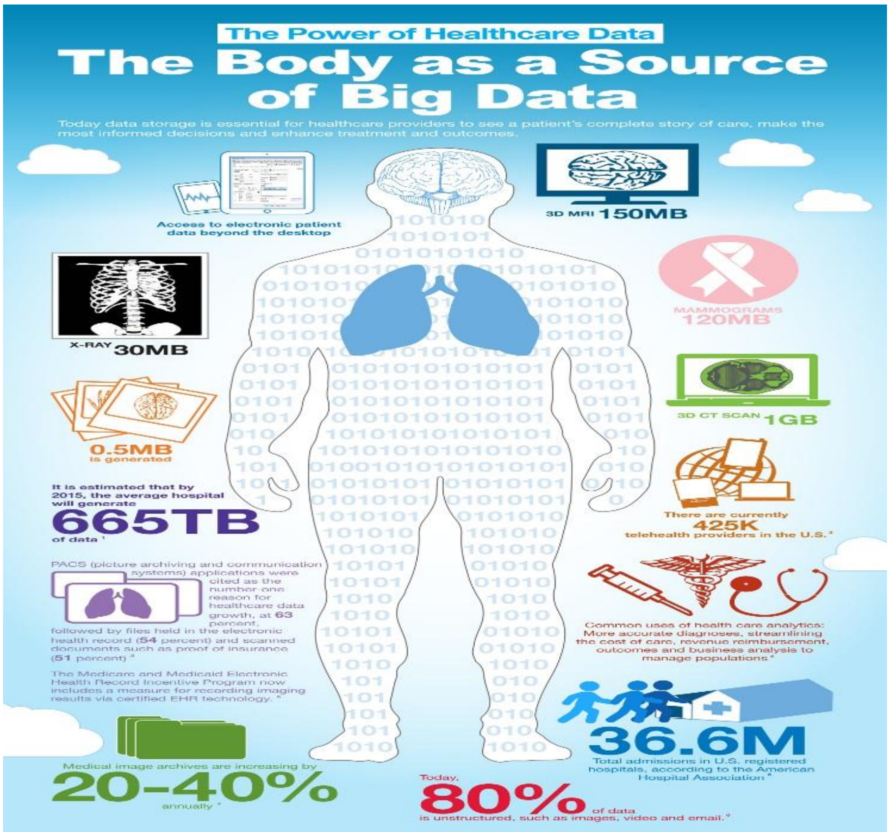

The Power of Healthcare Data
The Body as a Source of Big Data
Today data storage is essential for healthcare providers to see a patient's complete story of care, make the most informed decisions and enhance treatment and outcomes.
Access to electronic patient data beyond the desktop
3D MRI 150MB
X-RAY 30MB
MAMMOGRAMS 120MB
3D CT SCAN 1GB
0.5MB is generated
It is estimated that by 2015, the average hospital will generate 665TB of data [1]
There are currently 425K telehealth providers in the U.S. [3]
PACS (picture archiving and communication systems) applications were cited as the number one reason for healthcare data growth, at 63 percent, followed by files held in the electronic health record (54 percent) and scanned documents such as proof of insurance (51 percent) [2]
Common uses of health care analytics: More accurate diagnoses, streamlining the cost of care, revenue reimbursement, outcomes and business analysis to manage populations [4]
The Medicare and Medicaid Electronic Health Record Incentive Program now includes a measure for recording imaging results via certified EHR technology. [5]
36.6M
Total admissions in U.S. registered hospitals, according to the American Hospital Association [8]
Medical image archives are increasing by 20-40% annually [7]
Today, 80% of data is unstructured, such as images, video and email. [9]

Fig. 2. Human body as a source of Big Data

Human body is a key source of Big Data as shown in Fig. 2. Health care industry is now digitizing an adequate amount of information to readily take advantage of big data. Currently, more than 80 percent of hospitals and more than half of all doctors use electronic medical records to collect data and organize care. Hospitals, insurance companies, labs, providers are gathering information on patients, procedures, costs and medications and warehousing it in databases. Even though businesses are leading big data applications, the public sector has begun to be on the go, particularly in the search for effective uses of big data, with the aim of serving citizen's better and overcoming national challenges such as skyrocketing health care costs, job creation, natural disasters, terrorism, and other concerns. Researchers have often argued that it would be somewhat difficult to ensure that big data plays a central role in a health system's ability to secure improved health for its users. In particular, they are concerned that big data involves many new challenges regarding its complexity, security, and privacy risks, as well as the need for new technologies and human skills [5, 7, 8]. Given the implications, Big Data on health care is not a choice but a compulsion. At the level of individual patients, it will enable faster and more holistic decision-making, blending personal information with collective trends [9]. At the macro level, data aggregation across regions and geographies will deliver larger samples for more statistically accurate clinical studies. And it will enhance the overall quality of patient care while simultaneously reducing costs associated with under or over treatment [5, 11, 13].

II. IMPACT OF BIG DATA ON HEALTH CARE SYSTEM

The popularity of big data is transforming the discussion of what is appropriate or right for a patient and right for the healthcare environment. In keeping with these changes, we have created a holistic, patient-centered framework that considers five key pathways to value, based on the concept that value is derived from the balance of healthcare spend and patient impact [10, 14, 15].

A. RIGHT LIVING

Patients can build value by taking an active role in their own treatment, including disease prevention. The right-living pathway focuses on encouraging patients to make lifestyle choices that help them remain healthy, such as proper diet and exercise, and take an active role in their own care if they become sick.

B. CORRECT CARE

It involves ensuring that patients get the timely and appropriate treatment available. In addition to relying heavily on protocols, right care requires a coordinated approach: across settings and providers, all caregivers should have the same information and work toward the same goal to avoid duplication of effort and suboptimal strategies.

C. ACCURATE PROVIDER

It proposes that patients should always be treated by high-performing professionals that are best coordinated to the task and will achieve the best outcome. "Right provider" therefore has two meanings: the right match of provider skill set to the complexity of the assignment— for instance, nurses or physicians' assistants performing tasks that do not require a doctor—but also the specific selection of the provider with the best proven outcomes.

D. PRECISE VALUE

To fulfill the goals of precise value, providers will continuously enhance healthcare value while preserving or improving its quality. This pathway could involve multiple measures for ensuring cost-effectiveness of care, such as tying provider reimbursement to patient outcomes, or eliminating fraud, waste, or abuse in the system.

E. RIGHT INNOVATION

It involves the identification of new therapies and approaches to delivering care, across all aspects of the system, and improving the innovation engines themselves. For instance, by advancing medicine and boosting R&D productivity. To capture value in this pathway, stakeholders must make better use of prior trial data. They could also use the data to find opportunities to improve clinical trials and traditional treatment protocols, including those for births and inpatient surgeries [6, 12].

IV. MAP-REDUCE IN MANEUVER

Algorithm for Map-Reduce [19, 20] via pictorial representation

- The incoming data can be alienated into n number of modules which depends upon the amount of input data and processing power of the individual unit (Fig. 3).
- All these fragmented modules are then passed over to mapper function where these modules undergo simultaneous parallel processing (Fig. 3).

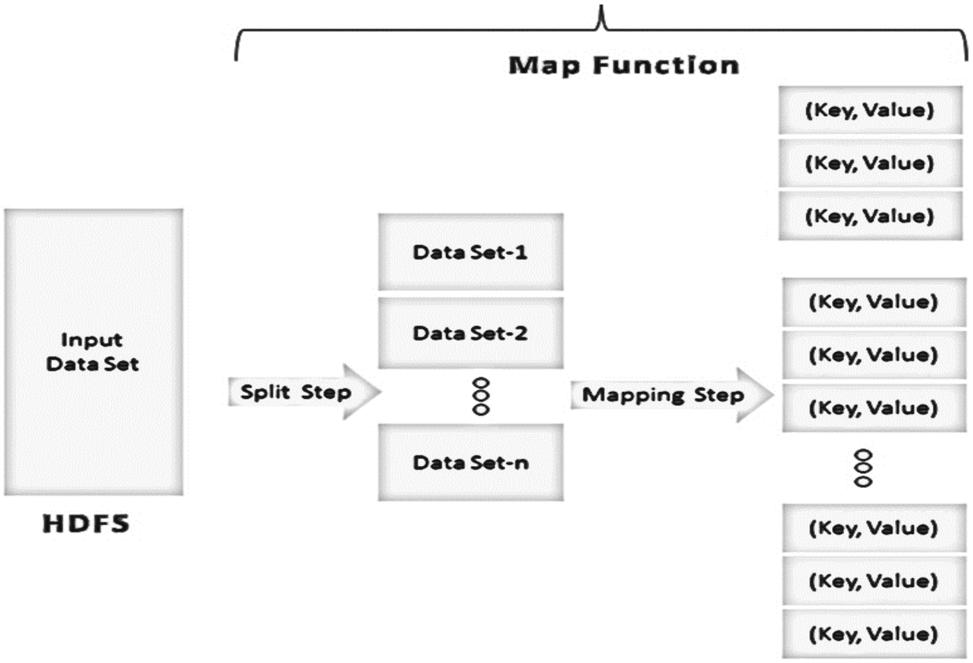


Fig. 3 MapReduce – Mapping Function

- Thereafter, shuffling is conducted in order to gather similar looking patterns (Fig. 4).

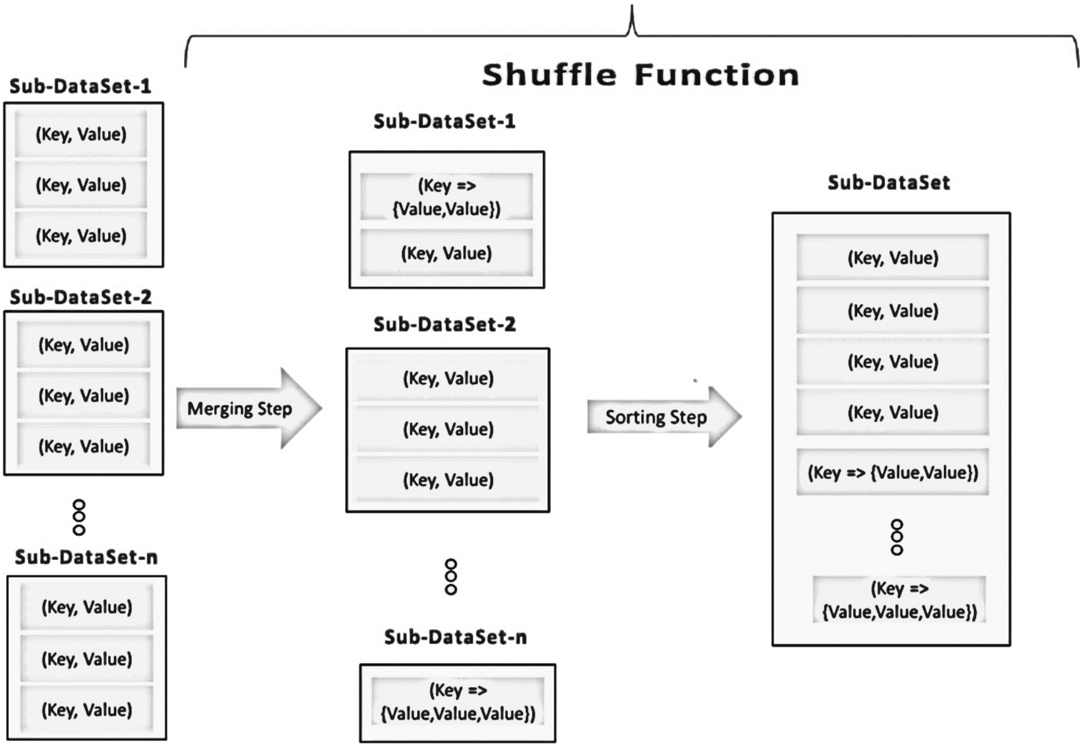


Fig. 4 MapReduce – Shuffle Function

- Finally, reducer function is called which is responsible for getting the ultimate output in a reduced form (Fig. 5).

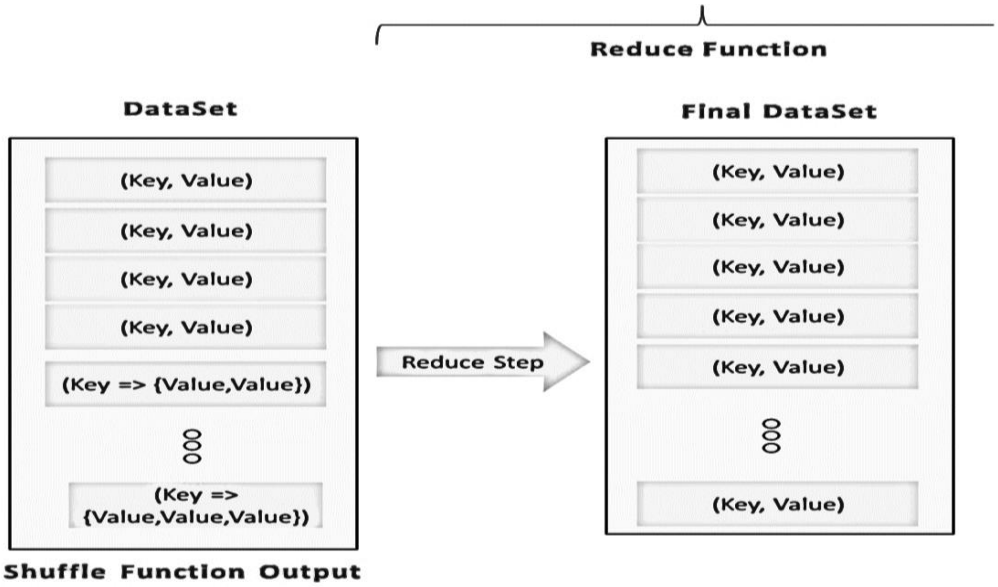


Fig. 5 MapReduce – Reduce Function

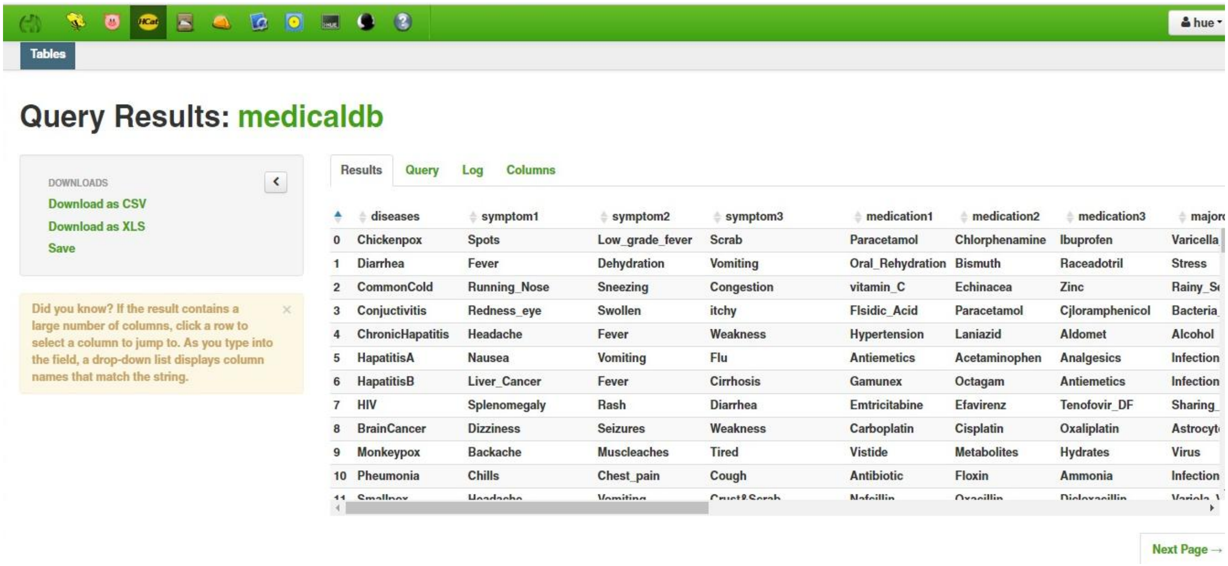
- Moreover, this technique is scalable and depending upon increase in the data to be processed, the processing units can be further extended.

V. IMPLEMENTATION

A self-constructed database has been created considering the most common diseases which prevails in society.

The tool used for uploading and mining database is Hortonworks [9].

A table is created using HCatalog (in this case table name is “medicaldb”) as shown in Fig. 6 [16, 17].



Did you know? If the result contains a large number of columns, click a row to select a column to jump to. As you type into the field, a drop-down list displays column names that match the string.

	diseases	symptom1	symptom2	symptom3	medication1	medication2	medication3	majoi
0	Chickenpox	Spots	Low_grade_fever	Scrab	Paracetamol	Chlorphenamine	Ibuprofen	Varicella
1	Diarrhea	Fever	Dehydration	Vomiting	Oral_Rehydration	Bismuth	Raceadotril	Stress
2	CommonCold	Running_Nose	Sneezing	Congestion	vitamin_C	Echinacea	Zinc	Rainy_S
3	Conjuctivitis	Redness_eye	Swollen	itchy	Flsidic_Acid	Paracetamol	Cjloramphenicol	Bacteria_
4	ChronicHapatitis	Headache	Fever	Weakness	Hypertension	Laniazid	Aldomet	Alcohol
5	HapatitisA	Nausea	Vomiting	Flu	Antiemetics	Acetaminophen	Analgesics	Infection
6	HapatitisB	Liver_Cancer	Fever	Cirrhosis	Gamunex	Octagam	Antiemetics	Infection
7	HIV	Splenomegaly	Rash	Diarrhea	Emtricitabine	Efavirenz	Tenofovir_DF	Sharing_
8	BrainCancer	Dizziness	Seizures	Weakness	Carboplatin	Cisplatin	Oxaliplatin	Astrocyt
9	Monkeypox	Backache	Muscleaches	Tired	Vistide	Metabolites	Hydrates	Virus
10	Pheumonia	Chills	Chest_pain	Cough	Antibiotic	Floxin	Ammonia	Infection
11	Smallpox	Headache	Vomiting	Crust&Scrab	Nafcillin	Oxacillin	Dicloxacillin	Variola_V

Fig. 6 Creating Table using HCatalog

Appropriate Hive Script is written to get desired results as shown in Fig. 7.

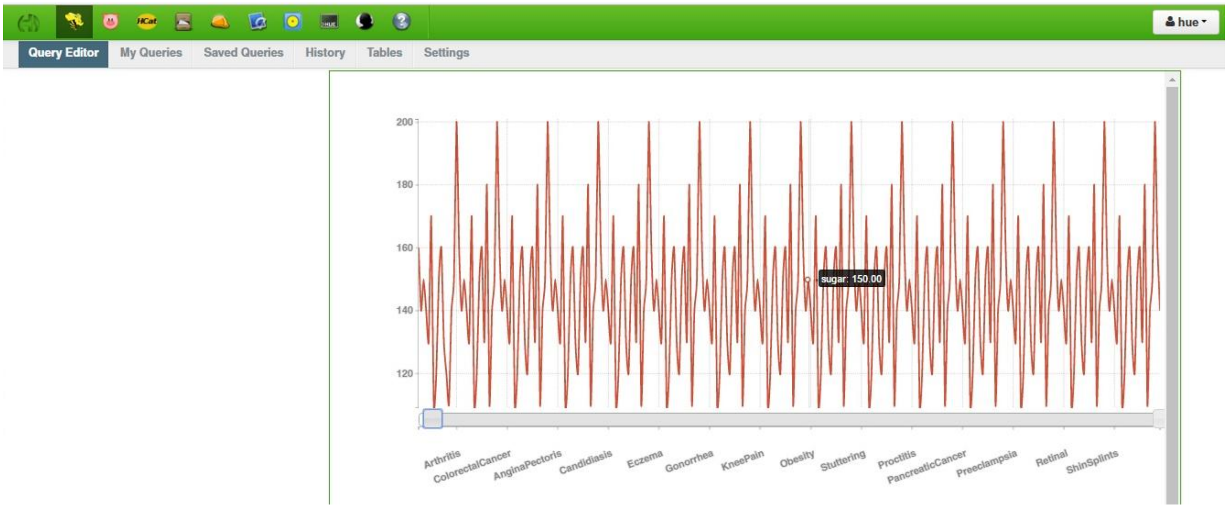


Fig. 7 Results displayed in visualization

VI. CONCLUSION

Today, a significant proportion of the cost and time spent in the drug development process is attributable to unsuccessful formulations. By enabling researchers to identify compounds with a higher likelihood of success, Big Data can help reduce the cost and the time to market for new drugs. Also, by integrating learning from medical data in the early stages of development, researchers will now be able to customize drugs to suit aggregated patient profiles.

Currently, information privacy concerns are the single biggest obstacle to Big Data adoption in health care. Another is the absence of an analytics solution powerful enough to gather massive volumes of largely unstructured health data, perform complex analyses quickly, and trigger meaningful solution, for instance, gather all the data from ICU monitors, which today goes un-stored, put it on the Cloud, decipher significant medical patterns that are yet undiscovered, and trigger a medical action instead of merely an alarm.

REFERENCES

[1] http://radar.oreilly.com/2012/01/what-is-big-data.html

[2] www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/

[3] http://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/the-role-of-big-data-in-medicine

[4] http://dashburst.com/infographic/big-data-volume-variety-velocity/

[5] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3848064/

[6] http://www.informationweek.in/informationweek/cio-blog/175124/healthcare-compulsion-choice

[7] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3717441/

[8] http://www.datanami.com/2013/07/19/big_data_emerges_in_indian_health_care/

[9] http://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

[10] http://www.livemint.com/Home-Page/7evdkEgC07i96HzyDF2OTJ /Long-way-to-go-for-benefits-from-Big-Data-in-health care-ind.html

[11] http://eandt.theiet.org/magazine/2013/03/journey-to-the-centre-of-big-data.cfm

[12] http://itbussinessconsulting.jimdo.com/2014/08/11/how-big-data-and-analytics-can-reshape-healthcare/

[13] http://www.hissjournal.com/content/2/1/3

[14] Dr.Gagandeep Jagdev, Sukhpreet Singh, “Applications of Big Data in Medical Science brings revolution in managing health care of humans”, IJEEE, Vol. 2, Spl. Issue 1(2015), e-ISSN:1694-2310.

[15] Information Week. 2012. “Big Data Widens Analytic Talent Gap.” Information Week April.

[16] R. C. Taylor, “An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics,” BMC Bioinformatics, vol. 11, no. 12, article S1, 2010. View at Publisher · View at Google Scholar · View at Scopus

[17] M. Krzywinski, I. Birol, S. J. Jones, and M. A. Marra, “Hive plots-rational approach to visualizing networks,” Briefings in Bioinformatics, vol. 13, no. 5, pp. 627–644, 2012. View at Publisher · View at Google Scholar · View at Scopus.

[18] Heudecker, Nick. 2013. “Hype Cycle for Big Data.” Gartner G00252431

[19] Edala, Seshu. 2012. “Big Data Analytics: Not Just for Big Business Anymore.” Forbes.

[20] Dean, Jeffery, and Ghemawat Sanjay. 2004. “MapReduce: Simplified Data Processing on Large Clusters.” Google.

[21] Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big Data: Issues and Challenges Moving Forward. International Conference on System Sciences (pp. 995-1004). Hawaii:IEEE Computer Soceity.

[22] Katal, A., Wazid, M., & Goudar, R. H. (2013). Big Data: Issues, Challenges, Tools and Good Practices. IEEE, 404-409.

About the author

Dr. Gagandeep Jagdev is a faculty member in Dept. of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib (PB). His total teaching experience is above 10 years and has above 90 international and national publications in reputed journals and conferences to his credit. He is also a member of editorial board of four international peer reviewed journals. His field of expertise is Big Data, ANN, Biometrics, RFID, Cloud Computing and VANETS.