ISSN (ONLINE): 2395-695X
ISSN (PRINT): 2395-695X

International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Volume 2, Special Issue 19, October 2016

BigDatainHealthcareApplications

Krishnakumar V
DepartmentofSoftware Technology
SACRED HEART COLLEGE (AUTONOMOUS)
Tirupattur, Vellore DT, Tamil Nadu, India
kichuveera@gmail.com

Abstract—**Nowadays "Bigdata"isquite common term tostore and retrieve the collection ofgiganticdatasets. In a school/college, every year we interact with all the students and doing complete medical check-up individually. After thatherewe implement student digital healthreport (SDHR)whichishighly capableofstoringcapaciousdataindatabaseand thisincludes student's previous medical data details,laboratorytestreport,presenttreatmentdetails giventostudent's,doctor's prescription/advice, diet details,diagnosticrecord's,pharmacy/medical shop information,healthinsurancerelateddata,medicaljournalsareused to properinvestigateand analysis.Objective of this paper is,todiscusstheimportance and characteristicsbig data with its key challengesof Healthcare domain.**

Keywords—**Big DataAnalysis(BDA), Student Digital HealthReport (SDHR),Data Node(D-node), Name Node(N-node)Hadoop,MapReduce.**

I. INTRODUCTION

'Thinking big' means being open-minded, positive, creative and seeing opportunity in the big picture. The digital revolution is re-shaping the way the people live, work and interact. As a digital provider and an acknowledged sustainability leader, people have an opportunity to contribute to the future by through our ability to help with change. People convinced of technology's ability to support the future by offering new solutions and play a part in tackling issues for society, improving social mobility, keeping people safe online, protecting the environment and combating climate change. Each and every day, people are produce more than 2.5 quintillion bytes of data, so much that 90% of the data in the world today has been produced in the last two years only. This data comes from all over the place, some devices used to gather environment/atmosphere information, posts to social media sites, health report,digital pictures and videos, purchase transaction records, and mobile phone GPS signals to name a few. Therisingcostofhealthcareisoneofthe world's mostimportantproblem.Thesedatabasesare designedfor maintaining individual clinical data/report.

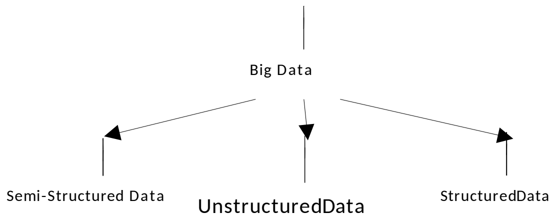
A. Characteristics of Big Data

TableI- Different types of V's in Big Data

Value	Clinically relevant data and Longitudinal Studies
Volume	High-throughput technologies Continuous monitoring of vital signs
Velocity	High-speed processing for fast clinical decision support Increasing data generation rate by the health infrastructure
Variety	Heterogeneous and unstructured data sources Differences in frequencies and taxonomies
Veracity	Data quality is unreliable Data coming from uncontrolled environments
Variability	Seasonal health effects and disease evolution Non-deterministic models of illness and health
Visualization	Graphical representation of complex data.

B. Different format of Big Data anditsSources:

BigDataand itssources canbe categorized into followingcategories as shown figure 1.



II.BIGDATALIFECYCLE

A.DataCollection -ItinvolvesthecollectionofdatafromvarioussourcesandstoringitinHDFS.Data canbeanythingsuchascasehistory,medicalimages,sociallogs,sensordataetc.

B.DataCleaning -Itinvolvestheprocessofverifyingwhetherthereisany junkdata orany datathathas missedvalues.Suchdataneedstoberemoved.

C.DataClassification -Itinvolvesthefilteringofdatabasedontheirstructure.ForexampleMedical Big dataconsistsofmostlyunstructureddatasuchas handwrittenphysiciannotes.Structured,semi-structuredand unstructureddatashouldbeclassifiedin ordertoperformmeaningfulanalysis.

D.DataModelling -Itinvolvesperforminganalysisontheclassifieddata.ForexampleG overnment may requirethelistofmalnourishedchildreninaparticularlocation.Firstit hastoclassifythedatabasedonthe specificlocation,needtotriggerthehealthreportof children,needtoidentify thechildrenwhosefamily areunder povertylineandthesedatashouldbeprocessed.

E.DataDelivery -Itinvolvesthegenerationof reportbasedonthedatamodellingdone.Basedonthe exampleafterthedatais processeditwillgeneratea reportbasedonmalnourishedchildrenin aparticularlocation. Thiswillhelpthegovernmenttotakenecessarymeasurestoavoidany furthercomplications.AttheallthestagesofBDLC(BigDataLifecycle)itrequiresdatastorage,dataintegrityanddataaccesscontrol.

The types ofdata anticipated tobeofuse in BDA include

Previous Medical data details - upto80 per centof health datais unstructuredasdocuments,images,clinical or prescribednotes.

Laboratorytestreport - Scanning report, bloodtest, sugar statement, salt report, urine testto be collected andmaintain the laboratory based health information.

Presenttreatment detail.

Doctor's prescription/advice and Diet details -Text-based practice guidelinesandhealthproduct (e.g., drug information) data.

Diagnosticrecord.
Pharmacy/medical shop information.
Healthinsurancerelateddata.

Medicaljournals -Clinical research and medical referencematerial.

Genomicdata - Representssignificant amountsof new gene sequencingdata.

Streamed data - Homemonitoring, Tele-health,handheld and sensor- based wireless or smart devicesarenewdata sourcesand types.

Webandsocialnetworkingdata - Consumeruse of internetdatafrom searchenginesandsocial networking sites.

Business,organizationalandexternal data- Administrativedata suchas billing, Schedulingand other non-health data.

III.BIGDATAINHEALTHCARE

A.BusinessGoals and Objective Addressedby Analytics
Improveclinicaleffectivenessand member/patient

Satisfaction -

Improveclinicalqualityofcare
Improve patient safely and reduce medical errors
Improvewellness,preventionanddisease Management
Understandphysicianprofilesand clinical performance
Improvecustomersatisfaction,acquisitionand retention

Improveoperational effectiveness -

Reducecostsand increaseefficiency
Optimize catchmentareaandnetworkmanagement
Improvepayforperformanceandaccountability

Increaseoperatingspeedandadaptability
Improve financial and administrativeperformance -

IncreaserevenueandROI
Improveutilization
Optimize supplychain andhuman capital management
Improveriskmanagement andregulatory compliance
Reducefraudandabuse

IV. ImpactofBig Data inHealthcare

A.Big data can change healthcare

After 20 years of steady increases, health-care expenses now represent 17.6 percent of GDP nearly $600 billion more than the expected benchmark for a nation of the United States' size and wealth .The report outlines five ways data will enable the healthcare industry to cut costs and improve quality.

Right living - Data can help patients to take an active role in their own health such as diet, exercise, and medication adherence to take control of their health.

Right care - Data can improve outcomes, reducing medical errors. Application of big data tools will facilitate evidence-based care that is personalized to the specific patient.

Right provider - Proven outcomes for patients to receive the best medical care based on data that helps us better match the provider's skill set with the needs of the patient and allow assessment of specific providers.

Right value - Cost-effective healthcare through different methods, such as patient-outcome reimbursement and eliminating fraud, waste, and abuse in the system utilizing big data.

Right innovation - Innovators will be able to address all aspects of therapeutic innovation discovery, development, and safety utilizing data from past trials as well as analyzing trends from current data. Healthcare providers can analyze patient history data, real-time data from monitors, clinical factors, lifestyle choices and social determinants to provide a holistic view of the patient and develop the most effective care plans. IBM has helped healthcare providers:

Identify crises before they happen and treat patients proactively by analyzing data in real time as it streams from monitoring equipment.

Predict patient health risks using predictive analytics to understand underlying clinical or social factors, and design more effective care plans.Improve healthcare outcomes by providing timely and meaningful insights to care providers, who can then administer the most effective treatments.

V. HDFSARCHITECTURE

Hadoopeffectively handlesthelargedataset.Thebelowfigurerepresentshowaclientcontacts namenodeforprocessingthedata.Namenodecommunicates toJobTrackerandassignthetaskgivenbytheclient foregtofindoutthelistofpatientswhoareintheriskofgettingdiabetes.Mapreduceprogram performsthe analysisonthedataandreturnstheresultstojobtracker.Italsoreturnstheblockwheretheclientcanstoreits data.HiveQLisusedto performthedata-warehousingtaskanditcanalsobecombinedwithmap-reduceprogram.PIG providers the platform for analyzing large data sets through parallel computations.

The daemonsinhdfsare following such as,

Namenode-Itisthemasternodewhichreceivestherequestfromtheclient(examplepatientmonitoring system).ItlooksuptheMetadatatofindoutwhich isthesuitable datanodeforstoring thedatarelatedto theclient.It selectsdatanodebasedonthelocalityandavailablefreeslots.

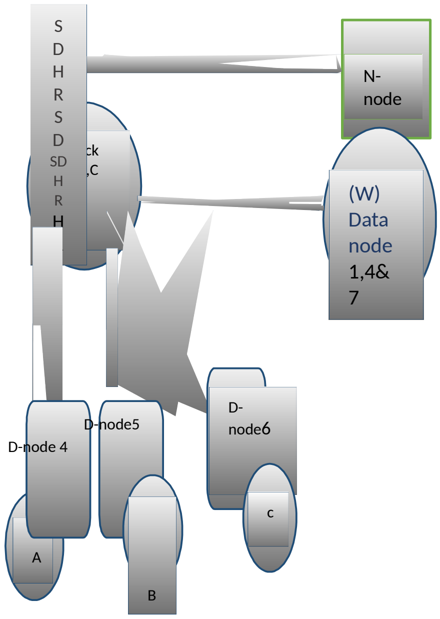


FIG.4.HDFSFileSystemArchitecture

SecondaryNamenode -Itisthebackupnodeforthenamenode.Itstoresthefsimagefilewhichcontains thedetailsaboutthedatanode.Fsimagehas to be restoredfromthesecondarynamenodewhennamenode fails.

JobTracker -Mapreduceprogramrunninginjobtrackerassignsjobtothedatanodeandtasktracker.Datanodestorestheactualdataanditperiodicallysendsheartbeattothenamenodeaboutthedatastored. Tasktrackerperformsthetaskassignedby jobtracker.

VI.PLATFORMS&TOOLSFOR BIGDATA ANALYTICS IN HEALTHCARE

A.Platform/ToolDescription

The*HadoopDistributedFileSystem (HDFS)*-HDFSenablestheunderlyingstoragefortheHadoopcluster. Itdividesthedataintosmallerpartsanddistributesitacrossthe variousservers/nodes.

MapReduce-MapReduceprovides theinterfaceforthedistributionofsub-tasks andthegathering ofoutputs. Whentasksareexecuted,MapReducetrackstheprocessingofeachserver/node.

PIGandPIGLatin(PigandPigLatin) : Pigprogramminglanguageisconfigured toassimilate alltypesofdata (structured/unstructured,etc.).Itiscomprised oftwokeymodules: thelanguageitself,calledPigLatin,andthe runtimeversioninwhichthePigLatincodeisexecuted.

Hive -HiveisaruntimeHadoopsupportarchitecturethatleveragesStructureQueryLanguage (SQL)with the Hadoopplatform.ItpermitsSQLprogrammers todevelopHiveQueryLanguage(HQL)statementsakinto typicalSQLstatements.

Jaql-Jaqlisafunctional,declarativequerylanguagedesignedtoprocesslargedatasets.Tofacilitateparallelprocessing,Jaqlconverts"„high-level"queriesinto„low-level"queries"consistingofMapReducetasks.

Zookeeper-Zookeeperallowsacentralizedinfrastructurewithvariousservices,providing synchronization acrossacluster ofservers.Bigdataanalyticsapplicationsutilizetheseservicestocoordinateparallelprocessing acrossbigclusters.

HBase-HBaseisa column-orienteddatabase managementsystemthatsitsontopofHDFS.It usesa non-SQLapproach.

Cassandra -Cassandra isalsoadistributed databasesystem.Itisdesignated asatop-level projectmodeledto handlebigdatadistributedacrossmanyutilityservers.Italsoprovidesreliableservicewithnoparticularpoint offailure(http://en.wikipedia.org/wiki/Apache_Cassandra)anditisa NoSQLsystem.

Oozie-Oozie,anopensourceproject,streamlinestheworkflowandcoordinationamongthe tasks.

Lucene-TheLuceneprojectis used widelyfortextanalytics/searchesandhasbeen incorporatedintoseveralopensourceprojects.Itsscopeincludesfulltextindexing andlibrarysearchfor use withina Javaapplication.

Avro-Avrofacilitatesdataserializationservices.Versioningandversioncontrolare additionalusefulfeatures.

Mahout-MahoutisyetanotherApacheprojectwhosegoalisto generatefreeapplicationsofdistributedandscalablemachinelearningalgorithmsthat supportbigdataanalyticsontheHadoopplatform.

VII. CHALLENGESINHEALTHCAREAPPLICATION INBIG DATA

Leveraging the patient/data correlations in longitudinal records.Understandingunstructuredclinicalnotesinthe right context.Efficiently handling large volumes of medical imagingdataandextracting potentially useful informationand biomarkers.Analyzing genomic data is a computationally intensive task and combining with standard clinical data adds additional layers of complexity.Capturing the patient's behavioral datathrough several sensors; their various social interactions and communications.

A.Processing Challenges

DataCollection,

Resolvingsimilarities,

ModificationOfdata,

Data Analysis,

outputrepresentation

B.ManagementChallenges

DataPrivacy,

DataSecurity,

Governance andethicalissues

VII. FUTURE DEMANDS OFANALYTICS

- *Focus on the biggest and highest value opportunities*
- *Within each opportunity,start with questions not data*
- *Embed insights to drive actions and deliver value*
- *Keep existing capabilities while adding new ones*
- *Use an information agenda to plan*

VIII. CONCLUSION

Lack ofstudent's medical recordsisthe significant challengethat government,doctorsandresearchers undergo from, which almostmotivates healthcaredomaintobuildstudent'sto helpstakeholdersintheirbusiness.Student's digital healthreportsystemscouldhelpto improvethecommunicationbetweengovernment,doctorsand student's ontheother handtoimprovethequalityofcarewhichmayleadtoreducemedicalerrorsandcosts.Inordertogetbetterreviewso fstudent'sin interestwithshare,save,manage,andretrievetheirmedicaldata,suchastheirmedicalhistory,medications,allergies,x-raysandtest results. Accordingly building these student digital healthreportis a big repositories give them an opportunity to interact with doctors, physicians and pharmacists,but IT expertsshouldtakeinmindstudentsprivacyandpolicesrisk.Inthis proposedresearchwork,theobjectiveistodevelop thesystem for student's digital healthreport,whichtheresult shouldbedepends ondataweinput ifthedataiscorrectthenitshouldgive consistentresult. Through the final result we can able to predict or identify the complete health problem and cure quick manner without difficulties.Student's digital healthreport research work will help us to create a healthy body and healthy mind among the students.

REFERENCES

[1].DemboskyA:"DataPrescriptionforBetterHealthcare."Financial Times,December12,2012,

[2]. FeldmanB,MartinEM,SkotnesT:"BigDatainHealthcareHype andHope."October2012.Dr.Bonnie.

[3]. FernandesL,O'ConnorM,WeaverV:Bigdata,biggeroutcomes.J AHIMA2012,38-42.

[4]. Transforming Health Care through Big Data Strategies for leveragingbigdatainthehealthcareindustry.2013.

[5]. WullianallurRaghupathiandVijuRaghupathiBigdataanalyticsin healthcare:promiseandpotential

[6].Korsten,PeterandChristian Seider."The world's4trillion dollar challenge.Usingasystem-of-systemsapproachto buildasmarter planet."IBMInstituteforBusinessValue. January2010. http://www-935.ibm.com/services/us/gbs/bus/html/ibv-smarter-planet-system-of-systems.htm

[7]Manning,Harley."Hotoffthepress:Forrester'sCustomer ExperienceIndex,2011."January11,2011.ForresterBlogs.

[8].Adams,Jim.PaulGrundy, MD,MartinS.Kohn,MDand EdgarL. Mounib."Patient-centeredmedicalhome:What, whyandhow?" IBMInstitute forBusinessValue.May2009.

[9].Ibid.

[10]YanglinRen,MonitoringpatientsviaaSecureandmobile healthcaresystem,IEEESymposiumonwireless communication,2011

[11]DaiYuefa,WuBo,GuYaqiang,DataSecurityModelforCloud Computing,InternationalWorkshoponInformationSecurityand Application,2009.

[12]Jeffrey Dean and SanjayGhemawat,MapReduceSimplified Data Processing on LargeClusters,ACM,2008

[13]CongWang,Privacy-PreservingPublicAuditingforSecureCloud Storage,IEEE,2010

[14]Konstantin Shvachko,HairongKuang, Sanjay Radia, Robert Chansler,TheHadoopDistributed FileSystem,IEEE,2010.

[15]BillHamilton,BigDataIstheFutureofHealthcare,Cognizantwhite paper,2010.

[16]WhitePaper bySAS,HowGovernmentareusingthePowerofHigh PerformanceAnalytics,2013.