

VELEUČILIŠTE U ŠIBENIKU

ODJEL MENADŽMENT

STRUČNI STUDIJ INFORMATIČKI MENADŽMENT

Stipe Batinica

PRIMJENA BIG DATA TEHNOLOGIJE U

ZDRAVSTVU

Završni rad

Šibenik, 2018.

VELEUČILIŠTE U ŠIBENIKU
ODJEL MENADŽMENT
STRUČNIJ STUDIJ INFORMATIČKI MENADŽMENT

PRIMJENA BIG DATA TEHNOLOGIJE U
ZDRAVSTVU

Završni rad

Kolegij: Elektroničko poslovanje

Mentor: mr.sc. Danijel Mileta, dipl.ing., v.pred

Student: Stipe Batinica

Matični broj studenta: 1219054131

Šibenik, lipanj 2018.

SADRŽAJ

1. UVOD.....	1
2. OPĆENITO O VELIKIM PODACIMA.....	2
3. TEHNOLOGIJE VELIKIH PODATAKA.....	6
3.1. HADOOP TEHNOLOGIJA.....	6
3.2. KAKO KORISTITI HADOOP TEHNOLOGIJU.....	8
3.3. U KOJE SE SVRHE KORISTI HADOOP?.....	8
3.4. KAKO HADOOP FUNKCIONIRA?.....	9
3.5. MAP REDUCE TEHNOLOGIJA.....	11
3.6. RAZUMIJEVANJE MAP REDUCE TEHNOLOGIJE.....	12
3.7. BIGTABLE TEHNOLOGIJA.....	15
4. POTREBA ZA BIG DATA TEHNOLOGIJOM U ZDRAVSTVU.....	16
5. RAZLIKA AMERIČKOG I EUROPSKOG ZDRAVSTVENOG SUSTAVA S ASPEKTA BIG DATA-E.....	17
6. PRIMJENA BIG DATA TEHNOLOGIJE U ZDRAVSTVU.....	18
7. ARHITEKTONSKI OKVIR BIG DATA-E U ZDRAVSTVU.....	21
8. BIG DATA TEHNOLOGIJA I 5V U ZDRAVSTVENOM SUSTAVU.....	24
8.1. VOLUMEN.....	24
8.2. VARIETY (RAZNOLIKOST).....	24
8.3. VELOCITY (BRZINA).....	24
8.4. VERACITY (ISTINITOST).....	25
8.5. VALUE (VRIJEDNOST).....	25
9. OGRANIČENJA I IZAZOVI PODATKOVNE ANALIZE U ZDRAVSTVENOM SUSTAVU.....	27
10. ZAKLJUČAK.....	28
11. LITERATURA I IZVORI.....	29

PRIMJENA BIG DATA TEHNOLOGIJE U ZDRAVSTVU

STIPE BATINICA

stipebatinica@icloud.com

U radu se prezentira utjecaj tehnologije velikih podataka na primjeru zdravstvenog sustava. Detaljnije se objašnjava sam koncept tehnologije kako bi se kasnije lakše shvatila primjena iste. Osim teoretskog dijela, u radu su prikazani i konkretni primjeri zdravstvenih aplikacija, te se detaljnije, s tehničkog aspekta objašnjava na koji način prezentirana tehnologija funkcionira. Prvotni rezultat rada bio bi prikazati na koji način suvremeni sustavi mogu unaprijediti svoje poslovanje koristeći aktualne tehnologije velikih podataka poput onih navedenih u radu (Map Reduce, Hadoop, Bigtable,...), a finalni da s pomoću tih istih tehnologija maksimiziraju profit i minimaliziraju troškove, uz što veći naglasak na kvalitetu usluge i što bolju skrb prema krajnjem korisniku, pacijentu.

(28 stranica / 7 slika / 11 literaturnih navoda / jezik izvornika: hrvatski)

Rad je pohranjen u: Knjižnici Veleučilišta u Šibeniku

Ključne riječi: Tehnologija velikih podataka, zdravstveni sustav, implementacija, arhitektura sustava, Map Reduce, Hadoop

Mentor: mr.sc.Danijel Mileta, dipl.ing., v.pred

Rad je prihvaćen za obranu:

Polytechnic of Šibenik

Final Paper

Department of Management

Professional Undergraduate Studies of Informatics Management

IMPLEMENTATION OF BIGDATA TECHNOLOGIES IN HEALTHCARE SYSTEMS

STIPE BATINICA

stipebatinica@icloud.com

The thesis presents the impact of Big data technologies in the healthcare systems. In addition to the theoretical part, the thesis presents concrete examples of healthcare applications, and detailed explains the technical aspect of technology itself. The first reason of the thesis is to “put accent“ on which way can modern systems improve their business by using current Bigdata technologies such as Map Reduce, Hadoop, Bigtable, etc, and the final one is usage of the same technologies to maximize system’s profit and minimize system’s costs, with a bigger emphasis on the quality of service, and to provide as best as possible care for the end user, the patient.

(28 pages / 7figures / 11 references / original in Croatian language)

Paper deposited in: Library of Polytechnic in Šibenik

Keywords: Bigdata technologies, healthcare system, implementation, system architecture, Map Reduce, Hadoop

Supervisor:mr.sc.Danijel Mileta, dipl.ing., v.pred

Paper accepted:

1. UVOD

Povijesno gledano, evidentno je da zdravstvena industrija generira velike količine podataka vođenih evidencijom i usklađenošću regulatornim zahtjevima i pacijentima. Dok je većina podataka pohranjena u obliku tiskanog materijala, trenutni trend je usmjeren na brzu digitalizaciju tih velikih količina podataka. Potaknute obveznim zahtjevima i potencijalom za poboljšanje kvalitete isporuke zdravstvene zaštite, u međuvremenu vodeći računa o smanjenju troškova, ove ogromne količine podataka (poznate pod nazivom "veliki podaci") "obećavaju" podršku širokom rasponu medicinskih i zdravstvenih funkcija, uključujući podršku kliničkim odlukama, nadzoru bolesti i upravljanju zdravljem populacije. Izvješća kažu da su podaci iz američkog zdravstvenog sustava sami dosegli 150 exabyte-a. Ovim rastom, veliki podaci za američku zdravstvenu skrb uskoro će dosegnuti ljestvicu zettabyte-a (1021 gigabajta) i nedugo nakon toga yotabyte-a (1024 gigabajta).¹

U prvom dijelu rada detaljnije se objašnjava Big data tehnologija kako bih se kasnije lakše shvatila njena primjenu u zdravstvenom sustavu. U ostatku rada se govori o implementaciji Big data tehnologije u zdravstvenom sustavu. Uspoređuju se svjetski zdravstveni sustavi s aspekta primjene tehnologije, te se navode primjeri iz prakse i pobliže sagledava sam koncept tehnologije.

¹Wullianallur Raghupathi, Viju Raghupathi: Big data analytics in healthcare: promise and potential, Health Information Science and Systems 2014, 2:3, str.1

2. OPĆENITO O VELIKIM PODACIMA

Budući da se informacijska tehnologija inovira na način na koji živimo, naša je zbirka digitalnih podataka počela brzo rasti. Danas postoji ogromna količina podataka koji se svakodnevno generiraju u sektorima proizvodnje, poslovanja, znanosti i našeg osobnog života. Pravilna obrada podataka mogla bi otkriti nova znanja o našem tržištu, društvu i okolišu i omogućiti nam pravovremeno reagiranje na nove mogućnosti i promjene. Međutim, rast obujma podataka u digitalnom svijetu je unaprijedio naše računalne infrastrukture. Konvencionalne tehnologije obrade podataka, poput baza podataka i skladišta podataka, postaju neadekvatne za količinu podataka s kojima se bavimo.

Ovaj novi izazov naziva se velikim podacima (Big data). Zbog svoje važnosti i uobičajenosti, posljednjih su godina stekli ogromnu pozornost. Nije bilo uobičajeno prihvaćene definicije o velikim podacima, iako ljudi obično vjeruju da veliki podaci trebaju uključivati skupove podataka s veličinama izvan sposobnosti uobičajenih softverskih alata za spremanje, upravljanje i obradu podataka u prihvatljivom proteklom vremenu. Na temelju tog koncepta, istraživači su saželi tri važna aspekta velikih podataka koji nadilaze sposobnost naše “tekuće” tehnologije obrade podataka. Oni su volumen, brzina (velocity) i raznolikost (variety), također poznat kao 3V.²

Izazov koji predstavlja volumen podataka je najznačajniji. U znanosti, kao što su biologija, meteorologija, astronomija itd., znanstvenici se stalno susreću s računalnim ograničenjima zbog povećanja volumena podataka. Unatoč različitim problemima velikih volumena, još uvijek nema sporazuma o kvantificiranju velikih podataka. Takva kvantifikacija ovisi o različitim čimbenicima. Prvo, složenost strukture podataka važan je čimbenik. Relativni skup podataka od nekoliko petabayta ne može se nazvati velikim podacima, budući da se njima lako može upravljati pomoću DBMS-a (Data Base Management sustava).³

Izazov brzine dolazi s potrebom za rješavanjem brzine kojom se stvaraju novi podaci, ili se ažuriraju postojeći. Ovo se pitanje posebno odnosi na podatke generirane na stroju, poput onih koje generiraju senzori ili mobilni uređaji. U tim aplikacijama, velika količina

²Jinchuan Chen, Yueguo Chen, Xiaoyong Du, Cuiping Li, Jiaheng Lu: Big data Challenge: a data management perspective, Key Laboratory of Data Engineering and Knowledge Engineering, School of Information, Renmin University of China, Beijing 100872, February 22, 2013, Front. Comput. Sci., 2013, 7(2): str. 157

³ Ibid., str 157

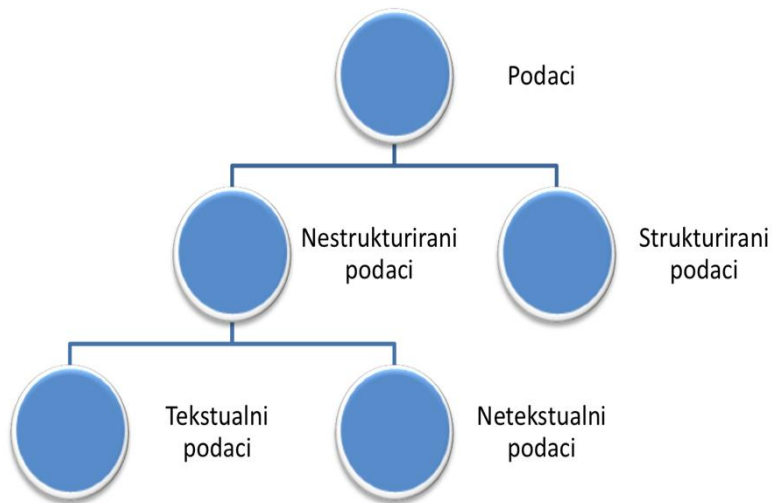
novih iažuriranih podataka neprestano odlazi u sustav, dok mi zahtijevamo od tog istog sustava smisao tih podataka odmah nakon što su stvoreni. Brzina podataka donosi izazove svakom sloju platforme za upravljanje podacima. I sloj za pohranu i sloj za obradu moraju biti iznimno brzi i skalabilni. Tehnologija prijenosa podataka nekoliko je godina istraživana za obradu velike brzine, međutim, kapacitet postojećih strujnih sustava i dalje je ograničen, pogotovo kada se bave povećim količinama ulaznih podataka.⁴

U aplikacijama u stvarnom svijetu podaci često ne dolaze iz jednog izvora. Velike implementacije podataka zahtijevaju rukovanje podacima iz različitih izvora, u kojima podaci mogu biti različitih formata i modela. To izaziva pojavu pojma količina podataka. Raznolikost podataka pruža više informacija za rješavanje problema ili za bolju uslugu. Pitanje je kako “uhvatiti” različite vrste podataka na način koji omogućuje povezivanje njihovih značenja. Tipično, podaci se mogu klasificirati u tri opće vrste podataka: strukturirane, polustrukturirane i nestrukturirane podatke. Bilo je sofisticirane tehnologije za rješavanje svake od tih vrsta podataka, kao što su one u bazi podataka ipronalaženje informacija. Ipak, besprijeckorna integracija ovih tehnologija i dalje ostaje izazov.⁵

⁴Jinchuan Chen, Yueguo Chen, Xiaoyong Du, Cuiping Li, Jiaheng Lu: Big data Challenge: a data management perspective, Key Laboratory of Data Engineering and Knowledge Engineering, School of Information, Renmin University of China, Beijing 100872, February 22, 2013, Front. Comput. Sci., 2013, 7(2): str. 158

⁵ Ibid., str.158

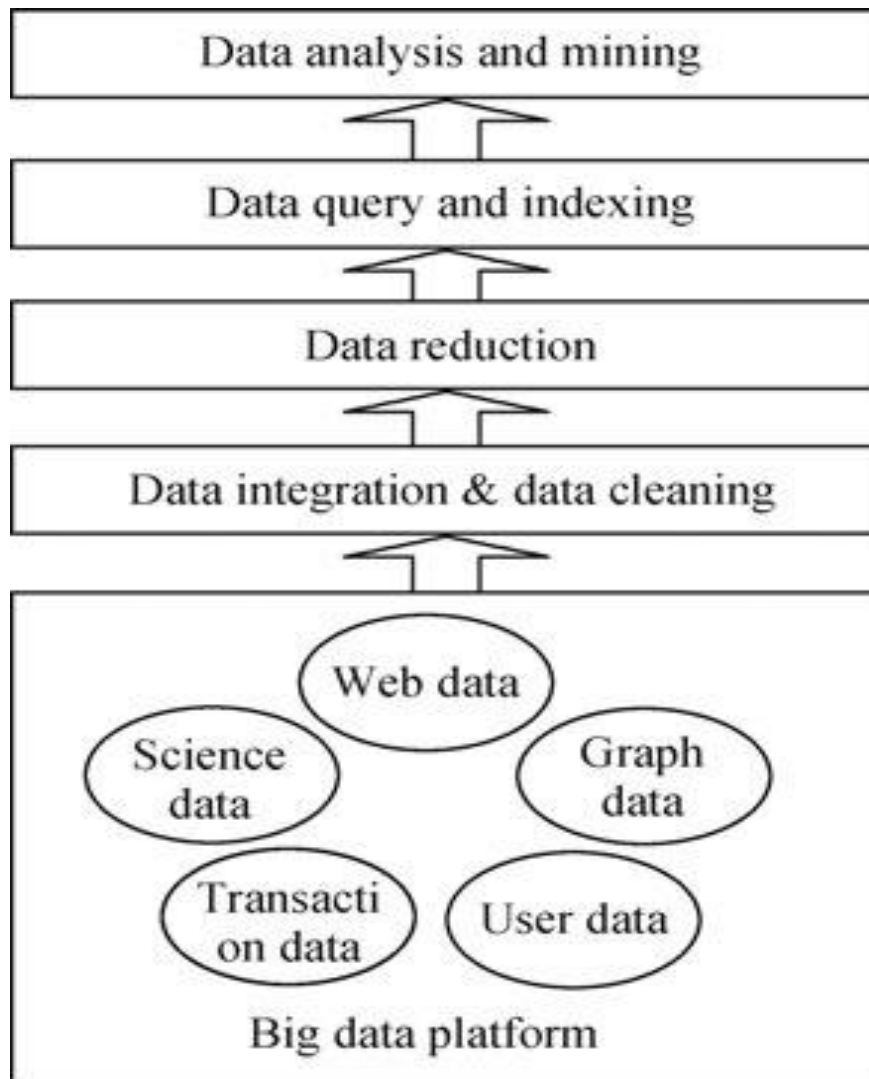
Slika 1. Prikaz vrsta podataka



“Na slici iznad se može vidjeti konceptualna podjela podataka s tim da je u ovom slučaju , zanimljiva grana nestrukturniranih podataka i to tekstualnih podataka. Osim tekstualnih, možese vidjeti da postoje netekstualni nestrukturnirani podaci , a oni predstavljaju grafike i slike (fotografije, ilustracije, X-zrake, MRI (Magnetska rezonanca)...).”

Izvor: Živko Krstić: Big Data i semantička analiza: Iskorištavanje vrijednosti nestrukturniranih podataka u poslovanju, Ekonomski fakultet Split, Sveučilište u Splitu, Svibanj 2014., str.15

Slika 2. Koraci procesuiranja velikih podataka



Izvor: Jinchuan Chen, Yueguo Chen, Xiaoyong Du, Cuiping Li, Jiaheng Lu: Big data Challenge: a data management perspective, Key Laboratory of Data Engineering and Knowledge Engineering, School of Information, Renmin University of China, Beijing 100872, February 22, 2013, Front. Comput. Sci., 2013, 7(2):

str. 158

3. TEHNOLOGIJE VELIKIH PODATAKA

Implementirati tehnologiju velikih podataka ne predstavlja postavljanje baze podataka i u nju spremanje podataka, jer to na kraju rezultira produkt investicije koja nije prvotno planirana, a za razumijevanje tehnološke podloge je ipak potrebno malo više prostora. Ako se prisjetite kako vas proizvođači ciljano obasipaju ponudama i znaju što vam treba baš onda kad vam treba - ispod takvih aktivnosti leže tehnologije velikih podataka. Jedne od najznačajnijih tehnologija koje ćemo objasniti u nastavku rada su Hadoop tehnologija, Map Reduce tehnologija i Bigtable tehnologija.⁶

3.1. Hadoop tehnologija

Hadoop je skup open source tehnologija koji se pojavljuje u bilo kojem značajnijem analitičkom proizvodu pri upravljanju velikim podacima. Hadoop se nalazi u 37% svih Open Source Big Data tehnologija, te u 47% komercijalnih Big Data tehnologija. Hadoop danas distribuiraju u svojim proizvodima Cloudera, Hortonworks, MapR, IBM, Microsoft, Amazon Web Service i mnogi drugi.

Osnovni razlog korištenja Hadoop-a za većinu organizacija je sposobnost skladištenja i procesiranja velike količine podataka, ali na brz način. Što se više računalnih čvorova koristi, to je veća brzina procesiranja na raspolaganju. Pod takvim čvorovima smatraju se računala koji se koriste u distribuiranoj mreži. Naime, ukoliko jedan čvor prestane raditi, zadaci se automatski prebacuju na druge čvorove. Također, čuvaju se višestruke kopije podataka. Hadoop je besplatna platforma, a s obzirom da je sposobna koristiti hardver koji neka organizacija već posjeduje, takva platforma zaista i je jeftina. Istovremeno je i fleksibilna, tako da se distribuirana mreža može širiti dodavanjem novih računalnih čvorova, a za koju nije potrebna velika količina administracije.⁷

Da bi jednostavnije shvatili poviše napisano, jednostavno zamislite da imate datoteku čija je veličina veća od ukupnog kapaciteta pohranjivanja vašeg sustava. Ne bi bilo moguće pohraniti tu datoteku u taj jedinstveni prostor za pohranu. Hadoop je okvir koji korisnicima omogućuje pohranu više datoteka velike veličine (veće od kapaciteta računala). Pojednostavljeno rečeno, Hadoop je zbirka knjižnica, odnosno open source biblioteka, za

⁶Ratko Mutavdžić: Big Data – izazovi implementacije, <http://www.infotrend.hr/clanak/2013/11/big-data---izazovi-implementacije,78,1029.html>, 21.02.2018.

⁷Ibid., 21.02.2018.

obradu velikih skupova podataka (pojam "velik" ovdje se može povezati s 4 milijuna upita za pretraživanje po minuti na Googleu) preko tisuća računala u skupinama. U ranijim danima organizacije su morale kupiti skupi hardver za postizanje visoke dostupnosti. Hadoop je prevladao ovu ovisnost jer se ne oslanja na hardver, već postiže visoku dostupnost i otkriva točku kvarova kroz sam softver.

Kako bi još jednostavnije shvatili sam koncept Hadoop tehnologije, metaforički ćuprikazati Hadoop tehnologiju kroz primjer koncepta blockbuster filma. Kao što svi znamo, blockbuster film zahtijeva snažnu glavnu ulogu, alitakođer zahtijeva i obećavajuće sporedne glumce. Dakle, pogledajmo četiri važne knjižnice Hadoopa, koje su postale super heroji –

1. Hadoop Common –uloga ovog karaktera je pružiti zajedničke alate koji se mogu koristiti u svim modulima.
2. Hadoop MapReduce - desna ruka našeg glumca, obavljajući sav posao koji mu je dodijeljen, tj. provodi raspoređivanje i obradu posla prekoklastera. Hadoop jepoput sustava za skladištenje podataka, tako da je potrebna knjižnica kao što je MapReduce za obradu podataka.
3. Hadoop distribuirani datotečni sustav (HDFS) - lijeva ruka glumca koja održava sve zapise, tj. upravljanje datotečnim sustavom preko klastera.
4. Hadoop YARN - Ovo je novija i poboljšana verzija programa MapReduce, od verzije 2.0 i radi isto.⁸

⁸Hadoop explained – How does Hadoop work and how to use it?, DeZyre:
<https://www.dezyre.com/article/hadoop-explained-how-does-hadoop-work-and-how-to-use-it-/237>, 23.02.2018.

3.2. Zašto koristiti Hadoop tehnologiju?

Hadoop se koristi kada se stvara velika količina podataka i kada organizacija zahtijeva uvide iz tih podataka. Snaga Hadoopa leži u okviru, jer se gotovo svi programi mogu priključiti i mogu se koristiti za vizualizaciju podataka. Može se proširiti iz jednog sustava na tisuće sustava u klasteru, a ti sustavi mogu biti nisko-krajni robni sustavi. Hadoop ne ovisi o hardveru za visoku dostupnost. Dva osnovna razloga za podršku pitanju "Zašto koristiti Hadoop":

- Ušteda troškova s Hadoopom su dramatične u usporedbi s naslijeđenim sustavima.
- Ima snažnu podršku zajednice koja se s vremenom razvija s novim napredovanjem.

3.3. U koje se svrhe koristi Hadoop?

Hadoop je postao velika podatkovna tehnologija zbog svoje snage za obradu velikih količina polustrukturiranih i nestrukturiranih podataka. Hadoop nije popularan zbog svoje brzine obrade u rješavanju malih skupova podataka. Ako razmišljate o tome za što se koristi Hadoop, ili okolnosti pod kojima je Hadoop korisno koristiti, evo odgovora -

Hadoop se koristi u velikim aplikacijama podataka koji prikupljaju podatke iz različitih izvora podataka u različitim formatima. HDFS (Hadoop Distributed File System) je fleksibilan u pohranjivanju različitih vrsta podataka, bez obzira na činjenicu sadrže li vaši podaci audio ili video datoteke (nestrukturirani podaci), ili sadrže podatke o razini zapisa, kao u ERP (Enterprise Resource Planing) sustavu (strukturirani podaci), log datoteke, ili XML (Extensible Markup Language) datoteke (polu-strukturirani podaci). Hadoop se koristi u velikim aplikacijama podataka koji se moraju spojiti i pridružiti podacima - klikovnim podacima, podacima društvenih medija, transakcijskim podacima ili bilo kojim drugim formatom podataka. Nemojte pogriješiti koristeći Hadoop kada su vaši podaci premali, recimo u MB-ima ili GB-ima. Da biste postigli visoku skalabilnost i uštedjeli novac i vrijeme, Hadoop treba koristiti samo kada su skupovi podataka u petabajtima ili terabajtima, inače je bolje koristiti Postgres ili Microsoft Excel.⁹

⁹Tavish Srivastava, What is Hadoop? – Simplified!, <https://www.analyticsvidhya.com/blog/2014/05/hadoop-simplified/>, 23.02.2018.

3.4. Kako Hadoop funkcionira?

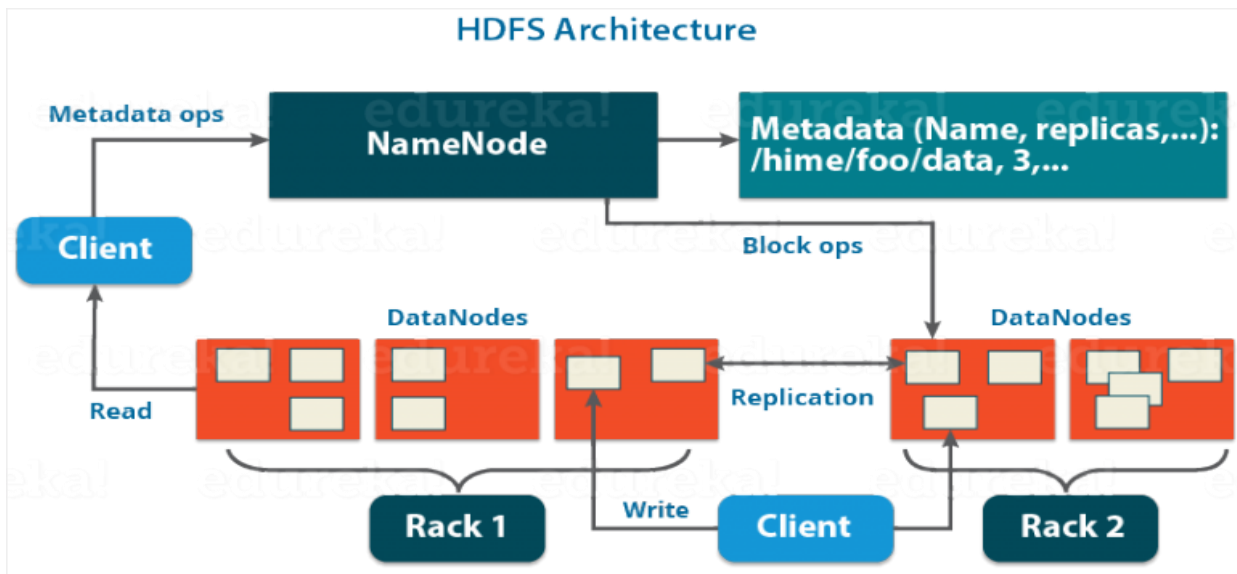
Hadoop je ekosustav knjižnica, a svaka knjižnica ima svoje zadatke za izvođenje. HDFS zapisuje podatke jednom poslužitelju, a zatim ih čita i ponovno upotrebljava mnogo puta. Kada ga uspoređujete s kontinuiranim višestrukim čitanjem i pisanjem akcija drugih datotečnih sustava, HDFS pokazuje brzinu kojom Hadoop radi i stoga se smatra savršenim rješenjem za obradu velikih količina podataka.

Job Tracker je glavni čvor koji upravlja svim čvorovima robnih zadataka praćenja i izvršava poslove. Kad god je potreban neki podatak, zahtjev se šalje na NameNode koji je glavni čvor (pametni čvor klastera) HDFS-a i upravlja svim čvorovima robnih čvorova podataka. Zahtjev se prenosi na cijelu podatkovnu mrežu koja služi traženim podacima.

MapReduce ili YARN, koriste se za zakazivanje i obradu podataka. Hadoop MapReduce izvršava niz poslova, gdje je svaki posao Java aplikacija koja se pokreće na podacima.¹⁰

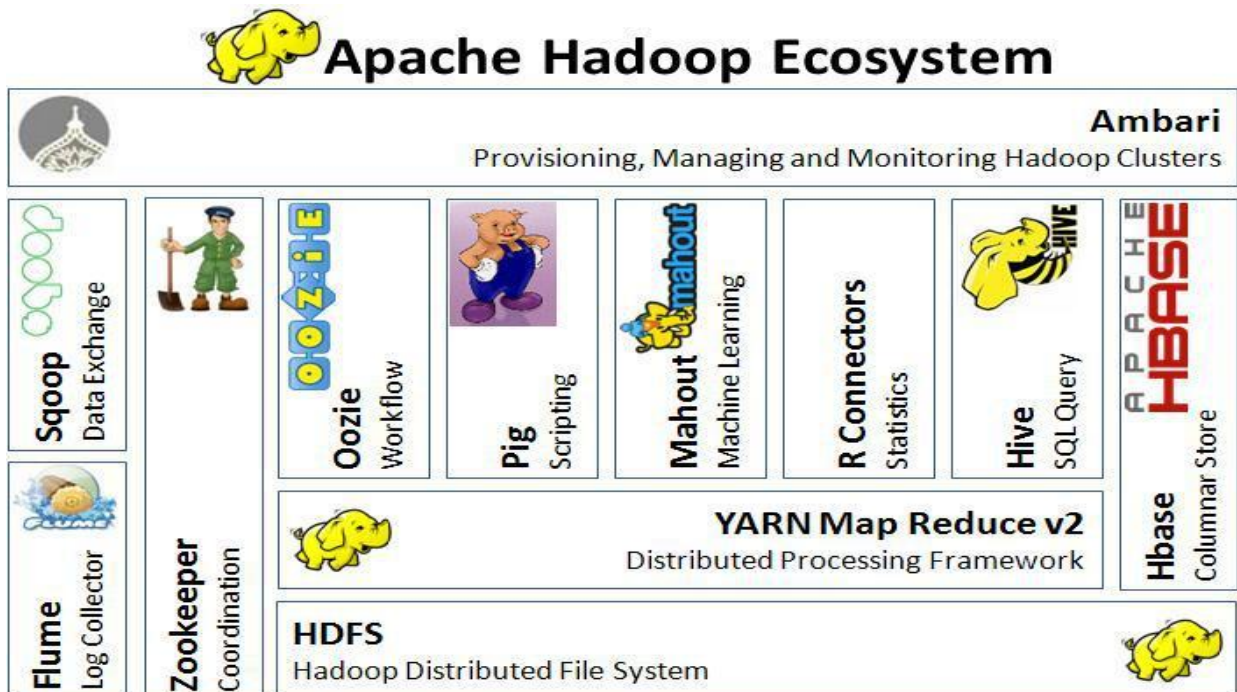
¹⁰Tavish Srivastava, What is Hadoop? – Simplified!, <https://www.analyticsvidhya.com/blog/2014/05/hadoop-simplified/>, 23.02.2018.

Slika 2. Arhitektura HDFS-a (Hadoop distribuiranog datotečnog sustava)



Izvor: <https://www.edureka.co/blog/apache-hadoop-hdfs-architecture/>

Slika 3. Apache Hadoop ekosustav



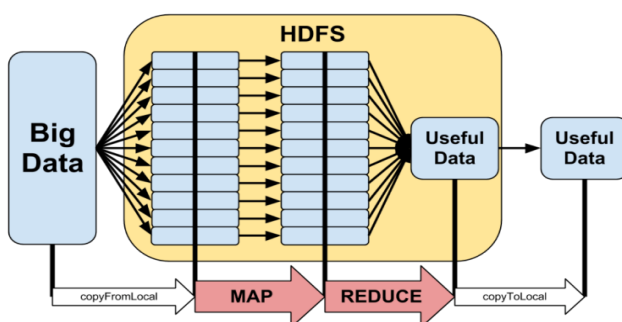
Izvor: <http://blog.agroknow.com/?p=3810>

3.5. Map Reduce tehnologija

MapReduce je postao najpoznatiji alat za obradu velikog broja podataka u analitičke svrhe. Njegov programski model je jednostavan, ali opet vrlo izražajan. Iako pruža samo dvije funkcije: map i reduce, pomoću njih je moguće izvršiti veliki broj analitičkih zadataka poput upita za baze podataka, dubinske analize podataka i procesuiranje grafova. Programski model neovisan je o sustavu pohrane podataka i omogućava obradu različitih tipova podataka, strukturiranih i nestrukturiranih. Ovakav dizajn koji je neovisan o sustavu pohrane podataka pokazao se jako korisnim u produkcijskom okruženju gdje nerijetko postoje različiti sustavi za pohranu podataka. MapReduce postiže elastičnu skalabilnost kroz planiranje na razini bloka podataka. Sustav automatski pretvara ulazne podatke u blokove podataka iste veličine i dinamički računa blokove koji će biti dostupni računalnim čvorovima za obradu. MapReduce također pruža funkcionalnost oporavka od pogreške gdje se samo zadatci koji su izvršavani na srušenim čvorovima trebaju ponavljati. No niti to sve ne znači da su performanse tehnike MapReduce idealne u kontekstu baza podataka.

MapReduce sustav upravlja distribuiranim serverima i općenito cijelim procesom. Sustav izvršava različite zadatke paralelno, upravlja svim komunikacijama kao i prijenosom podataka između različitih dijelova sustava, u isto vrijeme osiguravajući sustav od prekomjernosti podataka i grešaka. Inspiracija za model je proizašla iz map i reduce funkcija koje se često koriste u funkcionalnom programiranju iako njihova uloga u MapReduce sustavu nije ista kao što je u njihovom izvornom obliku.¹¹

Slika 4. Konceptualni pregled Map Reduce sustava



Izvor: <https://www.glennklockwood.com/data-intensive/hadoop/overview.html>

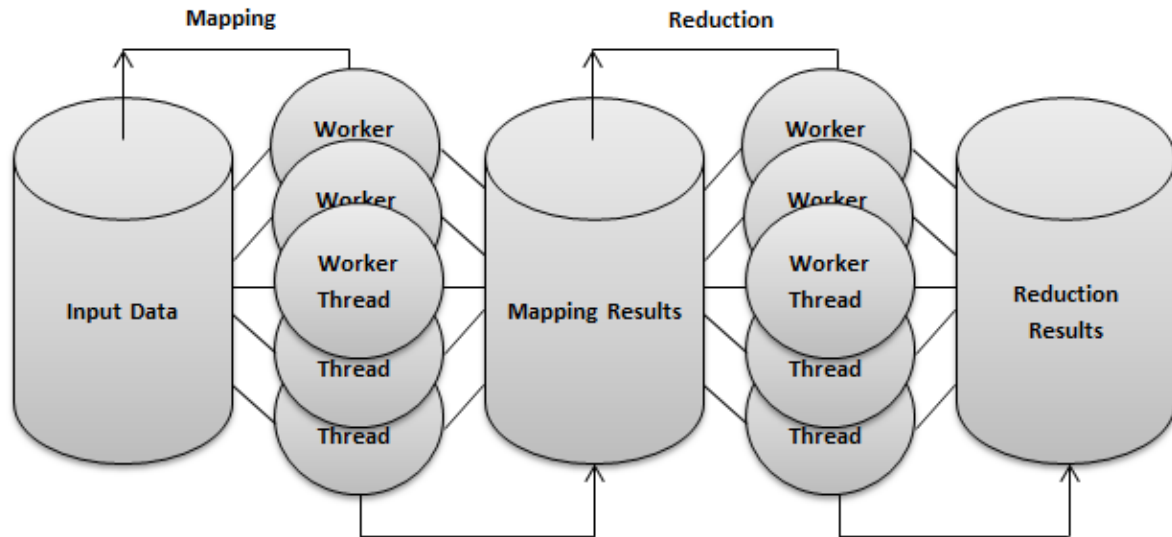
¹¹ Jake Drew, Map Reduce/Map Reduction Strategies using C#, <https://www.codeproject.com/Articles/524233/MapReduceplus-fplusMapplusReductionplusStrategies>, 24.02.2018.

3.6. Razumijevanje Map Reduce tehnologije

U složenijim oblicima, radna mjesta za MapReduce tehnologiju su razbijena u pojedinačne, neovisne jedinice rada i rasprostranjene na mnogim poslužiteljima, obično robnim hardverskim jedinicama, kako bi se pretvorio vrlo velik i složen zadatak obrade u nešto što je puno manje komplicirano i lako za upravljanje putem računala koji su zajedno povezani u klasteru. Kada je zadatak koji je pri ruci prevelik za jedno računalo, treba se formirati "tim" da dovrši posao. "Tim" za map reducing bi se sastojalo od jednog ili više višeprosorskih čvorova (računala) i neke vrste glavnih čvorova ili programa koji upravlja naporom dijeljenja posla između čvorova (mapiranje) i agregacije konačnih rezultata na svim radnim čvorovima (smanjenje). Glavni čvor ili program mogao bi se smatrati "voditeljem tima" za Map Reducing. U tim većim sustavima moglo bi se zahtijevati mnogo dodatnih indeksacija i drugih slojeva upravljanja podacima, ovisno o pojedinačnim projektnim zahtjevima. Međutim, prednosti Map Reducing-a mogu se ostvariti i na jednom multi-procesorskom računalu za manje projekte.

Osnovne prednosti bilo kojeg sustava za Map Reducing dolaze od podjele rada preko mnogih procesora i čuvanje što više podataka u memoriji što je više moguće tijekom obrade. Čitanje podataka i pisanje podataka na disk predstavlja najveću priliku za poboljšanje performansi u većini tipičnih sustava. Međutim, kada se klaster distribuira, uvodi se kompleksno programiranje. Ulazni podaci moraju biti podijeljeni (mapirani) preko klasterovih čvorova (računala) na jednak način koji i dalje proizvodi točne rezultate i lako se posvećuje agregaciji konačnih rezultata (smanjenja). Mapiranje ulaznih podataka na određene čvorove klastera uz mapiranje pojedinih jedinica unosa podataka na pojedine procesore unutar jednog čvora. Smanjenje na više čvorova klastera također zahtijeva dodatnu kompleksnost programiranja. U svim sustavima za Map Reducing neki oblik paralelne obrade mora se koristiti kada se koristi više procesora. Budući da je paralelna obrada uvijek uključena tijekom Map Reducing-a, sigurnost nije primarna briga za bilo koji sustav. Podaci za unos moraju se podijeliti na pojedinačne neovisne jedinice rada koje se u bilo kojem trenutku tijekom različitih faza mapiranja i redukcije mogu obrađivati bilo kojom radnom niti. Ponekad to zahtijeva dodatno razmišljanje tijekom faza projektiranja budući da ulazni podaci nisu nužno obrađeni linearno. Prilikom obrade tekstualnih podataka na primjer, posljednja rečenica dokumenta može biti obrađena prije prve rečenice dokumenta budući da više radničkih niti istodobno rade na svim dijelovima ulaznih podataka.

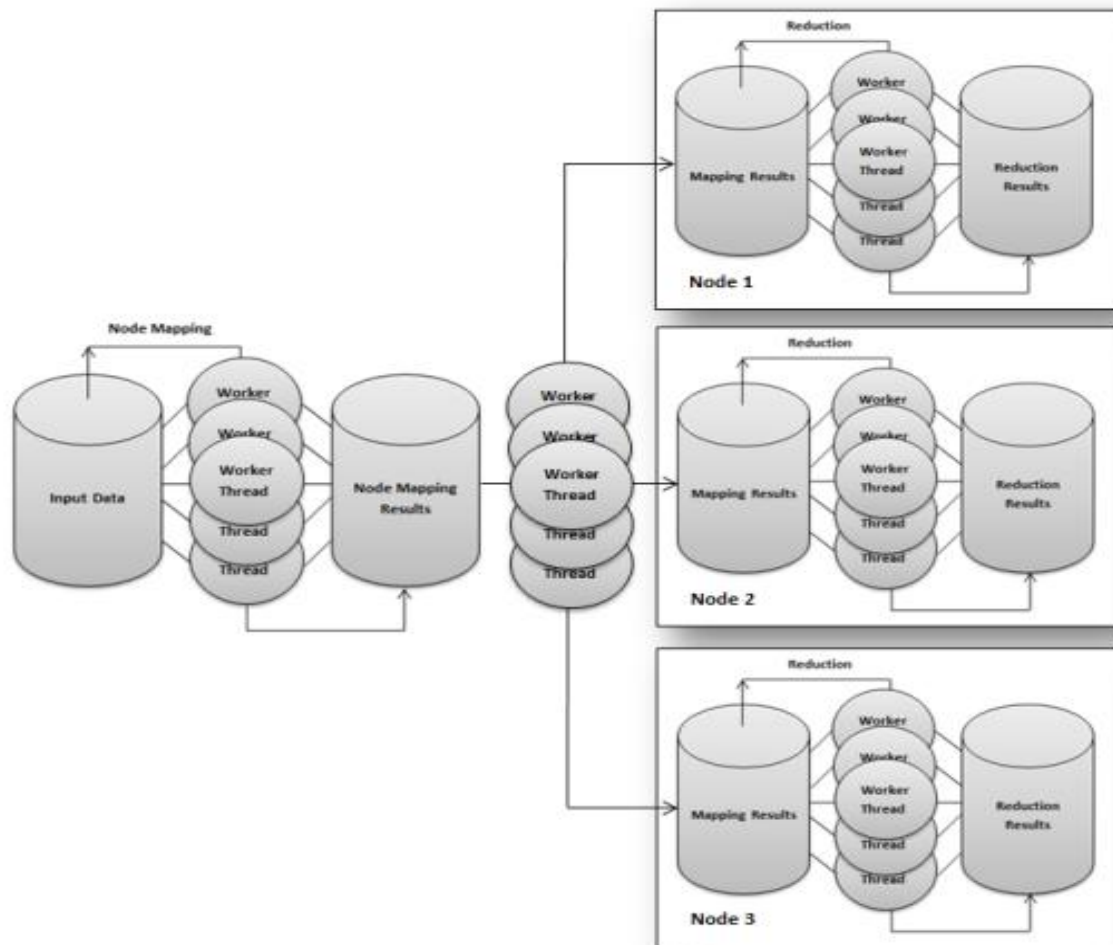
Slika 5. Prikaz Map Redukcije na jednom višeprocorskom računalu



Izvor: <https://www.codeproject.com/Articles/524233/MapReduceplus-fplusMapplusReductionplusStrategies>

Slika poviše prikazuje proces Map Redukcije koji se izvodi na jednom višeprocorskom računalu. Tijekom tog procesa višestruke radne niti istovremeno se prenose na različite dijelove ulaznih podataka postavljajući rezultate mapiranja u centralizirano mjesto za daljnju obradu pomoću drugih redukcijских radničkih niti. Budući da se taj proces događa na jednom stroju, mapiranje je manje složeno jer se ulazni podaci dijele samo između radničkih niti i procesora koji svi leže na istom računalu i obično unutar istog spremišta podataka.

Slika 6. Map Reducing na četiri višeprocorska računala



Izvor: <https://www.codeproject.com/Articles/524233/MapReduceplus-fplusMapplusReductionplusStrategies>

Slika poviše prikazuje Map redukciju koja se izvodi na četiri stroja. U ovom scenariju, jedan glavni čvor koristi se za dijeljenje (mape) ulaznih podataka između tri čvora za obradu podataka za eventualno smanjenje. Jedan od zajedničkih izazova prilikom projektiranja klastera je da se svi podaci o smanjenju ne mogu nalaziti na jednom fizičkom stroju. U primjernom sustavu Map Redukcije koji obrađuje tekstualne podatke kao ulaz i broji jedinstvene riječi. Sve riječi koje počinju s A-I mogu biti pohranjene u čvoru 1, J-R u čvoru 2 i S-Z u čvoru 3. To znači da se moraju koristiti dodatna logička usmjeravanja kada svaku riječ unesete na ispravan čvor za konačno smanjenje na temelju prvog slova svake riječi.

3.7. Bigtable tehnologija

Bigtable je dizajniran za podršku aplikacijama koje zahtijevaju veliku skalabilnost. Od svoje prve interakcije, tehnologija je namijenjena za korištenje s petabajt podacima. Baza podataka je dizajnirana za implementaciju na klasterirane sustave i koristi jednostavan model podataka koji je Google opisao kao "rijetku, distribuiranu i postojanu višedimenzionalnu sortiranu kartu". Podaci su sastavljeni redosljedom po retku ključa, a indeksiranje karte je uređeno prema redosljedu, tipkama stupaca i vremenskim oznakama. Algoritmi kompresije pomažu u postizanju velikog kapaciteta.

Google Bigtable služi kao baza podataka za aplikacije kao što su Google App Engine Datastore, Google personalizirano pretraživanje, Google Earth i Google Analytics. Google je zadržao softver kao vlasnička, in-house tehnologija. Bez obzira na to, Bigtable je imao veliki utjecaj na dizajn SQL (Structured Query Language) baze podataka. Googleovi programeri softvera javno su objavili detalje Bigtablea u tehničkom radu predstavljenom na USENIX simpoziju o operativnim sustavima i implementaciji dizajna 2006. godine.

Googleov temeljit opis unutarnjih funkcioniranja Bigtable-a omogućio je drugim organizacijama i razvojnim timovima open source-a da stvore derivate Bigtable, uključujući bazu Apache HBase, koja je izgrađena za pokretanje na vrhu Hadoop distribuiranog datotečnog sustava (HDFS). Drugi primjeri uključuju Cassandra, koja je nastala na Facebook Inc i Hypertable, tehnologiji open source koja se prodaje u komercijalnoj verziji kao alternativa HBase-u.¹²

¹² Margaret Rouse: Google BigTable, <http://searchdatamanagement.techtarget.com/definition/Google-BigTable> , 24.02.2018.

4. POTREBA ZA BIG DATA TEHNOLOGIJOM U ZDRAVSTVU

Tijekom slijedećih 10 godina svijet će doživjeti duboke demografske promjene. Očekuje se povećanje svjetske populacije za milijardu ljudi od kojih će polovica biti starija od 50 godina. Što to predstavlja za zdravstvene sustave i za buduće naraštaje?

Kronične bolesti trenutno predstavljaju 70% svih bolesti, a taj se postotak povećava s povećanjem starosne dobi stanovništva. Dolazi do situacije u kojoj dominira starija populacija društva s povećanim zdravstvenim potrebama. Ukoliko ne dođe do prilagodbe zdravstvenog sustava demografskim promjenama, nagovještava se udvostručenje troškova u zdravstvu. Dolazi do potrebe za pomicanjem video kruga zdravstvenog sustava, s uskog fokusa na smanjenje troškova prema maksimiziranju vrijednosti, što znači dobivanje najboljeg ishoda iz utrošenih sredstava za sve sudionike zdravstvenih sustava.

Iz gore navedenih razloga je pokrenut program koji će dovesti do promjena u načinu liječenja, te istovremeno doprinijeti održivosti europskih zdravstvenih sustava. Riječ je o važnom programu u okviru inovativne medicinske inicijative (IMI) pod nazivom Big Data For Better Outcomes ("Veliki podaci za bolje ishode").¹³

¹³Pharmabiz, stručni magazin za lijekove i medicinske proizvode, broj magazina 17., izdanje magazina: studeni, 2015. Str.26.

5. RAZLIKA AMERIČKOG I EUROPSKOG ZDRAVSTVENOG SUSTAVA S ASPEKTA BIG DATA-E

Američki sustav zdravstvene skrbi ubrzano prihvaća elektroničke evidencije o zdravlju, što dramatično povećava količinu kliničkih podataka koji su elektronski dostupni. Istodobno, u kliničkoj analitičkoj tehnici napravljen je brz napredak za analizu velikih količina podataka i prikupljanje novih spoznaja iz te analize – bolje poznato pod nazivom "Big Data". Kao rezultat toga, postoje mogućnosti za korištenje velikih podataka kako bi se smanjili troškovi zdravstvene zaštite u SAD-u.

Raspravlja se o vrstama spoznaja koje će vjerojatno izaći iz kliničke analize, vrstama podataka potrebnih za dobivanje takvih uvida i analiza infrastrukture, algoritmi, registri, uređaji za praćenje, itd. Organizacije će morati provesti potrebne analize i promjene koje će poboljšati skrb uz smanjenje troškova.¹⁴

Zdravstveni sustavi u Europi suočavaju se sa starenjem stanovništva što povećava broj slučajeva kroničnih bolesti. Lako se da zaključiti da će financijski pritisak na zdravstvene sustave kontinuirano rasti. Potrebno je provesti određene promjene u svrhu održavanja sustava. Neke od promjena mogu biti prelazak sa sadašnjeg načina liječenja, orijentiranog na pojedinu bolest u kojem ishodi nisu sustavno mjereni, na potpuno nov pristup koji je holistički ustrojen i orijentiran na terapijske rezultate. Takav pristup nije u velikoj mjeri zastupljen u današnjim zdravstvenim sustavima u Europi.

Povećanjem dostupnosti "Big data" vezanih za zdravlje, koje posjeduju i kreiraju dionici u sustavu, omogućilo bi se poboljšanje sadašnjeg zdravstvenog sustava. U sklopu javno-privatnog partnerstva između Europske komisije i Europske udruge inovativnih proizvođača lijekova osigurana su značajna sredstva za ostvarenje ovog cilja.¹⁵

¹⁴David W. Bates, Suchi Saria, Lucila Ohno-Machado, Anand Shah and Gabriel Escobar: Big Data In Health Care: Using Analytics To Identify And Manage High-Risk And High-Cost Patients,*Health Affairs*, 33, no.7 (2014):1123-1131

¹⁵Pharmabiz, stručni magazin za lijekove i medicinske proizvode, broj magazina 17., izdanje magazina: studeni, 2015. Str.27.

6. PRIMJENA BIG DATA TEHNOLOGIJE U ZDRAVSTVU

Od bankarskog do maloprodajnog, mnogi sektori su već objeručke prihvatili Big Datu, neovisno dolaze li informacije iz privatnih, ili javnih izvora. Prodavaonice prehrambenih proizvoda, na primjer, istražuju podatke o lojalnosti kupaca kako bi identificirali trendove u prodaji i razvili posebnu vrstu ponude. Ne samo da na taj način poboljšavaju profit, nego i povećavaju zadovoljstvo svojih kupaca. Tradicionalno, zdravstvena industrija zaostaje za drugim industrijama u aspektu korištenja Big Date. Dio problema proizlazi iz otpora promjenama, a pružatelji usluga navikli su samostalno donositi odluke o liječenju koristeći vlastitu kliničku prosudbu, a ne oslanjajući se na protokole koji se temelje na Big Data-i. Ostale prepreke su više strukturne prirode. Mnogi sudionici zdravstvene skrbi nedovoljno su ulagali u informacijsku tehnologiju zbog neizvjesnih prinosa. Iako su njihovi stariji sustavi funkcionirali, imaju ograničenu mogućnost standardizacije i konsolidacije podataka. Priroda same zdravstvene industrije također stvara nove izazove. Iako postoji mnogo sudionika, ne postoji način jednostavnog dijeljenja podatak između različitih pružatelja usluga. Dijelom zbog zabrinutosti za privatnost. Čak i unutar jedne bolnice, ili farmaceutske tvrtke, važne informacije često ostaju unutar jedne grupe, ili odjela jer organizacije nemaju postupke za integraciju podataka i komunikaciju s nalazima.

Dostupnost Big Data u zdravstvu preoblikuje pogled na stanje o tome što je primjereno, ili pogodno za pacijenta i zdravstveni ekosustav. U korak s ovim promjenama, stvoren je "holistički okvir" usmjeren na pacijenta zasnovan na konceptu da vrijednost proizlazi iz ravnoteže potrošnje zdravstvene zaštite (troškova) i utjecaja na pacijenta (ishoda). Razmatraju se i neki od rizika povezanih s Big Data tehnologijom kao što je opasnost izlaganja povjerljivih informacija o pacijentu i pregled fundamentalnih promjena koje se mogu pojaviti u zdravstvenom sustavu.

Implementacija Big Data-e moglo bi inspirirati mnoge tvrtke za razvoj zdravstvenih aplikacija, ili sličnih inovacija. Da bismo ocijenili taj trend, pogledat ćemo profile poduzeća i poslovne modele sudionika na Forumu inicijativa o zdravstvenim podacima za 2011. i 2012. Slijedeće navedeni primjeri dokazuju kako je velika revolucija podataka stvorila nove inovacije na području zdravstvene zaštite.

Neki od primjera su:

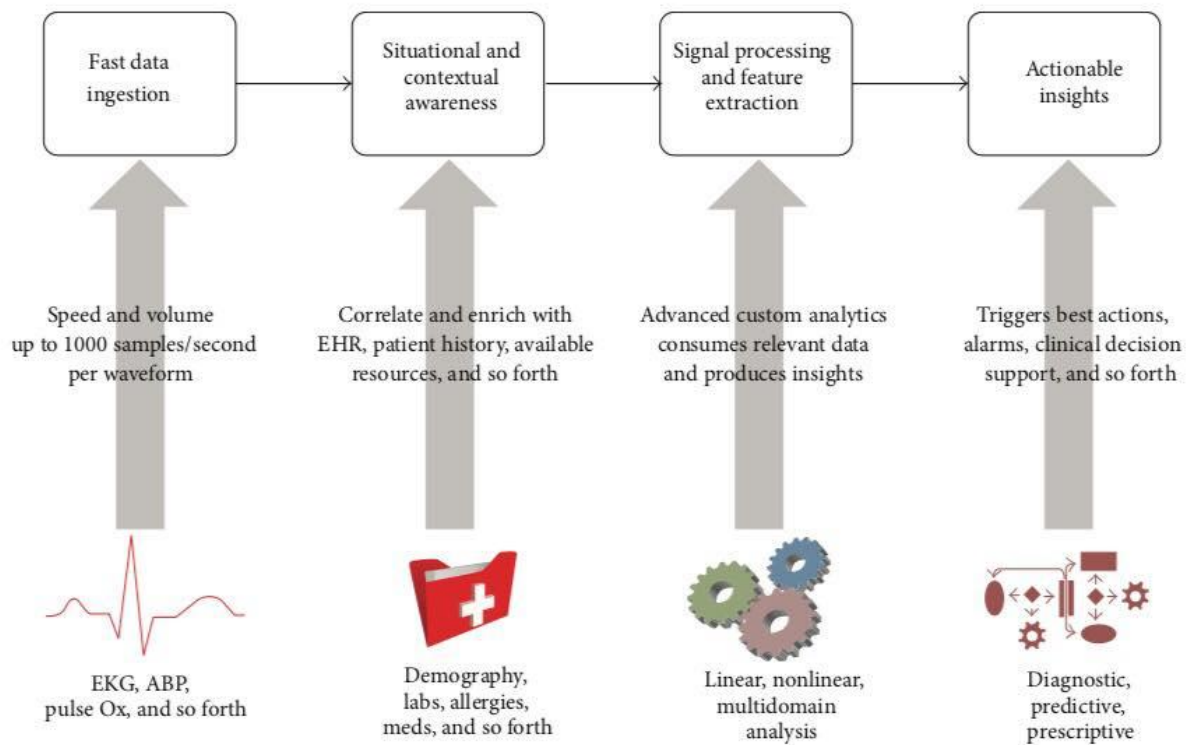
Asthampolis je stvorio GPS (Global Positioning System) tracker koji prati uporabu inhalatora astmatičarima. Informacije se prenose u središnju bazu podataka i spajaju se s informacijama CDC-a (Centres of Disease Control) o poznatim katalizatorima astme do pomoći liječnicima da razviju personalizirane planove liječenja i mogućnosti prevencije.

Ginger.io nudi mobilnu aplikaciju u kojoj se pacijenti (poput onih s dijabetesom) slažu, zajedno sa svojim pružateljima usluga, da se prate putem svojih mobilnih uređaja i pomažu u terapijama ponašanja. Praćenjem mobilnih senzora prisutnih upametnim telefonima, aplikacija bilježi informacije o pozivima, porukama, lokacije, pa čak i informacije o kretanju. Pacijenti također popunjavaju ankete preko svojih pametnih telefona. Program Ginger.io integrira navedene podatke s javnim istraživanjima iz NIH-a (National Institutes of Health) i drugih izvora podataka o zdravstvenom ponašanju. Dobiveni uvidi mogu biti otkriveni, npr., nedostatak kretanja, ili druge aktivnosti mogu signalizirati da se pacijent ne osjeća fizički dobro, a nepravilni uzroci spavanja mogu signalizirati da je jedan anksiozni napad neizbježan.

mHealthCoach podržava pacijente na kroničnoj njezi lijekova, osiguravajući obrazovanje i promicanje pridržavanja liječenja kroz interaktivni sustav. Aplikacija iskorištava podatke iz projekta Healthcare Cost and Utilization, koju sponzorira Agencija za istraživanje i kvalitetu zdravstvene zaštite, kao i rezultate i upozorenja iz kliničkih ispitivanja (preuzeta iz FDA-E (Food and Drug Administration/Američka agencija za hranu i lijekove) kliničke web stranice). mHealthCoach se također mogu koristiti od strane pružatelja usluga i platitelja kako bi identificirali bolesnike s većim rizikom i dostavili im ciljane poruke i podsjetnike.¹⁶

¹⁶Ashwin Belle,Raghuram Thiagarajan,S. M. Reza Soroushmehr,Fatemeh Navidi,Daniel A. Beard,and Kayvan Najarian: Big Data Analytics in Healthcare, Hindawi Publishing Corporation, BioMed Research International Volume 2015, Article ID 370194, str.1-2

Slika 7. Opći tijek rada Bigdata analitike koristeći se podacima zdravstvenog sustava



Izvor: Ashwin Belle, Raghuram Thiagarajan, S. M. Reza Soroushmehr, Fatemeh Navidi, Daniel A. Beard, and Kayvan Najarian: Big Data Analytics in Healthcare, Hindawi Publishing Corporation, BioMed Research International Volume 2015, Article ID 370194, str.6

7. ARHITEKTONSKI OKVIR BIG DATA-E U ZDRAVSTVU

Konceptualni okvir za Big data analitiku u zdravstvu sličan je onom tradicionalnog projekta zdravstvene informatike, ili analitike. Ključna razlika leži u tome kako se vrši obrada podataka. U redovitom projektu zdravstvene analize, analiza se može izvesti pomoću alata za poslovnu inteligenciju instaliranog na samostalni sustav, kao što je sustav stolnih ili prijenosnih računala. Budući da su veliki podaci po definiciji veliki, obrada se razrađuje i izvršava na više čvorova. Konceptija distribuiranog procesa postoji već desetljećima. Ono što je relativno novo jest njegova upotreba u analiziranju vrlo velikih skupova podataka kada se pružatelji usluga zdravstvene zaštite počnu upuštati u velika spremišta podataka kako bi stekli uvid u donošenje bolje informiranosti o zdravstvenim uslugama. Nadalje, open source platforme kao što je Hadoop/MapReduce, dostupne kao cloud usluge, potiču primjenu analize podataka u zdravstvu.

Iako su algoritmi i modeli slični, korisnička sučelja tradicionalnih analitičkih alata i onih koji se koriste za velike podatke su potpuno različiti; tradicionalni alati za analizu zdravlja postali su veoma korisni i transparentni. S druge strane, veliki alati za analizu podataka su iznimno složeni, programski intenzivni i zahtjevaju primjenu različitih vještina.

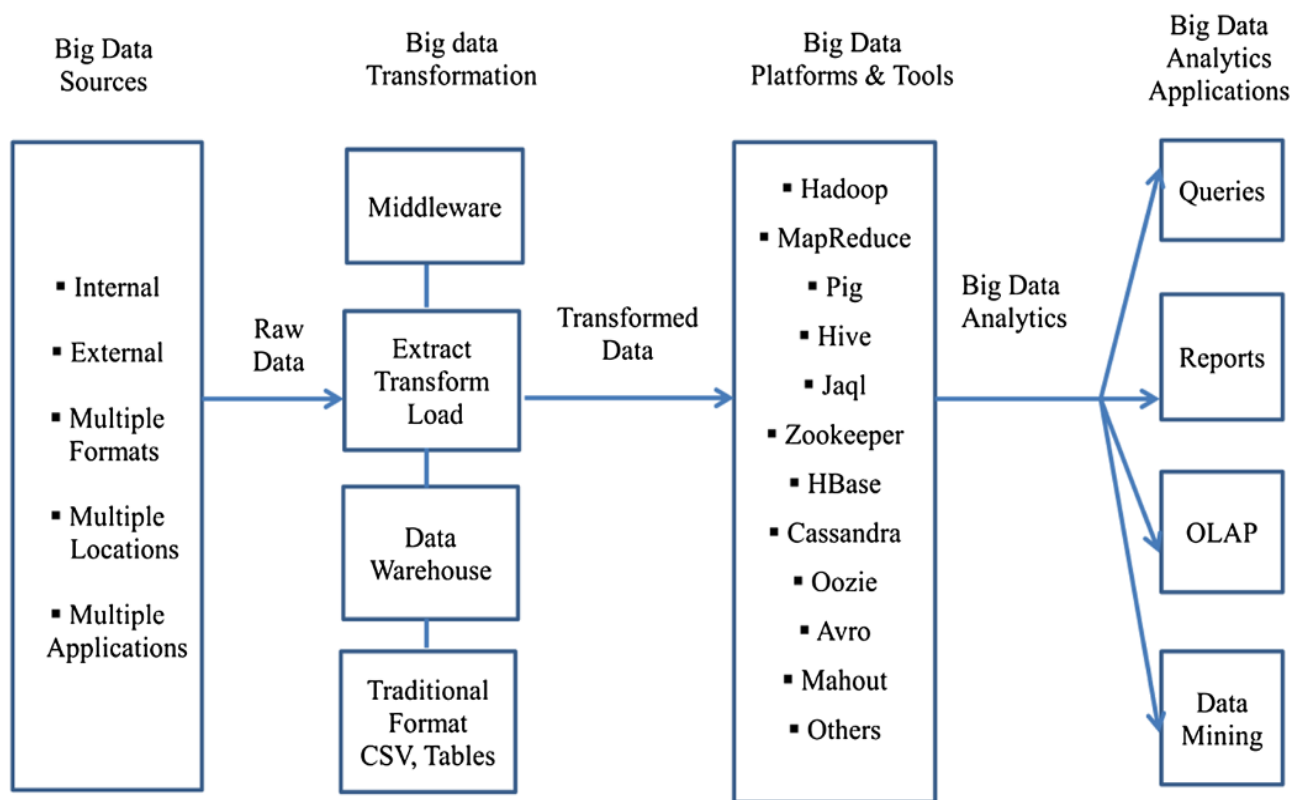
Veliki podaci u zdravstvenoj skrbi mogu doći iz internih (npr. Elektronskih zdravstvenih evidencija, sustava za podršku kliničkim odlukama, itd.) i vanjskih izvora (vladini izvori, laboratoriji, ljekarne, osiguravajuća društva, itd.). Izvori i vrste podataka uključuju:

1. Podatke na webu i društvenim mrežama (Facebook, Twitter, LinkedIn,...),
2. Strojni podaci na stroju (očitanja daljinskog senzora),
3. Veliki podaci o transakcijama (zahtjevi za zdravstvenu zaštitu i ostali zapisi o naplai sve su više dostupni u polustrukturiranim i nestrukturiranim formatima),
4. Biometrijski podaci (otisci prsta, genetika, rukopis, pregled mrežnice, rengetnske snimke i druge mdicinske snimke, krvni tlak, očitavanja impulsa i pulsioksimetrije i drugih sličnih vrsta podataka)
5. Ljudski generirani podaci (nestrukturirani i polustrukturirani podaci poput bilježaka liječnika, elektroničke pošte i papirnatih dokumenata)

U svrhu velike analize podataka, podaci se trebaju združiti. U drugoj komponenti podaci su u "širokom" stanju i trebaju biti obrađeni i transformirani, pri čemu je dostupno nekoliko opcija. Arhitektonski pristup orijentiranom prema usluzi u kombinaciji s web uslugama (middleware) je jedna od mogućnosti. Podaci ostaju neobrađeni, a službe se koriste za pozivanje, preuzimanje i obradu podataka. Drugi je pristup skladištenje podataka u kojem se podaci iz različitih izvora agregiraju i pripremaju za obradu, iako podaci nisu dostupni u realnom vremenu.¹⁷

¹⁷Wullianallur Raghupathi, Viju Raghupathi: Big data analytics in healthcare: promise and potential, Health Information Science and Systems 2014, 2:3

Slika 8. Primijenjena konceptualna arhitektura Bigdata analitike



Izvor: Wullianallur Raghupathi, Viju Raghupathi: Big data analytics in healthcare: promise and potential, Health Information Science and Systems 2014, 2:3

8. BIG DATA TEHNOLOGIJA I 5V U ZDRAVSTVENOM SUSTAVU

Postoji pet glavnih "dimenzija" za tehnologiju velikih podataka, obično nazvane Pet V: Volume, Velocity, Variety, Veracity i Value. Razumijevajući 5V Big Data tehnologiju, možemo je upotrijebiti u istraživanju i rješavanju stvarnih problema. Dakle, razgovarajmo o tome kako se te dimenzije mogu primijeniti na zdravstvenom sustavu.

8.1. Volumen

Podaci o zdravstvenoj zaštiti su u terabyte-ima i petabyte-ima. Ti sustavi uključuju informacije kao što su: osobni podaci, radiološke slike, osobne medicinske zapise, 3D snimke, genomike i očitavanja biometrijskog senzora. Sustav zdravstvenih usluga sada može imati potencijal za upravljanjem i analizom tih složenih podataka. Prema izvješću KPMG-a (Klynveld Peat Marwick Goerdeler), podaci o zdravstvenoj zaštiti dosegili su 150 Exabyte-a u 2013. godini, a povećava se istaknuta stopa od 1,2 do 2-4 Exabyte-a godišnje.

8.2. Variety (raznolikost)

Raznolikost podataka strukturirana je, nestrukturirana i polustrukturirana. Strukturirane informacije, kao što su klinički podaci, lako se manipuliraju, pohranjuju i analiziraju strojevima. Većina podataka o zdravstvenoj skrbi, kao što su uredski medicinski zapisi, bilješke liječnika, propisi o papiru, slike i radiografski filmovi, nestrukturirani su ili polustrukturirani.

8.3. Velocity (brzina)

Većina zdravstvenih podataka tradicionalno su statičke papirnate datoteke, rendgenski filmovi i skripte. Brzina povećanja podataka povećava se s podacima koji predstavljaju redovno praćenje, poput višestrukih dnevnih dijabetičkih mjerenja glukoze (ili kontinuiranog nadzora pomoću inzulinske pumpe), očitavanja krvnog tlaka i EKG-a (elektrokardiograma). U međuvremenu, u mnogim medicinskim situacijama, konstantni podaci u stvarnom vremenu (praćenje trauma za krvni tlak, monitor operacijskih prostorija za anesteziju, noćna srčana monitori itd.) mogu značiti razliku između života i smrti.

8.4. Veracity (istinitost)

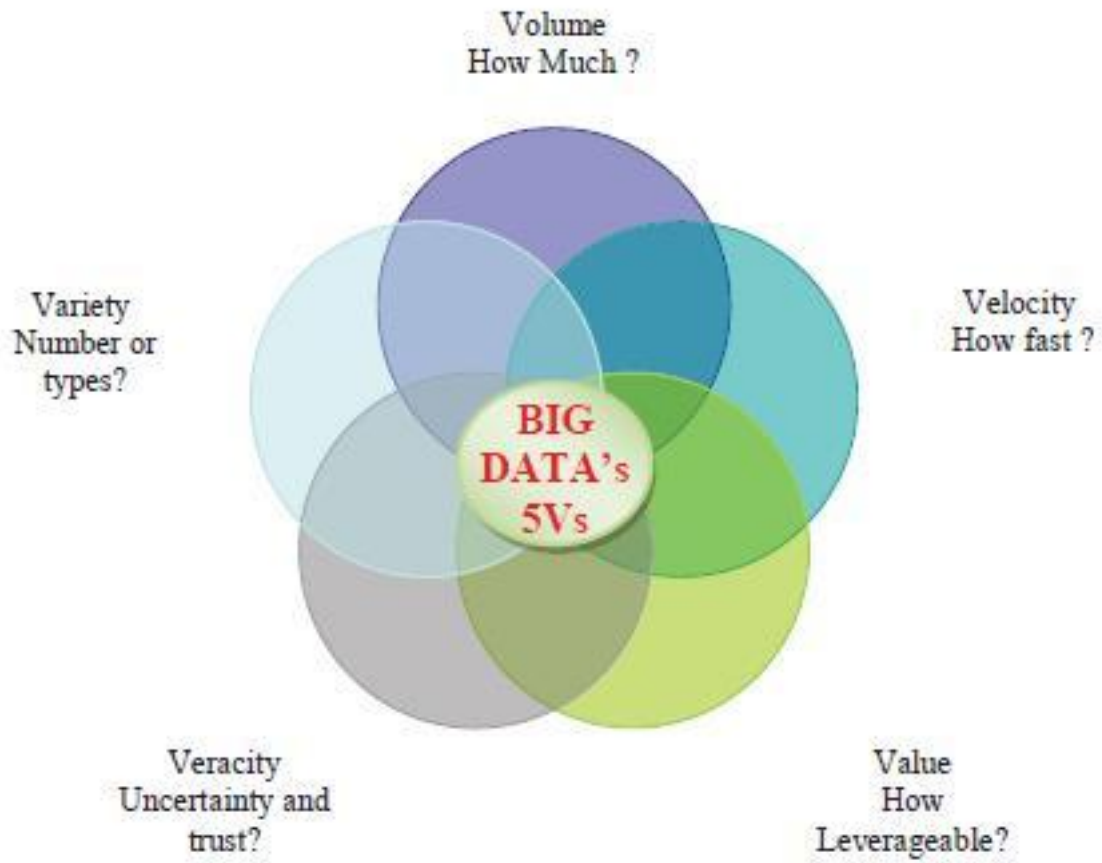
Neki praktičari i istraživači uveli su četvrto svojstvo, istinitost ili "sigurnost podataka". To jest, veliki podaci, analitika i rezultati su bez pogrešaka i vjerodostojni. Naravno, istinitost je cilj, a ne (još) stvarnost. Pitanja o kvaliteti podataka su od značajne važnosti u zdravstvu iz dva razloga: odluke o životu ili smrti ovise o tome da imaju točne podatke, a kvaliteta podataka o zdravstvenoj zaštiti, naročito nestrukturiranih podataka, vrlo je promjenjiva i prečesto je netočna (neodgovarajući "prijevodi" siromašnih rukopisa na recepte možda su najočitiji primjer).

8.5. Value (vrijednost)

Uvijek nas zanima prikupljanje i vađenje maksimalne stvarne vrijednosti iz velikih podataka s kojima radimo. Zapravo, moramo tražiti ulaganja u pohranu podataka jer vrijednost podataka ovisi o kvaliteti strategije i mehanizma upravljanja. Na primjer, pohranjivanje kliničkih podataka novih bolesti na nepouzdanu pohranu može značiti uštedu novčanih sredstava danas, no sutra može utjecati na podatke. Važna stavka koja se često zanemaruje jest ta da istinska vrijednost leži u očima kupaca poslovnih podataka. Također važna stavka za razmatranje je da neki podaci u vremenu prikupljanja imaju različite vrijednosti od rizika, ali ovaj rizik se može razviti na vrijeme.¹⁸

¹⁸Vibha Ganjir, Dr. B.K. Sarkar, Ravi Ranjan Kumar: Big Data Analytics for Healthcare, International Journal of Research in Engineering, Technology and Science, Volume VI, Special Issue, July 2016, str.2

Slika 9. 5V tehnologije velikih podataka



Izvor: Vibha Ganjir, Dr. B.K. Sarkar, Ravi Ranjan Kumar: Big Data Analytics for Healthcare, International Journal of Research in Engineering, Technology and Science, Volume VI, Special Issue, July 2016

9. OGRANIČENJA I IZAZOVI PODATKOVNE ANALIZE U ZDRAVSTVENOM SUSTAVU

Platforma za analizu velikih podataka u zdravstvu mora podržavati ključne funkcije potrebne za obradu podataka. Kriteriji za procjenu platforme mogu uključivati dostupnost, kontinuitet, jednostavnost korištenja, skalabilnost, sposobnost manipuliranja na različitim razinama granularnosti, omogućavanje privatnosti i sigurnosti, te osiguranje kvalitete. Ovo su neki od ograničenja, te ujedno i izazova kada govorimo o analizi podataka u zdravstvu:

1. Izvori podataka organizacija (bolnica, ljekarne, tvrtke, medicinski centri) su u različitim formatima. Ove organizacije imaju podatke u različitim sustavima i postavkama. Da biste koristili ovu ogromnu količinu podataka, te organizacije moraju imati zajednički skladište podataka kako bi dobili homogene informacije i mogli istima upravljati. Međutim, takvi sustavi zahtijevaju ogromne troškove.
2. Kvaliteta podataka je ozbiljno ograničenje. Prikupljeni podaci su, u nekim slučajevima, nestrukturirani, nepravilni i nisu standardizirani. Dakle, industrija mora primijeniti dodatne napore kako bi informacije pretvorila u korisne i smislene podatke.
3. Velika investicija je potreba organizacije kod stjecanja osoblja, resursa i za kupnju tehnologije za analizu podataka. Osim toga, tvrtke moraju uvjeriti medicinske organizacije o korištenju analizavelikih podataka.
4. Korištenje tehnologija za rudarenje podacima i analizu velikih podataka zahtijevaju visoku razinu stručnosti i znanja. Skupna je stvar za tvrtke da zaposle takve osobe.
5. Zbog ozbiljnih ograničenja u pogledu kvalitete prikupljenih podataka, nisu isključene varijacije i pogreške u rezultatima.¹⁹

¹⁹Vibha Ganjir, Dr. B.K. Sarkar, Ravi Ranjan Kumar: Big Data Analytics for Healthcare, International Journal of Research in Engineering, Technology and Science, Volume VI, Special Issue, July 2016, str.5

10. ZAKLJUČAK

Nekoć su zaposlenici organizacija važne dokumente i podatke čuvali u registratorima, spremnicima i oprezno pazili na njih kako bi ih u svakom trenutku mogli ponovo upotrijebiti. Danas je to stvar prošlosti, jer zahvaljujući Big Data tehnologiji, sve podatke enormne količine možemo spremati na jednom mjestu, bez straha da ćemo neke izgubiti i da ih nećemo moći iskoristiti baš u trenutku kada nam budu odključne važnosti. Da nema Big Data-e, ljudi bi i dalje čuvali podatke u spremnicima, ali ne i zadugo jer je nemoguće imati prostorijsku dimenziju u koju bi mogli stati svi podaci prošlih desetljeća. Može se reći da je nemoguće zamisliti jedan dan života bez Interneta, pa tako i bez Big Data-e. Godinama su organizacije skupljale strukturirane podatke i koristile batch obradu da unesu reprezentativne uzorke u relacijsku bazu podataka. Analiza ovakvih podataka je retrospektivna i istraživanja se vrše na skupovima podataka. Organizacije sada mogu sakupiti više podataka iz mnogo više izvora. Opcije za optimalno skladištenje i obradu podataka su se drastično proširile i tehnologije, kao što su MapReduce i in-memory computing, osiguravaju visoko optimizirane mogućnosti za različite poslovne svrhe. Analiza podataka može biti izvršena u realnom vremenu ili veoma blizu realnog vremena obrađujući cijeli skup podataka, a ne reprezentativne uzorke. Kada govorimo o Big Data-u u smislu zdravstvenog sustava, Big data analiza podataka ima mogućnost transformirati način na koji pružatelji zdravstvene zaštite koriste sofisticirane tehnologije kako bi stekli uvid u svoje kliničke i druge podatke i donijeli informirane odluke. U budućnosti ćemo vidjeti brzu, raširenu implementaciju i upotrebu Big Data analize podataka u zdravstvenoj organizaciji i zdravstvenoj industriji. Kako Big Data analitika postaje sveobuhvatnija, pitanja poput garancije privatnosti, očuvanje sigurnosti, uspostavljanje standarda i upravljanja, te kontinuirano poboljšanje alata i tehnologija postaju od sve većeg značaja. Velika analiza podataka i aplikacije u zdravstvu su u novoj fazi razvoja, ali brzi napredak u platformama i alatima može povećati njihov proces donošenja.

11.LITERATURA I IZVORI

[1] Pharmabiz, stručni magazin za lijekove i medicinske proizvode. Broj magazina 17., izdanje magazina: studeni, 2015. Str.26-27.,

Akademski članci:

[2] Ashwin Belle,Raghuram Thiagarajan,S. M. Reza Soroushmehr,Fatemeh Navidi,Daniel A. Beard,and Kayvan Najarian: Big Data Analytics in Healthcare, Hindawi Publishing Corporation, BioMed Research International Volume 2015, Article ID 370194

[3] David W. Bates, Suchi Saria, Lucila Ohno-Machado, Anand Shah and Gabriel Escobar: Big Data In Health Care: Using Analytics To Identify And Manage High-Risk And High-Cost Patients, Health Affairs, 33, no.7 (2014):1123-1131

[4] Jinchuan Chen, Yueguo Chen, Xiaoyong Du, Cuiping Li, Jiaheng Lu: Big data Challenge: a data management perspective, Key Laboratory of Data Engineering and Knowledge Engineering, School of Information, Renmin University of China, Beijing 100872, February 22, 2013, Front. Comput. Sci., 2013, 7(2)

[5] Vibha Ganjir, Dr. B.K. Sarkar, Ravi Ranjan Kumar: Big Data Analytics for Healthcare, International Journal of Research in Engineering, Technology and Science, Volume VI, Special Issue, July 2016

[6] Wullianallur Raghupathi, Viju Raghupathi: Big data analytics in healthcare: promise and potential, Health Information Science and Systems 2014, 2:3

Internet stranice:

[7] Hadoop explained – How does Hadoop work and how to use it?, DeZyre, <https://www.dezyre.com/article/hadoop-explained-how-does-hadoop-work-and-how-to-use-it-/237>, 23.02.2018.

[8] Jake Drew, Map Reduce/Map Reduction Strategies using C#, <https://www.codeproject.com/Articles/524233/MapReduceplus-fplusMapplusReductionplusStrategies>, 24.02.2018.

[9] Margaret Rouse: Google BigTable <http://searchdatamanagement.techtarget.com/definition/Google-BigTable> , 24.02.2018

- [10] Ratko Mutavdžić: Bigdata – Izazovi implementacije, <http://www.infotrend.hr/clanak/2013/11/big-data---izazovi-implementatione.78,1029.html>, 21.02.2018.
- [11] Tavish Sarivastava, Whats is Hadoop? Simplified!, <https://www.analyticsvidhya.com/blog/2014/05/hadoop-simplified/>, 23.02.2018.