



*PRIRODNO MATEMATIČKI FAKULTET U BANJOJ LUCI  
STUDIJSKI PROGRAM MATEMATIKA I INFORMATIKA  
SMJER INFORMATIKA*

# ANALITIKA VELIKIH PODATAKA

---

Student:  
Jovana Mika

Mentor:  
doc dr. Dragan Matić

## Sadržaj

Uvod.....	2
Definicija i domen Big Data .....	3
Big Data analitika .....	5
Tehnologije specifične za Big Data.....	7
Primjena Big Data.....	11
Big Data rješenje .....	14
Zaključak .....	14
Literatura .....	16

# Uvod

Razvoj informacionih tehnologija i interneta unaprijedio je generisanje velikog obima podataka i informacije kao što su: sadržaj na društvenim medijima (Twitter, Facebook, You Tube itd. ), državna administracija, poslovne aktivnosti ( industrija, bankarstvo.. ) , namjenska industrija, internet prodaja, mobilni telefoni itd.

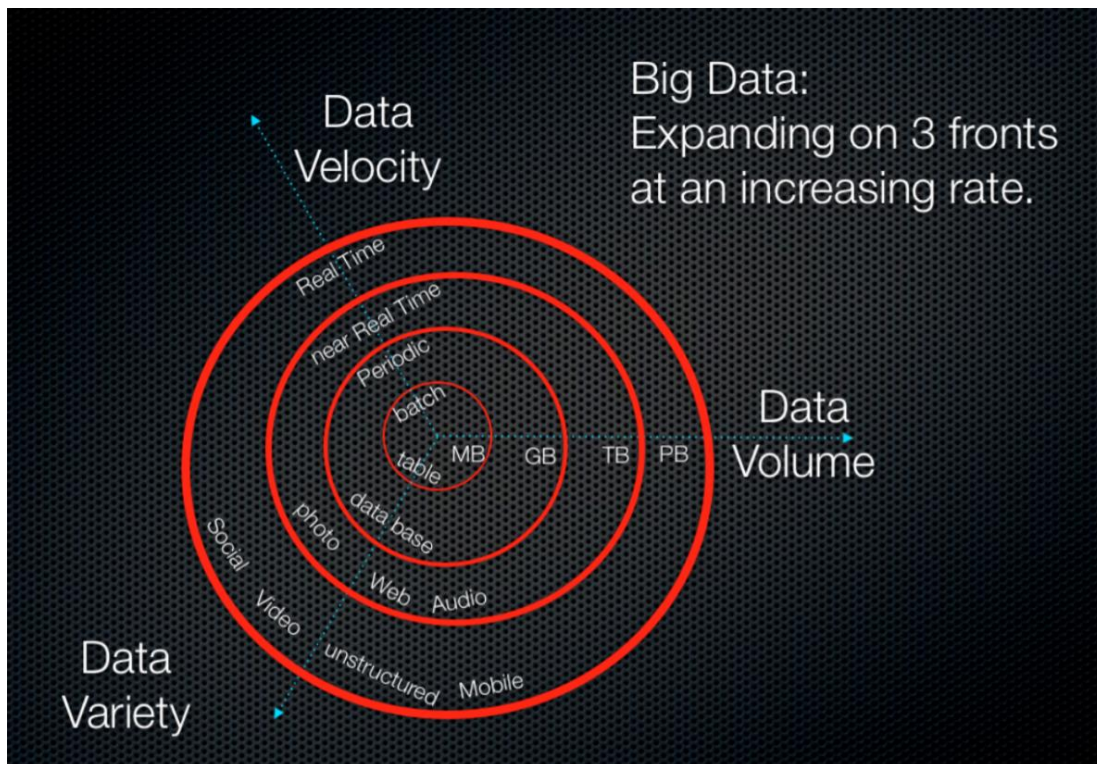
Analizom velikih količina podataka (eng. Big Data) omogućava se predviđanje ishoda događaja, procesa, posebno ljudskog ponašanja. Time se pospješuje proces donošenja odluka, otkrivaju se prevare i prijetnje, pronalaze se novi izvori prihoda, pronalaze se novi patentni, usavršavaju procesi i operacije, generišu nove informacije.

# Definicija i domen Big Data

**Big Data** (srp. veliki podaci) predstavlja sistem zasnovan na određenoj informacionoj tehnologiji. Odnosi se na obimne strukturisane ili nestrukturisane podatke. Označava velike skupove podataka čiji volume prevazilazi tradicionalne načine skladištenja.

Karakteristike velikih podataka:

- Obim (engl. volume)
- Brzina optičaja (engl. velocity)
- Raznovrsnost (engl. variety)
- Složenost (engl. complexity)
- Vjerovatnost (engl. probability)
- Osjetljivost (engl. sensibility)
- Kvalitet (engl. quality)
- Vrijednost (engl. value)



Slika 1.1 Grafički prikaz osnovnih dimenzija Big Data

Big Data je počela da transformiše većinu područja biznisa, razne inustrije i druge dijelove našeg života. Ne odnosi se samo na što veće količine podataka koje možemo analizirati nego i na brzinu kojom se kreiraju i koriste, a takođe i podrazumijeva drugačije strukture podataka koje se mogu analizirati.

### VELIČINA

Veličina je jedna od komponenti Big Data. Podaci se generišu kao nikad do sad. Ako bi se uzeli svi podaci koji su nastali u svijetu od početka mjerenja 2000. godine , možemo zaključiti da se ista količina podataka sada generiše svakih nekoliko minuta. Više od 90% podataka na svijetu je nastalo u posljednjih nekoliko godina.

### BRZINA

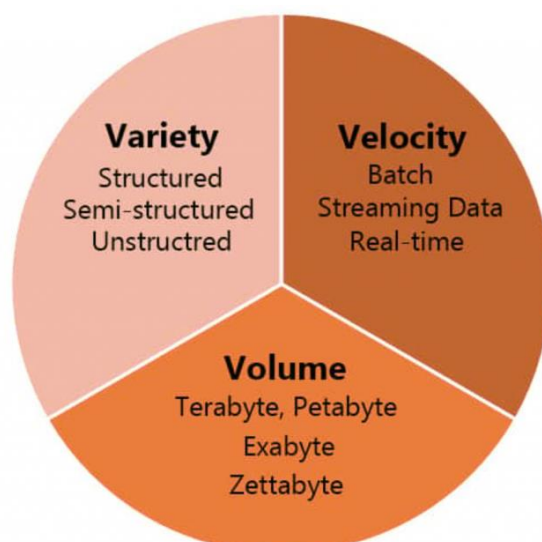
Brzina je takođe jedna od bitnijih komponenti Big Data. Svakog minuta pošalje se preko 200 miliona e-mailova, klikne se skoro 2 miliona lajkova na Facebooku, pošalje se preko 300 hiljada poruka na nekoj društvenoj mreži i slično.

Takođe samo preko Google-a se pretraži oko 3.5 milijardi pretraga dnevno.

### STRUKTURA

Struktura je treća komponenta Big Data. U prošlosti se uglavnom oslanjalo na struktuirane podatke, tipove podataka koje smo mogli da ubacimo u tabele I organizujemo brzo. Manje struktuirani podaci kao što su tekstualni fajlovi, fotografije, video sadržaji i slično, većinom su ignorisani.

Danas, imamo mogućnost da koristimo i analiziramo veliki izbor podataka uključujući i pisani tekst, izgovorene riječi, čak i ton u našem glasu, kao i biometrijske podatke i video sadržaj.



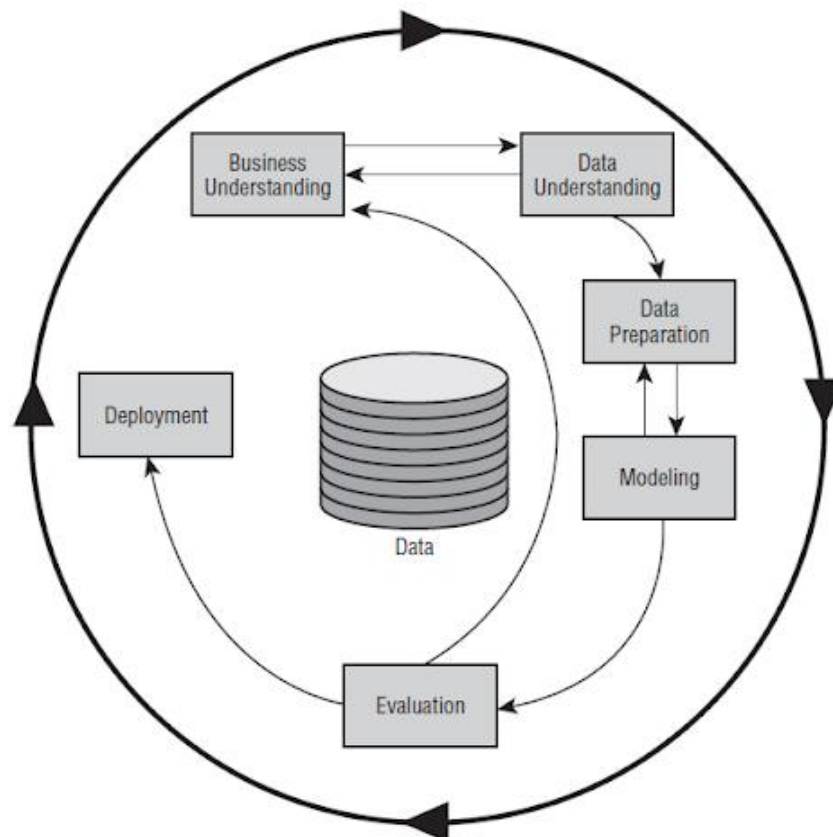
Slika 1.2. 3V model Big Data

# Big Data analitika

Analitika velikih baza podataka predstavlja analiziranje, predviđanje odnosno inteligentno donošenje odluka na osnovu prikupljenih podataka. Termini Big Data i Big Data analitika nisu isključivi i formiraju jednu komplementarnu cjelinu.

Analitika velikih podataka jeste postupak predviđanja ponašanja objekata ili uočavanja trendova na osnovu serija prikupljenih strukturiranih i nestruktuiranih podataka. Čini sastavni dio industrijskih operacija u mnogim privrednim granama, kao na primer: maloprodaja, veleprodaja, zdravstvo, bankarstvo, osiguranje, transport itd.

Analitika može da se odnosi na osnovne aplikacije poslovne inteligencije ili naprednije, prediktivne analitike poput onih koje koriste naučne organizacije. Među najnaprednijim tipovima analitike podataka je rudarenje podataka (engl. data mining), gdje Data analitičari procjenjuju velike skupove podataka kako bi identifikovali obrasce, odnose i trendove.



Slika 2.1. Rudarenja podataka

Analitika podataka može uključivati istraživačku analizu podataka (ili analizu numeričkih podataka koji imaju kvantifikovane promjenljive koje se mogu statistički uporediti) za razliku od kvalitativne analize podataka (koja se fokusira na nenumeričke podatke kao što su video, slike i tekst).

Postupak analiziranja analitike se konceptualno dijeli na dvije cjeline:

1. **oblikovanje/modeliranje uz analizu**
2. **interpretacija**

To su tehnike kojima se stiče iskustvo na osnovu velikih baza podataka. Među osnovnim tehnikama Big Data analitike mogu se izdvojiti:

- **Analitika teksta** – dobijanje informacija iz natpisa u društvenim medijima, e-pošti, blogovima, online forumima, poslovnim dokumentima, vijestima i dr. U upotrebi su tehnike poput algoritamske ekstrakcije podataka, sažetog sumiranja teksta, servisa za odgovore na pitanje (npr. Siri), ispitivanje stavova.
- **Audio analitika** – prikupljanje informacija iz snimljenih razgovora (npr. pozivni centri kompanija, bolničko dijagnostikovanje mentalnih poremećaja ili utvrđivanje stanja novorođenčeta na osnovu boje, visine i jačine glasa).
- **Video analitika** – generisanje informacija na osnovu video sadržaja (npr. sigurnosno nadgledanje objekata, prisluškivanje).
- **Analitika društvenih medija** – sticanje informacija na osnovu podataka objavljenim na društvenim medijima (npr. identifikovanje zajednica, društveni uticaj natpisa na pojedince).
- **Predviđanje** – donošenje odluka na osnovu prikupljenih i obrađenih podataka i iz njih izvedenih informacija upotrebom statističkih tehnika poput kretanja prosječnih vrijednosti i linearna regresija (npr. predviđanje sledeće kupovine potrošača).

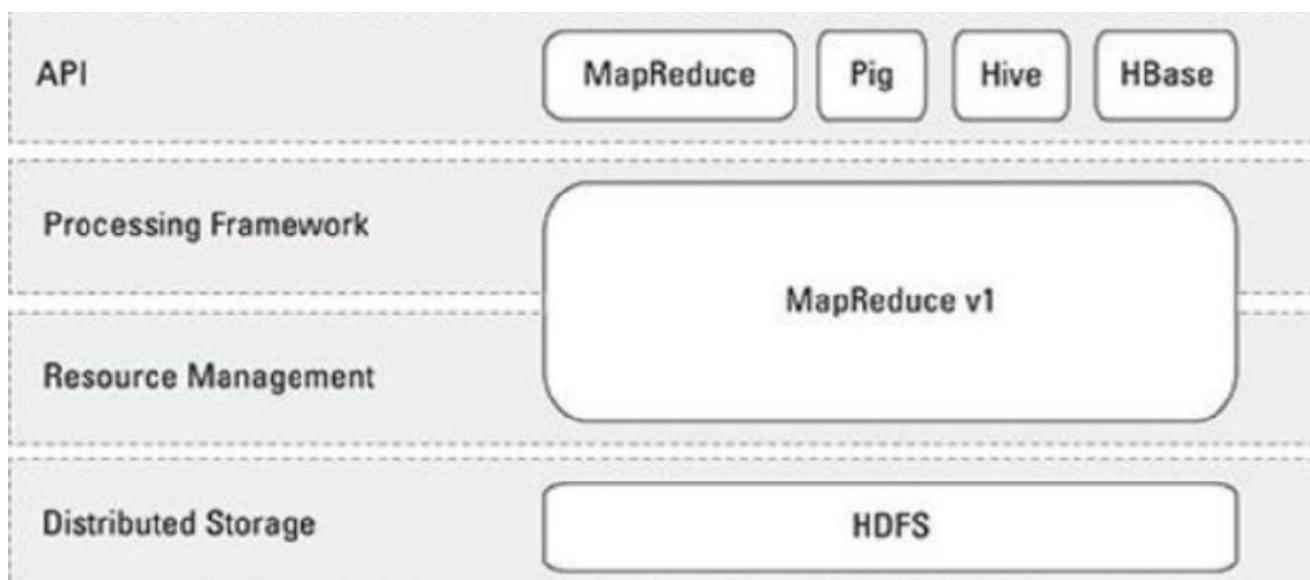
# Tehnologije specifične za Big Data

## EKOSISTEM HADOOP

Hadoop je jedna od tehnologija koja je blisko povezana sa Big Data. Projekat Apache Hadoop razvija softver otvorenog koda za skalabilno, distribuirano računarstvo.

Softverska biblioteka Hadoop predstavlja okvir koji omogućava distribuiranu obradu velikih skupova podataka preko klastera računara pomocu jednostavnih programskih modela.

Dizajniran je da skalira od jednog servera pa sve do hiljade server, od kojih svako nudi lokalno računanje i skladištenje.



Slika 3.1. Ključni elementi Hadoop platforme

Hadoop najbolje radi sa srednje velikim brojem ekstremno velikih datoteka (većih od 500MB) . Poželjan model upotrebe je *“piši jednom čitaj često”*.

Dozvoljene operacije su kreiranje nove datoteke, dodavanje na kraj datoteke, brisanje, preimenovanje.

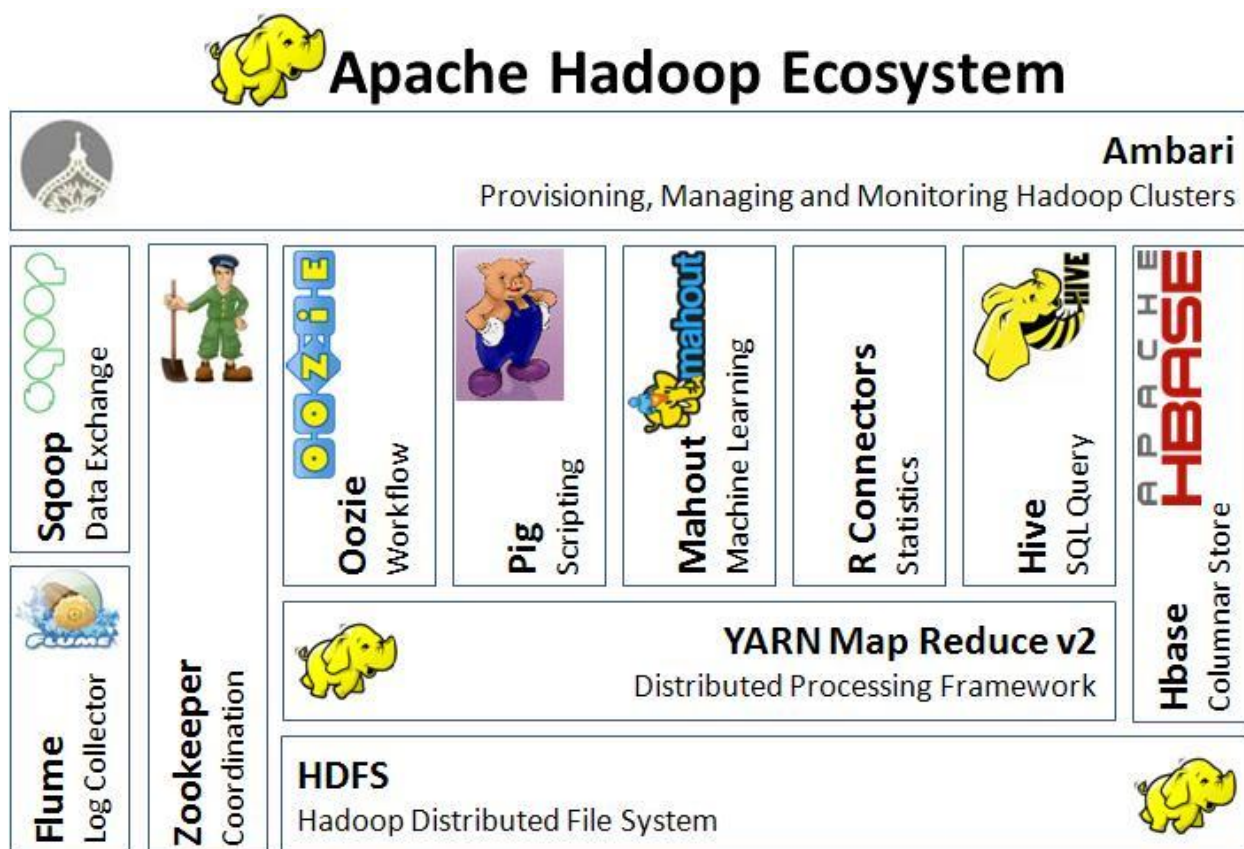


## Kako Hadoop ekosistem funkcionira?

Hadoop je ekosistema knjižica, a svaka knjižica ima svoje zadatke za izvođenje. Hadoop distribuirani datotečni sistem (HDFS) zapisuje podatke jednom poslužitelju, a zatim ih čita i ponovo upotrebljava mnogo puta. Kada ga usporedimo sa kontinuiranim višestrukim čitanjem i pisanjem akcija drugih datotečnih sistema, 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.

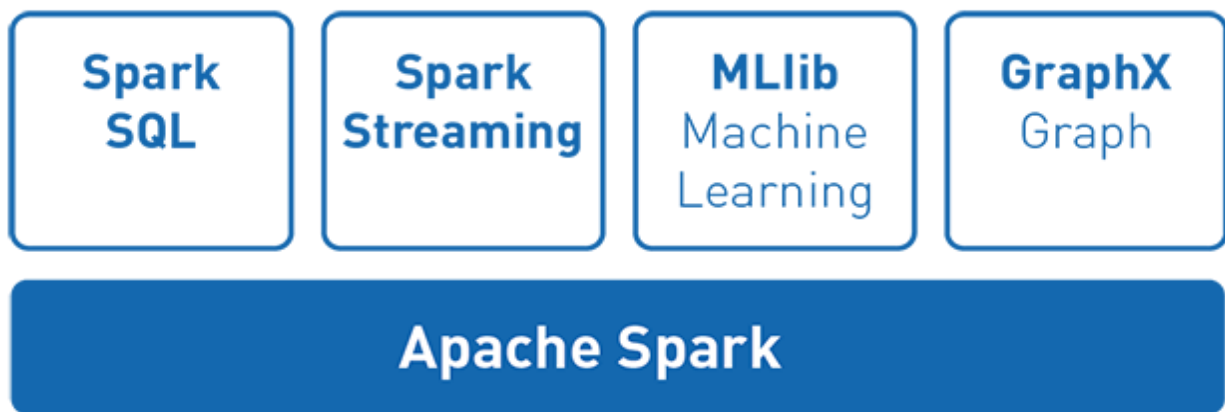
MapReduce ili YARN, koriste se za zakazivanje i obradu podataka. Hadoop MapReduce izvršava niz poslova, gdje je svaki posao Java aplikacija koja se pokreće na podacima.



Slika 3.2. Hadoop ekosistem

## APACHE SPARK

Dio ekosistema Hadoop, Apache Spark je radni okvir otvorenog koda za klastersko računarstvo koji služi kao motor za obradu Big Data u okviru Hadoop-a. Spark je postao jedno od ključnih okruženja za distribuiranu obradu podataka i može da se rasporedi na različite načine. Pruža izvorno povezivanje za Java, Scala, Python i R programske jezike, a podržava SQL, podatke u strimovanju, mašinsko obučavanje i obradu grafova.



Slika 3.2. Spark razvojni okvir

Osnovne funkcionalnosti:

- upravljanje zadacima
- upravljanje memorijom
- oporavak od otkaza
- interakcija sa sekundarnom memorijom

Gledano iz perspektive programera Spark program pokreće dva tipa programa:

1. Glavni program (engl. Driver program) koji pokreće različite paralelne operacije nad klasterom i vrši sažimanje parcijalnih računara.
2. Izvršni program (engl. workers) koji izvršavaju iste operacije nad različitim particijama podataka.

## **JEZERA PODATAKA**

Jezera podataka su skladišta koja čuvaju izuzetno velike količine sirovih podataka u izvornom formatu dok podaci ne budu potrebni kod poslovnih korisnika.

Pomoć u podsticanju rasta jezera podataka predstavljaju inicijative digitalne transformacije i rast IOT-a. Dizajnirana su tako da korisnicima olakšaju pristup velikim količinama podataka kada se javi potreba za tim.

## **NOSQL BAZE PODATAKA**

Konvencijalne SQL baze podataka su dizajnirane za pouzdane transakcije i ad hoc upite, ali imaju ograničenja kao što su rigidna šema koja ih čini manje pogodnima za neke vrste aplikacija. NoSQL baze podataka rješavaju ta ograničenja tako što čuvaju podatke i upravljaju njima na način koji omogućava veliku brzinu rada i veliku fleksibilnost. Mnoge su razvijene u kompanijama koje su tražile bolje načine za čuvanje sadržaja ili obradu podataka za velike veb stranice. Za razliku od SQL baza podataka, mnoge NoSQL baze podataka mogu da se skaliraju horizontalno preko stotina ili hiljada servera.

## **BAZE PODATAKA U MEMORIJI**

Baza podataka u memoriji (IMDB - in-memory data base) je sistem za upravljanje bazama podataka koji se za skladištenje podataka prvenstveno oslanja na glavnu memoriju, a ne na disk. Baze podataka u memoriji su brže od baza podataka optimizovanih za diskove, što je važno za analize Big Data i stvaranje skladišta podataka (engl. warehouse) i centara podataka (engl. data mart).

## **VJEŠTINE ZA BIG DATA**

Big Data i analitika za Big Data zahtjevaju specifične vještine, bilo da su unutar organizacije ili preko konsultanata. Mnoge od ovih vještina su povezane sa ključnim komponentama tehnologije Big Data, kao što su Hadoop, Spark, NoSQL baze podataka, baze podataka u memoriji i softver za analitiku. Druge su specifične za discipline kao što su nauka o podacima, rudarenju podataka, statistička i kvantitativna analiza, vizuelizacija podataka, opšte programiranje i struktura podataka i algoritmi. Takođe postoji potreba za ljudima koji imaju sveobuhvatne menadžerske sposobnosti da vode velike projekte podataka od početka do kraja (Project Management).

S obzirom na to kako su projekti analize Big Data postali uobičajeni i na nedostatak ljudi na tržištu rada sa ovakvim vrstama vještina, pronalazak iskusnih profesionalaca može biti jedan od najvećih izazova za organizacije.

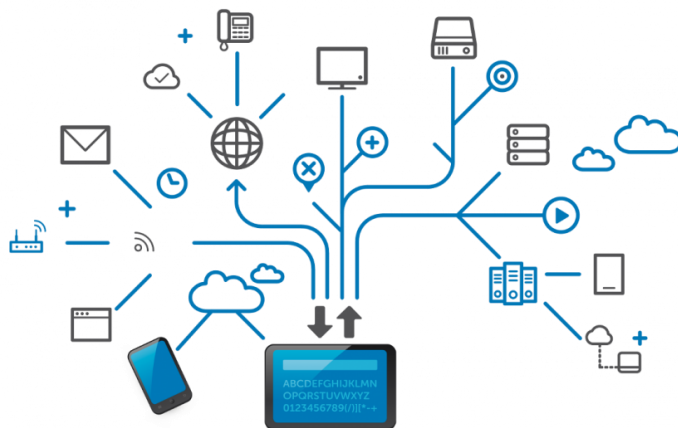
# Primjena Big Data

Uobičajeno je mišljenje da se Big Data analitika u najvećoj mjeri primjenjuje u oblasti informacione tehnologije. Međutim, činjenica je da se Big Data analitika danas primjenjuje i u drugim oblastima, i to ne samo u oblasti poslovanja, nego i u oblasti svakodnevnog života. Osnovnu primjenu ipak nalazi u oblasti informacionim tehnologijama odnosno oblasti obrade digitalnih podataka. Digitalni svijet predstavlja najdinamičniji dio poslovnog svijata današnjice. U digitalnom svijetu se svakog trenutka generišu terabajti podataka, što kroz korištenje društvenih mreža i mreža za dijeljenje sadržaja, što kroz poslovnu razmjenu podataka. Svaka kompanija kao što su npr. apoteke, svakodnevnim korištenjem interneta ostavljaju trag u digitalnom svijetu i generišu velike količine podataka koji adekvatno mogu biti obrađeni jedino upotrebom Big Data analitike.

**Primjer:** Privatna apoteka može imati naloge na svim društvenim mrežama, može imati i svoj sajt koji ima određenu posjećenost, može na dnevnoj osnovi vršiti na stotine upita na Google pretraživaču i može imati aktivnu E-mail komunikaciju. Sve ove aktivnosti generišu velike serije digitalnih podataka koji ne govore samo o načinu upotrebe informacionih resursa nego i o karakteristikama osoba sa kojima se vrši interakcija i karakteristikama samih zaposlenih.

Upotreba informacionih tehnologija i rad u digitalnom svijetu otvaraju mogućnosti analize podataka u cilju unapređenja poslovanja kompanija i svih učesnika na tržištu.

## PRIMJENA U TELEKOMUNIKACIJI



Slika 4.1. Telekomunikacija - umrežavanje

Big Data veliku primjenu nasla je i u oblasti telekomunikacije, koje se često nazivaju “mlađom sestrom” informacionih tehnologija. Pored toga što je informatički uvezan, svijet današnjice je uvezan i telekomunikacijama. Danas se ne govori o dva sektora pojedinačno nego se sve češće koristi izraz “informaciono-komunikacione tehnologije”. Najvažniji aspekt telekomunikacije su mobilne

komunikacije. Svakog dana operatori mobilnih komunikacija generišu ogromne količine podataka koji se odnose na telefonske razgovore, SMS i MMS poruke, korištenje mobilnog interneta, lokaciju korisnika mobilne telefonije i njihovo kretanje (geolokaciju), i dr.

Kao i kod informacionih tehnologija i telekomunikacije omogućavaju velikim farmaceutskim kompanijama, kao i drugim kompanijama da prikupe i analiziraju velike serije podataka vezane za kupce, potrošače, konkurente i ostale subjekte, i to samo na osnovu korištenja mobilnih komunikacija. I u ovom slučaju potrebni su kapaciteti Big Data analitike da bi se ovi podaci pretvorili u upotrebljive informacije i preporuke.

## **PRIMJENA U FINACIJSKOM SEKTORU**

S obzirom da su skoro sve finacijske transakcije danas automatizovane, podaci iz finacijskog sektora su veoma upotrebljivi u Big Data analitici. Samo u okviru jednom minuta u svijetu se obave preko 3.500 transakcija na berzama i nekoliko miliona bankarskih transakcija. Takođe svaka firma dnevno najčešće generiše na stotine finacijskih transakcija prodajom i nabavkom robe i usluga i re- lizacijom plaćnja za razne namjene. Big Data analitika omogućava da se svi podaci obrade u cjelosti, u najkraćem mogućem roku i da na osnovu njih vizuelizuje dobijene rezultate koji se mogu koristiti na dnevnoj osnovi za unapređivanje poslovanja.

## **BIG DATA U VOJNOM SEKTORU**

Širok spektar upotrebe velikih baza podataka se ogleda u sledećim vojno-organizacionim podsistemima:

- Vazduhoplovne jedinice
- Sajber- bezbjednost
- Komandovanje/kontrola
- Podsistem raketnog navođenja
- Pomorski sistem
- Podsistem radara i senzora
- Obuke personala

Američka mornarica razvija program NTCRI čiji je cilj dostupnost velikog broja kvalitetnih informacija u toku operacija na taktičnom nivou.

**Primer:** Tokom jednog celog dana misije jednostavne složenosti, bespilotna letelica dostavlja centrali 10 terabajta (TB) podataka od čega je samo 5% predmet analize dok se ostatak skladišti.<sup>37</sup> Nepostojanje uslova za analizu preostalog dela od 95% podataka minimizuje ukupan kvalitet donetih odluka na taktičkom nivou odlučivanja. Kvalitetna video veza, slanje fotografija visoke rezolucije, prenos tekstualnih sadržaja poput koordinata, očitavanja senzora itd. zajedno sačinjavaju tako veliki kapacitet transmisije. Bespilotna letelica prikuplja video-audio podatke u toku leta i pri nailasku na neprijateljske objekte ili jedinice preko video snimka dolazi do algoritamskog „isčitavanja“ čime se identifikuje vrsta naoružanja, brojnost i položaj neprijatelja, te prosleđuje predlog rešenja komandnoj jedinici na osnovu analiziranih faktora rizika.

Istovremeno, oslanjanjem na Big Data analitička rešenja povećava se sigurnost pilota u toku naleta usled poboljšanog predviđanja rizika u toku samog leta. Potrošnja goriva aviona i letelica je smanjena.

## **ULOGA ANALITIKE VELIKIH BAZA PODATAKA PRI SAJBER PRIJETNJAMA**

Digitalni format, obim, veličina i frekventnost podataka predstavljaju izvor informacije ali i izazov njihove obrade. Primjeri su podaci iz medicinskih kartona ( krvna grupa, visina, težina, boja očiju i dr. ), finansijski podaci (broj računa, poreski broj, sredstva na računu...), turistička putovanja (učestalost, dužina ), komercijalni podaci, podaci o plaćenim ili pretplatničkim računima...

Veličina baza podataka varira u opsegu od stotina do miliona unosa koji se mogu međusobno povezati preko određene varijable i time brzo steći uvid u sve podatke vezane za konkretan predmet analize. Sprečavanje kriminalnih djela, terorističkih ili subverzivnih(razornih) postupaka podrazumijeva praćenje mnoštva pojedinaca radi pronalaženja nekolicine odgo -vornih. Godišnji troškovi sajber kriminala poput krađe informacijskih dobara, onemogućavanja usluga, su procenjeni na 118 milijardi dolara godišnje.

Big Data analitika omogućava učinkovitije djelovanje vojno - obavještajnih agencija i time što se smanjuje vrijeme pretrage voluminoznih podataka, povećava brzinu procesuiranja prikupljenih nestruktuiranih podataka, čime se uočavaju skriveni šabloni i spoznaju anomalije ili slične informacije koje mogu služiti boljem i bržem odlučivanju pri kriminalnim ili subverznim aktivnostima.

Prednost je omogućena što analitičari na početku analize ne moraju znati šta je cilj analize na samom početku , već se do toga dolazi tokom postupka analize. U svom radu Horsman i Connis ukazuju na stepem anonimnosti koji uhodilac stiče korištenjem pripejd mobilnih kartica i aplikacija na telefonu. Dalje se ukazuje na značaj digitalnih forenzičara koji mogu da uđu u trag sajber progoniteljima putem korištenja aplikacija.

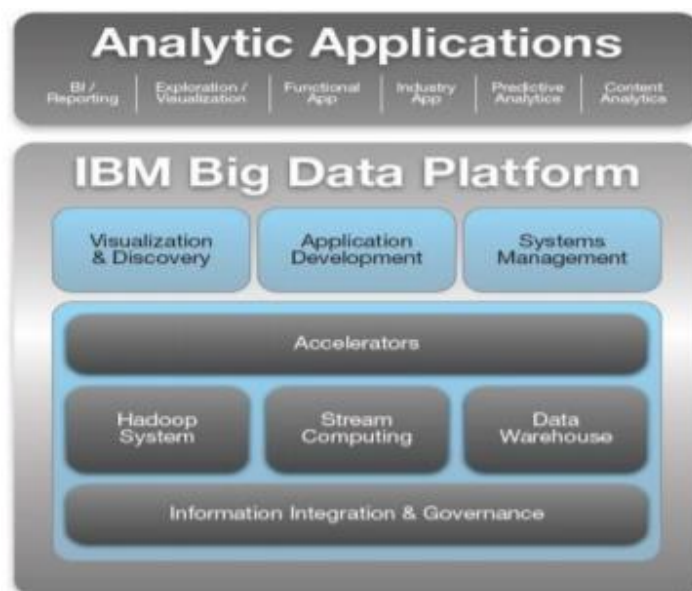
**Primer:** Njemačka Obavještajna služba u kojoj 67% radnih mjesta čine radna mjesta za informatičare i specijaliste sajber infrastrukture.

Prednost Big Data analitike je pravovremenost javljanja neovlaštenog pristupa u privatnoj mreži korisnika putem praćenja komunikacionih kanala kompjuterskih jedinica u jedinici vremena. Kriminalni delikti poput izbjegavanja poreskog plaćanja, korištenje lažnih bankarskih računa malverzacija u penzijsko-invalidskom osiguranju se kontrolišu primenom analitike voluminoznih podataka. Neovlašćeni pristupi internet sadržajima i bazama podataka ministarstava i državnih agencija su učestali postupci kojima se narušava bezbednost zemlje. Natpisi i sadržaji na društvenim mrežama, najave u medijima država uzimaju se kao vrednosne promenljive na osnovu kojih se vrše predviđanja verovatnoća dešavanja u budućnosti poput protesta, pokreta,

sklapanja ugovora, vojnih intervencija. Oslanjanje na softver i algoritme u analitičkom radu stvara automatizaciju u obaveštajnom radu, pomeranje fokusa na pretpostavke o budućim ishodima stavlja pretpostavljanje u prvi plan analitičkog odlučivanja, tendencija prilagođavanja se odnosi na promenu metoda obaveštajnog rada shodno oblastima u kojima se definiše predmet analize.

## Big Data rješenje

Tehnologija Big Data i prednosti koje ona donosi prepoznata je od vodećih softverskih kompanija koje isporučuju komercijalni softver. Talas kojeg nosi Big Data pokušavaju da iskoriste, te su dali svoje viđenje ovog koncepta i nadogradnju na Hadoop okruženje i iskoristili su ovu ideju kako bi proširili svoje poslovanje, odnosno isporučili nova rješenja. U zavisnosti od



njihove ponude softvera zavisilo je i mjesto i način implementacije Big Data koncepta.

IBM je u oktobru 2010 objavio svoju IBM InfoSphere BigInsights platformu sa ciljem da Hadoop bude dostupan njihovim poslovnim softverskim rješenjima sa tri glavna aspekta analitike, integracije podataka i upravljanja ovim sistemima. Platforma je postavljena na način da analitika velikih podataka koegzistira sa analitikom na drugim vrstama podataka, a Hadoop klasteri nesmetano rade svoj posao paralelno sa IBM serverima. BigInsights

osigurava interoperabilnost s DB2, Netezza, IBM Smart Analytics sistemima skladištenja podataka. IBM rješenje omogućava čvrstu integraciju gdje se upiti mogu automatski distribuirati između DB2 i Hadoop po potrebi.

## Zaključak

Big Data koncept možemo posmatrati kao novi element koji se može ugraditi u različite tipove informacionih sistema. Ne može da bude zamjena za sisteme poslovne inteligencije i skladište podataka, ali može kroz postojeće sisteme da donese novu vrijednost. Ako ga posmatramo u kontekstu donošenja odluka, Big Data je dodatni gradivni element u sistemu poslovne inteligencije koji može donijeti novo znanje, a sa njim i novu vrijednost. Takođe, ubudućnosti očekuje se veliki porast količine podataka.



# Literatura

1. Velike Baze Podataka – Big Data, Primena u Vojno-bezbednosnom sistemu  
[http://www.odbrana.mod.gov.rs/odbrana-stari/vojni\\_casopisi/arhiva/VD\\_2018-3/70-2018-3-17-Milojevic.pdf](http://www.odbrana.mod.gov.rs/odbrana-stari/vojni_casopisi/arhiva/VD_2018-3/70-2018-3-17-Milojevic.pdf)
2. Big Data i poslovna inteligencija - <https://infoteh.etf.ues.rs.ba/zbornik/2014/radovi/RSS-3/RSS-3-10.pdf>
3. <https://infoteh.etf.ues.rs.ba/zbornik/2014/radovi/RSS-3/RSS-3-10.pdf>
4. Uvod u analitiku velikih podataka – Aleksandar Kartelj
5. Big Data Analytics - [https://www.sas.com/en\\_us/insights/analytics/big-data-analytics.html](https://www.sas.com/en_us/insights/analytics/big-data-analytics.html)