



BIG DATA



Univerzitet u Beogradu
Saobraćajni fakultet

Prof. dr Gordana Radivojević
Prof. dr Dražen Popović

Copyright © 2025 Radivojević

Sva autorska prava su zaštićena i prezentacije se mogu koristiti samo za nastavu studenata odseka za logistiku, Saobraćajnog fakulteta Univerziteta u Beogradu u školskoj 2024/2025. Prezentacije i njihovi delovi se ne mogu koristiti za druge svrhe bez pismene saglasnosti autora.



-

(cloud.google.com, 2024)

(www.oracle.com, 2024)



-



► Opis Big Data – BD

- Primena ICT i digitalizacija su doveli do automatizacije svih poslovnih procesa i aktivnosti.
- Sa svakom transakcijom se generiše skup podataka a transakcije se izvršavaju svakog trenutka.
- Količina podataka koja se čuva u bazama podataka je sve veća i prevazilazi mogućnosti tradicionalnih pristupa za obradu podataka.
- Kompanije poseduju velike količine podataka o svom poslovanju i nameće se pitanje:
Kako iskoristiti podatke i pretvoriti ih u korisne informacije i znanje?





- Ekstremne količine podataka koje danas nastaju stvaraju izazove u oblasti upravljanja podacima (engl. Data Management).
- Sa druge strane, velike količine podataka pružaju nove mogućnosti u pogledu njihove obrade i transformacije u korisne informacije.
- *Slepa zona* je razlika koja postoji između generisanih i obrađenih podataka.



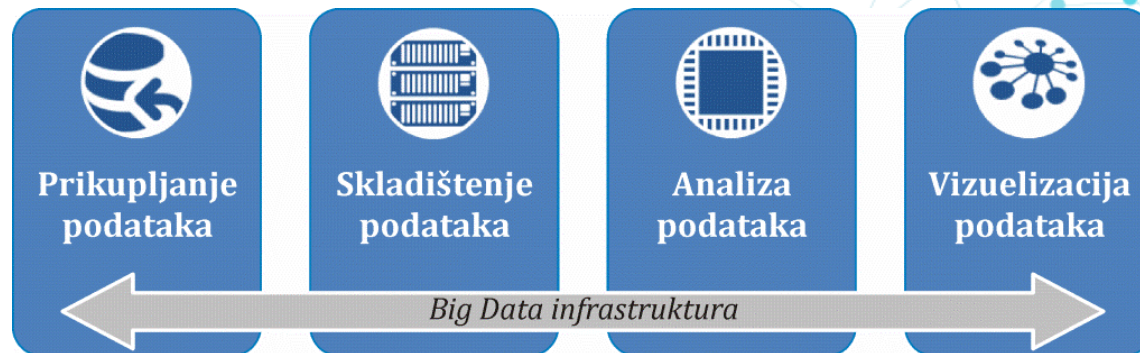
- Big Data obuhvata nov pristup i metode obrade velikih količina podataka.
- Ne postoji definisana granica koja kaže da podaci jesu ili nisu veliki.
- Danas se u većini sistema može govoriti o Big Data pristupu za obradu poslovnih podataka s obzirom na rastući trend generisanih podataka.



► BD infrastruktura

BD infrastruktura treba da mogući realizaciju četiri osnovna procesa:

- Prikupljanje podataka,
- Skladištenje podataka,
- Analiza podataka i
- Vizuelizacija podataka.



(www.forbes.com, 2016)



Prikupljanje podataka

- Podaci se prukupljaju iz internih i eksternih izvora.
- Infrastruktura treba da obezbedi prikupljanje svih podataka u sistemu.
- Uvođenje novih podataka zahteva ulaganja u infrastrukturu.
- Najveća ulaganja su hardversku i softversku podršku koja omogućava prikupljanje podataka sa IoT uređaja.

Skladištenje podataka

- Infrastruktura treba da obezbedi skladištenje svih prikupljenih podataka.
- Podaci se prikupljaju u realnom vremenu i njihov obim stalno raste.
- Opcije skladištenja podataka su baze podataka, skladišta podataka (engl. Data Warehouse – DWH), jezera podataka (engl. Data Lake), Data centar, sistemi skladištenja koji se zasnivaju na CC, itd.



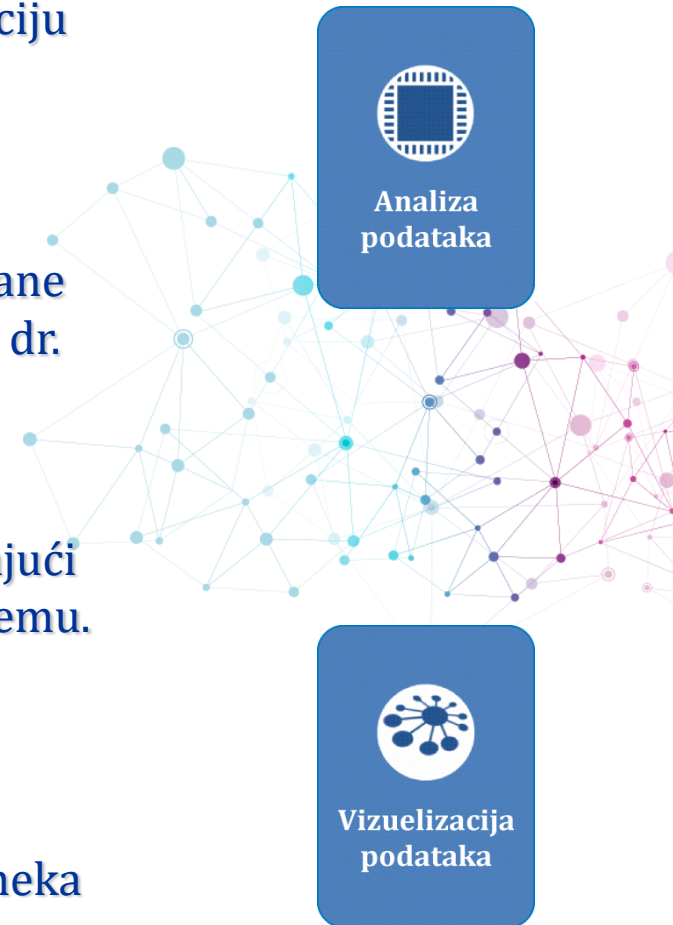


Analiza podataka

- Infrastruktura za analizu podataka obuhvata tehnologije i softverske platforme koje omogućavaju transformaciju podataka u informacije i nova znanja.
- Za analizu podataka se uglavnom koriste softverske platforme poznatih softverskih kompanija.
- Druge opcije za analizu podataka su dostupne od strane Microsoft HDInsight, AmazonWebService, Cloudera i dr.

Vizuelizacija podataka

- Rezultati dobijeni analizom podataka se na odgovarajući način moraju prezentirati donosiocima odluka u sistemu.
- Načini prezentacije mogu biti kontrolne table (engl. Dashboard), kratki izveštaji, grafikoni, numerički pokazatelji, ključne preporuke itd.
- Infrastruktura za vizuelizaciju podataka je najčešće neka do komercijalnih platformi za vizuelizaciju podataka.





► Izvori BD

Postoji devet osnovnih izvora podataka:

1. Senzori,
2. Arhive,
3. Automatski sistemi,
4. Dokumenta,
5. Mediji,
6. Baze podataka,
7. Poslovne aplikacije,
8. Internet i
9. Društvene mreže.



Neke karakteristike izvora podataka su: mesto i brzina nastanka, promena podataka, količina, raznolikost, postojanje standardnih Internet servisa ili posebnih metoda integracije i dr.

(datafloq.com, 2013)

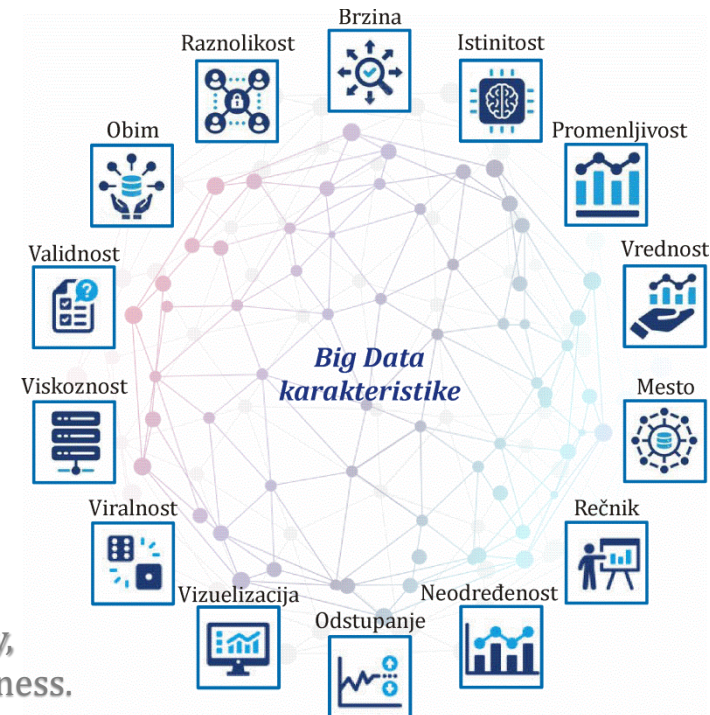


► BD karakteristike

- BD karakteristike su 2001. godine opisivane sa 3V – obim, raznolikost i brzina (engl. Volume, Variety, Velocity).
- Kasnije je skup karakteristika proširen na 4V, 5V, 7V, 10V i 14V.
- 14V karakteristika BD su:

1.	Obim	8.	Rečnik
2.	Raznolikost	9.	Neodređenost
3.	Brzina	10.	Odstupanje
4.	Istinitost	11.	Vizuelizacija
5.	Promenljivost	12.	Viralnost
6.	Vrednost	13.	Viskoznost
7.	Mesto	14.	Validnost

14 Vs – Volume, Velocity, Value, Variety, Veracity, Validity, Volatility, Visualization, Virality, Viscosity, Variability, Venue, Vocabulary, Vagueness.





- *Obim* i količina podataka su osnovna karakteristika BD, prikupljaju se u realnom vremenu iz različitih izvora i odnose se na različite objekte i procese.
- Podaci su *raznoliki* po formi, izvoru i strukturi (struktuirani, polustruktuirani i nestruktuirani).
- Podaci se generišu velikom *brzinom* u realnom vremenu.
- *Istinitost* podataka obuhvata njihov kvalitet i verodostojnost.
- *Promenljivost* podataka opisuje variranje vrednosti u skupu podataka.
- Podaci imaju *vrednost* jer omogućavaju da se obradom stvori vrednost za sistem i korisnike.
- Podaci se mogu razlikovati i po *mestu* nastanka, čuvanja i obrade.





- *Rečnik* podataka podrazumeva upotrebu zajedničke terminologije i semantike da bi se opisali i definisali modeli i strukture podataka u sistemu.
- Podaci mogu biti *neodređeni* i imati *odstupanje* od očekivanih vrednosti što zahteva dodatnu obradu i analizu.
- *Vizuelizacija* podataka podrazumeva prikazivanje podataka na način koji će omogućiti kvalitetnije i efikasnije zaključivanje i odlučivanje.
- *Virálnost* podataka opisuje pozitivan značaj i efekte obrađenih podataka kod korisnika.
- *Viskoznost* podataka je koncept koji obuhvata različite aspekte prenosa, obrade i analize podataka.
- *Validnost* je mera tačnosti i pouzdanosti koja pokazuje da li su podaci u skladu sa definisanim standardima, pravilima i ograničenjima.





► Tehnologije značajne za BD

- ICT imaju značajan uticaj na generisanje, prenos, čuvanje, obradu, analizu, prezentaciju i korišćenje podataka.
- ICT se koriste u celom životnom ciklusu podataka i imaju važnu ulogu u lancu stvaranja vrednosti iz podataka.
- Neke od tehnologija koje su povezane sa primenom BD su:
 - ❖ Internet stvari (engl. Internet of Things – IoT),
 - ❖ Data centri (engl. Data Center – DC),
 - ❖ Računarstvo u oblaku (engl. Cloud Computing – CC),
 - ❖ Apači Hadup (engl. Apache Hadoop – AH) i
 - ❖ Apači Spark (engl. Apache Spark – AS).





► BD izazovi

U oblasti prikupljanja, obrade i prezentacije velike količine podataka postoje brojni izazovi sa kojima se suočavaju poslovni sistemi.

Neki od najvažnijih izazova su:

- Skladištenje podataka,
- Obrada podataka,
- Kvalitet podataka,
- Integracija podataka,
- Izbor tehnologije,
- Analitika podataka,
- Bezbednost podataka,
- Vizuelizacija podataka,
- Definisanje politika i standarda.



(www.geeksforgeeks.org, www.simplilearn.com/challenges-of-big-data-article, 2024)



- *Skladištenje podataka* je najveći izazov s obzirom na količinu podataka koja se generiše svakoga dana. Rešenje je u korišćenju CC tehnologije i skladišta u oblaku.
- *Obrada podataka* treba da odgovara korisnicima koji zahtevaju odgovore u realnom vremenu.
- *Kvalitet podataka* podrazumeva integritet, tačnost i pouzdanost podataka.
- *Integracija podataka* obuhvata prikupljanje i organizovanje heterogenih podataka koji dolaze iz različitih izvora.
- *Izbor tehnologije* omogućava efikasniju obradu i analitiku podataka i kvalitetnije rezultate.
- *Analitika podataka* u realnom vremenu podrazumeva obradu i analizu podataka istovremeno sa procesom njihovog prikupljanja.
- *Bezbednost podataka* obuhvata sve procese od prikupljanja do čuvanja i prezentacije obrađenih podataka.
- *Vizuelizacija podataka* je izazov jer veliku količinu podataka treba u odgovarajućoj formi i na pravi način predstaviti korisniku.
- *Definisanje politika i standarda* ima za cilj zaštitu podataka koji su postali jedan od najvažnijih resursa poslovnih sistema.

(www.geeksforgeeks.org, www.simplilearn.com/challenges-of-big-data-article, 2024)





► Napredna analitika

- Napredna analitika obuhvata softverske platforme, tehnike i alate koji se koriste za analizu i tumačenje podataka.
- Napredna analitika obuhvata metodologiju koja se zasniva na statističkoj analizi, prediktivnom modeliranju, AI, ML i istraživanju podataka (engl. Data Mining – DM).
- U osnovi napredne analitike su čisti, visokokvalitetni podaci tj. BD koji imaju 14V karakteristike.

Proces BD analitike obuhvata četiri faze:

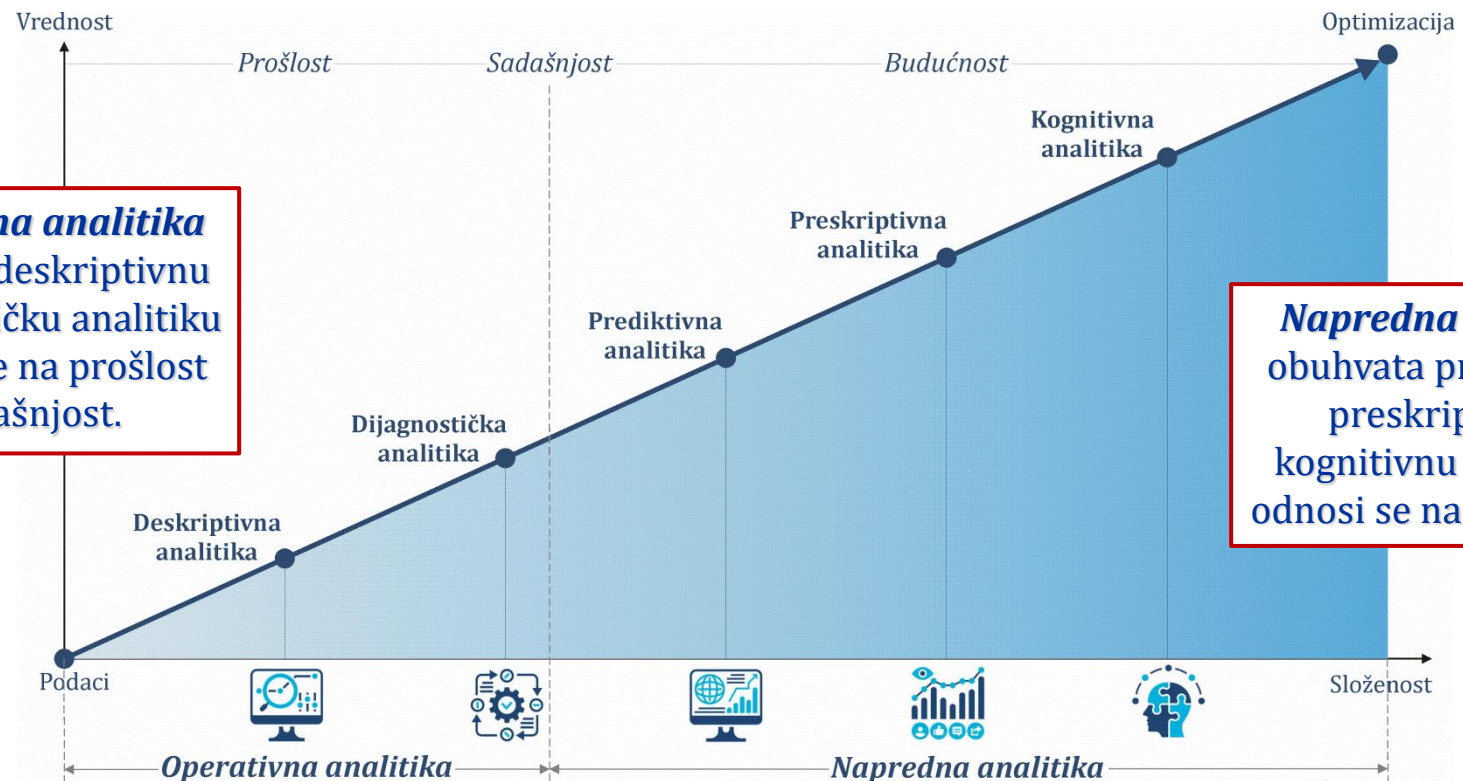
- I. Prikupljanje podataka,
- II. Prethodna obrada podataka,
- III. Čišćenje podataka i
- IV. Analiza podataka.





BD analitika se može podeliti u pet grupa:

- ❖ Deskriptivna analitika,
- ❖ Dijagnostička analitika,
- ❖ Prediktivna analitika,
- ❖ Preskriptivna analitika i
- ❖ Kognitivna analitika.





- **Deskriptivna analitika** analizira istorijske podatke, identifikuje realizovani učinak, probleme i efekte donetih odluka. Deskriptivna analitika daje odgovore na pitanja: *šta se i kada dogodilo, kakav je imalo uticaj, šta je problem.*
- **Dijagnostička analitika** se bavi uzrocima – *zašto se nešto dogodilo.* Cilj dijagnostičke analize je određivanje faktora koji su doveli do prethodnih događaja i (ne)željenih ishoda.
- **Prediktivna analitika** se zasniva na otkrivenim istorijskim obrascima kako bi se predvideli budući trendovi i događaji. Prediktivna analitika koristi statističke algoritme i procedure za prognozu budućih događaja i trendova (*šta će biti*).
- **Preskriptivna analitika** daje predloge i preporuke na osnovu prediktivne analitike i istorijskih trendova. Preskriptivna analitika daje odgovor na pitanje *šta treba da se radi.*
- **Kognitivna analitika** se zasniva na primeni različitih metoda AI (NLP, neuronske mreže) i ML za obradu nestruktuiranih podataka. Kognitivna analitika omogućava modeliranje ljudskog misaonog procesa kako bi učila iz podataka i otkrila skrivene obrasce i pravila.





► BD u logistici

- Velike količine podataka postoje o svim logističkim procesima i aktivnostima.
- Primena savremenih ICT je omogućila generisanje podataka o svakoj transakciji.
- U literaturi ima dosta primera primene BD analitike u logistici.
- Neki ciljevi BD analize u oblasti logistike su:



- ❖ Vidljivost i transparentnost,
- ❖ Predviđanje potražnje i optimizacija zaliha,
- ❖ Upravljanje skladištem i realizacija naloga,
- ❖ Analiza korisnika,
- ❖ Optimizacija ruta i efikasnost isporuke,
- ❖ Upravljanje rizikom i otpornost LS,
- ❖ Upravljanje snabdevačima i saradnja u LS i
- ❖ Kontinuirano poboljšanje i prediktivna analitika.



Vidljivost i transparentnost

BD analiza omogućava vidljivost svih aktivnosti u lancu vrednosti. Integracijom podataka iz različitih izvora kompanija u realnom vremenu prati nivo zaliha, status isporuke, obrasce potražnje itd.

Predviđanje potražnje i optimizacija zaliha

BD analitika se zasniva na internim i eksternim podacima o prodaji, tržišnim trendovima, ponašanju kupaca itd. Precizno predviđanje potražnje je preduslov optimizaciju zaliha, smanjenje troškova i ispunjavanje zahteva kupaca.

Upravljanje skladištem i realizacija naloga

BD analitika omogućava optimizaciju rasporeda u skladištu, raspodelu radnika na zadatke i efikasniju obradu naloga. To dovodi do smanjenja troškova, poboljšanja tačnosti komisioniranja, povećanja zadovoljstva korisnika itd.

Analiza korisnika

BD analitika se zasniva na različitim podacima o korisnicima – prethodne kupovine, demografski podaci, tržišni uslovi i konkurencija, evidentirani zahtevi korisnika. Rezultati BD analitike obezbeđuju planiranje proizvodnje, nabavke i isporuke kako bi se odgovorilo očekivanjima kupaca.



Optimizacija ruta i efikasnost isporuke

BD analitika omogućava bolje projektovanje ruta vozila u skladu sa uslovima na transportnoj mreži, vremenskim obrascima i prethodno realizovanim rutama. Primena naprednih algoritama obezbeđuje najefikasnije rute, smanjenje potrošnje goriva i kašnjenja, i povećanje zadovoljstva korisnika.

Upravljanje snabdevačima i saradnja u LS

Analiza podataka o učesnicima u LS olakšava njihovu saradnju i koordinaciju. Obradom podataka u realnom vremenu mogu se identifikovati potencijalni problemi i preduzeti proaktivne akcije.

Upravljanje rizikom i otpornost LS

BD analitika obezbeđuje identifikaciju potencijalnih rizika, proaktivno reagovanje, razvoj različitih scenarija ako se rizik realizuje. Analiza različitih podataka u realnom vremenu ukazuje na potencijalne poremećaje i preventivne akcije.

Kontinuirano poboljšanje i prediktivna analitika

BD analitika omogućava otkrivanje novih informacija, obrazaca u podacima, trendova i korelacija koji ranije nisu bili poznati. Proaktivni pristup obezbeđuje da se odlučuje na osnovu podataka, primenjuju strategije prediktivnog održavanja i optimizuju procesi i aktivnosti.



Pitanja...

