

MOGUĆNOSTI UNAPREĐENJA POSLOVANJA PROVAJDERA SERVISA PRIMENOM KONCEPTA BIG DATA

Marko Đogatović, Vesna Radonjić Đogatović
Univerzitet u Beogradu - Saobraćajni fakultet
m.djogatovic@sf.bg.ac.rs, v.radonjic@sf.bg.ac.rs

Rezime: *Big data tehnologije i servisi omogućavaju rešavanje novih izazova sa kojima se suočavaju provajderi servisa, pre svega u pogledu obima, heterogenosti podataka, kao i pravovremenosti i privatnosti podataka. Big data analitikom provajderi servisa mogu iskoristiti puni potencijal svojih podataka pomoću korelacije, obrade i dešifrovanja skrivenih informacija. Uobičajeni alati za analitiku bez korišćenja Big data postaju neadekvatni kako se trendovi pomeraju ka distribuiranoj obradi podataka u realnom vremenu. Korišćenjem koncepta Big data provajderi servisa mogu uticati na poboljšanje poslovnih procesa, korisničkog iskustva, pouzdanosti i kvaliteta servisa, što sve utiče na unapređenje poslovanja i profitabilnosti.*

Ključne reči: *Big data, poslovna analitika, provajder, Hadoop*

1. Uvod

Intenzivan razvoj informacionih i komunikacionih tehnologija (*Information and Communications Technology, ICT*), kao i Internet tehnologija i servisa, doprineo je tome da se generišu, prenose i čuvaju velike količine podataka sa izraženim trendom rasta. Podaci se generišu od strane mnogo različitih izvora, ne samo senzora, kamera ili mrežnih uređaja, već i veb stranica, sistema elektronske pošte, društvenih mreža itd. Datoteke postaju toliko velike i složene ili pristižu tako brzo da tradicionalne metode i alati za obradu podataka postaju neadekvatni. U takvim okolnostima potrebno je izvršiti efikasnu analizu podataka u kratkom vremenskom roku i brojne heterogene podatke iskoristiti za detaljnije analize. Paradigma koja se razvija radi rešavanja ovih izazova naziva se *Big data*.

Big data omogućava prikupljanje, skladištenje i obradu velikih količina podataka različitih struktura u realnom vremenu. Postoji široki spektar delatnosti koje mogu da koriste prednosti *Big data* tehnologija, pri čemu se obrada podataka karakteriše pomoću obima, raznolikosti, brzine, kredibiliteta, poslovne vrednosti i drugih karakteristika [1]. Realizaciju koncepta *Big data* omogućavaju druge tehnologije kao što su: NoSQL (*Not only Structured Query Language*) baze podataka, *Apache Hadoop*, IoT (*Internet of Things*), *Cloud computing* servisi, mašinsko učenje i veštačka inteligencija.

U ovom radu istraživanje je usmereno na *Big data* tehnologije i tehnike za analizu podataka, kao i potencijalne primene koje kompanijama i provajderima servisa omogućavaju transformisanje poslovnih procesa i unapređeno poslovno odlučivanje. Rad

je organizovan na sledeći način. Nakon uvoda, u drugom poglavlju razmatrane su karakteristike *Big data* okruženja, kao i uloge i relacije između učesnika u realizaciji *Big data* servisa. Treće poglavlje je posvećeno značajnim tehnologijama u pogledu *Big data* infrastrukture sa naglaskom na *Hadoop* i *MapReduce*. *Big data* analitika, koja omogućava provajderima da izvode zaključke i donose odluke na osnovu podataka o poslovanju, obrađena je u četvrtom poglavlju. U petom poglavlju su obrazložene mogućnosti primene koncepta *Big data* u telekomunikacijama i opisane su faze primene ovog koncepta u opštem slučaju. U šestom poglavlju su data zaključna razmatranja.

2. *Big data* okruženje

Za razumevanje *Big data* okruženja potrebno je razjasniti uloge i potrebne aktivnosti za učesnike u obezbeđivanju i korišćenju *Big data* servisa, kao i odnose između njih. *Big data* okruženje obuhvata tri kategorije učesnika: *Big data* provajderi podataka, *Big data* provajderi servisa i *Big data* korisnici servisa.

Provajder podataka može imati dvostruku ulogu u *Big data* okruženju, odnosno može obavljati aktivnosti distributera podataka i brokera podataka [1]. Distributer podataka pruža podatke iz različitih izvora brokeru podataka, kojima može pristupiti *Big data* provajder servisa. Njegove aktivnosti uključuju:

- generisanje podataka,
- formiranje informacija o metapodatku, odnosno opis izvora podatka i relevantne atributa digitalnog objekta i
- objavljivanje informacije o metapodatku za pristup metapodatku (meta-informacija).

Broker podataka služi za povezivanje distributera podataka i *Big data* provajdera servisa. Aktivnosti koje obavlja broker podataka su:

- pružanje registra meta-informacija distributeru podataka radi objavljivanja njihovih izvora podataka,
- pronalaženje *on-line open-data* izvora podataka i registracija odgovarajućih meta-informacija,
- obezbeđivanje kataloga servisa *Big data* provajderu servisa za pretraživanje korisnih podataka.

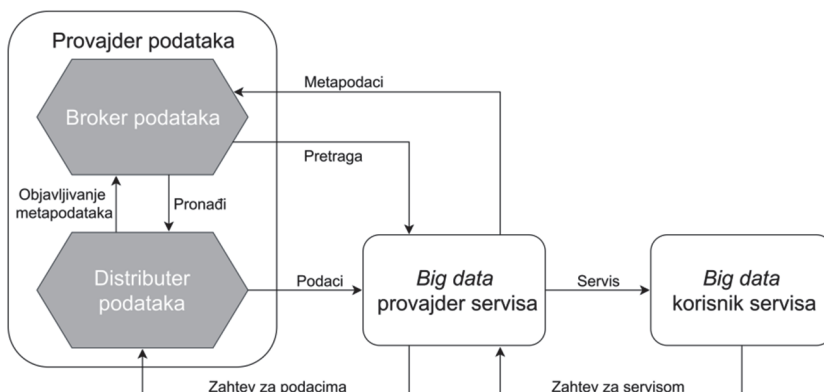
Big data provajder servisa podržava mogućnosti za analitiku i infrastrukturu *Big data* okruženja. Može delovati u formi *Big data* platforme ili proširenja postojeće platforme za analizu podataka. Aktivnosti *Big data* provajdera servisa uključuju:

- traženje izvora podataka (od brokera podataka) i prikupljanje podataka na osnovu zahteva,
- smeštanje podataka u repozitorijum podataka,
- integrisanje podataka,
- obezbeđivanje alata za analizu i vizuelno predstavljanje podataka,
- podršku za upravljanje podacima, kao što su: provera porekla podataka, privatnost podataka, bezbednost podataka, politika zadržavanja podataka, vlasništvo nad podacima itd.

Big data korisnik servisa je krajnji korisnik ili sistem koji koristi rezultate ili servise koje obezbeđuje *Big data* provajder servisa. *Big data* korisnik servisa može da generiše nove servise ili znanje o aktivnostima korisnika i da ih dostavlja izvan *Big data*

okruženja. Njegove aktivnosti uključuju: zahtevanje servisa od *Big data* provajdera servisa i korišćenje rezultata *Big data* servisa.

Na slici 1 prikazane su relacije između učesnika u *Big data* okruženju.



Slika 1. *Big data* okruženje [1]

Na tržištu *Big data* može postojati manji ili veći broj provajdera podataka i servisa i u zavisnosti od toga razlikuju se tri osnovne tržišne strukture: monopol, oligopol i tržište savršene konkurencije. Na tržištima nesavršene konkurencije, nedostatak konkurencije (u slučaju monopola), odnosno mali broj konkurenata (u slučaju oligopola) dovodi do sporijeg rasta i razvoja tržišta. Za razliku od monopolskog i oligopolskog tržišta, na tržištu savršene konkurencije provajderi podataka i servisa nemaju direktnu kontrolu nad cenama i profit koji ostvaruju zavisi od njihovih udela na tržištu, kao i od brojnih drugih tržišnih faktora [2]. Na tržištu savršene konkurencije *Big data* korisnici servisa u kratkom roku mogu ostvariti pogodnosti, poput nižih cena i boljeg kvaliteta servisa. Ipak, dugoročno, ova struktura se može transformisati u oligopol zbog toga što niže cene znače i manji profit za provajdere podataka i servisa, koji se zbog niske profitabilnosti mogu povući sa tržišta, a samim tim se smanjuje broj konkurenata.

3. *Big data* tehnologije

Značajne tehnologije u pogledu *Big data* infrastrukture su:

- Tradicionalni upravljački sistemi relacionih baza podataka (*Relational database management system*, RDBMS),
- NoSQL sistemi baza podataka,
- *Hadoop*, *MapReduce*.

Tradicionalni RDBMS sistemi su dugo bili standard u pogledu upravljanja bazama podataka. Ovakve baze podataka su organizovane po principu velike struktuiranosti prateći relacioni model podataka. Međutim, danas se RDBMS uglavnom ne smatra dovoljno skalabilnim rešenjem koje bi bilo u stanju da zadovolji potrebe *Big data*.

NoSQL predstavlja potpuno drugačiji vid baza podataka koji obezbeđuje visoke performanse i agilnu obradu velikih količina informacija. Drugim rečima, to je infrastruktura baze podataka koja je veoma dobro prilagođena zahtevima *Big data*. NoSQL baze podataka su efikasne zato što su za razliku od relacionalnih baza podataka

nestrukturirane po prirodi, umanjujući zahteve u pogledu konzistencije podataka, a povećavajući brzinu i agilnost. NoSQL se usredsređuje na koncept distribuiranih baza podataka, gde se nestrukturirani podaci mogu čuvati na više čvorova za obradu, a često i na više servera.

3.1 Hadoop

Hadoop je jedan od najznačajnijih *Apache* projekata. To je implementacija razvojnog okvira otvorenog koda za pouzdano, skalabilno, distribuirano računarstvo i skladištenje podataka. *Hadoop* predstavlja fleksibilnu i raspoloživu arhitekturu za računarstvo i obradu podataka u klaster mrežama koje se sastoje od velikog broja standardnih servera [3]. Takođe, obezbeđuje softverski razvojni okvir za distribuirano skladištenje i obradu *Big data* koristeći programski model *MapReduce*.

HDFS (*Hadoop Distributed File System*) je distribuirani, skalabilni i prenosivi fajl-sistem koji se koristi za skladištenje i obradu velikih količina podatka smeštenih na velikim klasterima servera sa standardnim komponentama (*commodity server*). HDFS je zasnovan na GFS-u (*Google File System*). Kao i ostali distribuirani fajl-sistemi, HDFS može da uskladišti velike količine podataka i da obezbedi transparentni pristup velikom broju distribuiranih korisnika. HDFS se odlikuje mogućnošću da skladišti velike datoteke na pouzdan i skalabilan način.

HDFS je namenjen za skladištenje ogromnih količina podataka, najčešće reda veličine terabajta, čak i petabajta. Ovo se postiže korišćenjem fajl-sistema struktuiranog u blokove. Individualne datoteke su podeljene na blokove fiksne veličine koje su uskladištene na klaster računara. Datoteke koje se sastoje od nekoliko blokova najčešće nemaju sve blokove uskladištene na jednom računaru [3]. HDFS obezbeđuje pouzdanost replikacijom blokova i distribuiranjem replika duž klastera. Replikacija na nivou bloka obezbeđuje raspoloživost podatka čak i u slučaju kvara računara u klasteru.

HDFS ima *master-slave* arhitekturu. Ona se sastoji od dva tipa procesa:

1. *NameNode* procesa koji sadrži metapodatke o fajl-sistemu,
2. *DataNode* procesa koji skladišti blokove od kojih su sačinjene datoteke.

NameNode i *DataNode* procesi mogu se izvršavati na jednom računaru, međutim HDFS klasteri se najčešće sastoje od posvećenog servera na kome se izvršava *NameNode* proces i nekoliko stotina ili hiljada mašina na kojima se izvršava *DataNode* proces.

NameNode je najznačajniji računar u HDFS-u, koji skladišti metapodatke za celokupni fajl-sistem:

- imena datoteka,
- dozvole za pristup datotekama i
- lokaciju svakog bloka svake datoteke.

Da bi dozvolio brzi pristup ovim podacima *NameNode* čvor skladišti celokupnu strukturu metapodataka u memoriji.

3.2 MapReduce

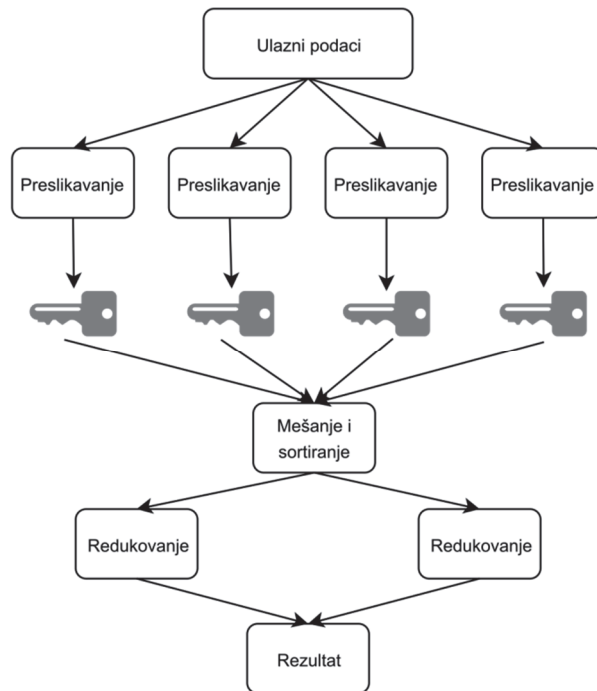
MapReduce je programski model koji obezbeđuje obradu i generisanje velikih količina podataka deljenjem poslova u nezavisne zadatke i paralelnim izvršavanjem zadataka na klasteru. *MapReduce* je inspirisan konstruktima *map* i *reduce* funkcionalnog programiranja koji se uobičajeno koriste za obradu lista podataka. Na višem nivou, svaki

MapReduce program transformiše listu elemenata ulaznih podataka u listu elemenata izlaznih podataka dva puta, jednom u *map* fazi i drugi put u *reduce* fazi [3, 4].

Razvojni okvir *MapReduce*-a sastoji se od tri glavne faze (slika 2):

- Preslikavanje (*map*)
- Mešanje i sortiranje (*shuffle and sort*) i
- Redukovanje (*reduce*)

Da bi iskoristili pogodnost paralelne obrade koju *Hadoop* obezbeđuje, neophodno je formirati upit u vidu procedure *MapReduce*.



Slika 2. Proces *MapReduce* [4]

Preslikavanje je prva faza procedure *MapReduce*. U okviru faze preslikavanja neophodno je obezbediti funkciju koja obrađuje seriju parova ključ-vrednost (*mapper*). *Mapper* sekvencijalno obrađuje pojedinačno svaki par ključ-vrednost i kao rezultat daje nijedan, jedan ili više izlaznih parova ključ-vrednost.

Druga faza procesa *MapReduce* je mešanje i sortiranje. Kada *mapper* završi sa generisanjem izlaza, međurezultati se iz faze preslikavanja prevode u fazu redukovanja. Pre nego što se to uradi vrši se mešanje i sortiranje vrednosti. Za mešanje je zadužena funkcija deljenja (*partitioner*). *Partitioner* se koristi da upravlja tokom parova ključ-vrednost od *mapper*-a ka fazi redukovanja. *Partitioner* funkciji se prosleđuje *mapper*-ov izlazni ključ i ukupni broj *reducer*-a i na osnovu te dve vrednosti on vraća indeks izabranog *reducer*-a. *Partitioner* treba da obezbedi da se svi parovi koji imaju isti ključ šalju istom *reducer*-u. To se postiže tako što se na osnovu ključa vrednost dodeljuje odgovarajućoj grupi. Vrednosti svake grupe se sortiraju pre nego što se dalje proslede u fazu redukovanja.

Redukovanje je poslednja faza procesa *MapReduce*. Unutar ove faze svaka grupa dobijena iz *partitioner*-a se prosleđuje funkciji *reducer*-a. Bitno je napomenuti da sve vrednosti unutar jedne grupe imaju isti ključ. Funkcija *reducer*-a agregira vrednosti za svaki jedinstveni ključ i stvara nijedan, jedan ili više izlaznih parova ključ-vrednost.

4. *Big data* analitika

Big data analitika je složen proces ispitivanja velikih i raznovrsnih skupova podataka radi otkrivanja informacija, poput skrivenih obrazaca (*patterns*), nepoznatih korelacija, tržišnih trendova i preferencija kupaca, a sve u cilju omogućavanja kompanijama da donose odgovarajuće poslovne odluke.

Tehnologije i tehnike za analizu podataka omogućavaju izvođenje zaključaka o podacima koji pomažu organizacijama da donose informisane poslovne odluke. Upiti softvera poslovne inteligencije (*Business Indigence*, BI) daju samo odgovore na osnovna pitanja o poslovanju i performansama. To je napredna analitika koja uključuje složene aplikacije poput: prediktivnih modela, statističkih algoritma i *what-if* analiza izvršenih na analitičkim sistemima vrhunskih performansi [4].

Vođena specijalizovanim analitičkim sistemima i softverom, kao i računarskim sistemima, *Big data* analitika omogućava prednosti u poslovanju poput:

- novih mogućnosti za prihod,
- efikasnijeg marketinga,
- bolje usluge korisnicima/kupcima,
- poboljšane operativne efikasnosti i
- konkurentske prednosti.

Aplikacije *Big data* analitike omogućavaju analitičarima i drugim stručnjacima da ispitaju rastuće obime strukturiranih podataka i druge oblike podataka koje uobičajena poslovna inteligencija i analitički programi često ne koriste. Ovo obuhvata kombinaciju polustrukturiranih i nestrukturiranih podataka. Na primer, podaci o aktivnosti korisnika na Internetu (*clickstream*), *log* podaci sa servera, sadržaji društvenih mreža, elektronske pošte korisnika i odgovori na ankete, podaci mobilnih telefona i podaci zabeleženi senzorima povezanim na IoT.

Big data analitiku je moguće podeliti na: deskriptivnu analitiku, dijagnostičku analitiku, prediktivnu analitiku i propisanu analitiku [4].

Deskriptivna analitika bavi se pitanjem „Šta se dogodilo?“ Ovaj oblik analitike uglavnom se bavi razumevanjem prikupljenih podataka. To uključuje upotrebu alata i algoritama za razumevanje unutrašnje strukture *Big data* i pronalaženje kategoričkih ili vremenskih obrazaca ili trendova u njima. Klaster analiza se najčešće koristi za davanje odgovora na pitanje postavljeno deskriptivnom analitikom.

Dijagnostička analitika daje odgovor na pitanje “Zašto se to dogodilo?”. Jednom kada se utvrdi unutrašnja struktura podataka, sledeći zadatak je tražiti razloge koji stoje iza takve strukture. Za dijagnostičku analitiku se najčešće koriste tehnike analize podataka, naročito tekstualnih, kao i različite statističke metode poput korelacije.

Imajući u vidu trenutna kretanja u podacima identifikovanih alatima deskriptivne analitike, “Šta će se možda desiti?” je presudno pitanje. Kompanije koje nisu sposobne da se prilagode budućim zahtevima svojih korisnika/kupaca ne mogu opstati na tržištu. Alat za prediktivnu analitiku pruža uvid u moguće buduće scenarije. Tehnike koje se koriste za

prediktivnu analizu su: regresija, vremenske serije, kao i različite metode mašinskog učenja.

Današnje kompanije žele da se na najbolji način pripreme za budućnost. Sa tim ciljem potrebno je da znaju koje su različite mogućnosti upravljanja poslovanjem i koja od njih je najprikladnija, obzirom na prognoze i ograničenja. Ova pitanja spadaju u domen preskriptivne analitike. U preskriptivnoj analitici se koriste sledeće tehnike: obrada slike i videa, mašinsko učenje, obrada signala, primenjena statistika, itd.

Za *Big data* analitiku razvijen je čitav niz komercijalnih i besplatnih *dekstop* i *cloud* aplikacija, kao i programskih jezika i specijalizovanih biblioteka. Neke od njih su: Xplenty, Microsoft HDInsight, Skytree, Spark, Plotly, R, IBM SPSS, itd.

5. Faze primene koncepta *Big data*, prednosti i izazovi

Big data ima potencijal da transformiše ili unapredi skoro svaku privrednu delatnost. U pojedinim implementacijama postavlja se pitanje na koji način prikupiti potrebne podatke, dok u drugim slučajevima već dostupne podatke treba integrisati u novo *Big data* okruženje. Nezavisno od konteksta primene, podaci u velikoj meri pomažu donošenju odluka zasnovanih na znanju. Stoga se *Big data* tehnologije uspešno koriste u mnogim oblastima gde su tradicionalne metode i alati neefikasni, pri čemu se obrada podataka karakteriše pomoću sledećih kriterijuma [1]:

- Obim - odnosi se na količinu prikupljenih, uskladištenih, analiziranih i vizualizovanih podataka, koje obrađuju *Big data* tehnologije;
- Raznolikost - odnosi se na različite tipove i formate podataka koji se obrađuju *Big data* tehnologijama;
- Brzina - odnosi se na brzinu prikupljanja podataka i brzinu obrade podataka *Big data* tehnologijama kako bi se ostvarili očekivani rezultati.

Takođe se mogu koristiti i drugi kriterijumi, kao što su kredibilitet, poslovna vrednost itd.

Generalno, primena *Big data* može se predstaviti kroz 5 faza [2]:

1. U prvoj fazi vrši se prikupljanje podataka. Razvoj pametnih uređaja i IoT-a omogućio je jednostavnije prikupljanje korisnih podataka, koje se sprovodi u tri koraka:
 - a) Različite vrste podataka se prikupljaju različitim metodama, dok sve neobrađene podatke skladište provajderi podataka;
 - b) Nakon prikupljanja, provajder podataka mora prethodno obraditi neobrađene podatke, ukloniti buku i sortirati različite vrste podataka u odgovarajuće grupe. Kao rezultat, nestruktuirani i struktuirani podaci biće razdvojeni za dalju obradu;
 - c) Verifikacija podataka se vrši da bi se osigurala upotrebljivost originalnih podataka.
2. U drugoj fazi vrši se analitika podataka. Nakon prikupljanja i obrade sirovih skupova podataka, analitika predstavlja najvažniju fazu za izvlačenje komercijalne vrednosti iz skupova podataka. Većina kompanija u ICT sektoru koristi bar jednu od naprednih metoda analitike.
3. U trećoj fazi vrši se formiranje cena. U ovoj fazi, provajderi podataka za svaki skup podataka određuju odgovarajuću cenu kako bi te podatke eksploatisali na digitalnim tržištima. Faktori koji utiču na cenu uključuju veličinu podataka i zahteve korisnika. Provajderi podataka mogu koristiti različite modele određivanja cena sa ciljem maksimiziranja profita.
4. Nakon što se odrede cene, u četvrtoj fazi vrši se trgovanje podacima. Trgovanje podacima se značajno razlikuje od tradicionalnog trgovanja robom. Za bezbednu i

korektnu trgovinu podacima, kao vrstom digitalnih proizvoda, i ostvarivanje zadovoljavajućeg profita, bitno je izabrati efikasan koncept za trgovanje podacima. U te svrhe se često primenjuju aukcije.

5. Poslednja faza primene *Big data*, podrazumeva zaštitu podataka, koja je neophodna radi zaštite autorskih prava i zakonskih prava svih učesnika u *Big data* okruženju.

Primena *Big data* tehnologija i servisa treba da doprinese rešavanju brojnih izazova i stvaranju mnoštva novih mogućnosti, od kojih su najznačajniji:

- Heterogenost i nepotpunost. Podaci obrađeni korišćenjem *Big data* mogu prevideti neke atribute ili uvesti šum u prenos podataka. Čak i nakon pročišćavanja podataka i korekcije grešaka, mogu ostati neke nepotpunosti i greške u podacima. To se može rešiti tokom analize podataka.
- Obim. Obrada sve većih obima podataka je veliki izazov jer resursi ne mogu da podrže toliku količinu podataka, zbog čega se koriste tehnologije kao što su paralelne baze podataka, memorijske baze podataka, No-SQL baze podataka i analitički algoritmi.
- Pravovremenost. Step en akvizicije i blagovremenost predstavljaju nove izazove sa kojima se suočava obrada podataka, radi efikasnog identifikovanja elemenata u ograničenom vremenu, koji zadovoljavaju određeni kriterijum u velikom skupu podataka. Drugi novi izazovi odnose se na vrste navedenih kriterijuma i postoji potreba za izradom novih indeksnih struktura i odgovora na upite koji imaju stroga vremenska ograničenja.
- Privatnost. Podaci o pojedincima, kao što su demografske informacije, Internet aktivnosti, obrasci razmene podataka, društvene interakcije, potrošnja energije ili vode, prikupljaju se i analiziraju u različite svrhe. *Big data* tehnologije i servisi suočavaju se sa problemom zaštite ličnih identiteta i osetljivih atributa podataka tokom čitavog ciklusa obrade podataka uz poštovanje važeće politike zadržavanja podataka.

Positivno rešavanje svih navedenih izazova otvara nove mogućnosti za otkrivanje novih odnosa podataka, skrivenih obrazaca ili nepoznatih zavisnosti.

U telekomunikacijama, zbog brze ekspanzije pametnih telefona i ostalih povezanih mobilnih uređaja, provajderi servisa moraju brzo da obrađuju, skladište i prikupljaju raznovrsne podatke koji se prenose kroz telekomunikacionih mreže. Procenjuje se da efektivnom primenom *Big data* tehnologija, telekomunikacione kompanije imaju značajan potencijal ostvarivanja većeg profita [5]. *Big data* analitika može pomoći provajderima servisa da poboljšaju profitabilnost optimizacijom mrežnih servisa, odnosno korišćenja resursa mreže, kao i unapređenjem korisničkog iskustva i bezbednosti, kao što je prikazano u tabeli 1. Primena *Big data* u telekomunikacijama takođe može obuhvatiti [6, 7]:

- optimizaciju i upravljanje kvalitetom servisa i rutiranjem na osnovu analize mrežnog saobraćaja u realnom vremenu,
- prognoziranje perioda vršnog opterećenja mreže i uskih grla i utvrđivanje koraka za ublažavanje zagušenja,
- analizu u realnom vremenu podataka o korišćenju servisa, podataka specifičnih za lokaciju i podataka o korisničkim računima, kako bi se izvršilo modeliranje uobičajenog ponašanja korisnika i sprečile zloupotrebe,
- tačnije i relevantnije informisanje korisnika u realnom vremenu (o novim servisima, tarifama, ciljanim ponudama itd.) na osnovu analiza evidencija

- podataka o korišćenju servisa i zadovoljstvu korisnika u kombinaciji sa podacima društvenih mreža da bi razumeli preferencije i ponašanje korisnika,
- prepoznavanje novih karakteristika servisa i prilagođavanje servisa zahtevima različitih tipova korisnika ili korisnika u različitim regionima, a sve na osnovu različitih podataka koji se generišu na mobilnim uređajima i
 - identifikovanje korisnika za koje postoji najveća verovatnoća nezadovoljstva i utvrđivanje koraka za sprečavanje odliva korisnika.

Tabela 1. Oblasti primene Big data od strane provajdera telekomunikacionih servisa

Stimulisanje korisničke lojalnosti, pridobijanje i zadržavanje korisnika	Unapređenje mrežnih servisa	Bezbednost
<ul style="list-style-type: none"> – Analiza odliva korisnika – Analiza (ne)zadovoljstva korisnika – Analiza aktivnosti korisnika na veb stranicama – Ciljni marketing: <ul style="list-style-type: none"> ○ profilisana segmentacija korisnika ○ preporuke – Optimizovani planovi 	<ul style="list-style-type: none"> – Optimizacija mrežnih performansi: <ul style="list-style-type: none"> ○ optimizacija rutiranja ○ analiza CDR (<i>Call Detail Record</i>) u realnom vremenu – Optimizacija mrežnih resursa na osnovu prognoziranja tražnje – Predikcija grešaka – Prilagođavanje servisa uvođenjem novih funkcionalnosti – Ostvarivanje profita na osnovu podataka vezanih za lokaciju – Proaktivno održavanje – Optimizovani planovi 	<ul style="list-style-type: none"> – Otkrivanje i sprečavanje zloupotreba: <ul style="list-style-type: none"> ○ detekcija grešaka ○ detekcija lažnog odgovora (u slučaju poziva koji se tarifira iako nije uspostavljen ili je završen, a tarifiranje se nastavlja) ○ detekcija neautorizovanih uređaja – Obrada finansijskih transakcija – Zaštita podataka – Ugovaranje

Neki od izazova sa kojima se suočavaju provajderi servisa prilikom primene koncepta *Big data* vezani su za velike kapitalne troškove, kvantifikaciju vrednosti skupova podataka, konvergenciju i integraciju sa drugim tehnologijama, tarifiranje i dizajn *Big data* aplikacija [7].

6. Zaključak

Za razliku od tradicionalnog pristupa, koji podrazumeva naknadnu obradu podataka, danas postoji sve veća potreba za obradom podataka u realnom vremenu, što je uslovalo razvoj koncepta *Big data*. *Big data* analitika je omogućila rad sa ogromnim količinama podataka. Zbog velikog obima, brzine pristizanja i raznolikosti podataka, *Big data* zahteva nove tehnike upravljanja podacima, koje se poput *Hadoop*-a, odlikuju skalabilnošću, paralelnom i distribuiranom obradom podataka, kao i ekonomičnošću.

Provajderi servisa primenom koncepta *Big data* mogu značajno unaprediti svoje poslovne procese u kontekstu poboljšanja servisa, optimizacije korišćenja resursa mreže, boljeg korisničkog iskustva i unapređene bezbednosti. U okviru budućih istraživanja trebalo bi razmotriti najznačajnije izazove sa kojima se suočavaju provajderi servisa prilikom primene koncepta *Big data*, poput kvantifikovanja vrednosti podataka, konvergenije sa drugim tehnologijama, troškova i tarifiranja.

Zahvalnica

Ovaj rad je deo istraživanja u okviru projekata TR32025 i TR44004 Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije.

Literatura

- [1] ITU-T Recommendation Y.3600, „Big data - Cloud computing based requirements and capabilities”, Geneva, 2015.
- [2] F. Liang, W. Yu, D. An, Q. Yang, X. Fu, W. Zhao, „A Survey on Big Data Market: Pricing, Trading and Protection”, *IEEE Access*, vol. 6, pp. 15132-15154, 2018.
- [3] Z. Radtka, D. Miner, *Hadoop with Python*, O'Reilly, 2016.
- [4] K. G. Srinivasa, Z. Ahmad, N. Siddiqui, A. Kumar „A Survey of Tools for Big Data Analytics”, *Big Data Analytics Tools and Technology for Effective Planning*, Eds. A. K. Somani, G. C. Deka, CRC Press, Taylor & Francis Group, pp. 75-92, 2018.
- [5] J. Bughin, „Reaping the benefits of big data in telecom”, *Journal of Big Data*, 3, 14, 2016.
- [6] F. C. Yayah, K. I. Ghauth, C. Y. Ting, „Adopting Big Data Analytics Strategy in Telecommunication Industry”, *Journal of Computer Science & Computational Mathematics*, vol. 7, no. 3, pp. 57-67, 2017.
- [7] C. M. Chen, „Use cases and challenges in telecom big data analytics”, *APSIPA Transactions on Signal and Information Processing*, vol. 5, no. 19, pp. 1-7, 2016.

Abstract: *Big data technologies and services allow addressing the new challenges for service providers and creating more opportunities than ever before, particularly in terms of data volume and heterogeneity, as well as timeliness and data privacy. Using Big data analytics, service providers can take full advantage of their data by correlating, processing, and decrypting hidden information. Conventional analytics tools without the use of Big data become inadequate as trends shift towards distributed real-time data processing. By using the Big data concept, service providers can improve business processes, customer experience, reliability and quality of service, all of which have the effect of improving business performance and profitability.*

Keywords: *Big data, business analytics, provider, Hadoop*

POSSIBILITIES OF IMPROVING A SERVICE PROVIDER'S BUSINESS BY APPLYING THE BIG DATA CONCEPT

Marko Đogatović, Vesna Radonjić Đogatović