

Big Data tehnologija u saobraćaju: Studija služaja automatskih brojača

SLAĐANA R. JANKOVIĆ, Univerzitet u Beogradu,
Saobraćajni fakultet, Beograd
SNEŽANA A. MLADENOVIĆ, Univerzitet u Beogradu,
Saobraćajni fakultet, Beograd
DUŠAN M. MLADENOVIĆ, Univerzitet u Beogradu,
Saobraćajni fakultet, Beograd
STEFAN S. ZDRAVKOVIĆ, Univerzitet u Beogradu,
Saobraćajni fakultet, Beograd
ANA R. UZELAC, Univerzitet u Beogradu,
Saobraćajni fakultet, Beograd

Stručni rad
UDC: 656.11:625.7/.8]:004.4
DOI: 10.5937/tehnika1602281J

Savremene informaciono-komunikacione tehnologije i inteligentni uređaji obezbeđuju saobraćajnim i transportnim sistemima neprekidan priliv ogromnih količina podataka. Danas u saobraćaju više nije teško prikupiti podatke, ali se nameće pitanje: kako skladištiti i obraditi sve veće količine podataka? U ovom radu istraživane su moguće nosite Big Data tehnologije na polju skladištenja i obrade podataka u oblasti saobraćaja. Pod pojmom Big Data podrazumeva se informacioni resurs velike količine, velike brzine uvećavanja i velike raznovrsnosti podataka, koji prevazilazi moguće nosite uobičajeno korišćenih softvera za skladištenje, obradu i upravljanje podacima. Realizovana je studija služaja u okviru koje su uz pomoć Apache™ Hadoop® Big Data platforme obrađivani podaci sa 10 automatskih brojača u saobraćaju postavljenih u gradu Novom Sadu i njegovoj okolini. Pokazatelji saobraćajnog opterećenja koji su izražavani na Big Data platformi tabelarno i grafički su prezentovani u Microsoft Office Excel alatu. Vizuelizacija i geolociranje dobijenih pokazatelja izvršeni su uz pomoć Microsoft Business Intelligence (BI) alata: Excel Power View i Excel Power Map. Studija služaja je pokazala da Big Data tehnologije u kombinaciji sa BI alatima mogu biti pouzdan oslonac u praćenju i upravljanju saobraćajnim sistemima.

Ključne reči: Big Data, senzori, brojanje saobraćaja, BI alati, geolociranje podataka

1. UVOD

EU je Direktivom 2010/40/EU iz 2010. godine definisala inteligentne transportne sisteme (ITS) kao napredne aplikacije koje imaju za cilj da obezbede inovativne servise za različite vidove transporta i različite metode upravljanja saobraćajem, i da omoguće različitim korisnicima da budu bolje informisani, sigurniji i da pametnije koriste transportne mreže [1]. ITS integrišu telekomunikacije, elektroniku i informacione tehnologije sa saobraćajnim inženjerstvom. Osnovni princip rada svake ITS aplikacije sastoji se u prikupljanju odgovarajućih podataka iz transportnog siste-

ma, obradi prikupljenih podataka i distribuciji informacionih podataka nadležnim za upravljanje transportnim sistemom, što dovodi do promene stanja transportnog sistema [2].

Neki od primera ITS aplikacija su: promenljiva saobraćajna signalizacija, dinamičko ograničenje brzine, adaptivna kontrola brzine kretanja vozila, inteligentno prilagođavanje brzine, sistem elektronske kontrole stabilnosti, sistem merenja brzine na određenoj deonici puta, digitalni nivo brzine, informisanje u vozilima poput „head – up display“ (HUD) [3], napredni sistemi za informisanje vozača (engl. Advanced Traveller Information System, ATIS), sistemi za voženje i informisanje korisnika o zauzetosti kapaciteta za parkiranje (engl. Parking Guidance and Information, PGI) [4], merenje osovinskog opterećenja vozila u pokretu (engl. Weigh-In-Motion, WIM) [5], i dr.

Adresa autora: Slađana Janković, Univerzitet u Beogradu, Saobraćajni fakultet, Beograd, Vojvode Stepe 305
Rad primljen: 14.03.2016.
Rad prihvaćen: 17.03.2016.

U prikupljanju podataka o saobraćajnim i transportnim sistemima, kako za potrebe aplikacija ITS, tako i u standardnim zadacima praćenja i upravljanja saobraćajem, značajnu ulogu imaju različite senzorske tehnologije [5]. One su bazirane na različitim vrstama detekcije vozila: video, bluetooth, audio, ultrazvučnoj detekciji, detekciji uz pomoć induktivnih petlji ili pneumatskih cevi, detekciji uz pomoć Radio-Frequency Identification (RFID) tehnologije, itd. Senzori generišu brojne podatke, kao što su: datum i vreme kada je vozilo registrovano, brzina vozila, dužina vozila, kategorija vozila, smer kretanja vozila, interval sleđenja vozila, rastojanje od susednog vozila, dinamika opterećenja vozila, zauzetost parking mesta, temperatura vazduha, itd. [6]. Da bi korišćenje senzorskih tehnologija dalo upotrebljive rezultate neophodno je skladištiti i u realnom vremenu obraditi ogromne količine heterogenih podataka, koji neprekidno pristižu. Poseban problem predstavlja obrada nestruktuiranih podataka, kao što su video zapisi i slike.

Prema definiciji koju je uvela kompanija IBM, za neki informacioni resurs kažemo da ima Big Data osobine ako ga karakteriše: velika količina podataka (engl. volume), velika raznovrsnost podataka (engl. Variety) i velika brzina pristizanja podataka (engl. Velocity) [7]. Očigledno je da savremeni informacioni resursi u saobraćaju imaju sve tri nabrojane osobine, što znači da se mogu okarakterisati kao Big Data.

Pojava Big Data tehnologije promenila je pogled na mnoge naučne oblasti, pa i na naučne oblasti iz domena saobraćaja [8]. Očekuje se da Big Data tehnologija u saobraćajnim sistemima omogući: naprednije praćenje ponašanja saobraćajnog sistema, prepoznavanje i analizu anomalija u njegovom funkcionisanju [9], automatsko preusmeravanje saobraćaja u slučaju predviđanja zagušenja, dinamiku promenu ograničenja brzine, otkrivanje opasnih uslova na putu, efikasniju obradu heterogenih podataka o saobraćajnim nezgodama i njihovim posledicama [10], izgradnju pametnih gradova, brže reagovanje na oštećenja na putevima, itd.

Mogućnosti i koristi ovog koncepta obrade podataka prepoznale su kompanije IBM i Volkswagen. Ove dve kompanije su nemačkom gradu Wolfsburgu, na osnovu geoprostornih podataka dobijenih iz vozila, omogućile prepoznavanje povećanog intenziteta saobraćaja, automatsko preusmeravanje saobraćaja i obaveštavanje vozača u realnom vremenu [11].

U okviru pametnih gradova cilj je svakako i implementacija sistema pametnog parkiranja. Vozači u Amsterdamu imaju priliku da putem mobilne ili web aplikacije MobyPark rezervišu svoje parking mesto u realnom vremenu [12]. Stanovnici Francuskog grada Tuluze u estvovali su u realizaciji projekta koji je

baziran na prikupljanju i analizi komentara na društvenim mrežama. Kao rezultat ovog projekta, Tuluz je ubrzao prosečno vreme sanacije i održavanja puteva za čak 93%, a vreme reagovanja na oštećenja na putu smanjeno je sa 15 dana na samo 1 dan [11].

U cilju upoznavanja sa mogućnostima Big Data tehnologije i njenih alata u domenu skladištenja i batch obrade saobraćajnih podataka, realizovali smo studiju slučaja. Na državnoj mreži puteva Republike Srbije instaliran je 391 automatski brojač saobraćaja (ABS) [6]. Mi smo prikupili podatke sa 10 ABS, za 2015. godinu, skladištili ih i obradili na jednoj od dostupnih Big Data platformi.

Obrada podataka sastojala se u izradi unavaznu pokazatelja saobraćajnog opterećenja na posmatranim brojačkim mestima. Drugi cilj studije slučaja bio je da pokaže mogućnosti jednog alata za poslovno informisanje (engl. Business Intelligence, BI) u analiziranju, vizuelizaciji i geolociranju pokazatelja saobraćajnog opterećenja. Tok studije slučaja opisan je u drugom delu rada. U ovom delu rada predstavljeni su rezultati korišćenja Big Data tehnologija u obradi podataka koje generišu ABS. Četvrti deo rada sadrži zaključak na razmatranja o mogućnostima korišćenja Big Data tehnologija i alata, u obradi i vizuelizaciji saobraćajnih podataka.

2. STUDIJA SLUČAJA

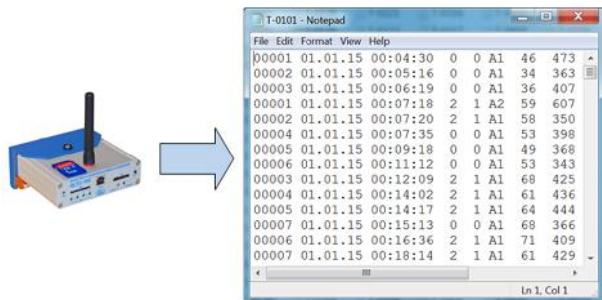
Razvoj studije slučaja tekao je u sledećim fazama:

- izbor kategorije podataka u domenu saobraćaja, koja ima osobine Big Data informacionog resursa;
- prikupljanje selektovanih podataka;
- definisanje operacija nad izabranim podacima, koje će dati upotrebljiv rezultat, a istovremeno predstavljati zahtevan test za Big Data softverske alate;
- izbor Big Data platforme i alata koji će se koristiti u istraživanju mogućnosti Big Data tehnologije u saobraćaju;
- izvršavanje definisanih operacija nad prikupljenim podacima, na odabranoj Big Data platformi;
- analiza performansi korišćenja Big Data platforme;
- analiza rezultata postignutih obradom podataka iz domena saobraćaja, uz pomoć Big Data tehnologije.

Faza 1: Za testiranje Big Data platforme izabrali smo podatke koji se dobijaju sa automatskih brojača saobraćaja, jer poseduju sve tri osobine koje neki informacioni resurs mora da ima, da bi bio okarakterisan kao Big Data. To su podaci velike količine (koja neprekidno i brzo raste), a ne spadaju u struktuirane podatke, jer se prvobitno skladište u tekstualnim datotekama.

Faza 2: U realizaciji studije služaja upotrebili smo podatke dobijene sa 10 broja a saobraćaja, postavljenih u gradu Novom Sadu i njegovoj okolini. Podaci predstavljaju rezultat brojanja saobraćaja tokom travnja 2015. godine, uz pomoć broja a serije QLTC-10C. QLTC-10C uređaji koriste se za detekciju i klasifikaciju vozila pomoću induktivnih petlji, koje se postavljaju u ulazne i izlazne trake na asfaltnom sloju kolovozne konstrukcije. Podaci koje generišu QLTC-10C broja a saobraćaja uvaju se u tekstualnim (txt) datotekama (slika 1). Za svako vozilo koje registruje broj a u txt datoteci kreira se jedan zapis. Zapis u txt fajlu sadrži sledeće podatke:

- index: dnevni redni broj vozila za saobraćajnu traku;
- datum i vreme: u formatu dd.mm.yy hh:mm:ss;
- kanal: može uzimati vrednosti: 0, 1, 2 ili 3, u zavisnosti od redosleda kojim vozilo nailazi na petlje;
- traka: može uzimati vrednosti: 0 (vozilo u saobraćajnoj traci 1) ili 1 (vozilo u saobraćajnoj traci 2);
- kategorija vozila: jedna od 11 oznaka;
- brzina vozila: izražena u [km/h];
- dužina vozila: bez faktora korekcije, izražena u [cm].



Slika 1 – QLTC-10C broja a saobraćaja i txt fajl koji on generiše

Na slici 1 prikazana je datoteka T-0101.txt, koja uvu podatke koje je generisao jedan od broja a 1.1.2015. godine. Ime txt fajla sadrži datum u tekućoj godini kada je to brojanje saobraćaja izvršeno. Svaki broj a u toku jednog dana “upisuje” podatke u jedan txt fajl, tako da se tokom godine dana, od strane svakog broja a generiše 365/366 fajlova. Veličina txt fajla određuje obim saobraćaja u jednom danu na posmatranom broja a kom mestu. U našoj studiji služaja, za svaki od 10 broja a imali smo po 365 fajlova, a svaki txt fajl uvu je izmeću u 4.000 i 14.000 zapisa. Prikupili smo više od 37 miliona zapisa, a pred nama se pojavilo pitanje: kako obraditi te zapise i iz te ogromne količine podataka izvući neke korisne informacije?

Faza 3: QLTC-10C broja a saobraćaja vrše snimanje sledećih parametara saobraćajnog toka: dnevni redni broj vozila, smer kretanja vozila, kategorija vozila,

brzina vozila, interval sledećeg izmeću u vozila, zauzetost saobraćajne trake i razvrstavaju vozila u 16 brzinskih klasa (razreda) [13]. Ovi uređaji klasifikuju vozila u 10 + 1 kategoriju: motocikli, automobili, kombi vozila, laka teretna vozila, kamioni, teška teretna vozila, kamioni sa prikolicom, teglja i s poluprikolicom, autobusi, zglobni autobusi i nekategorisana vozila. Klasifikacija je u skladu sa direktivom Evropske unije EEC 1108/70. U okviru ove studije služaja, za potrebe istraživanja mogući primene Big Data tehnologija u saobraćaju, izabrali smo da izraunavamo sledeće pokazatelje saobraćajnog opterećenja:

- PGDS (prosečan godišnji dnevni saobraćaj) ukupno za oba smera kretanja vozila,
- PGDS, za svaki smer kretanja vozila,
- PGDS, po kategorijama vozila, ukupno za oba smera kretanja vozila,
- PGDS, po kategorijama vozila, za svaki smer kretanja vozila,
- PMDS (prosečan mesečni dnevni saobraćaj) po mesecima, ukupno za oba smera,
- PMDS, po mesecima, za svaki smer,
- dnevni saobraćaj, ukupno za oba smera,
- dnevni saobraćaj, za svaki smer.

Definisali smo operacije koje je potrebno izvršiti nad prikupljenim podacima:

- izraunavanje izabranih pokazatelja saobraćajnog opterećenja, postavljanjem odgovarajućih upitnih nad prikupljenim podacima, na Big Data platformi,
- tabelarno i grafički prezentovanje izraunatih pokazatelja,
- geolociranje izraunatih pokazatelja.

Faza 4: Za skladištenje i obradu prikupljenih podataka izabrali smo open-source okruženje Apache™ Hadoop®. U izboru Big Data platforme opredelile su nas njene sledeće osobine: besplatno je dostupna, nudi upitni jezik HiveQL koji je sličan najmodernijem upitnom jeziku danas - SQL (Structured Query Language) i omogućava jednostavno povezivanje sa aplikacijom Microsoft Office Excel, kao jednim od najmodernijih savremenih alata za vizuelizaciju podataka. Za tabelarnu i grafičku prezentaciju rezultata obrade podataka, uključujući i njihovo geolociranje, izabrali smo alat Microsoft Office Excel 2013.

Faza 5: Da bismo omogućili izraunavanje pokazatelja saobraćajnog opterećenja na izabranoj Big Data platformi, neophodno je bilo da se prikupljeni podaci sa broja a saobraćaja postave na Hadoop Distributed File System (HDFS). Budući da je bilo nepraktično i neefikasno raditi sa 365x10=3650 txt fajlova, za svaki broj a saobraćaja objedinili smo sadržaj njegovih 365 txt fajlova u jedan txt fajl. Za tu namenu razvili smo

jednu Windows aplikaciju u razvojnom okruženju Microsoft Visual Studio 2015 i programskom jeziku Visual Basic.

Ova aplikacija je za svaki broj kreirala po jedan txt fajl i u njega upisala sve ispravne zapise iz 365 fajlova koje je on generisao. Na taj način smo od 3650 txt fajlova kreirali 10 velikih txt fajlova. Broj zapisa i veličina ovih 10 fajlova prikazani su u Tabeli 1. Važno je uočiti da veličina pojedinačnih txt fajlova iznosi i do 1.5 GB!

U sledećoj fazi, uz pomoć korisnog interfejsa Apache Ambari, uploadovali smo 10 velikih txt fajlova na HDFS. Zatim smo u istom okruženju, uz pomoć Apache Hive servisa i upitnog jezika HiveQL, na Hadoop platformi kreirali bazu podataka *BROJANJE SAOBRAĆAJA*. U bazi podataka *BROJANJE SAOBRAĆAJA*, za svaki txt fajl, tj. za svaki broj saobraćaja, kreirali smo po jednu tabelu.

Tabela 1. Datoteke obrađene na Big Data platformi

naziv datoteke	broj zapisa u datoteci	veličina datoteke [GB]
Alibegovac.txt	4088971	1.2
Bocke.txt	4363535	1.3
Bulevar Evrope 3.txt	2037270	0.67
Klisa.txt	3413757	0.99
Paragovo.txt	3674736	1.1
Sancevi.txt	4240577	1.2
Somborski bulevar.txt	4173994	1.4
Tekije.txt	5001035	1.5
Tunel.txt	2107966	0.65
Vojvode Stepe.txt	4784728	1.5
UKUPNO	37886569	11.5

```
CREATE TABLE ALIBEGOVAC (brojac_ID STRING, naziv_brojaca STRING, X_koordinata
STRING, Y_koordinata STRING, redni_broj_vozila STRING, datum STRING, vreme STRING,
kanal STRING, smer STRING, kategorija_vozila STRING, brzina_vozila STRING,
duzina_vozila STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
STORED AS TEXTFILE;
```

Slika 2 – Primer HiveQL upita kojim je u bazi podataka *BROJANJE SAOBRAĆAJA* kreirana tabela

Na slici 2 prikazan je jedan primer HiveQL upita kojim je kreirana jedna od tabela baze podataka – tabela ALIBEGOVAC. Tabela ALIBEGOVAC namenjena je za skladištenje svih podataka koje je generisao automatski broj saobraćaja na brojnom mestu „Alibegovac“, tokom 2015. godine. Da bismo omogućili geolokaciju izraza unatrag, svaka

od 10 tabela sadrži polja X_koordinata i Y_koordinata, koja predstavljaju GPS (Global Positioning System) koordinate broja saobraćaja. Tabele smo „napunili“ podacima iz txt fajlova koji su smešteni na HDFS.

Na slici 3 prikazan je primer HiveQL upita uz pomoć kojeg se svi podaci iz datoteke Alibegovac.txt „presipaju“ u tabelu ALIBEGOVAC.

```
LOAD DATA INPATH '/user/admin/Alibegovac.txt' OVERWRITE INTO TABLE ALIBEGOVAC
```

Slika 3 – Primer HiveQL upita kojim se podaci iz txt fajla „presipaju“ u tabelu baze podataka

Zatim smo u bazi podataka *BROJANJE SAOBRAĆAJA* kreirali tabelu *SVI_BROJAJI* u koju smo uz pomoć HiveQL upita „presuli“ sve podatke iz svih 10 tabela baze podataka. Dalje smo nad tabelom *SVI_BROJAJI* izvršavali brojne HiveQL upite čiji rezultati sadrže ranije izabrane pokazatelje saobraćajnog opterećenja. Da bismo trajno sačuvali rezultate tih upita, plasirali smo ih u nove tabele baze podataka *BROJANJE SAOBRAĆAJA*. Za ovakav način uvođenja

rezultata upita opredelili smo se jer je za izvršavanje jednog upita u proseku bilo potrebno oko 20 minuta, a osim toga, uvođenje rezultata upita u tabelama omogućava njihovo preuzimanje iz drugih alata, kao što je Microsoft Office Excel 2013.

Na slici 4, kao primer, prikazan je upit kojim je napunjena tabela *PGDS_PO_KATEGORIJA_VOZILA*, a na slici 5 upit kojim je napunjena tabela *PMDS*.

```
INSERT OVERWRITE TABLE PGDS_PO_KATEGORIJA_VOZILA
SELECT naziv_brojaca, x_koordinata, y_koordinata, smer, kategorija_vozila,
FLOOR(COUNT(*)/365) FROM svi_brojaji
GROUP BY naziv_brojaca, x_koordinata, y_koordinata, smer, kategorija_vozila;
```

Slika 4 – HiveQL upit kojim se tabela *PGDS_PO_KATEGORIJA_VOZILA* puni podacima

```
INSERT OVERWRITE TABLE PMDS
SELECT naziv_brojaca, x_koordinata, y_koordinata, smer,
CAST(SUBSTRING(TRIM(datum),4,2) AS TINYINT), FLOOR(COUNT(*)/31)
FROM svi_brojaji GROUP BY naziv_brojaca, x_koordinata, y_koordinata, smer,
CAST(SUBSTRING(TRIM(datum),4,2) AS TINYINT);
```

Slika 5 – HiveQL upit kojim se tabela *PMDS* puni podacima

Faza 6: U cilju jasnog sagledavanja eventualnih prednosti korišćenje Big Data platforme pokušali smo da Apache™ Hadoop® uporedimo sa jednim od najpopularnijih konvencionalnih relacionih sistema za upravljanje bazama podataka (RSubP) – Microsoft SQL Server-om. Iste operacije nad podacima pokušali smo da izvršimo u besplatno dostupnoj verziji ovog okruženja – Microsoft SQL Server 2012 Express. Međutim, kada smo pokušali da importujemo podatke iz 10 velikih txt fajlova, u 10 tabela SQL Server baze podataka, pojavio se problem. U SQL Server 2012 Express okruženju moguće je kreirati bazu podataka veličine do 10 GB. Naša baza sa 10 osnovnih tabela značajno prevazilazi taj limit. To znači da se u njoj svakako ne bi mogli sa uvatiti izračunati pokazatelji. Ovakvo ograničenje ne postoji u komercijalnim verzijama SQL Server-a, ali treba imati na umu da npr. RSubP SQL Server 2014 Enterprise danas košta preko 14000\$!

Faza 7: Zbog važnosti ove faze i obima poslova ko-

Tabela 2. PGDS za 2015. godinu

naziv broja a saobraćaja	GPS latitude koordinata broja a saobraćaja	GPS longitude koordinata broja a saobraćaja	smer kretanja vozila	PGDS po smerovima [vozila/dan]	PGDS [vozila/dan]
Alibegovac	45.225687	19.878489	0	5795	11202
			1	5407	
Bocke	45.221182	19.832336	1	6022	11954
			0	5932	
Bulevar Evrope 3	45.2697	19.8021	1	7192	22326
			0	15134	
Klisa	45.305863	19.825713	0	5185	9352
			1	4167	
Paragovo	45.187717	19.840307	1	10396	20134
			0	9738	
Šančevi	45.318827	19.834347	1	5670	11617
			0	5947	
Somborski bulevar	45.24012	19.81739	1	16763	22870
			0	6107	
Tekije	45.229549	19.89532	1	6785	13700
			0	6915	
Tunel	45.22837	19.860833	1	1516	5774
			0	4258	
Vojvode Stepe	45.25951	19.806591	0	6752	13108
			1	6356	

Na slici 6 prikazano je geolociranje podataka iz kolone „PGDS po smerovima“ iz tabele 2. Geolociranje je izvršeno na Microsoft Bing mapi, korišćenjem Excel-ovih dodataka Microsoft Power View i Microsoft Power Map.

Prethodni krugova kojima je na mapi predstavljen PGDS, proporcionalan je veličini PGDS-a, tako da se

na prvi pogled mogu uočiti brojačka mesta sa najvećim saobraćajnim opterećenjem.

3. REZULTATI

Izvršavanjem odgovarajućih HiveQL upita nad tabelama baze podataka BROJANJE SAOBRAĆAJA, na Big Data platformi, izračunali smo sve izabrane pokazatelje saobraćajnog opterećenja. Time je izvršena prva grupa operacija nad prikupljenim podacima, definisana u fazi 3 studije služaja. Da bismo prikazali i analizirali rezultate obrade podataka na Big Data platformi, sve rezultate HiveQL upita importovali smo u Excel 2013, uz pomoć Microsoft Query Wizard-a. Grupisanje, filtriranje, sortiranje i grafiku prezentaciju podataka obavili smo uz pomoć alata Microsoft Power View Excel 2013.

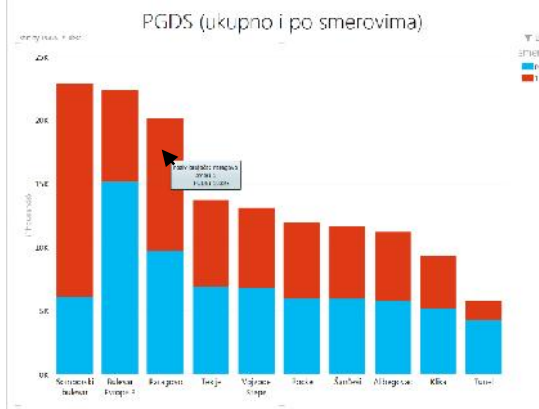
U tabeli 2 prikazane su izračunate vrednosti osnovnih pokazatelja saobraćajnog opterećenja, za svih 10 brojačkih mesta, za 2015. godinu: PGDS (ukupno) i PGDS po smerovima.

saobraćajnih tokova po smerovima: „Bulevar Evrope 3“, „Somborski Bulevar“ i „Tunel“.



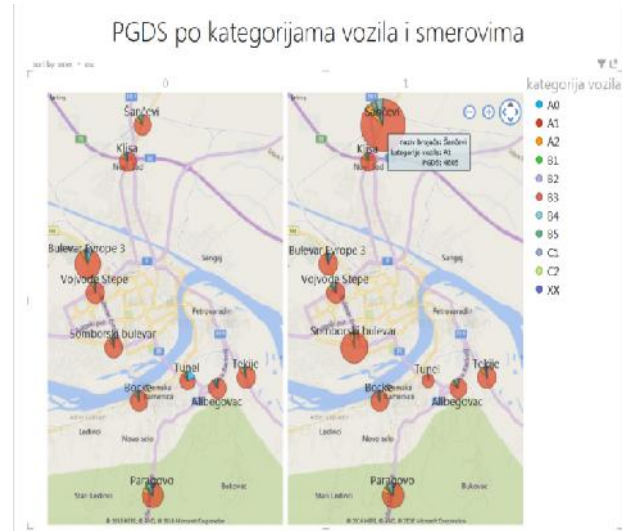
Slika 6 - PGDS po smerovima – prikaz na mapi

Na slici 7 prikazan je grafikon, kreiran u istom alatu, koji predstavlja podatke iz poslednje dve kolone tabele 2, sortirane prema vrednosti pokazatelja PGDS, u opadajućem poretku. Na grafikonu se vidi da je saobraćajno opterećenje najveće na brojačkim mestima „Bulevar Evrope 3“ i „Somborski Bulevar“, što je bilo i očekivano.



Slika 7 - PGDS ukupno i po smerovima, sortiran u opadajućem poretku

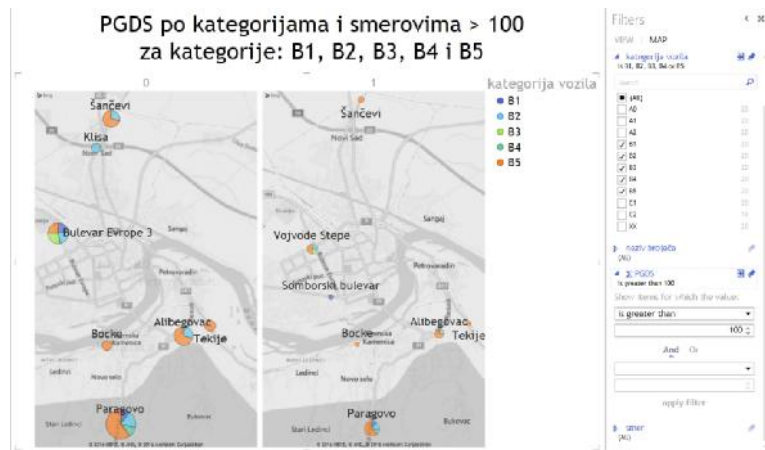
Na slici 8 prikazano je geolociranje pokazatelja PGDS po kategorijama vozila i smerovima. Kategorija vozila koja je na Slici 8, u legendi, označena kao xx, predstavlja nekategorisana vozila. Informacija koja se na prvi pogled izdvaja na Slici 8 je visok procenat vozila kategorije B4 (kamioni sa prikolicom) na brojačkom mestu „Tunel“.



Slika 8 - PGDS po smerovima i kategorijama vozila – prikaz na mapi

Na slici 9 na mapi je prikazan pokazatelj PGDS po kategorijama vozila i smerovima, ali je pri tom izvršeno i filtriranje podataka koji će biti prikazani. Filter je formiran od dva kriterijuma: kategorija vozila je iz skupa {B1, B2, B3, B4, B5}, koji predstavlja samo teretna vozila, i PGDS > 100 [vozila/dan].

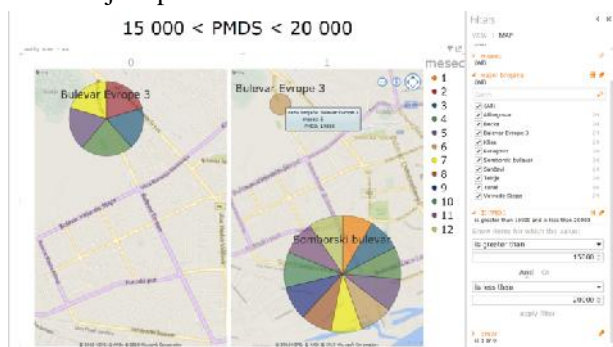
Interesantno je da se brojačko mesto „Bulevar Evrope 3“ za smer 0 javlja sa visokim intenzitetom teretnog saobraćaja, dok ga za smer 1 na mapi uopšte nema. Na prvi pogled se uočava da su na većini brojačkih mesta, prema zadatom filteru, najzastupljenija vozila iz kategorije B5 - teglja i sa poluprikolicom.



Slika 9 - PGDS po smerovima i kategorijama vozila, sa filterom – prikaz na mapi

Na slici 10 prikazan je pokazatelj *PMDS* na mapi, ali samo za brojačka mesta i mesece kada je bio veći od 15000 [vozila/dan] i manji od 20000 [vozila/dan]. Za brojačko mesto „Bulevar Evrope 3“ i smer 1, *PMDS* odgovara postavljenom kriterijumu za 5 meseci u godini, dok za smer 0 odgovara kriterijumu samo za mesec jun. Interesantno je da *PMDS* za brojačko mesto „Somborski bulevar“ i smer 1 odgovara postavljenom kriterijumu za čak 11 meseci, a za smer 0 ga uopšte nema na mapi.

Na slikama 6, 8 i 9 možemo primetiti da Excel Power Map nudi različite pozadine mapa. Na slikama 8 i 10 izabrana je pozadina Road Map Background, koja se pokazala kao najadekvatnija za geolociranje saobraćajnih podataka.



Slika 10 - *PMDS* po smerovima, sa filterom

4. ZAKLJUČAK

Praćenje i upravljanje saobraćajem danas je bazirano na prikupljanju podataka o saobraćaju u realnom vremenu. U prikupljanju podataka u estvuju različiti sistemi bazirani na: kamerama, magnetnim petljama, laserima, infracrvenim i ultrazvučnim detektorima, itd. Cilj rada bio je da pokaže da li Big Data tehnologije mogu da unaprede skladištenje, obradu i vizuelizaciju podataka u saobraćaju.

Realizovana je studija služaja u okviru koje je na Apache™ Hadoop® Big Data platformi izvršena batch obrada podataka dobijenih sa 10 automatskih brojača saobraćaja. Utvrdili smo da ova besplatno dostupna platforma omogućava skladištenje strukturiranih i nestruktuiranih podataka, kao i da raspolaže alatima za efikasnu obradu velikih količina heterogenih podataka. Studija služaja je pokazala da bi bilo opravdano na Hadoop Big Data platformi obraditi podatke sa svih 391 ABS, koliko ih je trenutno instalirano u Srbiji. Uverili smo se da ova Big Data platforma omogućava drugim alatima, kao što je Excel 2013, da na jednostavan način uvezu podatke sa nje.

Rad je obuhvatio i istraživanje mogućnosti primene Microsoft BI alata Microsoft Power View i Microsoft Power Map, za analizu, vizuelizaciju i geolociranje podataka u saobraćaju. Ustanovili smo da ovi

alati značajno olakšavaju analizu velikih količina podataka, a posebno su korisni u saobraćaju zbog mogućnosti prikazivanja izraza unatrag pokazatelja na saobraćajnim mapama.

5. ZAHVALNICA

Ovaj rad delimično je podržan od strane Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije, u okviru projekta pod brojem 036012. Podatke koje generišu automatski brojači saobraćaja obezbedila je firma MHM - projekat d.o.o. iz Novog Sada.

LITERATURA

- [1] EUR-Lex. [Internet]. European Union; 2010. [citirano 28.02.2016.]. dostupno na: <http://eur-lex.europa.eu/>
- [2] Ho W. K. Advanced Traveler Information Systems: Six Criteria to Facilitate Public Participation, University of Hong Kong, 2008.
- [3] Subotić J, Jović O, Simić M, Mitrović . Upravljanje brzinama na putevima, Tehnika, Broj 5, str. 844-852, 2014.
- [4] Simićević J, Vukanović S, Milosavljević N. Sistemi za vožnje i informisanje korisnika o zauzetosti kapaciteta za parkiranje, Tehnika, Broj 3, str. 475-486, 2014.
- [5] Todorović N, Subotić M. Mjerenje osovinskog opterećenja vozila u pokretu (WIM) weigh-in-motion, Tehnika, Broj 4, str. 669-676, 2014.
- [6] Lipovac K, Vujanić M, Ivanišević T, Rosić M. Efekti primene automatskih brojača saobraćaja u kontroli prekoračenja brzine na državnim putevima Republike Srbije, u Zborniku radova 10. Međunarodne konferencije „Bezbednost saobraćaja u lokalnoj zajednici“, Kragujevac, Srbija, str. 131-140, 22-25. april 2015.
- [7] IBM.2015.[citirano 20.02.2016.] dostupno na: www-01.ibm.com/software/au/data/bigdata/
- [8] Shi Q, Adel-Aty M. Big Data applications in real-time traffic operation and safety monitoring, Transportation Research Part C: Emerging Technologies, Vol. 58, No. B, pp. 380-394, 2015.
- [9] Hayes M. A. Contextual anomaly detection framework for big sensor data, Journal of Big Data, Vol. 2, No. 2, pp. 122, 2015.
- [10] Kumar S, Toshniwal D. A data mining framework to analyze road accident data, Journal of Big Data, Vol. 2, No. 26, pp. 1-18, 2015.
- [11] Traffic Management for a Smarter Planet [Internet]. IBM; 2016. [citirano 25.02.2016.]. dostupno na: <http://www.ibm.com>

- [12]Smart Parking. [Internet]. Amsterdamski grad; 2016. [citirano 24.02.2016.] dostupno na: <http://amsterdam-smartcity.com/projects/>
- [13]Traffic Agent System. [Internet]. Microbit d.o.o., 2009. [citirano 27.02.2016.] dostupno na: <http://www.zigns.rs/saus/Index.htm>

SUMMARY

BIG DATA TECHNOLOGY IN TRAFFIC: A CASE STUDY OF AUTOMATIC COUNTERS

Modern information and communication technologies together with intelligent devices provide a continuous inflow of large amounts of data that are used by traffic and transport systems. Collecting traffic data does not represent a challenge nowadays, but the issues remains in relation to storing and processing increasing amounts of data. In this paper we have investigated the possibilities of using Big Data technology to store and process data in the transport domain. The term Big Data refers to a large volume of information resource, its velocity and variety, far beyond the capabilities of commonly used software for storing, processing and data management. In our case study, Apache™ Hadoop® Big Data was used for processing data collected from 10 automatic traffic counters set up in Novi Sad and its surroundings. Indicators of traffic load which were calculated using the Big Data platforms were presented using tables and graphs in Microsoft Office Excel tool. The visualization and geolocation of the obtained indicators were performed using the Microsoft Business Intelligence (BI) tools such as: Excel Power View and Excel Power Map. This case study showed that Big Data technologies combined with the BI tools can be used as a reliable support in monitoring of the traffic management systems.

Key words: *Big Data, sensors, traffic counting, BI tools, geolocation of data*