

Elektronski fakultet Niš  
Sistemi za upravljanje bazama podataka

# Interna struktura i organizacija indeksa PostgreSQL baze podataka

Seminarski rad

Mentor:  
Doc. dr Aleksandar Stanimirović

Student:  
Julije Kostov 1026

# Sadržaj

1.Uvod.....	3
2.Interna organizacija PostgreSQL baze podataka.....	4
2.1 Logička struktura klastera baza .....	4
2.2 Fizička struktura klastera baza.....	6
2.3 Izgled datoteke tabele .....	8
3.Indeksi.....	9
3.1 Kako rade indeksi .....	9
3.2 PostgreSQL indeksi .....	10
3.3 Implementacija indeksa .....	11
4.Primer Korišćenja indeksa .....	15
5.Zaključak .....	22
Literatura.....	23

# 1. Uvod

Koncept baze podataka nastao je početkom 1970-ih godina kao sistem koji se koristi za organizovanje podataka u kolekciju kojoj se lako pristupa i brzo pretražuje. Baze podataka rešavaju problem upravljanja podacima, obezbeđuju pouzdanost, integritet prilikom vršenja operacija nad podacima i nezavisnost podataka od softvera.

PostgreSQL ili Postgres predstavlja jedan od najpopularnijih sistema za upravljanje bazama podataka koji je otvorenog koda. To znači da je svakome dozvoljeno korišćenje i modifikacija softvera za svoje potrebe. Osim besplatne verzije postoji i enterprise verzija obogaćena alatima koja je namenjena velikim sistemima.

Ovaj rad opisuje PostgreSQL bazu podataka i njene karakteristike. Pored toga opisuje internu strukturu sistema za upravljanje bazama podataka i internu reprezentaciju podataka.

Treće poglavlje rada opisuje indekse PostgreSQL baze podataka, prednosti i mane njihovog korišćenja, način kako oni funkcionišu i strukture podatka koje se koriste za njihovu implementaciju.

U četvrtom poglavlju rada dati su praktični primeri korišćenja indeksa kao i performanse izvršenja upita sa i bez njihovog korišćenja.

## 2. Interna organizacija PostgreSQL baze podataka

PostgreSQL je objektno-relaciona baza podataka otvorenog koda napisana na programskom jeziku C. Nastala je nadogradnjom projekta POSTGRES koji je započet još 1986 a koristio se za analizu i skladištenje različitih tipova podataka. Dolazi sa velikim brojem ugrađenih alata koji olakšavaju rad sa bazom bez obzira na količinu podataka koja se skladišti. Ima ACID svojstva i podržava veliki deo SQL standarda: kompleksno pretraživanje, strani ključevi, okidači (eng. *trigger*), pogledi (eng. *view*) koji se mogu ažurirati, integritet transakcije, itd. Pored toga omogućava korisnicima nadogradnju funkcionalnosti i bez ponovnog generisanja izvršnog fajla baze [1]. Moguće je dodati nove tipove podataka, funkcije, operatore, funkcije za agregaciju, načine za indeksiranje, itd.

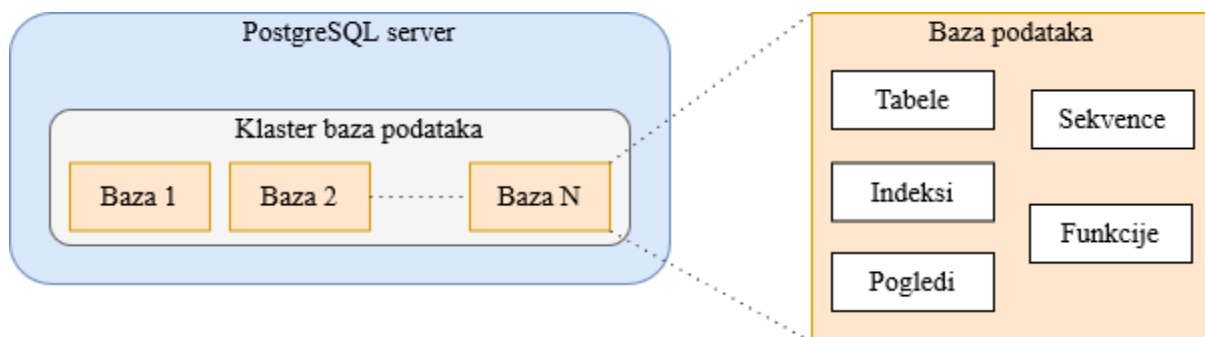
Karakteristike PostgreSQL baze podataka:

- Podrška za veliki broj različitih tipova podataka
- Integritet podataka
- Konkurentnost
- Pouzdanost i oporavak od greške
- Sigurnost
- Proširljivost

Pored toga što je Postgres relaciona baza podataka, ima i određene karakteristike objektno baze. To znači da ona ima objektno-orijentisane karakteristike kao što su: koncept nasleđivanja, mogućnost definisanja kompleksnih tipova podataka i funkcija za rad sa njima. Većina korisnika ne koristi ove funkcionalnosti ali u pojedinim slučajevima mogu da pomognu.

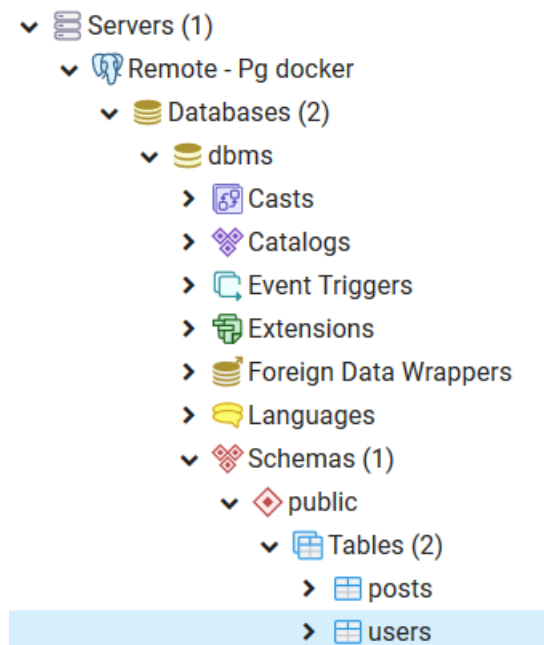
### 2.1 Logička struktura klastera baza

Klaster baza podataka je kolekcija baza podataka kojima upravlja Postgres server. Ovim terminom se ne opisuje grupa servera baza podataka već jedan server koji sadrži više baza podataka. Postgres server se izvršava na jednom hostu i upravlja jednim klasterom baza podataka [2]. Na slici 2.1 prikazana je logička struktura klastera baza podataka.



Slika 2.1 Logička struktura klastera baza podataka

Baza podataka predstavlja kolekciju objekata baze (eng. *database object*), a pod objektom se podrazumeva struktura podataka koja se koristi za čuvanje ili za referenciranje podataka. Primeri objekata su tabele, indeksi, sekvence funkcije, itd. Kod Postgres-a baze podataka kao i njihovi objekti su logički razdvojeni i svi objekti jedne baze pripadaju toj bazi. Svi objekti interno su predstavljeni pomoću identifikatora objekta **OID** koje je ceo broj od 4B. Veza između objekata i njegovog OID-a čuva se u odgovarajućem sistemskom katalogu u zavisnosti od njegovog tipa. Na primer OID-i baza se čuvaju u tabeli *pg\_database*, a ODI-i tabela se čuvaju u *pg\_class*. Na slici 2.2 prikazana je struktura klastera baza podataka na serveru, na slici 2.3 prikazano je preuzimanje informacija za bazu *dbms*, dok je na slici 2.4 prikazano preuzimanje informacije za teblu *users* koja je u okviru *dbms* baze.



Slika 2.2 Klaster baza podataka

1

```
SELECT datname, oid FROM pg_database WHERE datname = 'dbms';
```

Data Output

Explain

Messages

Notifications

	<div>datname</div> <div>name</div>	<div>oid</div> <div>oid</div>	
1	dbms	16384	

Slika 2.3 Preuzimanje podataka za *dbms* bazu podataka

1


**SELECT** relname, oid **FROM** pg\_class **WHERE** relname = 'users';

Data Output

Explain

Messages

Notifications

	<b>relname</b> name	 <b>oid</b> oid	
1	users	24919	

Slika 2.4 Preuzimanje podataka za tabelu *users*

## 2.2 Fizička struktura klastera baza

Klaster baza podataka fizički predstavlja jedan folder u kome se nalaze dodatni folderi i datoteke. Taj folder se zove **base** i nalazi se u folderu u koje je instaliran PostgreSQL[2]. Na slici 2.5 prikazan je sadržaj Postgres foldera.

```
/var/lib/postgresql/data # ls
PG_VERSION      pg_dynshmem      pg_multixact      pg_snapshots      pg_tblspc      postgresql.auto.conf
base             pg_hba.conf       pg_notify          pg_stat            pg_twophase     postgresql.conf
global          pg_ident.conf     pg_repslot        pg_stat_tmp        pg_wal          postmaster.opts
pg_commit_ts    pg_logical        pg_serial          pg_subtrans        pg_xact         postmaster.pid
```

Slika 2.5 Sadržaj Postgres foldera

```
/var/lib/postgresql/data/base # ls
1          13422      13423      16384      pgsql_tmp
```

Slika 2.6 Sadržaj base foldera

Na slici 2.6 prikazan je sadržaj base foldera. Folderi koji se nalaze unutar base foldera predstavljaju folder za odgovarajuću bazu podataka. Sa slike 2.3 vidi se da je OID dbms baze podataka 16384, pa folder 16384 odgovara folderu baze dbms. U tom folderu nalaze se objekti vezani za tu bazu podataka.

Svaki objekat baze ima svoj fajl koji se nalazi u folderu baze podataka. Sa slike 2.4 vidi se da je OID tabele *users* 24919 što zahteva da se u folderu 16384 nalazi datoteka 24919. Na slici 2.7 prikazan je fajl vezan za tabelu *users*.

```
/var/lib/postgresql/data/base/16384 # ls -lh 24919
-rw----- 1 postgres postgres 256.0K Mar 24 21:37 24919
```

Slika 2.7 Fajl koji sadrži podatke *users* tabele

Sa slike 2.7 može se videti da je fajl koji sadži podatke users tabele velik 256kB što se može potvrditi u klijentu koji je konektovan na dati Postgres server. Na slici 2.8 prikazane su informacije vezane za users tabelu iz PgAdmin4 klijenta.

Table size	256 kB
Toast table size	8192
Indexes size	128 kB

**Slika 2.8** Informacije vezane za users tabelu

Veličina indeksa za tabelu je 128kB što se može uočiti na slici 2.8. Kao i za tabelu, za indekse postoje fajlovi koji ih čuvaju. Za pronalaženje fajla na sistemu potrebno je prvo pronaći njegov OID. Kako tabela users ima kreirane indekse samo za kolonu **id** postoji jedna datoteka koja ih čuva.

```
1 SELECT relname, oid FROM pg_class WHERE relname = 'users_pkey';
```

Data Output		Explain	Messages	Notifications
	relname name	oid oid		
1	users_pkey	24927		

**Slika 2.9** Preuzimanje podataka za indekse id kolone users tabele

Na slici 2.10 prikazana je datoteka koja sadrži indekse za id kolonu tabelu users.

```
/var/lib/postgresql/data/base/16384 # ls -lh 24927
-rw----- 1 postgres postgres 128.0K Mar 24 21:37 24927
/var/lib/postgresql/data/base/16384 # |
```

**Slika 2.10** Fajl koji sadrži indekse id kolone users tabele

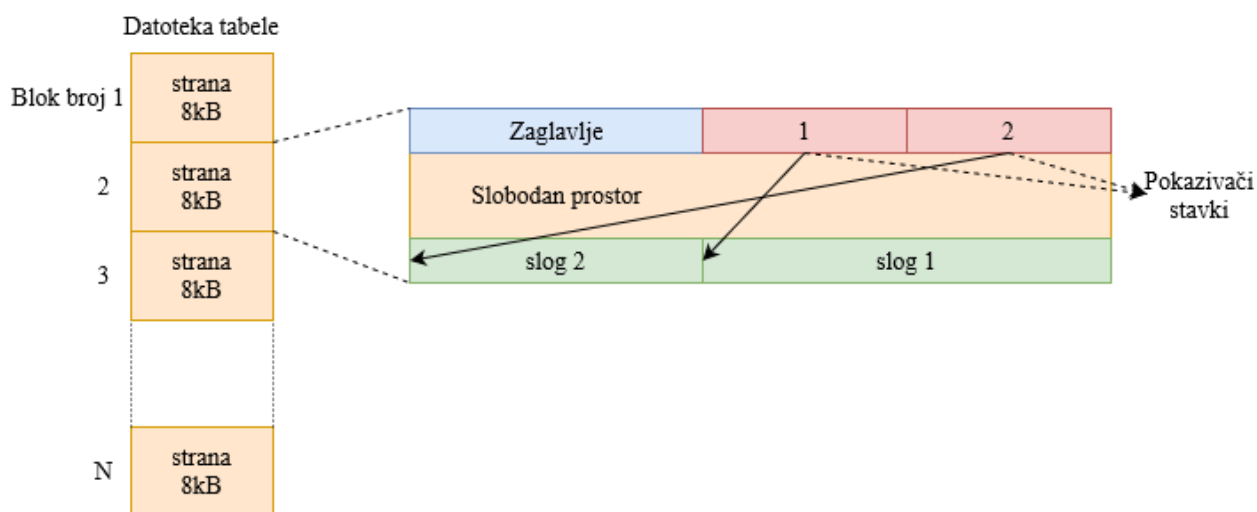
Objekti kao što su indeksi i tabele čuvaju se u jednom fajlu sve dok imaju veličinu manju od 1GB [2]. Kada ti podaci pređu veličinu veću od 1GB Postgres kreira novu datoteku gde na OID nadovezuje .1, .2, itd.

Pored datoteke koji sadži podatke vezane za tabelu postoje još dva fajla koja nakon OID-a imaju `_fsm` i `_vm` u imenu. Za indekse ne postoji fajl sa `_vm`. Mapa slobodnog prostora (eng. *free space map – fsm*) čuva informacije kapaciteta slobodnog mesta svake strane (eng. *page*) u okviru datoteke tabele. Prilikom dodavanja heap ili index sloga (eng. *tuple*), Postgres koristi fsm datoteku odgovarajuće tabele ili indeksa kako bi pronašao stranu u kojoj da izvrši dodavanje.

Mapa vidljivosti (eng. *visibility map –vm*) čuva vidljivost svake strane u okviru datoteke tabele. Vm datoteka se koristi u procesu čišćenja (eng. *vacuum processing*) gde se vrši brisanje mrtvih slogova. Vidljivost strane određuje da li strana ima mrtvih slogova i time se optimizuje proces čišćenja jer on može da preskoči strane koje nemaju mrtve slogove bez proveravanja svakog sloga.

## 2.3 Izgled datoteke tabele

Svaka datoteka sa podacima je podeljena na strane (blokove) fiksne dužine čija je dužina obično 8kB. Te strane su numerisane tako da svaka strana ima svoj broj i on predstavlja broj bloka. Kada se datoteka napuni Postgres dodaje novu praznu stranu na kraj datoteke kako bi joj se povećala veličina. Unutrašnji izgled datoteke podataka zavisi od tipa objekta za koji je ta datoteka namenjena [2]. Na slici 2.11 prikazana je struktura heap datoteke tabele.



**Slika 2.11** Izgled heap datoteke tabele

Podaci koji se nalaza u okviru datoteke tabele:

- Slogovi podataka – Predstavljaju same podatke
- Pokazivači stavki – Prestavlja pokazivač veličine 4B koji pokazuje na slog i zove se pokazivač stavke. Pokazivači su uređeni u niz i počinju sa indeksom 1. Posle dodavanja novog sloga dodaje se i novi pokazivač u niz.
- Zaglavlje – Postoji na početku svake strane i sadrži osnovne informacije o strani, dužine je 24B. Sastoji se od polja `pd_lsn`, `pd_checksum`, `pd_lower`, `pd_upper`, `pd_special`, `pd_pagesize_version` i `pd_prune_xid`. `Pd_lower` polje pokazuje na zadnji pokazivač, a `pd_upper` na prvi najbliži slog.

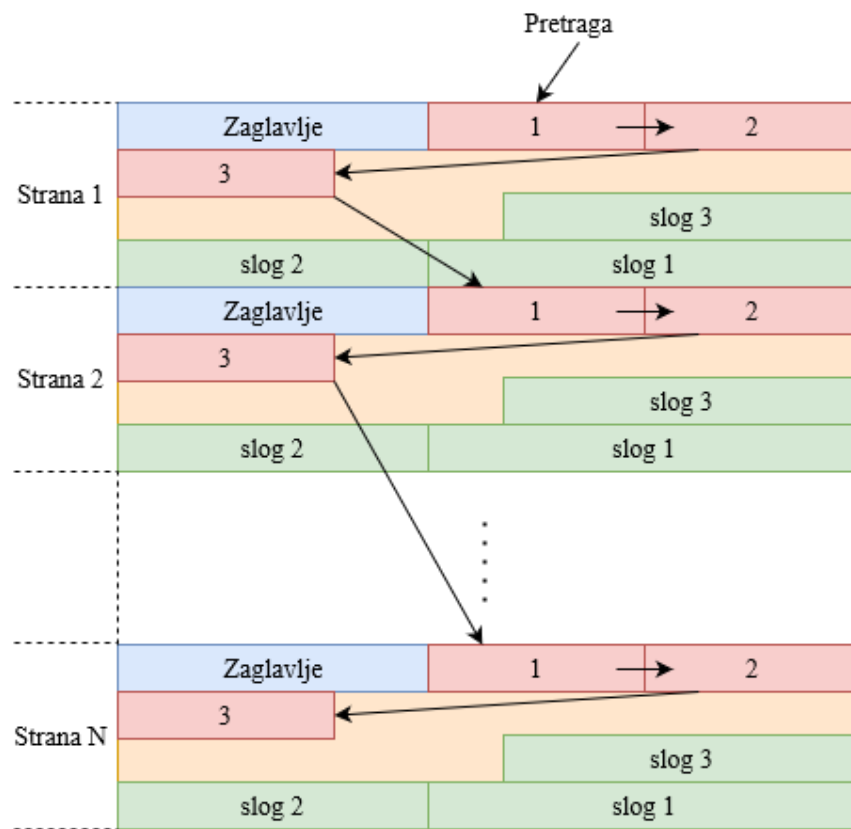


### 3. Indeksi

Indeksi u bazama podataka predstavljaju strukturu podataka koja se koristi kako bi se ubrzale određene operacije nad podacima koji se čuvaju u bazi. Korišćenjem indeksa dobija se ubrzanje ali cena korišćenja je povećanje veličine podataka i dodatne operacije koje se izvršavaju prilikom upisa novih podataka radi održanja strukture indeksa. Prilikom pretrage podataka koristi se metoda sekvencijalnog pretraživanja podataka i to je veoma skupo, korišćenjem indeksa izbegava se ovaj način traženja. Indekse je moguće kreirati za jednu ili više kolona tabele, a najbolje je kreirati indekse za one kolone koje se često pretražuju.

#### 3.1 Kako rade indeksi

Na slici 2.11 videli smo kako se tabela čuva u datoteku koja predstavlja niz strana. Prilikom pretrage podataka potrebno je proveriti da li svaka strana zadovoljava kriterijum pretrage. Svaki upit koji se izvršava u bazi ima određenu cenu koja je direktno povezana sa brojem strana koje se učitavaju i kojima se pristupa. Baze podataka mogu da rade i bez indeksa ali problem je brzina. Ako za kolonu po kojoj se vrši pretraga ne postoje indeksi tada se vrši sekvencijalna pretraga i prolazi se kroz sve strane i podatke [3]. Na slici 3.1 prikazana je sekvencijalna pretraga podataka.



Slika 3.1 Sekvencijalna pretraga podataka

Kako bi se izbegla sekvencijalna pretraga podataka moguće je definisati indekse nad kolonom. Indeksi u suštini čuvaju ključ-vrednost podatke, gde ključ predstavlja vrednost iz kolone, a vrednost je par broj strane i broj sloga. Cena indeksa koja se mora platiti je dodatna memorija za skladištenje ovih podataka. Pored toga indeksi se moraju održavati, tako da prilikom unosa ažuriranja ili brisanja podataka indeksi se takođe moraju ažurirati što zahteva dodatne operacije u zavisnosti od tipa indeksa. Indekse je moguće definisati za više kolona i samo za određene redove koji ispunjavaju neki uslov, takvi indeksi se nazivaju delimični (eng. *partial*).

## 3.2 PostgreSQL indeksi

Postoje različiti tipovi indeksa, PostgreSQL podržava sledeće: B-stablo, hash, GiST, SP-GiST, GIN i BRIN. Svaki od ovih tipova koristi različiti algoritam koji je optimizovan za određene pretrage. Indeksi koji se kreiraju komandom CREATE INDEX su tipa B-stablo i oni se najčešće koriste [2].

### B-stabla

Postgres uključuje implementaciju standardnog B-stabla koje predstavlja višesmerno balansirano stablo. Moguće ih je koristiti za sve podatke koji mogu da se sortiraju u linearni niz. B-stabla se koriste od strane PostgreSQL planera upita kada se vrši pretraga po indeksiranoj koloni sa jednim od sledećih operatora: <, <=, =, >= i >. Takođe B-stabla se koriste i sa operatorima BETWEEN, IN, IS NULL i IS NOT NULL. B-stable se mogu koristiti i sa operatorom LIKE u slučaju da je početak stringa poznat, npr. LIKE 'tes%' .

### Hash

Hash indeksi se koriste samo sa konkretnim poređenjem. Planer upita koristi hash indekse samo kada se vrši pretraga po indeksiranoj koloni sa operatorom =. Komanda za kreiranje indeksa je:

```
CREATE INDEX <ime_indeksa> ON <tabela> USING HASH (<kolona>)
```

### GiST

GiST je skraćenica za eng. *Generalized Search Tree* što znači generalizovano stablo pretrage. Kod njih su definisane metode koje se koriste za pristup strukturi stabla, a konkretnu implementaciju te strukture je moguće menjati, pa i definisati svoje tipove podataka sa tim metodama. GiST indeksi nisu jedinstvenog tipa i za njih mogu da se koriste različite strategije. Operatori sa kojima se GiST indeksi mogu koristiti zavise od strategije koja se koristi. Primer operatora za dvodimenzionalne geometrijske podatke su: <<, &<, &>, >>, <<|, &<|, |&>, |>>, @>, <@, ~= I &&. Takođe oni mogu da optimizuju pronalaženje najbližeg suseda u upitima za zadatu tačku.

## SP-GiST

SP-GiST je skraćenica za eng. *space-partitioned* GiST što znači podeljeni prostor. Oni koriste podeljena stabla pretrage koja imaju za cilj da izmapiraju čvorove stabla na stranice na disku tako da se prilikom pretrage smanji broj pristupa stranama. Kao i GiST indeksi oni pružaju infrastrukturu za različite tipove pretraga. Moguće ih je koristiti sa različitim strukturama podataka kao što su: quad stabla, k-d stabla i radix stabla.

## GIN

GIN je skraćenica za eng. *Generalized inverted index*. GIN indeksi predstavljaju invertovane indekse koji se koriste za tipove podataka koji sadrže više vrednosti, kao što su nizovi. Invertovani indeksi sadrži jedan unos za svaku vrednost i veoma je efikasan kod pretraga gde se proverava da li postoji vrednost u nizu. Kao i kod GiST i SP-GiST indeksa i GIN indeksi podržavaju više strategija za indeksiranje pa se tako operatori sa kojima oni koriste razlikuju u zavisnosti od izabrane strategije.

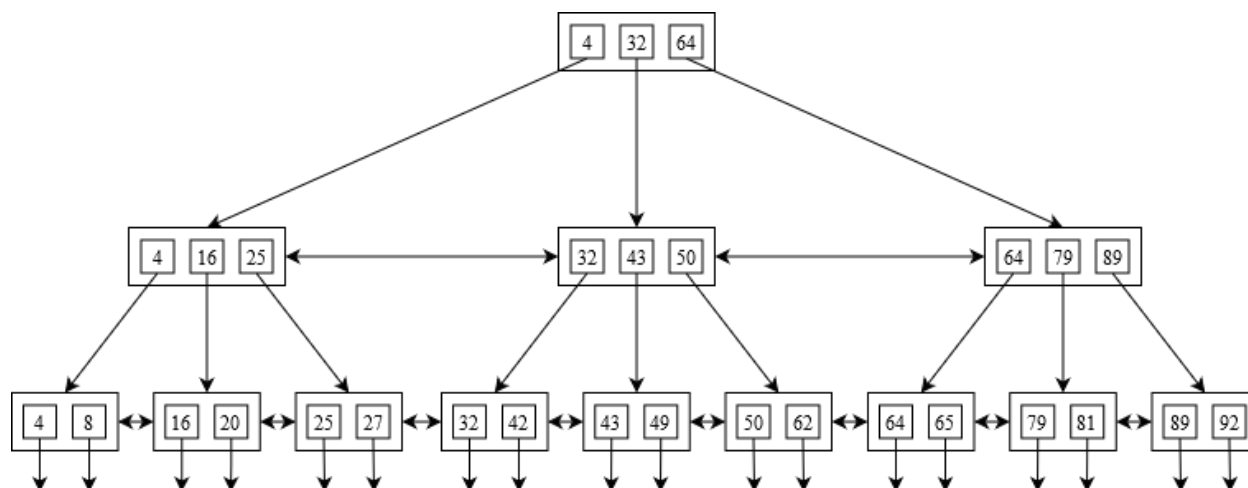
## BRIN

BRIN je skraćenica za eng. *Block range index*. Namenjeni su za velike tabele kod kojih određene kolone imaju vezu sa fizičkom lokacijom u tabeli. Blok u određenom opsegu predstavlja grupu strana koje su fizički susedne u tabeli. Brin indeksi čuvaju određene sumirizovane podatke o vrednostima koje se nalaze u strani. Primer predstavljaju minimalna i maksimalna vrednost u okviru strane. Kako je i ovde moguće koristiti različite strategije, operatori sa kojima se oni koriste zavise od same strategije.

## 3.3 Implementacija indeksa

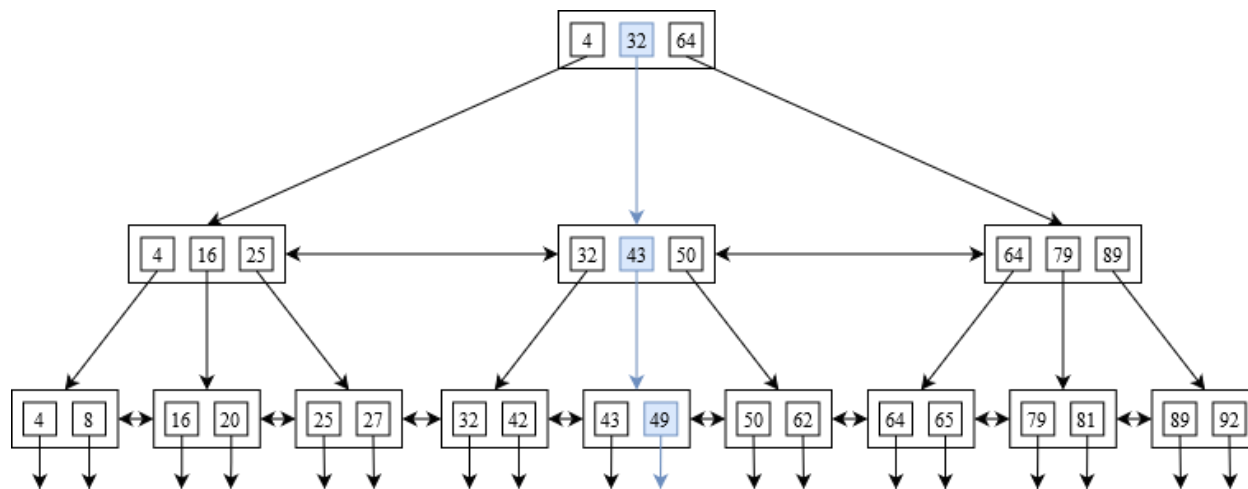
### B-stabla

Koriste se kod podataka koji mogu da se sortiraju. B-stabla predstavljaju balansirana stabla i na njihovim listovima nalaze se podaci indeksa koji pokazuju na redove u tabeli. Kod balansiranih stabala svaki list je podjedano udaljen od korena stabla. Zbog toga pretraga bilo koje vrednosti traje isto. Listovi su povezani tako da svaki list ima pokazivač na sledeći i naredni čvor, pa nije potrebno vraćati se na koren svaki put prilikom preuzimanja poređanih podataka. Kod unutrašnjih čvorova svaki čvor pokazuje na dete koje sadrži vrednosti iz roditelja i ona je minimalna za taj čvor. Na slici 3.2 prikazan je primer B-stabla.



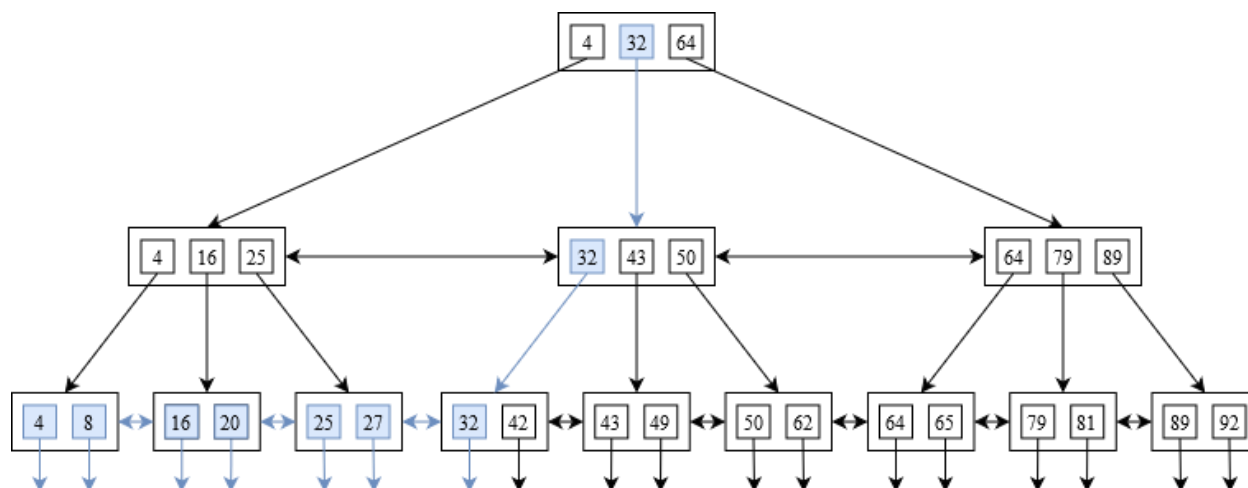
**Slika 3.2** Primer B-stabla

Ako je potrebno izvršiti pretragu podataka tako da je tražena vrednost 49, postupak je sledeći: kreće se od vrha stabla i za svaki čvor se traži naredni dok se ne stigne do lista. U svakom kako bi se našao naredni ispituje se tražena vrednost sa vrednostima unutar čvora. Pošto je  $32 < 49 < 64$  uzima se sledeći čvor na kog pokazuje vrednost 32. Ovaj postupak se radi rekurzivno sve dok se ne stigne do lista. List sadrži indeks koji pokazuje na broj strane i broj sloga koji sadrži traženi podatak i on se uzima. Na slici 3.3 prikazan je primer obilaska stabla za broj 49.



**Slika 3.3** Obilazak stabla za pretragu broj 49

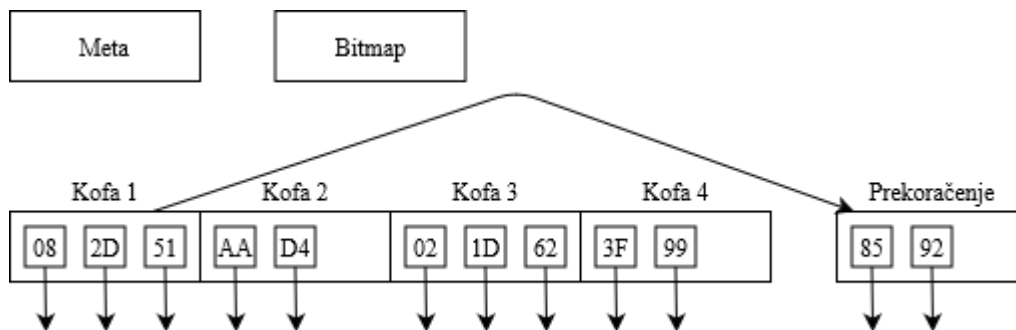
Na slici 3.4 prikazan je obilazak stabla za slučaj pretrage vrednosti manje od 35.



**Slika 3.4** Obilazak stabla za pretragu vrednosti manje od 35

### Hash indeksi

U mnogim programskim jezicima postoji tip heš tablica. Postgres koristi sličnu implementaciju heš tablica. Heš tablice predstavljaju niz od N vrednosti koje su indeksirane heširanom vrednošću a njihova vrednost pokazuje na par broj strane i broj sloga. Heš vrednost dobija se korišćenjem heš funkcije. Ova funkcija mora dobro da distribuira vrednosti i da se izvršava brzo. Problem je što se nekad dešava da se za različitu vrednost dobije isti heš. Zbog toga heš tablica je organizovana u kofe (eng. *bucket*) koje služe za čuvanje pokazivača na različite redove u tabeli koji imaju isti heš. Na slici 3.5 prikazan je primer heš tablice.



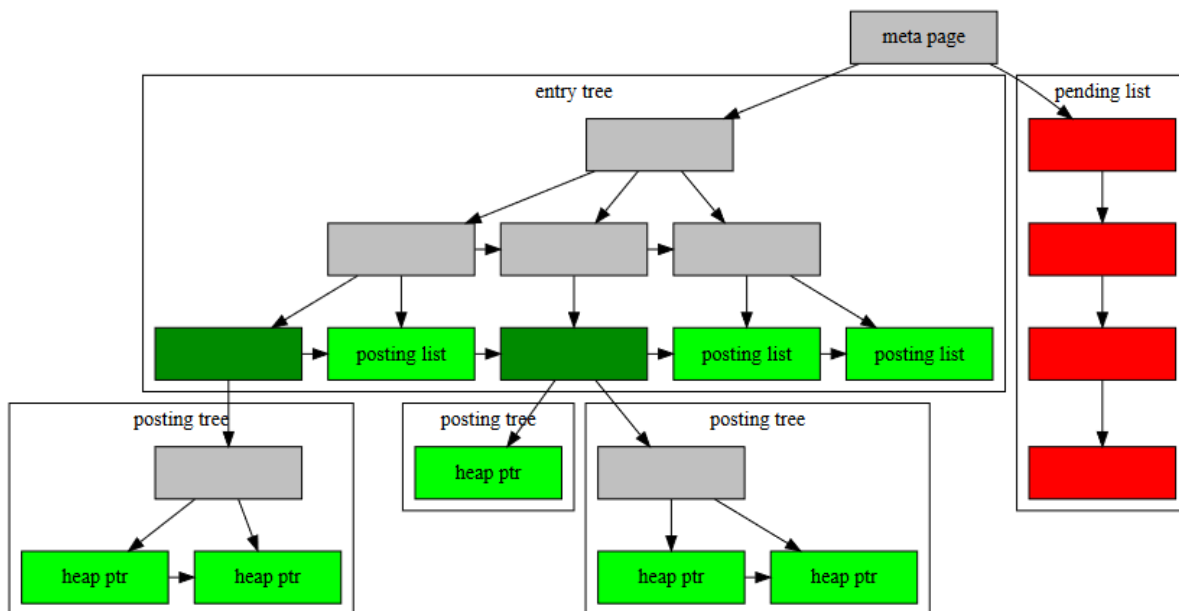
**Slika 3.5** Haš tablica

Kofa za prekoračenje se koristi kada jedna kofa nije dovoljna za sve vrednosti. Bitmapa se koristi kako bi se pratilo koje kofe za prekoračenje su prazne i mogu se koristiti za druge kofe. Heš tablice funkcionišu tako što se za određenu vrednost odradi heš funkcija. Dobija vrednost je broj kofe gde je potrebno smesiti indeksiranu vrednost. Kada se vrši pretraga po ključu potrebno je izvršiti haš funkciju na vrednost gde se dobija broj kofe a nakon toga potrebno je proveriti sadržaj kofe i vratiti samo onu vrednost koja je tražena [2].

## GIN indeksi

GIN indeksi se koriste kod slučajeva gde stavke koje se indeksiraju predstavljaju kompozitne vrednosti, a upiti koji se optimizuju pomoću ovih indeksa traže jedan element koji se nalazi u okviru kompozitne vrednosti. Na primer, moguće je indeksiranje više dokumenata, a upiti koji koriste ove indekse vrše pretragu dokumenata koji sadrži određenu reč. Kod GIN indeksa reč *stavka* koristimo za kompozitne vrednosti koje se indeksiraju, a reč ključ za vrednost elementa.

GIN indeksa čuva par eng. *key*, *posting list*, gde je *posting list* niz identifikatora redova koji sadrže ključ. Isti identifikator nekog reda može da se pojavi u više *posting list*-i. Svaka vrednost ključa se čuva samo jednom, pa su GIN indeksi efikasni za slučajeve gde se ključ pojavljuje više puta. GIN indeksi mogu da sadrže B-stablo definisano za odgovarajući ključ, gde je svaki ključ deo jedne ili više indeksiranih stavki i gde svaki slog u listu sadrži pokazivač na B-stablo indeksa ili jednostavnu listu indeksa [2]. Na slici 3.6 prikazane su komponente GIN indeksa.



Slika 3.6 Komponente GIN indeksa

## 4. Primer korišćenja indeksa

Poređenje performansi indeksa rađeno je na udaljenoj virtuelnoj mašini koja ima 1 CPU od 1.8 GHz sa 30720kB keša, 1GB RAM memorije i 25GB SSD. Na toj virtuelnoj mašini podignuta je PostgreSQL baza u okviru Docker kontejnera, a naredba za to je:

```
docker run -d --restart unless-stopped - POSTGRES_USER=root -e POSTGRES_PASSWORD=dmbstest1232 -e POSTGRES_DB=dbms -p 5000:5432 postgres:12.2-alpine
```

Na podignutoj bazi kreirane su dve tabele *users* i *posts*, naredba za kreiranje table je:

```
CREATE TABLE users (  
    id SERIAL PRIMARY KEY,  
    name TEXT NOT NULL,  
    created_at TIMESTAMP NOT NULL DEFAULT NOW()  
);  
  
CREATE TABLE posts (  
    id SERIAL PRIMARY KEY,  
    user_id INTEGER NOT NULL REFERENCES users (id),  
    public BOOLEAN NOT NULL DEFAULT TRUE,  
    text TEXT NOT NULL,  
    created_at TIMESTAMP NOT NULL DEFAULT NOW()  
);
```

Indeksi se automatski kreirani samo za kolone *id* jer su one primarni ključ. Posle kreiranja table potrebno je dodati podatke, što je učinjeno komandama:

```
INSERT INTO users (name, created_at)  
SELECT  
    ARRAY_TO_STRING(ARRAY(SELECT SUBSTR('abcdefghijklmnopqrstuvxyz', ((RANDOM()*(30-1)+1)::INTEGER),1) FROM GENERATE_SERIES(1,7 + b * 0)), ''),  
    NOW() + (RANDOM() * (NOW() + '365 days' - NOW()))  
FROM GENERATE_SERIES(1, 5000) as a(b);  
  
INSERT INTO posts (user_id, text, public, created_at)  
SELECT  
    RANDOM() * 4999 + 1,  
    ARRAY_TO_STRING(ARRAY(SELECT SUBSTR('abcdef ghi jklmno pqrst uvwxyz', ((RANDOM()*(34-1)+1)::INTEGER),1) FROM GENERATE_SERIES(1,120 + b * 0)), ''),  
    CASE WHEN RANDOM() > 0.3 THEN TRUE ELSE FALSE END,  
    NOW() + (RANDOM() * (NOW() + '365 days' - NOW()))  
FROM GENERATE_SERIES(1, 500000) as a(b);
```

U tabeli *users* dodato je 5000 redova, veličina datoteke koju ova tabela zauzima na disku je 256kB. Na slici 4.1 prikazan je primer umetnutih podataka u tabelu *users*.






```
1 SELECT * FROM users;
```

Data Output	Explain	Messages	Notifications
	<b>id</b> [PK] integer	<b>name</b> text	<b>created_at</b> timestamp without time zone
1	1	flzzj	2020-05-11 16:33:58.594603
2	2	pjnmy...	2020-06-26 14:32:09.164618
3	3	fqzyqrz	2021-03-14 02:12:44.204916
4	4	kjwiyv	2020-08-29 21:43:36.534963
5	5	lkgu	2021-01-17 12:15:48.639957
6	6	xwoqyz	2020-03-30 05:57:01.663116
7	7	snnzqm	2020-05-28 22:23:28.093653
8	8	zxrshj	2021-02-06 16:55:55.318386
9	9	rqzyue	2020-06-13 22:09:41.579976
10	10	weqtia	2020-10-13 05:17:00.259743

**Slika 4.1** Deo umetnutih podataka u tabelu *users*

U tabeli *posts* dodato je 500000 redova, veličina datoteke koju ova tabela zauzima na disku je 78MB. Na slici 4.2 prikazan je primer umetnutih podataka u tabelu *posts*.

```
1 SELECT * FROM posts;
```

Data Output		Explain	Messages	Notifications		
	 <b>id</b> [PK] integer 	<b>user_id</b> integer 	<b>public</b> boolean 	<b>text</b> text 	<b>created_at</b> timestamp without time zone	
1	1	4830	true	qnk wx...	2020-07-22 08:51:04.159649	
2	2	4900	false	t l nuk ...	2021-02-18 02:13:19.862107	
3	3	2191	false	kg bpn...	2020-10-11 07:50:53.892299	
4	4	3268	true	k ufvp...	2020-09-10 06:59:28.795342	
5	5	3894	false	utwzr o...	2020-05-06 23:22:27.014382	
6	6	957	true	wux w...	2020-09-01 23:39:02.881987	
7	7	712	false	gtluu s...	2020-05-28 12:04:46.460573	
8	8	3807	true	m qfzt...	2021-02-24 04:05:10.100405	
9	9	2514	true	ahh c...	2020-10-26 11:49:51.191283	
10	10	1608	true	bkrun k...	2021-01-25 16:54:36.705054	

**Slika 4.2** Deo umetnutih podataka u tabelu *posts*



Na slici 4.3 predstavljen je upit za preuzimanje podataka iz *users* tabele po imenu.

```
1 SELECT * FROM users WHERE name = 'fbxjnl';
```

	Data Output	Explain	Messages	Notifications
	<b>id</b> [PK] integer	<b>name</b> text	<b>created_at</b> timestamp without time zone	
1	305	fbxjnl	2020-11-03 18:10:17.115782	

**Slika 4.3** Upit za preuzimanje korisnika po imenu

Kako bi proverili cenu izvršenje ovog upita ispred komande potrebno je dodati naredbe EXPLAIN ANALYZE. Na slici 4.4 prikazan je plan i cena izvršenja prethodnog upita.

```
1 EXPLAIN ANALYZE SELECT * FROM users WHERE name = 'fbxjnl';
```

	Data Output	Explain	Messages	Notifications
	<b>QUERY PLAN</b> text			
1	Seq Scan on users (cost=0.00..94.50 rows=1 width=19) (actual time=0.078..0.767 rows=1 loops=1)			
2	Filter: (name = 'fbxjnl'::text)			
3	Rows Removed by Filter: 4999			
4	Planning Time: 0.089 ms			
5	Execution Time: 0.802 ms			


**Slika 4.4** Plan i cena izvršenja prethodnog upita

Sa slike 4.4 vidi se da kako bi se pronašao korisnik sa specificiranim imenom baza vrši sekvencijalnu pretragu podataka, što znači da prolazi kroz sve podatke i vrši filtriranje svih redova osim jednog. Pored toga vidi se da je cena izvršenja upita 94.50, a vreme izvršenja 0.802ms. Komandom:

```
CREATE INDEX index_users_name ON users USING HASH (name);
```

kreiraćemo indekse za kolonu *name* tabele *users*. Nakon dodavanja indeksa za kolonu *name* kreirana je datoteka na disku koja zauzima 272kB i sadrži indekse za tu kolonu. Na slici 4.5 prikazan je plan izvršenja i cena.

```
1 EXPLAIN ANALYZE SELECT * FROM users WHERE name = 'fbxjnl';
```


Data Output	Explain	Messages	Notifications
<b>QUERY PLAN</b>  text			
1	Index Scan using index_users_name on users (cost=0.00..8.02 rows=1 width=19) (actual time=0.028..0.029 rows=1 loops=1)		
2	Index Cond: (name = 'fbxjnl'::text)		
3	Planning Time: 0.082 ms		
4	Execution Time: 0.060 ms		

**Slika 4.5** Plan i cena izvršenja prethodnog upita sa indeksima

Sa slike 4.5 vidi se da se sada ne radi sekvencijalna pretraga podataka već se vrši pretraga indeksa što dovodi do toga da je cena upita 8.02, a vreme izvršenja upita 0.060ms što je više od 10 puta brže od slučaja bez indeksa. Izvršenje upita jeste brže i cena je manja ali dodata je datoteka koja sadrži indekse i njena veličina je u ovom slučaju 272kB dok je veličina indeksa za kolonu *id* 128kB. Sada prilikom svakom ažuriranja tabele potrebno je ažurirati i ove dve datoteke sa indeksima. Ako se često vrši pretraga podataka po koloni *name* ovo predstavlja malu cenu koju treba platiti.

Na slici 4.6 prikazan je upit i plan njegovog izvršenja za preuzimanje podataka iz tabele *posts* za korisnika sa određenim identifikatorom.

```
1 EXPLAIN ANALYZE SELECT * from posts WHERE user_id = 4831;
```

Data Output	Explain	Messages	Notifications
<b>QUERY PLAN</b>  text			
1	Gather (cost=1000.00..13569.17 rows=100 width=128) (actual time=0.566..96.495 rows=104 loops=1)		
2	Workers Planned: 2		
3	Workers Launched: 2		
4	-> Parallel Seq Scan on posts (cost=0.00..12559.17 rows=42 width=128) (actual time=3.208..81.111 rows=35 loops=3)		
5	Filter: (user_id = 4831)		
6	Rows Removed by Filter: 166632		
7	Planning Time: 0.088 ms		
8	Execution Time: 96.565 ms		

**Slika 4.6** Plan i cena izvršenja upita za preuzimanje postova određenog korisnika

Tabela *posts* ima 500000 redova i zbog toga se vidi da se u planu za pretraživanje podataka puštaju 2 radnika koja paralelno pretražuju sve podatke tabele. Cena izvršenja ovog upita je 12559.17, a vreme izvršenja 96.565ms. Komandom:

```
CREATE INDEX index_posts_user_id ON posts (user_id);
```

kreiraju se indeksi za kolonu *user\_id*. Veličina datoteke koja sadrži indekse za ovu kolonu je 10MB. Radi poređenja, veličina indeksa za kolonu *id* je 11MB, što znači da indeksi nad ovom tablom zauzimaju 21MB, a sama veličina tabele je 78MB. Na slici 4.7 prikazan je plan i cena izvršenja indeksa nakon kreiranja indeksa.

```
1 EXPLAIN ANALYZE SELECT * FROM posts WHERE user_id = 4381;
```

Data Output	Explain	Messages	Notifications
<b>QUERY PLAN</b>			
text			
1	Bitmap Heap Scan on posts (cost=5.20..376.38 rows=100 width=128) (actual time=0.063..0.577 rows=104 loops=1)		
2	Recheck Cond: (user_id = 4381)		
3	Heap Blocks: exact=103		
4	-> Bitmap Index Scan on index_posts_user_id (cost=0.00..5.17 rows=100 width=0) (actual time=0.042..0.042 rows=104 loops=1)		
5	Index Cond: (user_id = 4381)		
6	Planning Time: 0.284 ms		
7	Execution Time: 0.617 ms		

**Slika 4.7** Plan i cena izvršenja prethodnog upita sa indeksima


Sa slike 4.7 vidi se da je cena izvršenja ovog upita 376.38, dok je vreme izvršenja 0.617ms.

Osim indeksa za jednu kolonu moguće je kreirati i indekse za više kolona. Takvi indeksi se koriste za upite gde postoji bilo koji podskup indeksiranih kolona, međutim on je najefikasniji ako je prisutna kolona koja se nalazi na prvom mestu u indeksu. Ovo označava da je redosled nabiranja kolona u indeksu bitan i efikasnost upita zavisi od korišćenja kolona u njemu. Pored indeksa nad više kolona moguće je kreirati i parcijalne indekse. Kod njih se definiše indeks samo za određene podatke. Primer parcijalnog indeksa:

```
CREATE INDEX index_posts_user_id_public ON posts (user_id) WHERE public = TRUE;
```

Ovom komandom kreira se indeks nad kolonom *user\_id* samo za one postove koji su javni. Na slici 4.8 prikazan je upit gde se pretražuju postovi za određenog korisnika.


```
1 EXPLAIN ANALYZE SELECT* FROM posts WHERE user_id = 3821;
```

Data Output	Explain	Messages	Notifications
<b>QUERY PLAN</b>  text			
1	Gather (cost=1000.00..13569.17 rows=100 width=128) (actual time=1.721..94.041 rows=87 loops=1)		
2	Workers Planned: 2		
3	Workers Launched: 2		
4	-> Parallel Seq Scan on posts (cost=0.00..12559.17 rows=42 width=128) (actual time=0.367..81.145 rows=29 loops=3)		
5	Filter: (user_id = 3821)		
6	Rows Removed by Filter: 166638		
7	Planning Time: 0.444 ms		
8	Execution Time: 94.095 ms		

**Slika 4.8** Pretraga postova za određenog korisnika

Sa slike 4.8 vidi se da se ne koristi prethodno kreirani indeks, već da se vrši sekvencijalna pretraga svih podataka. Cena izvršenja upita je 13569.17, a vreme izvršenja 94.095ms. Na slici 4.9 prikazan je upit gde se traže javni postovi za određenog korisnika.

```
1 EXPLAIN ANALYZE SELECT* FROM posts WHERE user_id = 3821 AND public = TRUE;
```

Data Output	Explain	Messages	Notifications
<b>QUERY PLAN</b>  text			
1	Bitmap Heap Scan on posts (cost=4.97..268.23 rows=70 width=128) (actual time=0.045..0.356 rows=64 loops=1)		
2	Recheck Cond: ((user_id = 3821) AND public)		
3	Heap Blocks: exact=64		
4	-> Bitmap Index Scan on index_posts_public (cost=0.00..4.95 rows=70 width=0) (actual time=0.029..0.029 rows=64 loops=1)		
5	Index Cond: (user_id = 3821)		
6	Planning Time: 0.124 ms		
7	Execution Time: 0.384 ms		

**Slika 4.9** Pretraga javnih postova za određenog korisnika

Sa slike 4.9 vidi se da se u ovom slučaju koristi prethodno kreirani parcijalni indeks. Sada je cena ovog upita 268.23, a vreme izvršenja 0.384. Veličina datoteke koja sadrži parcijalni indeks za kolonu *user\_id* je 7MB, dok je veličina indeksa nad svim podacima ove kolone 10MB. Ukoliko imamo dosta upita sa određenim filtriranjima moguće je ta filtriranja izdvojiti i u

indeksu kako bi se smanjila veličina indeksa. Problem je što izostavljanjem dela za filtriranje se indeksi ne koriste za izvršenje upita.

## 5. Zaključak

U ovom radu opisani su koncepti, implementacija i korišćenje određenih tipova indeksa za PostgreSQL bazu podataka. Baza podataka za svoje funkcionisanje ne zahteva postojanje indeksa, ali njene performanse su drastično slabije bez postojanja istih. Cena koju moramo platiti prilikom kreiranja indeksa jeste dodatna memorija i operacija prilikom modifikacije podataka radi održavanja indeksa. U zavisnosti od količine podataka i broja upita nad određenim podacima potrebno je izvršiti analizu i pažljivo iskoristiti indekse.

## LITERATURA

- [1] „*Postres about*“ - <https://www.postgresql.org/about/>, Poslednji pristup: 02.04.2020.
- [2] „*PostgreSQL 12.2 Documentation*” - <https://www.postgresql.org/files/documentation/pdf/12/postgresql-12-A4.pdf>, Poslednji pristup: 02.04.2020.
- [3] „*Indexes in PostgreSQL*“ - <https://postgrespro.com/blog/pgsql/3994098>, Poslednji pristup: 02.04.2020.