



UNIVERZITET U NOVOM SADU  
PRIRODNO-MATEMATIČKI FAKULTET  
DEPARTMAN ZA MATEMATIKU I  
INFORMATIKU



# Data Warehouse baza podataka banke

projekat iz predmeta Softversko inženjerstvo za sisteme baza  
podataka

prof. dr. Jovana Vidaković

Miloš Marić

Novi Sad, avgust 2023.



## Sadržaj

Uvod.....	5
OLTP baza podataka.....	6
Model OLTP baze podataka i generisanje podataka.....	6
OLAP baza podataka.....	8
Model OLAP baze podataka.....	8
Transformacija podataka iz OLTP u OLAP bazu podataka.....	10
Izveštaji.....	13
Promet valuta po okruzima u 2022. godini.....	13
Kreditna viabilnost za dinarske kredite.....	13
Potrošnja po starosnoj dobi i nameni u prvom polugodištu 2023. godine.....	14
Broj transakcija po danu u nedelji i mestu transakcije u 2021., 2022., 2023. godini.....	14
Zaključak.....	15



# Uvod

Praktični projekat u okviru predmeta Softversko inženjerstvo za sisteme baza podataka fokusira se na analizu operacija (fiktivne) banke koja posluje na teritoriji Republike Srbije. Transakciona baza podataka je osmišljena tako da vodi evidenciju o svim novčanim transakcijama koje korisnici banke vrše.

Prikupljanje podataka u oblastima poslovanja kao što je bankarstvo je od presudnog značaja za uspeh. Svakodnevno poslovanje u banci generiše obimne količine podataka kao što su izvršavanje transakcija, otvaranje i zatvaranje računa, kao i vođenje evidencije o korisnicima. Na osnovu ovih podataka moguće je doći do zaključaka o efikasnosti poslovanja, finansijskim trendovima, kao i preferencijama naših (trenutnih i potencijalnih) korisnika.

Uzmimo za primer vreme izvršavanja transakcija. Podaci o njima su nam korisni za raspodelu resursa (novčanih i ljudskih) i optimizaciju procesa izvršavanja transakcija. Analizom ovih podataka, banka može preciznije utvrditi obrazac ponašanja korisnika i prilagoditi svoj poslovni raspored kako bi efikasno odgovorila na zahteve korisnika.

Podaci o korisnicima i njihovim transakcijama omogućavaju banci uvid u njihovu potrošnju i prihode. Služeći se ovim informacijama banka može da izračuna kreditnu viabilnost korisnika i oceni njegov kreditni rizik i na osnovu toga predloži kredit svome korisniku ili neke druge personalizovane usluge.

Podaci o transakcijama izvršenim na bankomatima i ekspoziturama nam omogućavaju da uvidimo koje su ključne lokacije za naše poslovanje, na kojim bi trebalo da otvorimo još ekspozitura i bankomata, a na kojim bi možda mogli da zatvorimo poslovne jedinice i izbegnemo nepotrebne rashode.

Na kraju, izveštaji o preferencijama korisnika za različite bankarske usluge mogu pružiti vredne uvide za usmeravanje marketinških napora. Ovo može poboljšati korisničko iskustvo i doprineti povećanju broja korisnika banke.

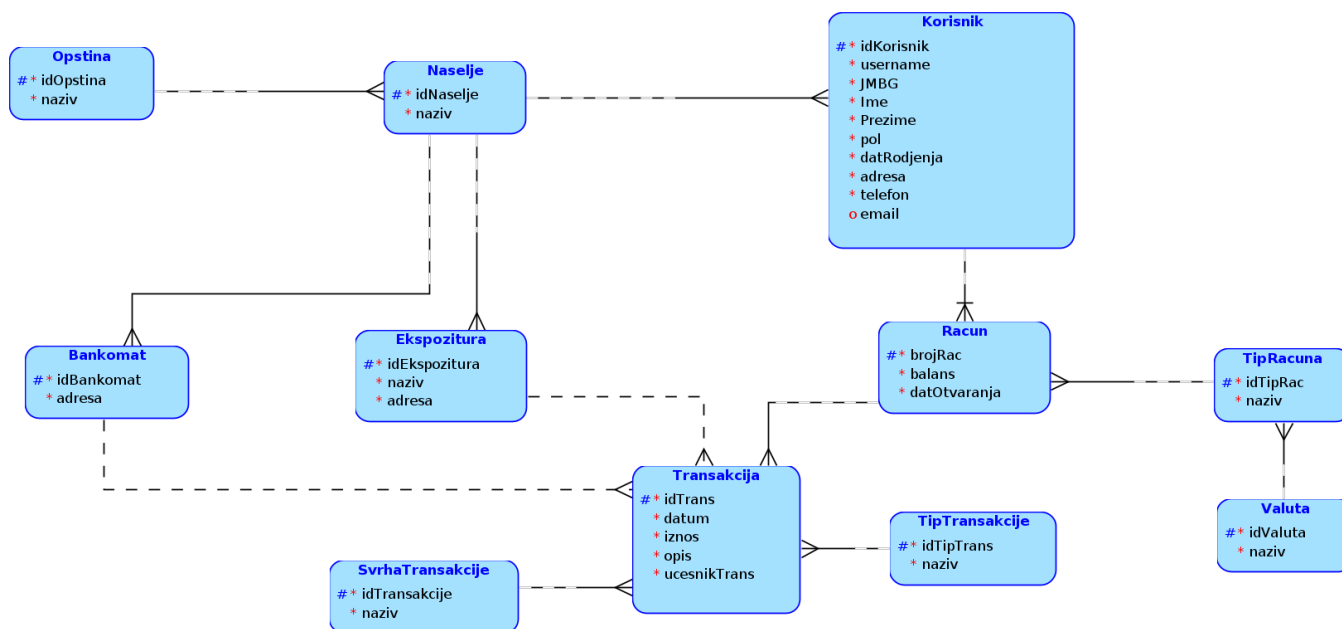
U izradi ovog projekta od velikog značaja je bio razgovor sa domenskim ekspertom, radnicom banke, dipl. ekonomistom Ivanom Marić, kojoj se ovim putem zahvaljujem.

# OLTP baza podataka

Online Transactional Processing (OLTP) baze podataka predstavljaju osnovu gotovo svih informacionih sistema današnjice. One služe za efikasno upravljanje i obradu svakodnevnih transakcija i operacija u stvarnom vremenu. Podržavaju brzu i preciznu obradu velikog broja kratkotrajnih transakcija, kao što su unos i izmena podataka, a koriste se za podršku operativnim procesima u organizacijama, kao što su prodaja, upravljanje inventarom, uplate i isplate, rezervacije i slično. U većini slučajeva one predstavljaju početnu tačku u razvoju OLAP baze podataka.

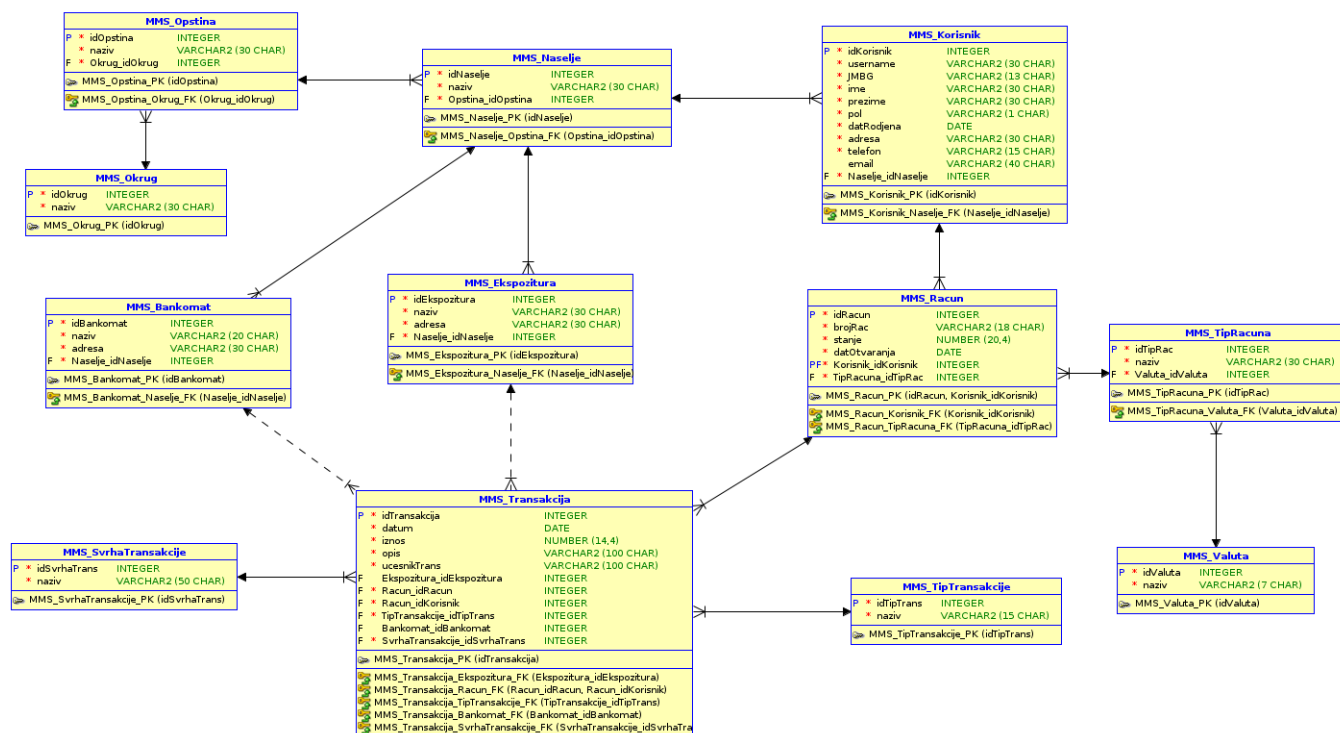
## Model OLTP baze podataka i generisanje podataka

Na slici 1 je prikazan ER model OLTP baze podataka koja služi za vođenje evidencije o transakcijama fizičkih lica u banci.



Slika 1: ER model OLTP baze

Na slici 2 prikazan je Relacioni model iste baze.



Slika 2: Relacioni model OLTP baze

Tabele Okrug, Opština i Naselje su table koje sadže podatke o odgovarajućim teritorijalnim jedinicama Republike Srbije. Ovi podaci kao i adrese (ulica i broj) koje se nalaze u tabelama Korisnik, Ekspozitura i Bankomat su relani podaci preuzetih sa zvaničnog sajta [Ministarstva za rad, zapošljavanje, boračka i socijalna pitanja \(MRZBSP\)](#).

Tabele Valuta, TipRacuna, TipTransakcije i SvrhaTransakcije ručno su unesene. Tabela TipRačuna sadrži podatke o vrstama računa koje banka nudi svojim korisnicima. TipTransakcije je šifarnička tabela koja određuje na koji način se izvršila transakcija, preko bankomata, u ekspozituri, preko Post terminala za bezgotovinsko plaćanje u nekoj od prodavnica ili za transakcije preko interneta.

Podaci za tabelu Korisnik i Račun izgenerisani su uz pomoć Python skripti. Imena i prezime za korisnike su preuzeti sa [Wikipedie](#), a na osnovu imena su izgenerisani i username i email. Pri generisanju datuma rođenja (1925 – 2006) vodilo se računa da raspodela ne bude uniformna, već da teži normalnoj raspodeli. Na osnovu datuma rođenja i pola korisnika generisan je i JMBG (od 13 cifara). Datum rođenja je takođe određivao i koju vrstu računa korisnik može da ima. Najmlađi korisnici mogu da imaju samo omladinski dinarski, mladi mogu da imaju sve, a stariji korisnici mogu da imaju sve račune sem omladinskog dinarskog. Takođe, balans na računu je izgenerisan tako da bude veći za dinarske račune, iz razloga što je dinar znatno slabija valuta od ostalih deviza (USD, EUR, GBP).

Na kraju, središte baze čini tabela Transakcija. Podaci za nju su takođe izgenerisani pomoću Python skripta gde se vodilo računa o tome da iznos transakcije zavisi od valute, svrhe i tipa transakcije, sa normalnom raspodelom verovatnoće i većom verovatnoćom za odlive, ali znatno većim prosekom za prilive. Opisi i ucesnici transakcije su izgenerisani pomoću veštačke

inteligencije (ChatGTP) tako da budu upareni sa svrhom transakcije (opsi) i tipom transakcije (učesnik transakcije). Takođe, vodilo se i računa da datumi transakcije ne budu pre datuma otvaranja računa (iz očiglednih razloga).

Svi generisani podaci su importovani u bazu preko CSV fajlova.

## **OLAP baza podataka**

Online Analytical Processing (OLAP) baza podataka je tip baze koja se koristi za analizu velikih količina podataka radi donošenja informisanih poslovnih odluka. Ova vrsta baze fokusira se na podršku analitičkim procesima, omogućavajući korisnicima da istražuju podatke iz različitih uglova, vrše dubinske analize i identifikuju uzorke. OLAP baza sadrži podatke koji su uglavnom istorijski i agregirani, što omogućava korisnicima da uoče razne obrasce i trendove.

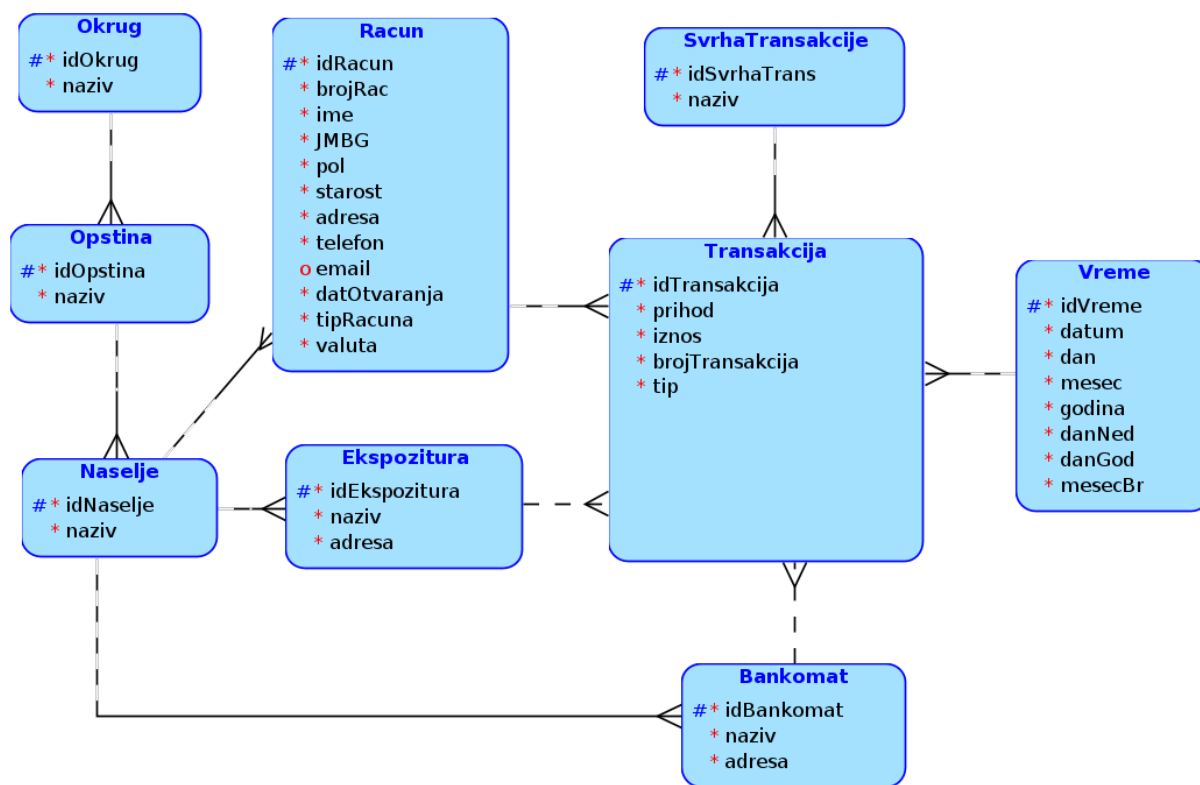
OLAP kocka, ili ti višedimenzionalna kocka, je ključna komponenta OLAP sistema. Svaka dimenzija u kocki predstavlja određenu karakteristiku podataka, kao što su vreme, geografska lokacija ili proizvodi. Mere su numerički podaci koje korisnici analiziraju, kao što su prodaja, troškovi ili profit. Kocka omogućava korisnicima da upitima dodatno manipulišu podacima kroz razne tehnike (agregacija podataka – roll-up, isecanje (slice & dice)), omogućavajući dublje razumevanje uzoraka i prirode poslovnog okruženja.

Razlika između OLAP i OLTP (Online Transaction Processing) baza podataka leži u njihovoj svrsi i strukturi. OLTP baze su usmerene na podršku operativnim transakcijama u stvarnom vremenu, kao što su unos, ažuriranje i brisanje podataka. Ove baze sadrže detaljne, tačne i najsvežije informacije o poslovnom procesu, a po strukturi su normalizovane. S druge strane, OLAP baze se fokusiraju na analizu i donošenje odluka, te sadrže istorijske i agregirane podatke. I njihova struktura je denormalizovana.

## **Model OLAP baze podataka**

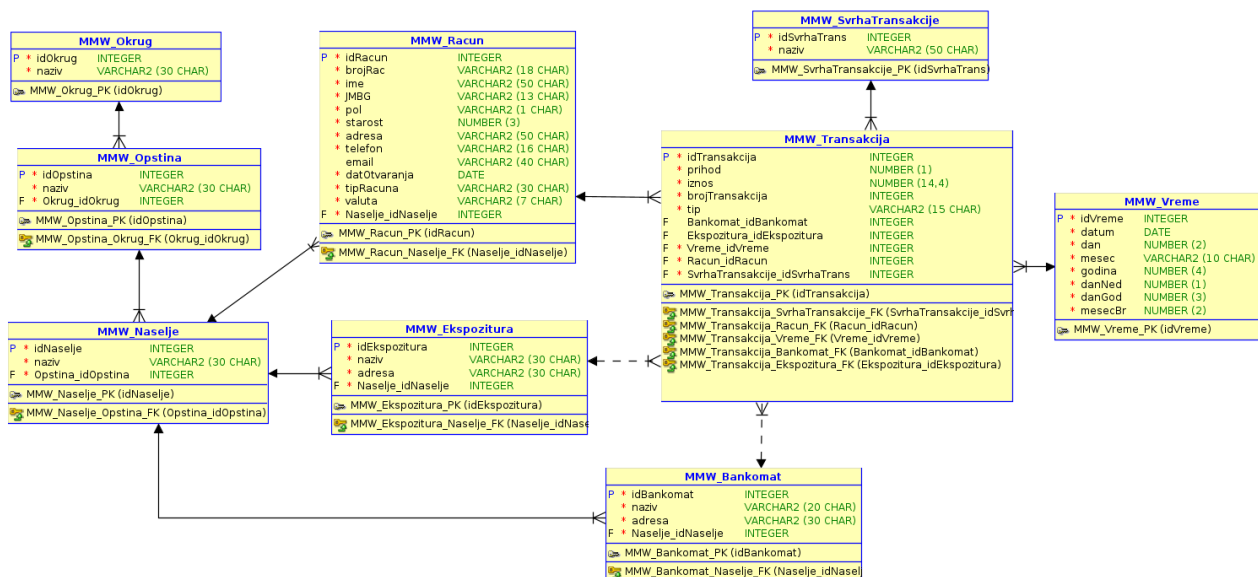
Na slici 3 je prikazan ER model OLAP baze podataka.





Slika 3: ER model OLAP baze podataka

Na slici 4 prikazan je relacioni model OLAP baze podataka.



Slika 4: Relacioni model OLAP baze podataka

U našem slučaju OLAP baza je tipa zvезде, sa glavnom tabelom činjenica Transakcija okružena tabelama dimenzija. Dimenzije su tabela Racun, zatim za tabele Ekspozitura i Bankomat

možemo reći da predstavljaju jednu dimenziju zajedno sa vrednostima iz tabele TipTransakcije koju smo ovde integrisali u tabelu činjenica Transakcija, SvrhaTransakcije predstavlja još jednu dimenziju i na kraju vremenska dimenzija u tabeli Vreme koja je granulirana na veličinu od jednog dana.

Tabele koje smo izgubili su tabele Korisnik, TipRačuna i Valuta koje su se spojile sa tabelom Račun, i već pomenuta tabela TipTransakcije koja se sada nalazi u tabeli Transakcija u vidu diskriminatorskog polja sa vrednostima: Bankomat, Ekspozitura, PostTerminal i Online.

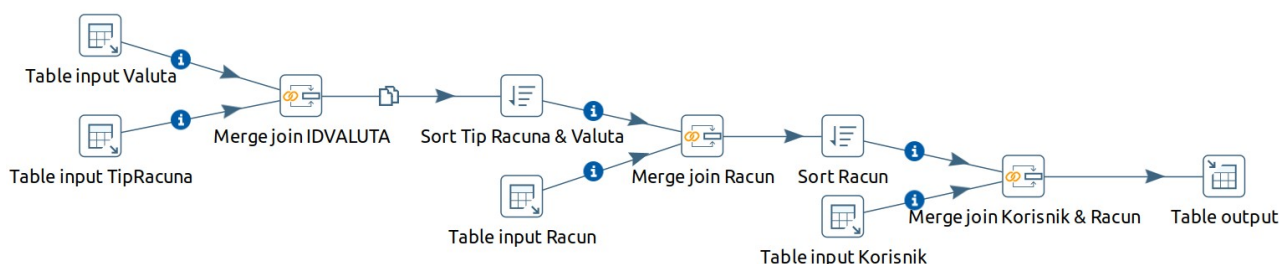
Još jedan uslov po kojem su granulirani podaci u tabeli činjenica jeste po tome da li je reč o prihodu ili rashodu. U OLTP tabeli Transakcija iznos je realan broj, dok je u OLAP tabeli sumiran iznos uvek pozitivan realan broj, a osobina toga šta je prihod a šta rashod se čuva u polju prihod (1 je prihod, a 0 je rashod). Takođe, još jedna agregirana osobina u tabeli činjenica jeste broj transakcija, koja nam daje još jedan pogled na naše podatke.

## Transformacija podataka iz OLTP u OLAP bazu podataka

Služeći se alatom *Spoon* iz programskog paketa *Pentaho Data Integration* izvršili smo transformaciju podataka iz OLTP u OLAP bazu podataka.

Tabele MMW\_Okrug, MMW\_Opstina, MMW\_Naselje, MMW\_Bankomat, MMW\_Ekspozitura i MMW\_SvrhaTransakcije su samo preslikane tabele MMS\_Okrug, MMS\_Opstina, MMS\_Naselje, MMS\_Bankomat i MMS\_Ekspozitura, i ne zavslužuju veću pažnju.

Dobijanje tabele MMW\_Racun koja u sebi sažima četiti tabele iz OLTP baze (MMS\_Korisnik, MMS\_Valuta, MMS\_TipRacuna, MMS\_Racun) je složeniji proces.



Slika 5: Transformacija za dobijanje tabele MMW\_Racun

Najpre smo sortirali tabele MMS\_Valuta i MMS\_TipRacuna po primarnom, odnosno stranom ključu IDVALUTA i spojili ih po istom koristeći transformaciju *Merge Join*. Zatim smo ih sortirali po ključu IDTIPRACUNA. Po istom ključu (IDTIPRACUNA) smo sortirali i tabelu MMS\_Racun i tako ih spojili u koraku *Merge join Racun*. Analogno prethodnim koracima, sortirali smo dosadašnje spojene tabele, i tabelu MMS\_Korisnik po ključu IDKORISNIK i spojili ih u koraku *Merge join Korisnik & Racun*, i tako ih ispisali u tabelu MMW\_Racun.

Dok su kod tabela MMS\_Valuta, MMS\_TipRacuna i MMS\_Racun upiti bili klasični SELECT – FROM – ORDER BY upiti, usled potrebe da se spoje ime i prezime i izvuče starost korisnika, kod tabele MMS\_Korisnik to nije bio slučaj, te ga ovde prikazujemo.

```
SELECT IDKORISNIK, IME || ' ' || PREZIME as IME, JMBG, POL,
CAST(EXTRACT(YEAR FROM CURRENT_DATE) - EXTRACT(YEAR FROM DATRODZENJA) AS INT) AS STAROST,
ADRESA, TELEFON, EMAIL, NASELJE_IDNASELJE
FROM MMS_KORISNIK
ORDER BY IDKORISNIK
```

Slika 6: Select upit za tabelu MMS\_Korisnik

Treba još napomenuti da su informacije o stanju na računu i username odbačeni jer nemaju analitičku svrhu. Stanje računa je trenutna stvar koja se menja iz dana u dan, dok username ima čisto funkcionalnu ulogu u tabeli MMS\_Korisnik, i koja u integrisanom sistemu banke ima ulogu poveznika sa LDAP-om ili sličnim vidom baze gde se čuvaju korisnički nalozi i lozinke.

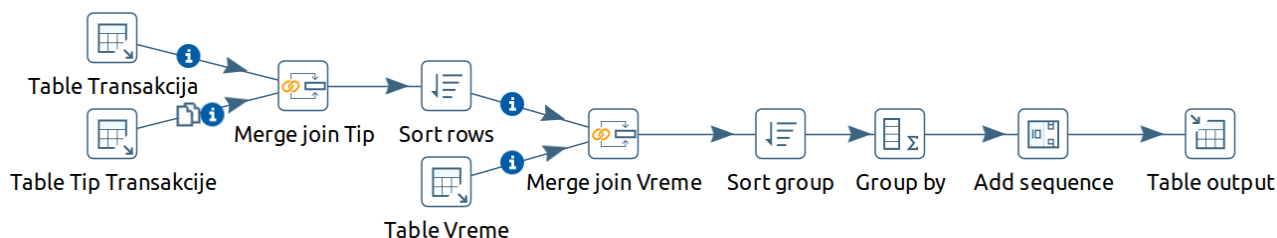
Dobijanje tabele za Vremensku dimenziju MMW\_Vreme izgleda pravolinijski, ali zapravo uključuje veliki broj transformacija. U našem slučaju generisani su datumi u periodu od 10 godina, počevši od 1.1.2018.



Slika 7: Generisanje tabele MMW\_Vreme

Generate time generiše 3653 (tačno 10 godina sa prestupnim godinama) datuma 31.12.2017. Zatim se ono u koraku Calculator sabira sa igenerisanom sekvencom brojeva 1,2,... U koraku Select values izabira se samo novodobijena sekvenca datuma iz koje se u koraku Calculator 2 ugrađenim operacijama izvlače elementi tog datuma: dan, mesec, godina, dan u nedelji, dan u godini, mesec (broj) i podešavaju se odgovarajući fomati za datum. Na kraju, dodaje se sekvenca brojeva (1,2,...) koji će predstavljati primarni ključ u tabeli MMW\_Vreme.

Na kraju generišemo tabelu MMW\_Transakcija. Slika 8.



Slika 8: Generisanje tabele MMW\_Transakcija

```

SELECT IDTRANSAKCIJA, DATUM, IZNOS, EKSPozitura_IDEKSPozitura,
RACUN_IDRACUN, TIPTRANSAKCIJE_IDTIPTRANS,
BANKOMAT_IDBANKOMAT, SVRHATTRANSAKCIJE_IDSVRHATRANS,
CAST (CASE
    WHEN IZNOS > 0 THEN 1
    ELSE 0
END AS INT) AS PRIHOD
FROM MMS_TRANSAKCIJA
ORDER BY TIPTRANSAKCIJE_IDTIPTRANS

```

Slika 9: *SELECT* upit za tabelu *MMS\_Transakcija*

Odmah na početku, u *SELECT* upitu za tabelu *MMS\_Transakcija* izvlačimo polje prihod. Zatim, analogno sa prethodnima, spajamo tabele *MMS\_Transakcija* i *MMS\_TipTransakcije* u koraku *Merge join Tip*. Nakon toga spojenu tabelu sortiramo po datumu kako bismo je povezali (ne spojili!) sa tabelom *MMW\_Vreme* i po datumu u koraku *Merge join Vreme*. Zatim obavljamo najpre sortiranje (važno za pravilno grupisanje), a zatim grupisanje i agregiranje podataka po obeležjima *EKSPOZITURA\_IDEKSPozitura*, *RACUN\_IDRACUN*, *BANKOMAT\_IDBANKOMAT*, *SVRHATTRANSAKCIJE\_IDSVRHATRANS*, *PRIHOD*, *TIP*, *IDVREME* redom. Agregirani podaci su *IZNOS* i *BROJTRANSAKCIJA*. Nakon toga dodaje se sekvenca brojeva koja će predstavljati primarni ključ i rezultati transformacije se upisuju u tabelu *MMW\_Transakcija*.

Time je transformacija podataka iz OLTP u OLAP bazu završena.

# Izveštaji

Zarad poboljšanja organizacije poslovanja i povećanja profita banke se neretko osvrću na nekoliko ključnih faktora u poslovanju. Sledeći izveštaji daju bolje uvide u podatke koji su najvažniji za uspešno poslovanje banke.

1. Promet valuta po okruzima u 2022. godini.
2. Kreditna viabilnost za dinarske kredite.
3. Potrošnja po starosnoj dobi i nameni u prvom polugodištu 2023. godine
4. Broj transakcija po danu u nedelji i mestu transakcije u 2021., 2022., 2023. godini.

Svi izveštaji su generisani pomoću *Pentaho Report Designer*-a.

## Promet valuta po okruzima u 2022. godini

U modernom svetu keš se dobrim delom više ne koristi. Štaviše, u razvijenim zemljama svega 5-10% novca je zapravo keš, ostalo je novac u digitalnoj formi. S toga, situacija u kojoj neko dolazi na šalter u ekspozituru i nije u mogućnosti da podigne svoj novac jer u ekspozituri u tom trenutku nema dovoljno gotovine da mu se traženi novac isplati je skroz moguća. Banke zato zahtevaju od korisnika da se veće svote novca (> 5 000 €) najavljuju unapred, kako ne bi došlo do ponestanka gotovine. Još jedan vid prevencije ove situacije jeste vođenje evidencije o valutnom prometu. Na osnovu njega, mogu se utvrditi obrasci valutnog prometa i predvideti promet u budućnosti.

U ovom [izveštaju](#) smo prikazali ukupan promet po valutama i okruzima u 2022. godini. [Upit za gerisanje izveštaja](#).

## Kreditna viabilnost za dinarske kredite

Glavni izvor prihoda za banku jeste kreditna kamata. Stoga je za banku vrlo važno da pravilno oceni potrebe, a još više mogućnosti svojih klijenata da uzmu kredit. U stvarnosti banka vrši niz provera i ocena pre nego što da ponudu za kredit. U te provere spadaju zaposlenost kao i tip ugovora po kojem je taj klijent zaposlen (određeno, neodređeno), koliko je dugo zaposlen na tom radom mestu, koliko dugo i koliko dobro posluje firma u kojoj je zaposlen (mora poslovati bar 2 godine), kolika su njegova prosečna mesečna primanja na tromesečnom, šestomesečnom i godišnjem nivou, kakav je kreditni rejting žiranta... Od svega navedenog mi smo u mogućnosti da

izračunamo samo pretposlednje, tj. prosečna mesečna primanja na tromesečnom, šestomesečnom i godišnjem nivou.

Za potrebe generisanja ovog [izveštaja](#) u [upitu](#) smo koristili materijalizovan pogled MMW\_KreditnaViabilnost.

## **Potrošnja po starosnoj dobi i nameni u prvom polugodištu 2023. godine**

Kako živimo u modernom, kapitalističkom društvu jasno nam je da onaj koji najbolje može da predvidi potrebe svojih klijenata i da im pruži najbolju uslugu može da ostvari najveći profit. Zato banku interesuje potrošnja svojih korisnika. Na osnovu nje, ona može da uoči obrasce ponašanja svojih korisnika, kategorizuje reklame na osnovu starnosne dobi i plasira svoje usluge ciljnim grupama.

U ovom [izveštaju](#) prikazana je dinarska potrošnja po starosnoj dobi i nameni ostvarena u prvoj polovini 2023. godine. Izveštaj je izgenerisan pomoću sledećeg [upita](#).

## **Broj transakcija po danu u nedelji i mestu transakcije u 2021., 2022., 2023. godini**

Pravilna organizacija poslovanja i raspodela resursa je važna u svakoj delatnosti. Prema rečima jednog poznatog srpskog ekonomiste (parafraziram) “Bogatstvo koje neko poseduje može biti ogromno, ma i najveće na svetu, ali ako nije sposoban za racionalizaciju troškova, taj je osuđen na propast”. S toga racionalizacija troškova je ključan faktor u opstanku poslovanja banke. U ovom izveštaju prikazan je rad bankomata i ekspozitura kroz broj transakcija koje su obavljene u njima po danima u nedelji i po godinama. Ovaj izveštaj nam daje uvid u to koji bankomati i ekspoziture su isplativi, a koje bi trebalo ugasiti, kao i koji dani su najfrekventniji po broju obavljenih transakcija. [Izveštaj](#) je generisan pomoću sledećeg [upita](#).

# Zaključak

OLAP baze podataka sve se više koriste u privredi zbog omogućavanja dubljeg razumevanja poslovne logike, uočavanja mogućnosti razvoja, predviđanja rizika, kao i prilagođavanja organizacija poslovanja. S toga, razvoj ovakvog vida baze podataka predstavlja jednu od vitalnih vještina koju jedan database programer mora da posjeduje.

Ovaj seminarski rad je pružio dubok uvid u suštinske koncepte OLAP (Online Analytical Processing) baza podataka i njihovu ključnu ulogu u modernom poslovanju. Projekat je pokrio procese kreiranja, popunjavanja i analize jedne proste OLAP baze, kao i pisanje odgovarajućih upita za dobijanje korisnih informacija. Kako je ovaj projekat odrađen u ograničenom okruženju sa relativno malom količinom proizvoljno generisanih podataka agregirani podaci se čine kao da su iz OLTP baze podataka, te je pun efekat OLAP baze podataka nažalost izostao.