Univerzitet u Novom Sadu
Prirodno-matematički fakultet
Departman za matematiku i informatiku

Informacione tehnologije



# Seminarski rad iz predmeta Softversko inženjerstvo za sisteme baza podataka

Student: Sanja Šimić

Broj indeksa: 47/19

#### Uvod i opis informacionog sistema

Informacioni sistem za rezervaciju avionskih karata bi bio softverski sistem koji omogućava putnicima da rezervišu avionske karte putem različitih kanala, kao što su veb stranice, mobilne aplikacije ili putem telefona. Ovaj sistem bi mogao da omogući putnicima da pregledaju dostupne letove, biraju željene datume, polazne i odredišne aerodrome, načine plaćanja, veličine prtljaga koje mogu da nose sa sobom, biraju avionska mesta i slično.

Ključna i polazna komponenta za kreiranje ovakvog informacionog sistema bi bila baza podataka. Informacioni sistem koristi bazu podataka za čuvanje informacija o letovima, raspoloživosti putnih mesta, podacima o putnicima, aerodromima i drugim bitnim informacija za njegovo funkcionisanje.

#### **ER Model sistema**

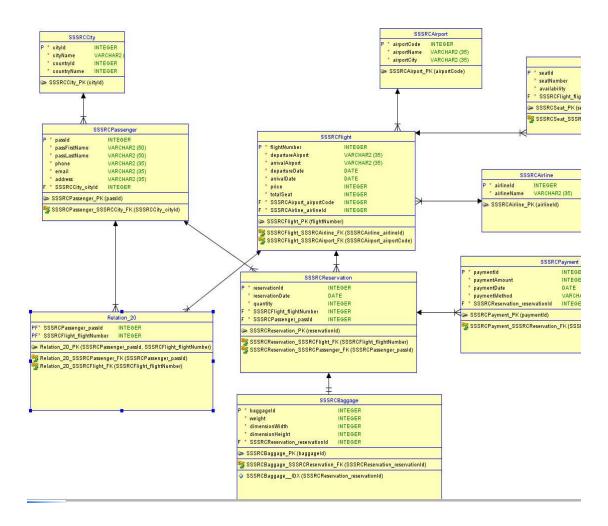
ER (Entity-Relationship) model za rezervaciju avio karata predstavlja vizuelni prikaz entiteta (objekata) i njihovih međusobnih odnosa u sistemu. Ovaj model pomaže u razumevanju strukture podataka i njihovog povezivanja u informacionom sistemu za rezervaciju avio karata. Sledi opis nekih ključnih entiteta i njihovih odnosa u ER modelu:

- 1.Entitet "Grad" u konkretnom projektu tabela pod nazivom SSSRCCity predstavlja konkretan grad sa poljima cityld za identifikaciju grada, cityName što je naziv grad, kao i polja countryld i countryName što su polja koja se odnose na državu kojoj konkretan grad pripada, a od tih polja ćemo u DataWarehouse šemu napraviti odvojenu tabelu.
- 2.Entitet "Putnik" u konkretnom projektu tabela pod nazivom SSSRCPassenger predstavlja putnika koji obavlja rezervacije i sadrži polja poput imena, prezime, email adrese, broj telefona i slično kako bi se identifikovao konkretan putnik.
- 3.Entitet "Let " u konkretnom projektu tabela pod nazivom SSSRCFlight predstavlja let koji se može rezervisati. Sadrži atribute poput broja leta, polaznog i odredišnog aerodroma, vreme dolaska, cenu leta.
- 4. Entitet "Aerodrom" tabela pod nazivom SSSRCAirport predstavlja aerodrom na kom se organizuju letove. Ima atribute kod aerodroma, ime aerodroma i grad u kom se aerodrom

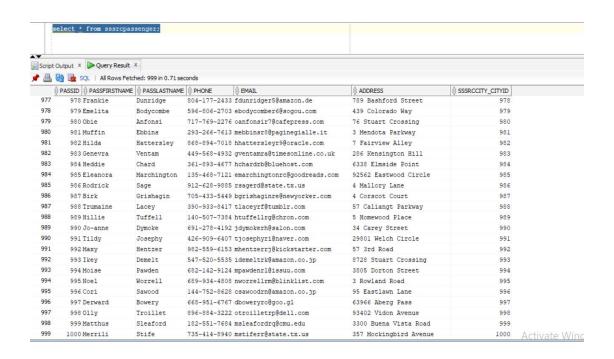
nalazi.

- 5. Entitet "Avionska kompanija" tabela pod nazivom SSSRCAirline predstavlja avionsku kompaniju koja obavlja letove. Sadrži id kompanije i naziv.
- 6. Entitet "Sedište" tabela SSSRCSeat predstavlja raspoloživa sedišta na nekom letu. Sadrži id sedišta i broj.
- 7. Entitet "Rezervacija" tabela SSSRCReservation predstavlja rezervaciju koju je obavio neki korisnik za neki let. Atributi su datum rezervacije, broj rezervacija i id rezervacije.
- 8. Entitet "Placanje" tabela SSSRCPayment predstavlja plaćanje za konkretan let i sadrži atribute kao što su datum plaćanja, način plaćanja.
- 9. Entitet "Prtljag" tabela SSSRCBaggage predstavlja prtljag koji putnik nosi sa sobom i sadrži atribute poput dimenzija prtljaga i težine.

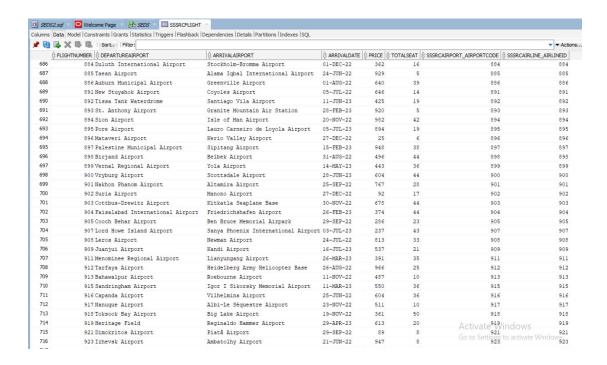
Prikaz ER modela za rezervaciju avionskih karata



ER model je zatim popunjen izgenerisanim podacima.



Na slikama su prikazane tabele koje predstavljaju putnika i letove, kao i deo podataka od kojih su sačinjene.



#### Datawarehouse šema baze podataka

Datawarehouse baza podataka je poseban tip baze podataka koji se koristi za skladištenje velike količine podataka iz različitih izvora. Ova vrsta baze podataka se koristi za podršku procesu donošenja odluka i analitičkim aktivnostima u organizacijama. Neke od glavnih karakteristika su da integrišu podatke iz različitih izvora, kao što su operativne baze podataka, eksterni sistemi, spoljni podaci i drugi relevantni izvori. Često sadrži istorijske podatke koji se koriste za analizu trendova, sezonalnosti i performansi tokom vremena. Ovi podaci se često organizuju u dimenzionalnom modelu sa činjenicama (merama) i dimenzijama (atributima) koji omogućavaju složene upite i analize.

Datawarehouse baza podataka za rezervaciju avionskih karata može biti korisna za analizu i donošenje odluka u vezi sa rezervacijama, praćenjem performansi i trendova, kao i optimizacijom poslovnih procesa.

- 1. Činjeničke tabele "Rezervacije" u konkretnom modelu SSDWReservations Sadrži podatke o rezervacijama, kao i podatke o putnicima i letovima. Nalazi se u centralnom delu modela. Omogućava analize rezervacija u različitim dimenzijama. Sadrži informacije o prihodima generisanim putem rezervacija avionskih karata.
- 2. Tabele dimenzija su "Vreme", "Putnik", "Let", "Grad", "Aerodrom" sve ove tabele omogućava analize rezervacija avionskih karata u odnosu na ono što same tabele predstavljaju. Recimo vremenska dimenzija omogućava da se analize vrše prema nekom datumu ili vremenskom periodu, dok se prostorna dimenzija bavi analizama koje se odnose na neko mesto.

SSDWFlight flightNumber departureAirport arrivalAirport arrivalDate countryld INTEGER VARCHAR2 (30) countryName airportC SSDWCountry\_PK (countryId) totalSeat INTEGER SSDWAir airline SSDWAirport\_airportCode INTEGER SSDWFlight\_PK (flightNumber) 👺 SSDWFlight\_SSDWAirport\_FK (SSDWAirport\_airportCode) SSDWCity cityld cityName SSDWCountry\_countryId INTEGER SSDWCity\_PK (cityId) 👺 SSDWCity\_SSDWCountry\_FK (SSDWCountry\_countryId) SSDWReservations SSDWPassenger quantity INTEGER INTEGER firstResDate year dayin Y lastResDate SSDWPassenger\_passId SSDWTime\_timeId SSDWFlight\_flightNumber email SSDWTi SSDWCity cityld INTEGER SSDWReservations\_PK (reservationId) SSDWPassenger\_PK (passId) SSDWReservations\_SSDWFlight\_FK (SSDWFlight\_flightNumber)
SSDWReservations\_SSDWPassenger\_FK (SSDWR) SSDWPassenger\_SSDWCity\_FK (SSDWCity\_cityId) SSDWReservations\_SSDWTime\_FK (SSDWTime\_timeld)

Datawarehouse model sistema za rezervaciju avionskih karata.

#### **ETL** procesi

ETL procesi su ključni koraci u pripremi i prenosu podataka iz izvora podataka u datawarehouse bazu podataka. Ovi procesi uključuju ekstrakciju podataka iz izvora, transformaciju podataka kako bi se prilagodili potrebama datawarehouse-a i učitavanje transformisanih podataka u samu bazu.

Ekstrakcija podataka podrazumeva izdvajanje podataka iz izvora, koji mogu biti operativne baze podataka i neki drugi izvori. Transformacija podataka uključuje čišćenje, agregiranje, preoblikovanje i obogaćivanje podataka kako bi se prilagodili zahtevima datawarehouse baze podataka.

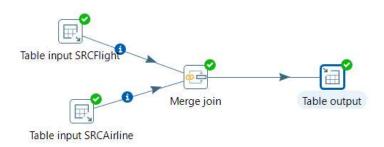
U konkretnom primeru koristili smo Pentahoo Data Integration alat kako bismo iz operativne baze podataka podatke transformisali i na taj način popunili datawarehouse bazu.

Prikazaćemo neke od korišćenih transformacija.



Na slici je prikazana transformacija gde su iz operativne baze, iz tabele koja predstavlja entitet "Grad", preuzeta njena dva atributa countryld i countryName, a odnose se na državu, i ubačeni u tabelu DWCountry. Na ovaj način smo transformacijom popunili datawarehouse tabelu country.

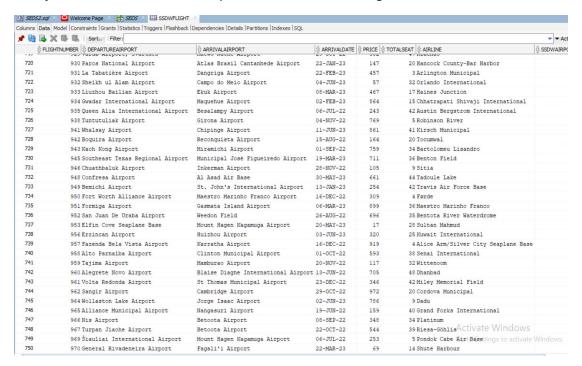
Na sledećoj transformaciji vidimo kako smo popunili tabelu DWFlight.



U ovom slučaju smo koristili dve tabele iz operativne baze kao ulazne izvore kako bismo ispunili zahtev datawarehouse baze podataka. U ovom slučaju, u operativnoj bazi tabela

Airline koja predstavlja avionsku kompaniju predstavljala odvojenu tabelu, ali se u dw modelu pojavljuje kao entitet u okviru tabele DWFlight koja predstavlja let zbog čega smo morali da ih spojimo. Za to smo koristili merge join funkciju koja je zahtevala da ove dve ulazne tabele budu sortirane po zajedničkom obeležju. Na ovaj način smo transformisali podatke i popunili ciljanu tabelu. Na sličan način su popunjene i ostale tabele u DW modelu koje su zahtevale jednostavnije transformacije.

Na taj način smo došli do sledećih podataka u tabeli SSDWFlight.

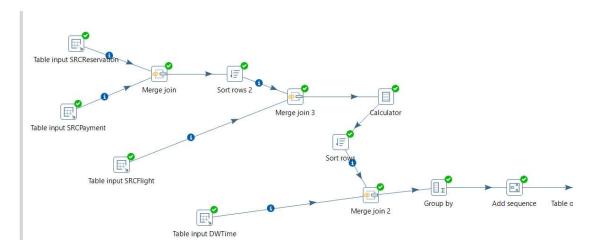


Popunjavanje tabele DWTime koja predstavlja vremensku dimenziju je bilo specifično jer nismo imali ulazni izvor već smo koristeći funkcije koje podržava alat Pentahoo Data Integration izgenerisali podatke i ubacili ih u ciljnu tabelu. Koristili smo funkcije generate rows kako bismo izgenerisali veliki broj redova za neko polje, u našem slučaju datum, zatim i add sequence kako bismo omogućili inkrementaciju konkretnih polja. Ključna funkcija je bila calculator koja nam je omogućila da iz izgenerisanih datum izvučem dan, mesec i godinu u različitim formatima. Na kraju smo opet upotrebili add sequence kako bismo izgenerisali primarni ključ u našoj ciljanoj tabeli.



Popunjavanje činjeničke tabele je zahtevalo četiri ulazne tabele. Prvo smo tabelu koja

predstavlja Rezervaciju i tabelu koja predstavlja Plaćanje sortirali po zajedničkom obeležju, a zatim pomoću merge join-a smo ih ponovo povezali po istom tom zajedničkom obeležju, što je u ovom slučaju bio id rezervacije. Zatim da bismo ta dva obeležja povezali sa trećom tabelom koja predstavlja Let, morali smo ih sortirati po zajedničkom obeležju pa opet primenili merge join. Dalje smo dodali funkciju kalkulator kako bismo mogli da popunimo polje amount koje je predstavljalo ukupnu cenu, a računalo se kao cena leta pomnožena sa poljem quantity koje je predstavljalo broj. Na kraju smo uključili i tabelu Vreme u transformaciju i na taj način popunili ciljnu, činjeničku tabelu.



#### Izveštaji

Kao što smo već spomenuli analitički izveštaji datawarehouse baze podataka pružaju korisnicima dublji uvid u podatke i omogućavaju donošenje informisanih odluka. Ovi izveštaji obuhvataju širok spektar analiza, agregacija i vizualizacija podataka iz datawarehouse-a. Datawarehouse baza podataka omogućava generisanje različitih izveštaja i nadzor nad rezervacijama, prihodima, trendovima putovanja, popunjenosti letova i drugim ključnim pokazateljima performansi.

Za generisanje izveštaja koristili smo Pentahoo Data Reporter alat. Izdvojila sam tri poslovna procesa koje ćemo posmatrati.

1. Analiza rezervacija letova koji su grupisani po danima - u ovom izveštaju smo spajanjem činjeničke tabele i vremenske dimenzije, sabrali polje quantity koje predstavlja broj rezervacija i polje amount koje predstavlja ukupnu cenu i te podatke grupisali po konkretnim datumima.

Koristili smo sledeći upit kako bi došli do željenih podataka.

```
Query Name

Rezervisani letovi po danima

Static Query Query Scripting

Query

1 select datum, sum(quantity), sum(amount)
2 from ssdwreservations r, ssdwtime t
3 where r.ssdwtime_timeid = t.timeid
4 group by datum

5 order by datum
```

Dobili smo sledeći izveštaj.

June 14, 2023 @ 07:17

# Rezervacija letova grupisana po danima

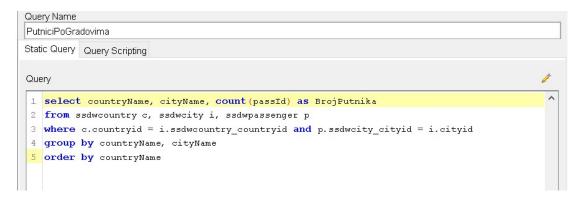
	Ukupan iznos	Kolicina
	1,250	Almitecanis
	1,250	81
DATUM14 Jun 2022, 00:00:00		
	Ukupan iznos	Kolicina
	1,390	141
DATUM16 Jun 2022, 00:00:00		
	Ukupan iznos	Kolicina
	410	13
DATUM17 Jun 2022, 00:00:00	VV	
	Ukupan iznos	Kolicina
	1,548	101
DATUM18 Jun 2022, 00:00:00		
	Ukupan iznos	Kolicina
	2,376	130
DATUM20 Jun 2022, 00:00:00		124
	Ukupan iznos	Kolicina
	765	31
DATUM21 Jun 2022, 00:00:00	0.578	
DAT GM21 GM1 2022, 00:00:00	Ukupan iznos	Kolicina
	1,036	203
DATUM22 I 2022 00:00:00	1,000	200
DATUM22 Jun 2022, 00:00:00	I DAMES I MARKET	Mallala
	Ukupan iznos	Kolicina
	2,229	142
DATUM23 Jun 2022, 00:00:00	Account of the control of the contro	
	Ukupan iznos	Kolicina
	3,053	227
DATUM24 Jun 2022, 00:00:00		
	Ukupan iznos	Kolicina
	973	89

Wed Jun 14 19:17:26 CEST 2023



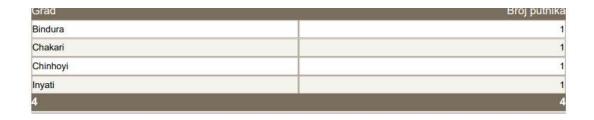
2. Analiza rezervacije letova koja je grupisana prema drzavama i gradovima - u ovom slučaju smo spajanjem tabela koje predstavljaju Grad, Državu i Putnike došli do sledećih analiza. Grupisali smo podatke prema imenu države i imenu grada i za iste izlistali broj putnika. Dodali smo još izbrojavanje gradova u konkretnoj državi, kao i na kraju mali grafikon koji vizuelno prikazuje ove analize.

Primenom sledećeg upita smo došli do željenih analiza.



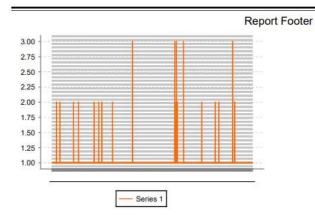
Dobili smo sledeći izveštaj.





Wed Jun 14 19:46:37 CEST 2023

June 14, 2023 @ 07:46



3. Ukupni broj i ukupna cena letova grupisana po kupcima i broju leta - pri ovoj analizi smo spajali tri tabele, činjeničku koja predstavlja Rezervacije i dve tabele dimenzija koje predstavljaju Putnika i Let. Grupisali smo dobijene podatke prema putniku i letu kako bismo dobili informacije ukupna količina rezervisanih letova i koja je ukupna cena letova koje je neki putnik rezervisao za let.

Upit koji smo primenili.

```
Cetvrtilzvestaj

Static Query Query Scripting

Query

1 select passName, flightNumber, sum(quantity) as UkupnaKolicina, sum(price * quantity) as ^
2 from ssdwreservations r, ssdwflight f, ssdwpassenger p
3 where r.ssdwflight_flightnumber = f.flightnumber and r.ssdwpassenger_passid = p.passid
4 group by passName, flightNumber
5 order by passName
6
```

### Dobili smo sledeći izveštaj.

		Broj leta£395.0
	Ukupna cena	Ukupni br
	15,405	
PutnikAbagaelBovingdon		
		Broj leta£700.
	Ukupna cena	Ukupni b
	9,400	
PutnikAbbaWynes		
		Broj leta£371.
	Ukupna cena	Ukupni b
	61,494	2
PutnikAbbeLusty		
		Broj leta£259.
	Ukupna cena	Ukupni b
	44,712	
PutnikAddieSamsonsen		
PutnikAddieSamsonsen		Broj leta£874.
PutnikAddieSamsonsen	Ukupna cena	Broj leta£874. Ukupni b
PutnikAddieSamsonsen	Ukupna cena 6,136	
PutnikAddieSamsonsen  PutnikAdelaidaClucas	The state of the s	
	The state of the s	
	The state of the s	Ukupni b

	Ukupna cena	Ukupni broj
	19,166	37
PutnikYulmaBugg		
		Broj leta£509.00
	Ukupna cena	Ukupni broj
	30,192	68
PutnikZachariahPiser		
		Broj leta£667.00
	Ukupna cena	Ukupni broj
	20,454	21
PutnikZebGregori		
		Broj leta£694.00
	Ukupna cena	Ukupni broj
	11,160	31
PutnikZedekiahLahrs		
		Broj leta£388.00
	Ukupna cena	Ukupni broj
	39,240	72
PutnikZelmaPennoni		
		Broj leta£391.00
	Ukupna cena	Ukupni broj
	71,000	71
PutnikZitellaDottrell		
		Broj leta£91.00
	Ukupna cena	Ukupni broj
	7,312	16

## Zaključak

Datawarehouse baza podataka pruža mnoge prednosti u odnosu na tradicionalne operativne baze podataka. Datawarehouse baza podataka integriše podatke iz različitih izvora i formata. To omogućava konzistentnost i doslednost podataka, eliminisanje duplikata i poboljšanje kvaliteta podataka. Integracija podataka takođe olakšava složene analize i izveštavanje preko više izvora podataka. Zaista je olakšala analize koje su bile neophodne za gore opisani informacioni sistemi.