logo-fesb

SVEUČILIŠTE U SPLITU

FAKULTET ELEKTROTEHNIKE, STROJARSTVA I BRODOGRADNJE

SKLADIŠTA PODATAKA

Izrada skladišta podataka za iznajmljivanje smještaja

Studenti:

Ani Stjepić

Duje Senta

Stipe Perković

Split, 2018.

Sadržaj

[UVOD 2](#_Toc518635379)

[1. Transakcijska baza podataka 3](#_Toc518635380)

[1.1. Entitet State 4](#_Toc518635381)

[1.2. Entitet User 4](#_Toc518635382)

[1.3. Entitet ApartmentOwnership 5](#_Toc518635383)

[1.4. Entitet Apartment 5](#_Toc518635384)

[1.5. Entitet Room 6](#_Toc518635385)

[1.6. Entitet Visit 6](#_Toc518635386)

[1.7. Entitet VisitPeriod 7](#_Toc518635387)

[1.8. Entitet Receipt 7](#_Toc518635388)

[1.9. Entitet PaymentDate 8](#_Toc518635389)

[2. Skladište podataka 9](#_Toc518635390)

[2.1. Osnovne informacije o skladištima podataka 9](#_Toc518635391)

[2.2. Skladište podataka za iznajmljivanje smještaja 9](#_Toc518635392)

[2.2.1. Dimenzija Receipt 10](#_Toc518635393)

[2.2.2. Dimenzija State 11](#_Toc518635394)

[2.2.3. Dimenzija PaymentDate 11](#_Toc518635395)

[2.2.4. Dimenzija VisitPeriod 11](#_Toc518635396)

[3. Prebacivanje podataka iz transakcijske baze u skladište podataka 13](#_Toc518635397)

[3.1. Prebacivanje podataka u dimenzije 13](#_Toc518635398)

[3.1.1. Odabir tablice u transakcijskoj bazi podataka 14](#_Toc518635399)

[3.1.2. Lookup 16](#_Toc518635400)

[3.1.3. Conditional Split 17](#_Toc518635401)

[3.1.4. Odabir dimenzije u skladištu u koju se prebacuju podaci 17](#_Toc518635402)

[3.2. Popunjavanje tablice činjenica u skladištu podataka 18](#_Toc518635403)

[3.2.1. Izbor tablica i atributa koji sudjeluju u punjenju tablice činjenica 18](#_Toc518635404)

[3.2.2. Lookup 19](#_Toc518635405)

[3.2.3. Left Join 20](#_Toc518635406)

[3.2.4. Upisivanje podataka u tablicu činjenica 21](#_Toc518635407)

[4. Izrada kocke 22](#_Toc518635408)

[5. Zaključak 31](#_Toc518635409)

# UVOD

U sklopu seminarskog rada iz kolegija Skladišta podataka napravili smo skladište podataka za iznajmljivanje smještaja. Za izradu navedenog skladišta koristili smo SQL Server 2017 sa popratnim alatima za dizajn i razvoj transakcijske baze podataka te skladišta, prebacivanja podataka iz tablica u transakcijskoj bazi u tablice dimenzija i činjenica u skladištu podataka, te u konačnici za razvoj kocke.

Prilikom izrade skladišta podataka prije svega je potrebna dobro organizirana relacijska baza podataka, u suprotnom se izrada skladišta može zakomplicirati. Nadalje, potrebno je dobro razumjeti poslovni proces koji želimo pratiti.

Cilj ovog seminarskog rada je u kratkom vremenu dobiti izvještaje o računima određenih posjeta na osnovu datuma plaćanja računa, periodu posjeta ili države u kojoj se korisnik nalazi.

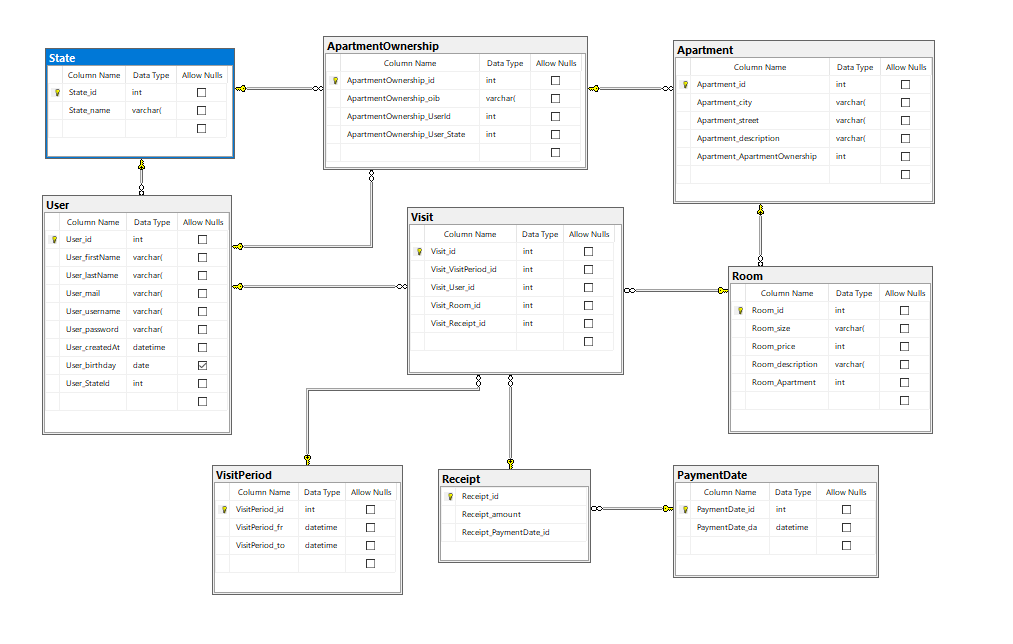
U nastavku rada biti će detaljno opisan postupak izrade transakcijske baze podataka, skladišta podataka, prebacivanja podataka iz transakcijske baze u skladište, izrada kocke te generiranje izvještaja iz iste.

# Transakcijska baza podataka

Transakcijska baza se sastoji od sljedećih entiteta:

* State
* User
* ApartmentOwnership
* Apartment
* Room
* Visit
* Visit Period
* Receipt
* Payment Date

Na slici ispod (Slika 1‑1) može se vidjeti dijagram baze s navedenim entitetima.

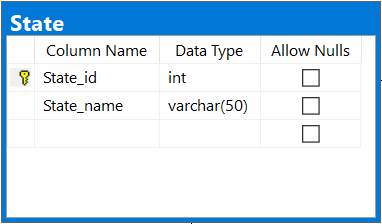


Slika 1‑1 Dijagram transakcijske baze podataka

## Entitet State

Ovaj entitet sadrži podatak o imenu države.

Svi njegovi atributi i tipovi podataka atributa prikazani su na slici (Slika 1‑2) ispod.

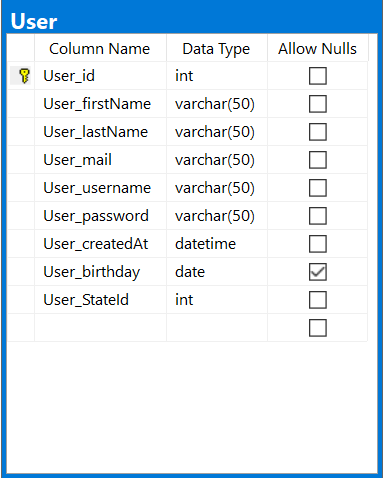


Slika 1‑2 Entitet State

## Entitet User

Ovaj entitet sadrži podatke o korisniku te strani ključ na entitet State.

Svi njegovi atributi i tipovi podataka atributa prikazani su na slici ispod (Slika 1‑3).

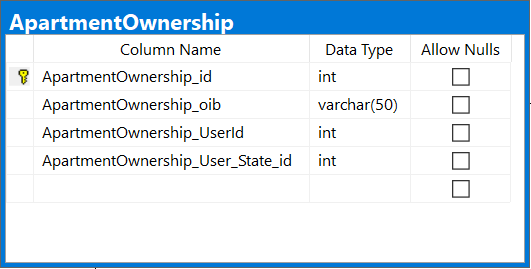


Slika 1‑3 Entitet User

## Entitet ApartmentOwnership

Ovaj entitet sadrži podatke o vlasništvu smještaja te strani ključ na entitet User. Koristi se kako bi se moglo ustanoviti je li određeni korisnik vlasnik nekog od smještaja.

Svi njegovi atributi i tipovi podataka atributa prikazani su na slici ispod (Slika 1‑4).

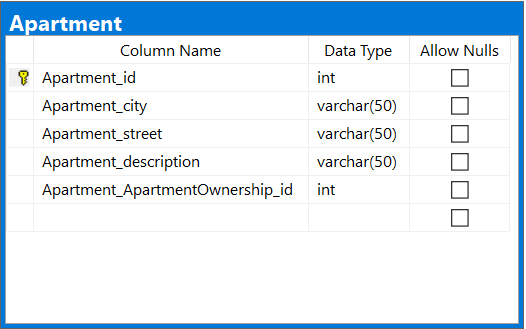


Slika 1‑4 Entitet ApartmentOwnership

## Entitet Apartment

Ovaj entitet sadrži podatke o smještaju te strani ključ na ApartmentOwnership kako bi se moglo ustanoviti tko je vlasnik određenog smještaja.

Svi njegovi atributi i tipovi podataka atributa prikazani su na slici ispod (Slika 1‑5).

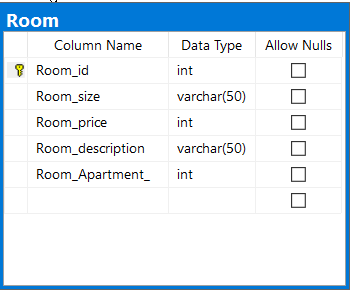


Slika 1‑5 Entitet Apartment

## Entitet Room

Ovaj entitet sadrži podatke o sobi unutar smještaja , također sadrži strani ključ na smještaj (apartman) kojemu pripada.

Svi njegovi atributi i tipovi podataka atributa prikazani su na slici ispod (Slika 1‑6).

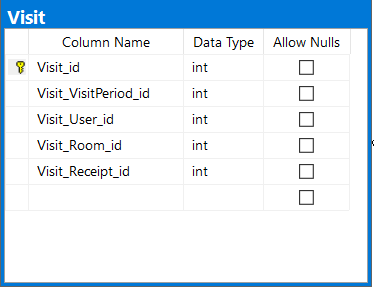


Slika 1‑6 Entitet Room

## Entitet Visit

Ovaj entitet sadrži podatke o posjeti određenog smještaja, s tim da ovisi isključivo o drugim entitetima, tj sadrži strane ključeve entiteta : VisitPeriod, User, Room, Receipt.

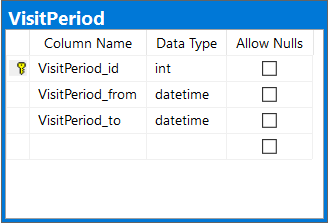
Svi njegovi atributi i tipovi podataka atributa prikazani su na slici (Slika 1‑7).



Slika 1‑7 Entitet Visit

## Entitet VisitPeriod

Ovaj entitet sadrži podatke o datumu početka i završetka neke posjete. Svi njegovi atributi i tipovi podataka atributa prikazani su na slici ispod (Slika 1‑8).

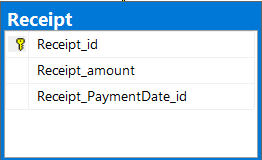


Slika 1‑8 Entitet VisitPeriod

## Entitet Receipt

Ovaj entitet sadrži podatke o ukupnoj cijeni te strani ključ na entitet PaymentDate kako bi se znalo je li račun plaćen ili ne.

Svi njegovi atributi i tipovi podataka atributa prikazani su na slici ispod (Slika 1‑9).

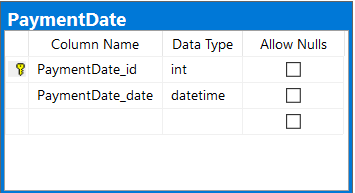


Slika 1‑9 Entitet Receipt

## Entitet PaymentDate

Ovaj entitet sadrži podatak o datumu plaćanja računa.

Svi njegovi atributi i tipovi podataka atributa prikazani su na slici ispod (Slika 1‑10).



Slika 1‑10 Entitet PaymentDate

# Skladište podataka

## Osnovne informacije o skladištima podataka

Osnovni razlog zbog kojeg se rade skladišta podataka je brzo generiranje izvještaja. To je posebno bitno za velike firme čije baze podataka sadrže ogromnu količinu informacija. U takvim firmama traženje nekog izvještaja iz transakcijske baze podataka može trajati duže vremena. Zbog toga se može dogoditi da se propusti neka dobra poslovna prilika ili da generiranje izvještaja optereti server pa ostali korisnici ne mogu normalno raditi. Zbog navedenih razloga, firmama koje rade sa velikom količinom podataka se isplati investirati u skladište podataka jer će im se ta investicija prije ili poslije isplatiti.

Svako skladište podataka sastoji se obično od više tablica dimenzija i jedne ili više tablica činjenica. U dimenzije se upisuju podaci koji su nam potrebni za praćenje poslovnog procesa.

Obično se ne prebacuju svi podaci iz transakcijske baze u dimenzije. Tablice činjenica sastoje se od stranih ključeva tablica dimenzija i mjera. Mjere su posebno značajne jer one čine osnovu svakog izvještaja.

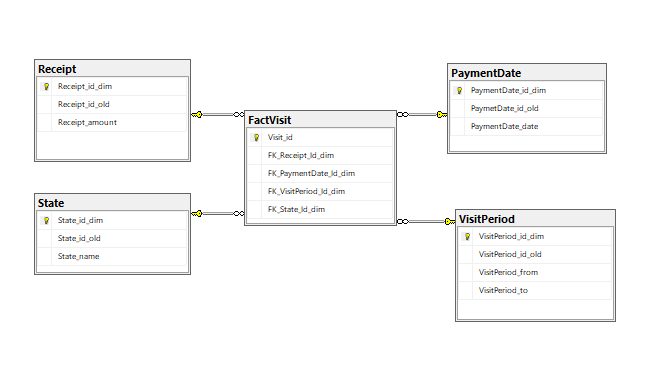
Za dizajn skladišta podataka mogu se koristiti dva pristupa: model zvijezde i model pahuljice. Model pahuljice je nešto kompliciraniji od modela zvijezde, te su kod takvog modela tablice dimenzija povezane sa drugim tablicama dimenzija. Kod modela zvijezde, svaka tablica dimenzija je povezana jedino sa tablicom činjenica.

## Skladište podataka za iznajmljivanje smještaja

Skladište podataka koje smo napravili sastoji se od 4 tablice dimenzija i 1 tablice činjenica. Koristili smo model zvijezde. Tablice dimenzija su:

* Receipt
* State
* PaymentDate
* VisitPeriod

Dijagram skladišta podataka za iznajmljivanje smještaja prikazan je na slici ispod (Slika 2‑1).



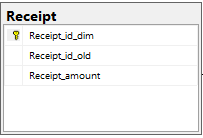
Slika 2‑1 Skladište podataka

U svaku dimenziju kao atribut se dodaje polje oldID u koje će se spremati primarni ključ tog unosa iz transakcijske baze u svrhu sprječavanja višestrukog kopiranja istih podataka iz transakcijske baze u skladište podataka.

### Dimenzija Receipt

Ova dimenzija sadrži osnovnu informaciju o računu, tj. cijenu.

Svi atributi koje ova dimenzija sadrži prikazani su na slici ispod (Slika 2‑2).

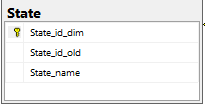


Slika 2‑2 Dimenzija Receipt

### Dimenzija State

Ova dimenzija sadrži podatak o imenu države.

Svi atributi i tipovi podataka koje sadrži ova dimenzija prikazani su na slici ispod (Slika 2‑3).

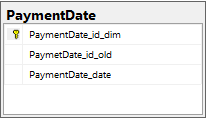


Slika 2‑3 Dimenzija State

### Dimenzija PaymentDate

Ova dimenzija sadrži podatak o datumu plaćanja određenog računa.

Svi atributi i tipovi podataka koje sadrži ova dimenzija prikazani su na slici ispod (Slika 2‑4).

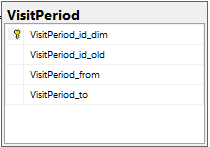


Slika 2‑4 Dimenzija PaymentDate

### Dimenzija VisitPeriod

Ova dimenzija sadrži podatke o datumu dolaska i odlaska iz određenog smještaja.

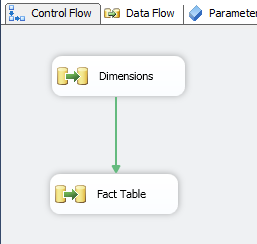
Svi atributi i tipovi podataka koje sadrži ova dimenzija prikazani su na slici ispod (Slika 2‑5).



Slika 2‑5 Dimenzija VisitPeriod

# Prebacivanje podataka iz transakcijske baze u skladište podataka

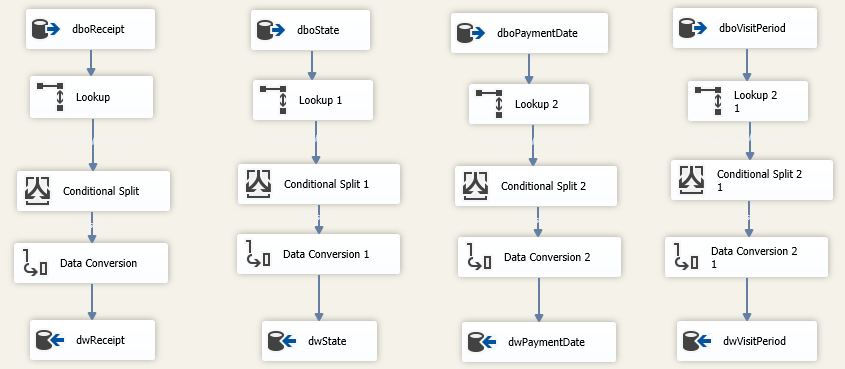
Nakon što smo napravili transakcijsku bazu podataka i skladište podataka, potrebno je podatke prebaciti iz tablica(entiteta) transkacijske baze podataka u odgovarajuće dimenzije skladišta podataka. Prebacivanje podataka iz transakcijske baze podataka u skladište podataka vršimo u integracijskom servisu SQL Server Data Tools. Podatke je potrebno prilagoditi na način da budu pogodni za prebacivanja. Također, vrlo je bitno odrediti koji podaci se kopiraju na koje mjesto. Postupak prebacivanja podataka iz tablica u transakcijskoj bazi podataka u skladište podataka prikazan je u nastavku korak po korak. Nakon prebacivanja podataka u tablice dimenzija, potrebno je odgovarajuće podatke prebaciti u tablicu činjenica. Na slici ispod (Slika 3-1) je prikazano kako to izgleda u osnovnom obliku.



Slika 3‑1

## Prebacivanje podataka u dimenzije

Postupak prebacivanja podataka u dimenzije u skladištu podataka iz tablica u transakcijskoj bazi sastoji se od nekoliko međusobno ovisnih i povezanih koraka. Svaki korak je bitan, i bez njega ovaj postupak ne bi funkcionirao. Generalni prikaz se nalazi na slici ispod (Slika 3-2).



Slika 3‑2 Prikaz prebacivanja podataka u dimenzije

Iz prethodne slike se može zaključiti da je potrebno odvojeno definirati skup koraka za prebacivanje podataka za svaku dimenziju. Također, moramo biti spojeni i sa transakcijskom bazom podataka i sa skladištem podataka.

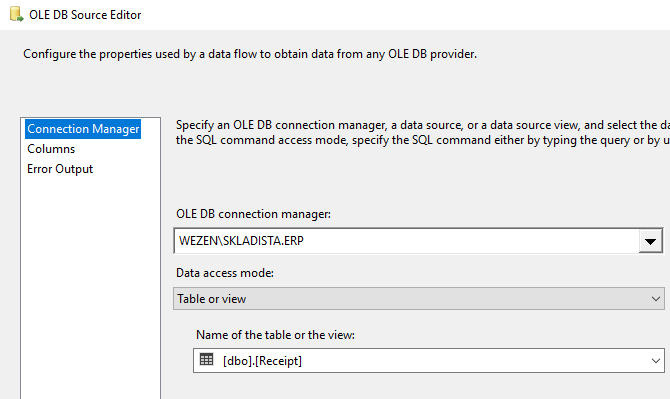
Koraci za prebacivanje podataka iz transakcijske baze podataka u skladište podataka su:

1. Odabir tablice u transakcijskoj bazi podataka
2. Lookup
3. Conditional Split
4. Odabir tablice dimenzija u koju se podaci prebacuju

Navedeni koraci su detaljno opisani u nastavku ovoga rada. Prikazat ćemo postupak za prebacivanje podataka iz tablice VisitPeriod u transakcijskoj bazi u dimenziju VisitPeriod\_DW u skladištu podataka. Navedeni postupak je analogan i za sve druge tablice, pa nema potrebe da za svaku tablicu detaljno prikazujemo postupak.

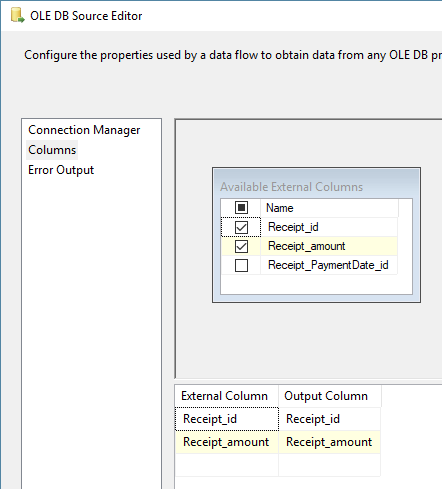
### Odabir tablice u transakcijskoj bazi podataka

U ovom koraku potrebno je odabrati i povezati se sa transakcijskom bazom podataka, te odabrati tablicu iz transakcijske baze podataka, to je prikazano na sljedećoj slici (Slika 3-3).



Slika 3‑3 Odabir tablice iz transakcijske baze

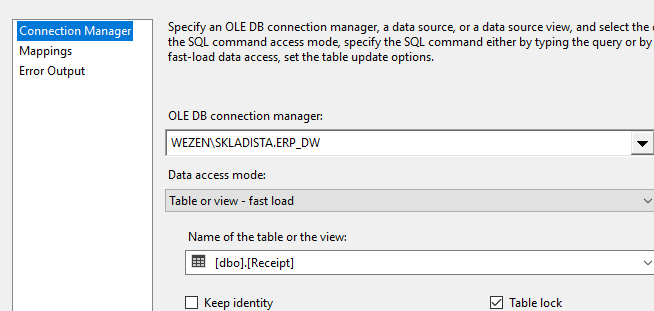
Nakon toga potrebno je označiti atribute u tablici koju smo odabrali. Označavanje atributa je prikazano na slici ispod (Slika 3-4).



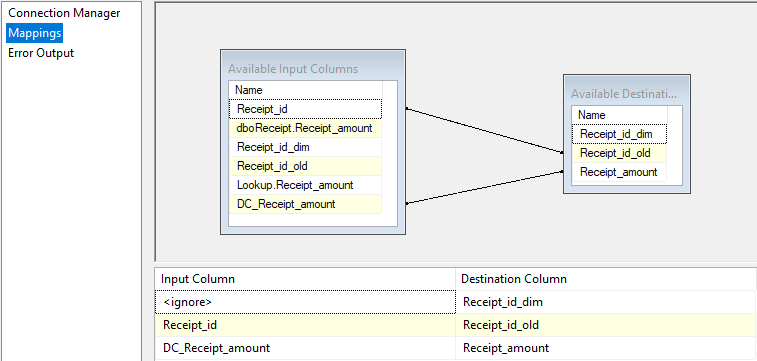
Slika 3‑4 Označavanje atributa

### Lookup

U ovome koraku povezujemo atribut oldId iz dimenzije sa ID-om koji su ti podaci imali u transakcijskoj bazi podataka. Ovaj korak je naročito značajan zbog toga što osigurava da se isti podaci ne prebacuju dva ili više puta u tablicu dimenzija u skladištu podataka. Prije svega, potrebno se povezati sa bazom gdje se nalazi skladište podataka i izabrati dimenziju u koju se podaci prebacuju. Na slijedećim slikama (Slika 3-5 i Slika 3-6) je prikazan način na koji se to radi.



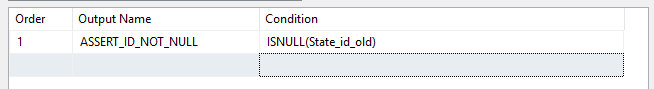
Slika 3‑5



Slika 3‑6

### Conditional Split

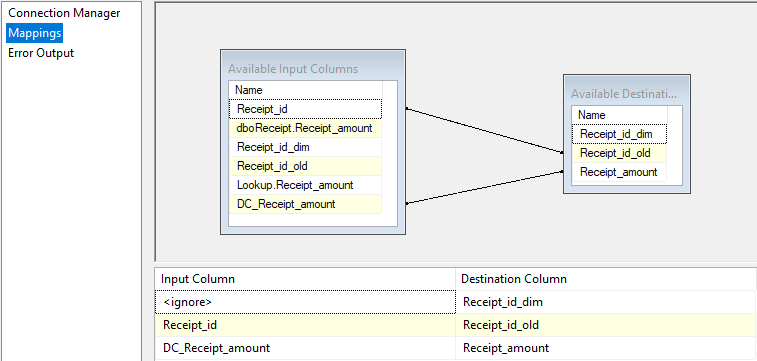
U ovome koraku provjerava se jesu li podaci već prije prebačeni u skladište podataka. Ne dozvoljava da se isti podaci dva ili više puta prebacuju u skladište podataka, prikazano je na slijedećoj slici (Slika 3-7).



Slika 3‑7 Conditional Split

### Odabir dimenzije u skladištu u koju se prebacuju podaci

Na kraju je potrebno odabrati dimeniju u skladištu u koju se prebacuju podaci. Potrebno je odrediti mapiranja podataka. Odnosno, odrediti koji atributi iz tablice u transakcijskoj bazi podataka se prebacuju u koje atribute u tablici dimenzija. Ovaj korak je prikazan na slici (Slika 3-8).



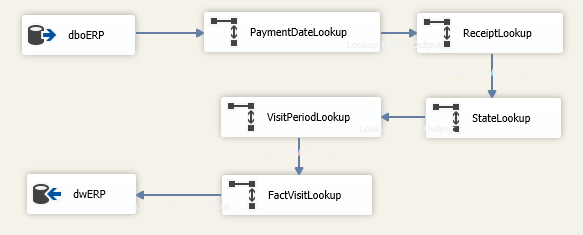
Slika 3‑8 Mapiranje podataka

Nakon što smo definirali korake za prebacivanje podataka u dimenzije, i pokrenuli postupak prebacivanja, podaci koje smo imali u tablicama u transakcijskoj bazi podataka su se prebacili u odgovarajuće tablice dimenzija. Nakon toga, potrebno je popuniti tablicu činjenica iz tablica dimenzija. Kao što smo ranije naglasili, tablica činjenica se sastoji od stranih ključeva i mjera. Dakle, sadrži samo brojčane vrijednosti.

## Popunjavanje tablice činjenica u skladištu podataka

U ovome dijelu prikazat ćemo na koji način se puni tablica činjenica. Budući da mi pratimo zaradu iznajmljivača, datum plačanja računa, period u kojem je apartman iznajmljen i državu iz koje dolaze gosti, naša tablica činjenica sadrži strane ključeve dimenzija Receipt, State, paymentDate i VisitPeriod. Mjera će biti iznos računa.

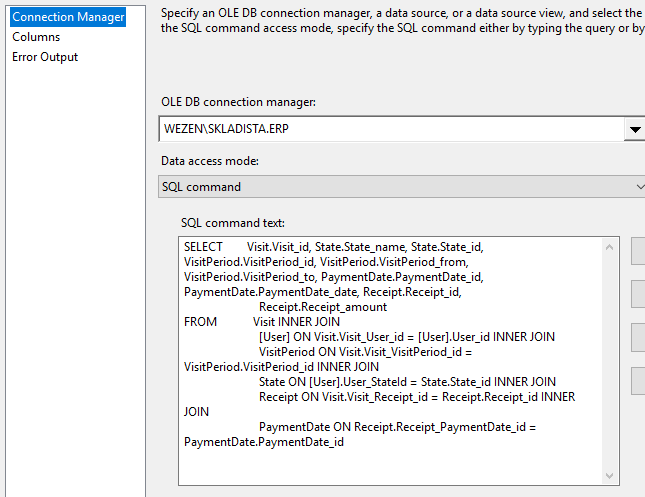
Postupak punjenja tablice činjenica se također sastoji od više koraka, koji su međusobno povezani i ovisni. U koracima se biraju podaci koji se prebacuju, i prilagođavaju za upis u tablicu činjenica. Postupak se sastoji od sljedećih koraka: Izbor tablica i atributa koji sudjeluju u punjenju tablice činjenica, Lookup, Left Join, Conditional Split, Upis u zablicu činjenica. Postupak je prikazan na slijedećoj slici (Slika 3-9).



Slika 3‑9 Postupak popunjavanja tablice činjenica

### Izbor tablica i atributa koji sudjeluju u punjenju tablice činjenica

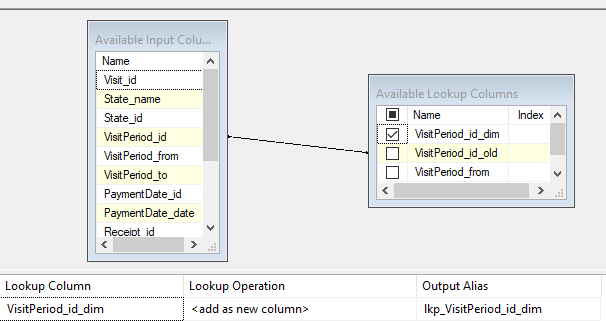
U ovome koraku se pomoću SQL upita definiraju tablice i atributi koji učestvuju u popunjavanju tablice činjenica. Prikaz toga se nalazi na slici (Slika 3-10).



Slika 3‑10 Definiranje tablica i atributa

### Lookup

U ovome koraku se povezuju ID iz tablice u transakcijskoj bazi sa oldId u dimenzijama. Također, primarnim ključevima u dimenzijama se dodaje prefiks lkp. Dalje se propuštaju samo oni ID-evi koji imaju prefiks lkp. Postupak je prikazan na slici (Slika 3-11).

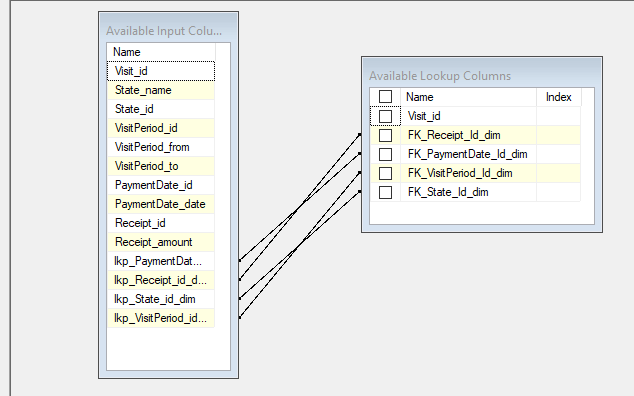


Slika 3‑11 Lookup

Ovaj korak je potrebno napraviti za sve dimenzije koje sudjeluju u popunjavanju tablice činjenica.

### Left Join

Sada se za mjere dodaje prefiks test. Također povezuju se atributi koji imaju prefiks lkp sa odgovarajućim atributima u tablici činjenica. Postupak je prikazan na slici (Slika 3-12).



Slika 3‑12

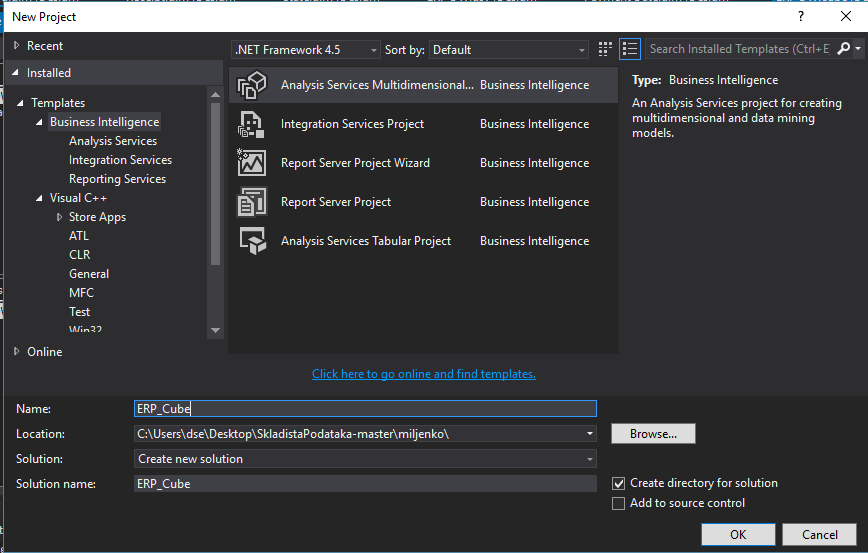
### Upisivanje podataka u tablicu činjenica

Sada se rade mapiranja podataka. Odnosno, označava se gdje se pojedini podaci upisuju. Postupak prikazan na prethodnoj slici (Slika 3-12).

Nakon što smo definirali korake, pokrećemo postupak upisivanja u tablicu činjenica. Kada se podaci upišu u tablicu činjenica, možemo generirati izvještaje koji su znatno brži od generiranja izvještaja nad transakcijskom bazom podataka. Međutim, ukoliko želimo izvještaje koji su još brži nego u skladištu, moramo napraviti kocku. Postupak izrade kocke prikazan je u sljedećem poglavlju.

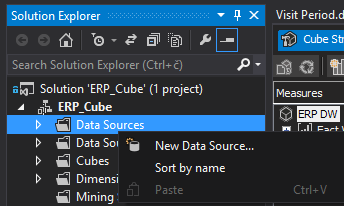
# Izrada kocke

Postupak pravljenja kocke počinje kreiranjem SSAS projekta u BIDS-u. Prikazano na slici (Slika 4-1).



Slika 4‑1 Kreiranje projekta

Zatim, potrebno je definirati izvorište podataka. Postupak prikazan kroz slijedeće slike (Slika 4-2, Slika 4-3, Slika 4-4, Slika 4-5).

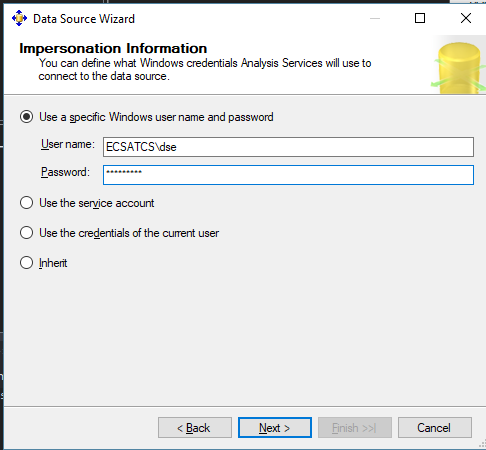


Slika 4‑2 Odabir Data Source-a

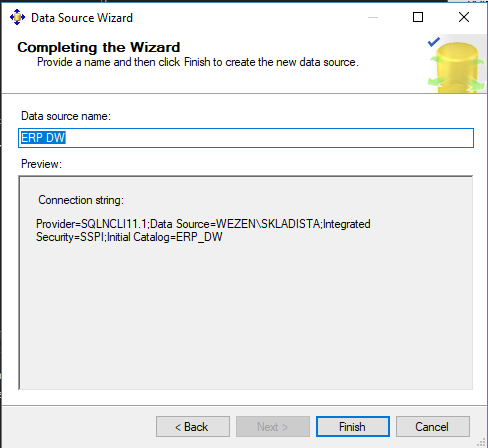
Treba odabrati bazu podataka u kojoj se nalazi naše skladište.

Slika 4‑3 Odabir baze podataka

Odabrali smo Windows autentifikaciju. Moramo upisati ime računala i password.

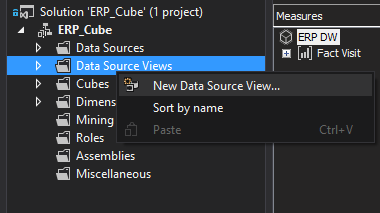


Slika 4‑4 Povezivanje sa Data Source-om

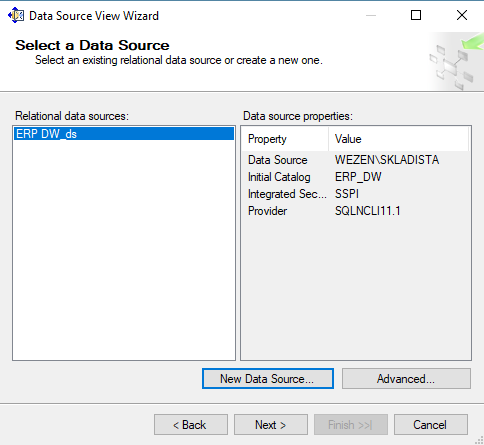


Slika 4‑5 Connection string za pristup serveru i bazi podataka u kojoj se nalazi skladište

Nakon što smo uspješno izvršili prethodni postupak, potrebno je definirati poglede. Način na koji se to radi prikazan je na slikama ( Slika 4-6, Slika 4-7).

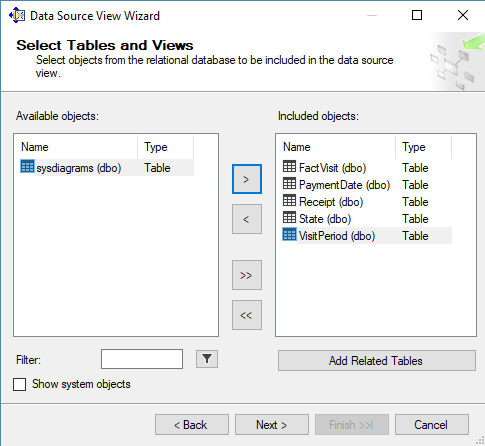


Slika 4‑6

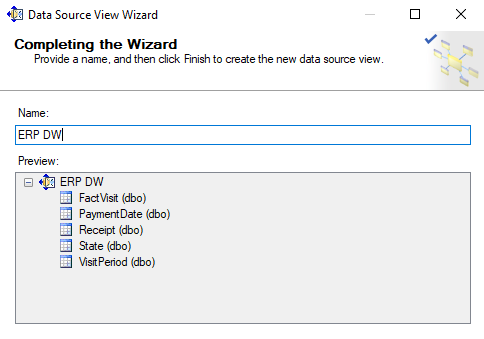


Slika 4‑7

Nakon toga je potrebno izabrati tablicu činjenica i tablice dimenzija koje se nalaze u skladištu. Postupak je prikazan na slikama (Slika 4-8, Slika 4-9).

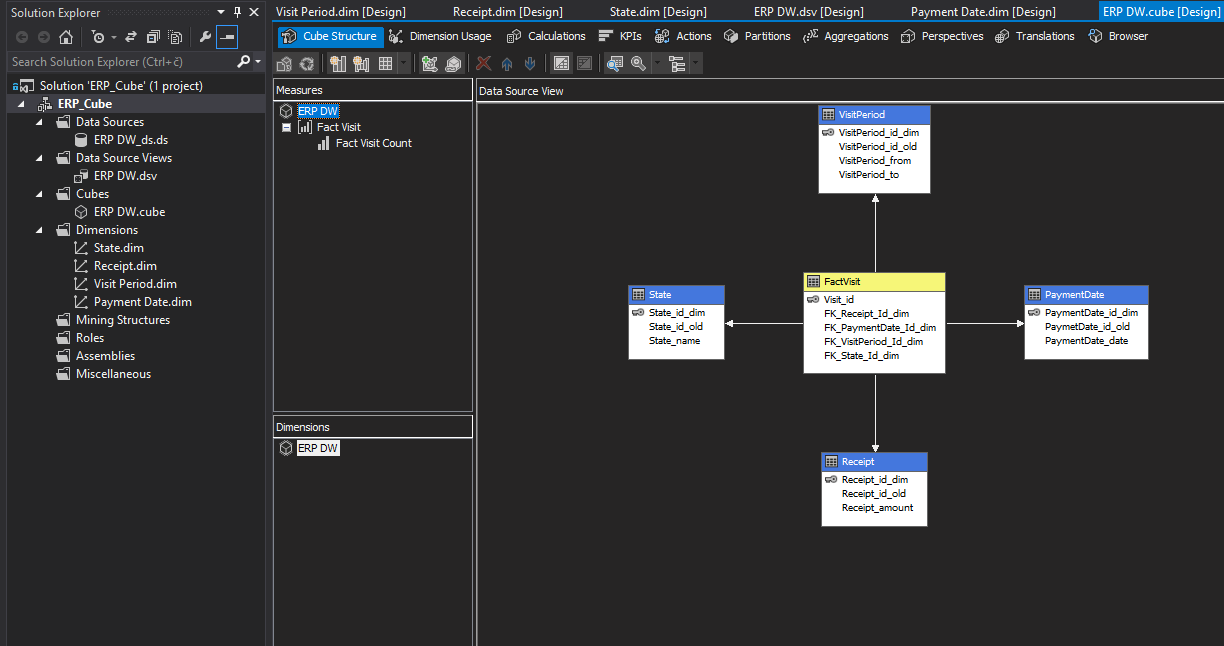


Slika 4‑8 Odabir tablice činjenica i tablice dimenzija



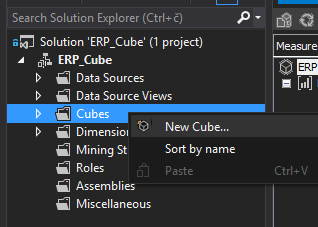
Slika 4‑9 Tablica činjenica i tablice dimenzija

Nakon svih navedenih koraka, pogled izgleda kao na slici (Slika 4-10).



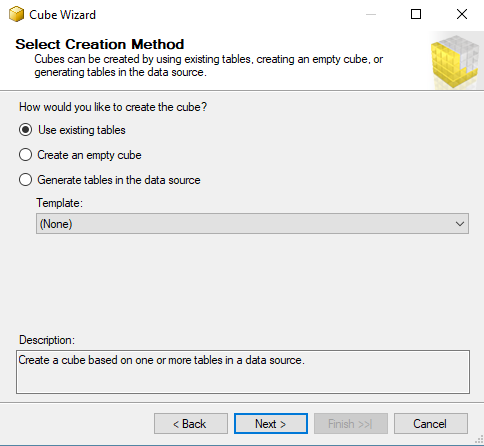
Slika 4‑10 Prikaz tablica

Potrebno je odraditi deploy kocke, način na koji se to radi nalazi se u nastavku. Prvi korak je prikazan na slici (Slika 4-11).



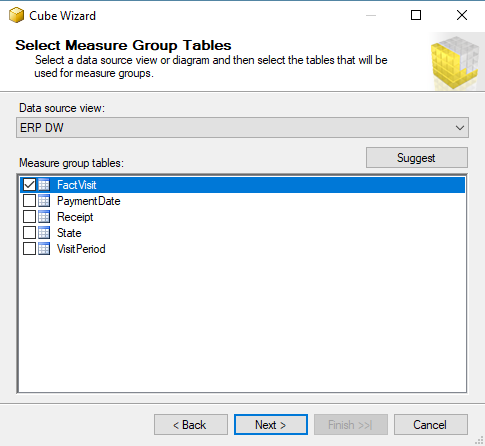
Slika 4‑11 Odabir nove kocke

Zatim je potrebno je odabrati već postojeće tablice. Postupak se nalazi na slici (Slika 4-12).

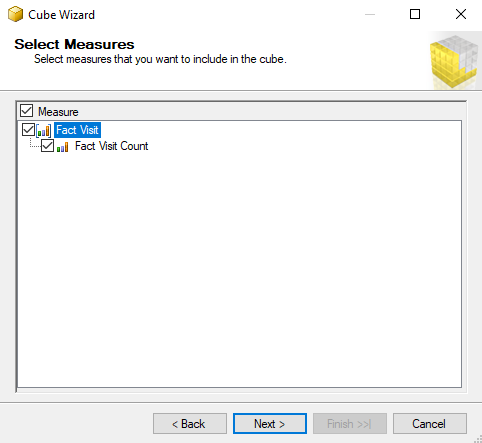


Slika 4‑12 Odabir postojećih tablica

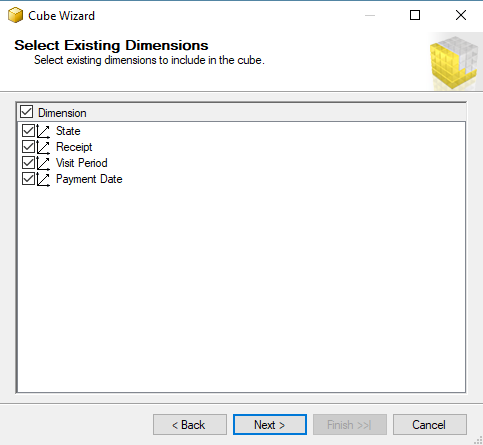
Nakon toga, trebamo odabrati mjere, tablicu činjenica i tablice dimenzija. Postupak je prikazan na slikama (Slika 4-13, Slika 4-14, Slika 4-15).



Slika 4‑13 Odabir tablica

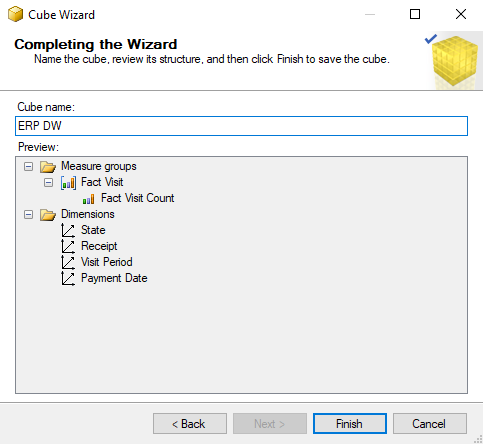


Slika 4‑14 Odabir podataka iz tablice



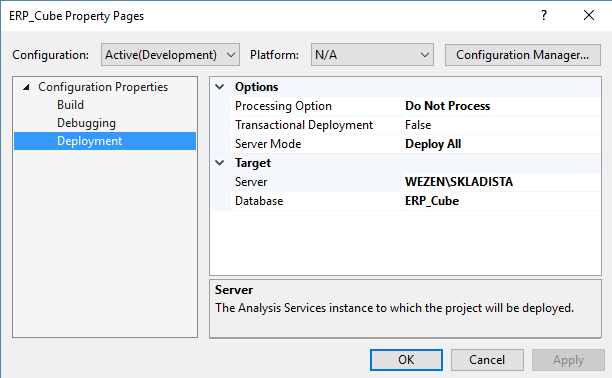
Slika 4‑15 Označavanje dimenzije

Na kraju dobijemo organizaciju kocke. Organizacija kocke je prikazana slici (Slika 4-16).



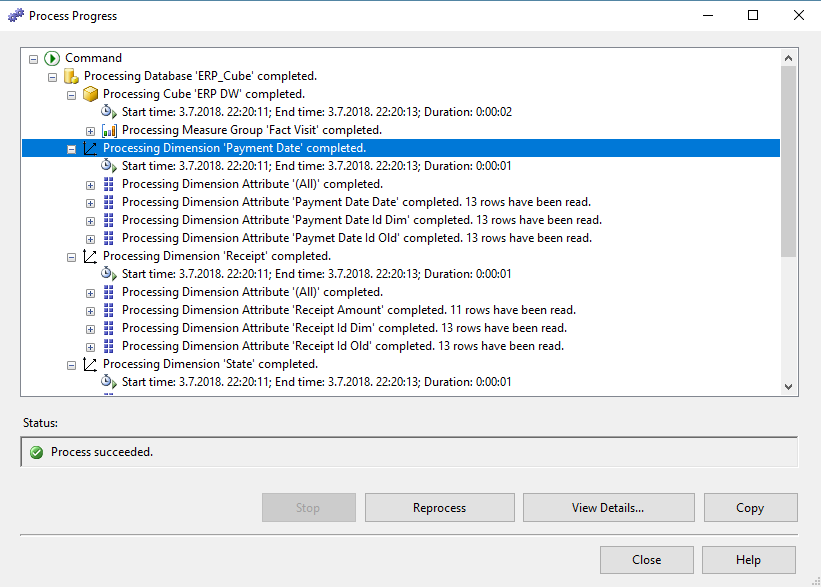
Slika 4‑16 Organizacija kocke

Kocku je zatim potrebno deployati, prikazano na slici (Slika 4-17) na kojoj vidimo koje su vrijednosti parametara. Potrebno je upisati ime servera na kojem se baza nalazi.



Slika 4‑17 Vrijednosti parametara

Na kraju je potrebno još i procesirati podatke u kocku. Smisao svega ovoga do sada je brzo generiranje izvještaja, koje nam je nekada vrlo važno. Rezultati u izvještajima moraju biti isti kao i rezultati koje bismo dobili ako izvještaje generiramo iz transakcijske baze podataka. Ukoliko izvještaji nisu jednaki onima iz transakcijske baze podataka, onda smo negdje u postupku pogriješili. Jedino što bi se trebalo razlikovati među izvještajima iz transakcijske baze podataka i iz kocke je brzina generiranja izvještaja. Na slici (Slika 4-18) je prikazano što bismo trebali dobiti ako su podaci ispravno procesuirani u kocku.



Slika 4‑18 Prikaz ispravno procesuiranih podataka u kocku

# Zaključak

U ovome radu smo detaljno korak po korak pokazali postupak prebacivanja transakcijske baze podataka u skladište podataka, a nakon toga i prebacivanje podataka iz skladišta u kocku. Na kraju smo generirali izvještaje koji su značajni za poslovni proces koji pratimo.

Velike firme, čije baze podataka raspolažu sa ogromnim brojem zapisa, vrlo često imaju problema sa generiranjem izvještaja. Razlog tome je što je generiranje izvještaja iz transakcijske baze podataka zahtjevan i spor proces. Treba pregledati veliki broj tablica, i to traje određeno vrijeme. Pri tome je i opterećenje na bazu veliko, pa se može dogoditi da ostali korisnici ne mogu normalno raditi ako se generira neko izvješće. Zbog toga, sve češće se koriste skladišta podataka.

Za razliku od transakcijskih baza podataka koje su normalizirane, skladišta podataka ne moraju biti normalizirana. U skladištu podataka podaci su organizirani koji je pogodan za generiranje izvještaja.

Skladište podataka daje brze izvještaje, međutim ako podatke procesuiramo u kocku, izvještaji će biti još brži. U kocki svaka ćelija ima svoje koordinate, te ih je na osnovu toga jednostavno pronaći i pročitati mjeru iz ćelije.