

**SVEUČILIŠTE U SPLITU**

**FAKULTET ELEKTROTEHNIKE, STROJARSTVA I BRODOGRADNJE**

*Kolegij: Skladišta podataka*

SEMINARSKI RAD

**Statistika u društvenim mrežama**

Miličić Antonio

Pavlović Karlo

Smjer: Računarstvo, 250

Sadržaj

[1 Uvod 1](#_Toc13071335)

[2 Transakcijska baza podataka 2](#_Toc13071336)

[2.1 Entiteti 3](#_Toc13071337)

[2.1.1 Entitet Post 3](#_Toc13071338)

[2.1.2 Entitet Timestamp 3](#_Toc13071339)

[2.1.3 Entitet PostTag 4](#_Toc13071340)

[2.1.4 Entitet Tag 4](#_Toc13071341)

[2.1.5 Entitet User 5](#_Toc13071342)

[2.1.6 Entitet Country 5](#_Toc13071343)

[2.1.7 Entitet Gender 6](#_Toc13071344)

[2.1.8 Entitet AgeGroup 6](#_Toc13071345)

[2.1.9 Entitet ActionTaken 7](#_Toc13071346)

[2.1.10 Entitet View 7](#_Toc13071347)

[2.1.11 Entitet Like 8](#_Toc13071348)

[2.1.12 Entitet Follow 8](#_Toc13071349)

[2.2 Veze među entitetima 9](#_Toc13071350)

[2.2.1 Post – Timestamp 9](#_Toc13071351)

[2.2.2 Post – User 9](#_Toc13071352)

[2.2.3 Post – Tag (PostTag) 9](#_Toc13071353)

[2.2.4 Post – (ActionTaken) – View/Like /Follow 10](#_Toc13071354)

[2.2.5 User – (ActionTaken) – View/Like/Follow 10](#_Toc13071355)

[2.2.6 User – Country 10](#_Toc13071356)

[2.2.7 User – Gender 11](#_Toc13071357)

[2.2.8 User – AgeGroup 11](#_Toc13071358)

[2.3 Popunjavanje transakcijske baze 11](#_Toc13071359)

[3 Skladište podataka 12](#_Toc13071360)

[3.1 Fact tablica FactTag 14](#_Toc13071361)

[3.2 Dimenzije 15](#_Toc13071362)

[3.2.1 Dimenzija DimTimeStamp 16](#_Toc13071363)

[3.2.2 Dimenzija DimGender 16](#_Toc13071364)

[3.2.3 Dimenzija DimCountry 17](#_Toc13071365)

[3.2.4 Dimenzija DimAgeGroup 17](#_Toc13071366)

[4 Poveznica OLTP-a i Skladišta Podataka 18](#_Toc13071367)

[4.1 Pravila za dimenzije Country 20](#_Toc13071368)

[4.1.1 OLEDB Source 20](#_Toc13071369)

[4.1.2 Lookup 21](#_Toc13071370)

[4.1.3 OLEDB Destination 22](#_Toc13071371)

[5 Definiranje pravila Fact tablice 24](#_Toc13071372)

[5.1 Pravila Fact tablice 25](#_Toc13071373)

[5.1.1 Postavke za OLEDB Source 26](#_Toc13071374)

[5.1.2 Definiranje LookUp-a 28](#_Toc13071375)

[5.1.3 OLEDB Destination 30](#_Toc13071376)

[6 SSAS kao povezinca DW i Kocke 32](#_Toc13071377)

[6.1 Izrada SQL Server Analysis Servis projekta (SSAS) 32](#_Toc13071378)

[6.2 Definiranje pogleda 33](#_Toc13071379)

[6.3 Stvaranje kocke 35](#_Toc13071380)

[6.4 Deploy Kocke 39](#_Toc13071381)

[7 Analiza podataka kocke 40](#_Toc13071382)

[7.1 SSAS 40](#_Toc13071383)

[7.2 Izrada izvješća 42](#_Toc13071384)

[8 Zaključak 43](#_Toc13071385)

# Uvod

Bilo koje veće poslovanje zahtjeva upotrebu, te obradu velikog broja akumuliranih podataka. Također, uporaba statistike u poslovnom procesu može napraviti razliku između suficitnog i deficitnog poslovanja, upravo zbog mogućnosti konkuriranja na tržištu, koje danas sve više napreduje i prelazi u modernije doba.

Stariji transakcijski sustavi ne mogu osigurati sve potrebne informacije iz više razloga, bilo da se radi o tom da su podatci na različitim tipovima baza podataka, medijima, mjestima. Ponekad loše strukturiranje same baze može uzrokovati problem pri povezivanju više izvora podataka, različite oznake/tipovi za iste podatke itd.

Na kraju svega, sami izračuni u relacijskim bazama traju od nekoliko sati ( prihvatljivo ), do nekoliko dana, što je pre sporo. Samim tim se dodatno opterećuje baza podataka i rad transakcijskog sustava. Tim više ne osigurava se čistoća podataka, iako bitna stavka sustava.

Upravo zbog tih problema danas se prelazi na moderniji pristup obradi podataka, skladišta podataka . Skladišta podataka prikupljaju transakcijske podatke za koje imamo interes, iz dostupnih izvora, te se integriraju u jedinstveni repozitoriji.

Takav sustav dodatno se proširuje s OLAP( On Line Analytical Processing ) pojmom. OLAP omogućava analizu i izradu izvještaja koji pomažu korisniku da selektivno odvaja i pregledava podatke s različitim uvjetima, koristeći strukturu kocke , u literaturi poznatom i kao CUBE.

Za ovaj seminar koristit ćemo navedeni sustav za cjelokupan proces izrade skladišta podataka i izradu izvještaja. Skladište podataka je izgrađeno za obradu podataka vezanih uz Instagram objave, odnosno reakcije na iste. Financijski ideja se temelji na statistici korisnika, koja nas dovodi do profita temeljem reklama, prosljeđivanja informacije tržištu robe, auta, zabave…

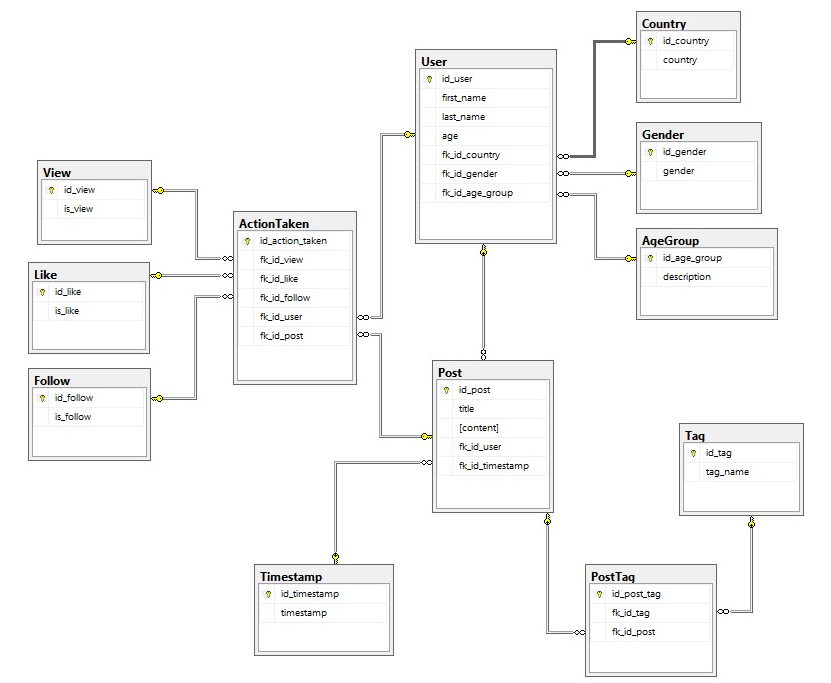
Kroz prvi dio obrađena je izrada transakcijske baze podataka s potrebnim atributima i entitetima za uspješnu obradu statističkih podataka, te izrada skladišta podataka iz te baze.

Nakon toga prikazano je punjenje skladišta podataka, temeljem koje se gradi kocka.

Temeljem te kocke možemo izraditi analizu podataka, odnosno izvještaj.

# Transakcijska baza podataka

Ovo poglavlje opisuje izradu transakcijske baze podataka. Iako bi bilo dobro napraviti papirnati prototip baze, zbog poznavanja rada u ovim sustavima taj korak je preskočen, te smo prešli odmah na digitalnu verziju transakcijske baze, vidljivo na Slika 1.



Slika 1 Transakcijska baza podataka

## Entiteti

Kroz ovo poglavlje prikazane su slike entiteta i tekstualno opisani primarni ključevi.

### Entitet Post

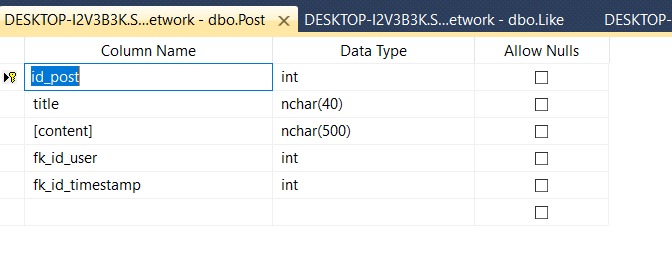
Entitet post sadrži sljedeće atribute:

* Primarni ključ: id\_post
* Ostali atributi: title – (nchar40)

[content] – (nchar500)

fk\_id\_user – (int)

fk\_id\_timestamp – (int)

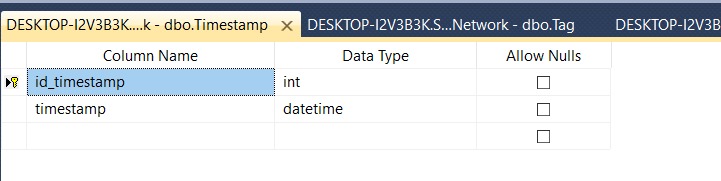


Slika 2 Post tablica

### Entitet Timestamp

Entitet post sadrži sljedeće atribute:

* Primarni ključ: id\_timestamp
* Ostali atributi: timestamp – (datetime)



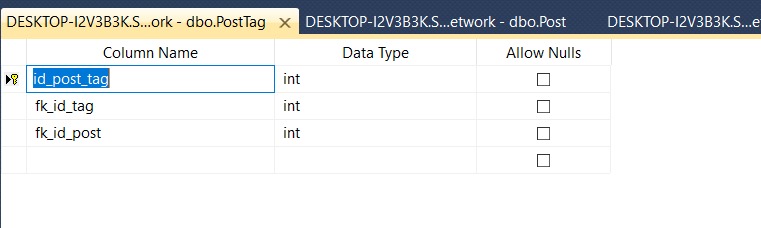
Slika 3 Timestamp tablica

### Entitet PostTag

Entitet post sadrži sljedeće atribute:

* Primarni ključ: id\_post\_tag
* Strani ključevi: fk\_id\_tag

fk\_id\_post

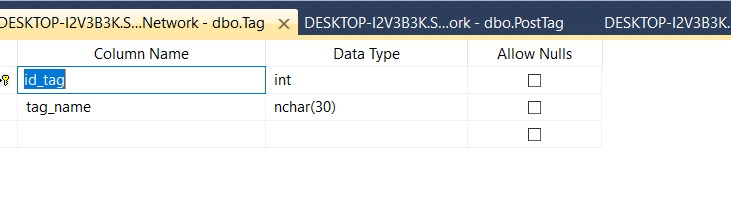


Slika 4 PostTag tablica

### Entitet Tag

Entitet post sadrži sljedeće atribute:

* Primarni ključ: id\_tag
* Ostali atributi: tag\_name – (nchar30)



Slika 5 Tag tablica

### Entitet User

Entitet post sadrži sljedeće atribute:

* Primarni ključ: id\_user
* Ostali atributi: first\_name – (nchar20)

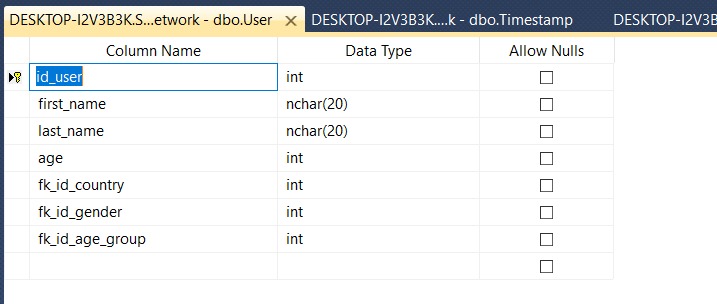
last\_name – (nchar20)

age

* Strani ključevi: fk\_id\_country

fk\_id\_gender

fk\_id\_age\_group

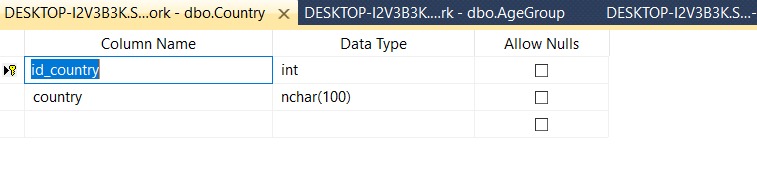


Slika 6 User tablica

### Entitet Country

Entitet post sadrži sljedeće atribute:

* Primarni ključ: id\_country
* Ostali atributi: country – (nchar100)

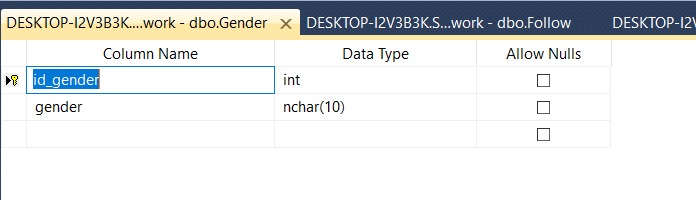


Slika 7 Country tablica

### Entitet Gender

Entitet post sadrži sljedeće atribute:

* Primarni ključ: id\_gender
* Ostali atributi: gender – (nchar10)

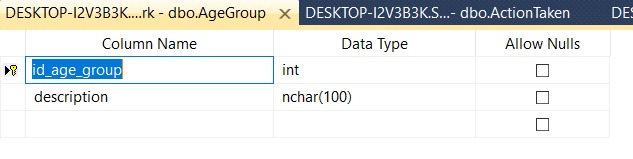


Slika 8 Gender tablica

### Entitet AgeGroup

Entitet post sadrži sljedeće atribute:

* Primarni ključ: id\_age\_group
* Ostali atributi: description – (nchar100)



Slika 9 AgeGroup tablica

### Entitet ActionTaken

Entitet post sadrži sljedeće atribute:

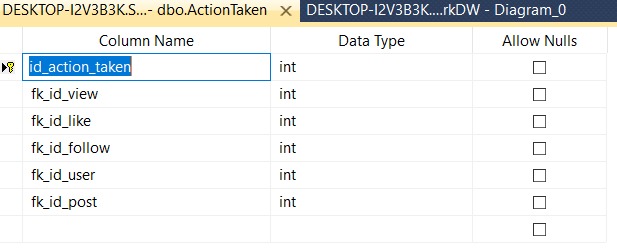
* Primarni ključ: id\_age\_group
* Strani ključevi: fk\_id\_view

fk\_id\_like

fk\_id\_follow

fk\_id\_user

fk\_id\_post

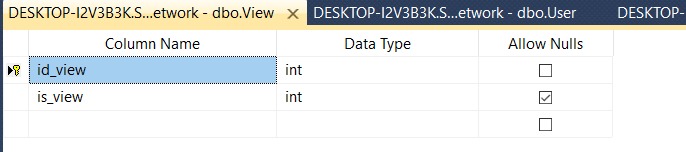


Slika 10 ActionTaken tablica

### Entitet View

Entitet post sadrži sljedeće atribute:

* Primarni ključ: id\_view
* Ostali atributi: is\_view – (int)

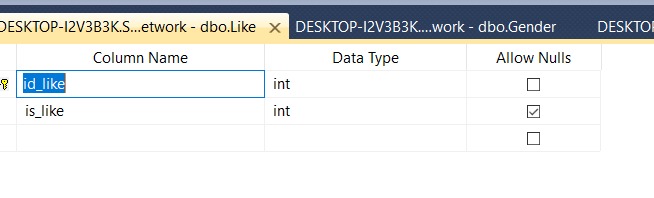


Slika 11 View tablica

### Entitet Like

Entitet post sadrži sljedeće atribute:

* Primarni ključ: id\_like
* Ostali atributi: is\_like – (int)

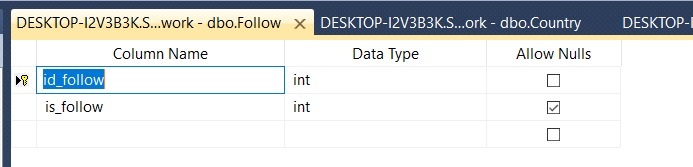


Slika 12 Like tablica

### Entitet Follow

Entitet post sadrži sljedeće atribute:

* Primarni ključ: id\_follow
* Ostali atributi: is\_follow – (int)



Slika 13 Follow tablica

## Veze među entitetima

Kroz ovo poglavlje opisane su veze među entitetima.

### Post – Timestamp

Veza: N : 1  
Entiteti Post i Timestamp[[1]](#footnote-1) su povezani vezom N : 1 , što znači da više Post-ova može imati jedan te isti Timestamp. Rezultat toga je atribut fk\_id\_timestamp koji je dio Post[[2]](#footnote-2) entiteta. Pravilo proizlazi iz logike da u točno to vrijeme više objava će imati tu oznaku vremena.

### Post – User

Veza: N : 1  
Entiteti Post i User[[3]](#footnote-3) su povezani vezom N : 1 , što znači da jedan User može imati više Post-ova. Rezultat toga je atribut fk\_id\_user koji je dio Post entiteta. Proizlazi iz toga da jedan korisnik može napraviti više objava, ali jedna objava se odnosi samo na tog korisnika.

### Post – Tag (PostTag)

Veza: 1 : N  
Entiteti Post i Tag[[4]](#footnote-4) su povezani vezom 1 : N , što znači da jedan Post može imati N Tag-ova.  
Logika proizlazi iz mogućnosti postavljanja više oznaka na jednu objavu.  
Kao rezultat kreirana je nova među-tablica PostTag, kao poveznica:

PostTag  
(Post) fk\_id\_post – fk\_id\_tag ( Tag )

### Post – (ActionTaken[[5]](#footnote-5)) – View[[6]](#footnote-6)/Like[[7]](#footnote-7) /Follow[[8]](#footnote-8)

Veza: 1 : N , 1 : 1  
Entiteti Post i ActionTaken su povezani vezom 1 : N , što znači da jedan Post može imati N akcija. Ali svaka akcija može biti jedinstven view/like/follow. Stoga je ActionTaken u vezi 1 : 1 s njima. Kao rezultat kreirana je nova među-tablica ActionTaken, kao poveznica:

ActionTaken  
(Post) fk\_id\_post – fk\_id\_View/Like/Follow ( View/Like/Follow )

### User – (ActionTaken) – View/Like/Follow

Veza: 1 : N , 1 : 1  
Entiteti User i ActionTaken su povezani vezom 1 : N , što znači da jedan User može imati N ActionTaken-a. Logično je da će biti jedan korisnik moći ostaviti više dojamova. Ali zato svaka akcija obuhvaća samo jedan view/like/follow. Stoga je ActionTaken u vezi 1 : 1 s njima.  
Kao rezultat kreirana je među-tablica ActionTaken kao poveznica:

ActionTaken  
(User) fk\_id\_user – fk\_id\_View/Like/Follow ( View/Like/Follow )

### User – Country

Veza: N : 1  
Entiteti User i Country[[9]](#footnote-9) su povezani vezom N : 1 , što znači da više korisnika može pripadati istoj državi. Rezultat toga je atribut fk\_id\_country koji je dio User entiteta. Pravilo dolazi iz realne situacije gdje jedan korisnik može imati jedno državljanstvo, ali jedna država više korisnika.

### User – Gender

Veza: N : 1  
Entiteti User i Gender[[10]](#footnote-10) su povezani vezom N : 1 , što znači da više korisnika može imati jedan rod. Rezultat toga je atribut fk\_id\_gender koji je dio User entiteta. Obratno pravilo, jednom rodu može pripadat više korisnika.

### User – AgeGroup

Veza: N : 1  
Entiteti User i AgeGroup[[11]](#footnote-11) su povezani vezom N : 1 , što znači da jedan korisnik može imati jednu dobnu skupinu. Ali jednoj dobnoj skupini može pripadati više korisnika. Rezultat toga je atribut fk\_id\_age\_group koji je dio User entiteta.

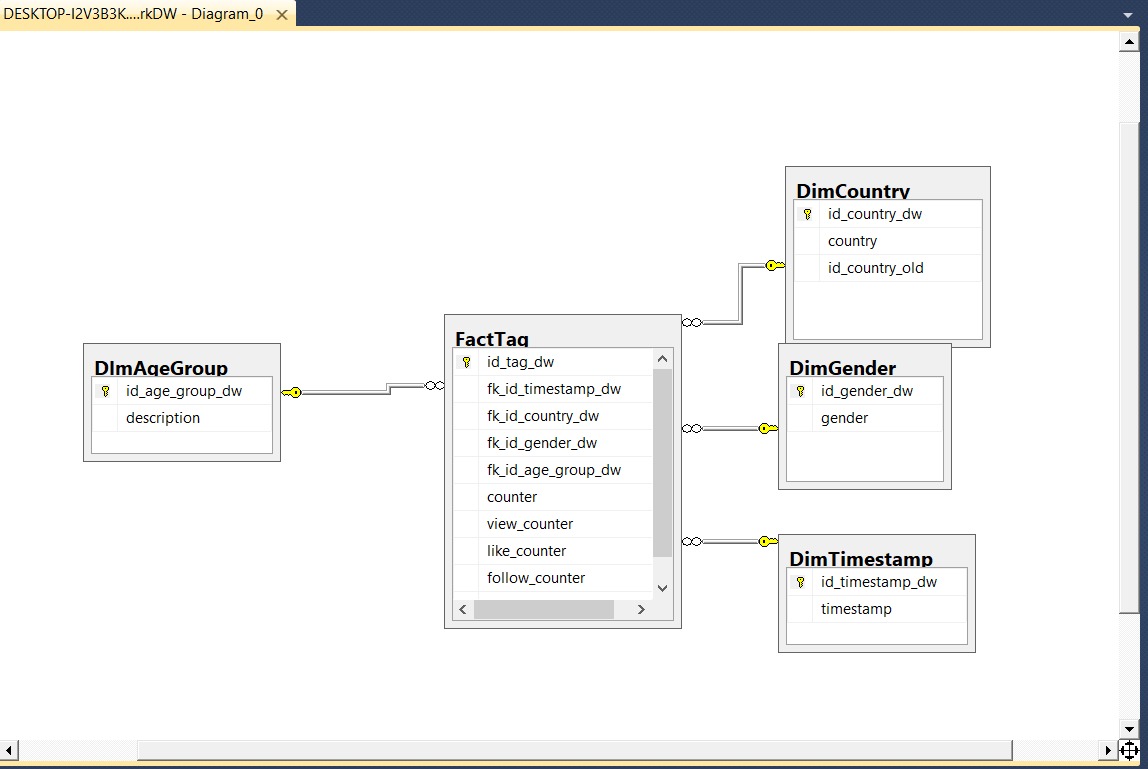
## Popunjavanje transakcijske baze

Nakon što smo izradili transakcijsku bazu podataka, potrebno je napuniti bazu podatcima koje koristimo za izvještaj i analizu.

# Skladište podataka

Skladište podatak, poznato i kao „Data Warehouse“, je sistem korišten za objavu i obradu/analizu podataka, te je osnova poslovnih sustava. DW[[12]](#footnote-12) je središnji repozitoriji za integriranje podataka iz više odvojenih izvora. Pohranjuju se trenutni i prethodni podatci na jedno mjesto, koji se kasnije koriste za analizu i izradu izvještaja. Svi podatci koji ulaze u DW prolaze kroz fazu pročišćavanja, a prije te faze nisu u korištenju za izvješće.

Podatci u DW su podijeljeni po dimenzijama[[13]](#footnote-13), te činjenicama[[14]](#footnote-14). Ta kombinacija se naziva shema zvijezde, te ona omogućava korisnicima pristup i dohvat podataka.



Slika 14 Skladište podataka

Fact tablica nam omogućava praćenje podataka, to je takozvana tablica činjenica te mora biti brojčane vrijednosti. Dimenzijske tablice pak sadrže podatke koje pratimo po Fact tablici, te su iz tog razloga direktno vezani uz Fact tablicu.  
Standard je u nazivu Fact tablice staviti riječ Fact, dok se dimenzije označava s Dim.

Primarna ideja projekta je praćenje podataka oznaka ( Tag ), odnosno svih podataka koji su vezani uz isto. Stoga je Fact tablica nazvana FactTag. Kroz nju pratimo statistiku koja se veže uz informacije koju korisnici svojevoljno ostavljaju prilikom postavljanja oznaka, kao što su njihova dobna skupina, rod, državljanstvo… Da bi mogli pratiti vremenski kako se ponaša statistika koristimo TimeStamp, kao vremensku oznaku pomoću koje možemo odrediti vrijeme, tj. raspon za koji nas zanima statistika.

Stoga uz Fact tablicu vežemo dimenzije:

* DimAgeGroup
* DimTimeStamp
* DimGender
* DimCountry

## Fact tablica FactTag

Kao što je prethodno opisano, Fact tablica sadrži brojčane podatke pomoću kojih pratimo broj ( counter ) koji je vezan uz oznake, poglede, sviđanja i praćenja.

Tablica sadrži:

* Primarni ključ: id\_tag\_dw
* Strani ključevi: fk\_id\_timestamp\_dw

fk\_id\_country\_dw

fk\_id\_gender\_dw

fk\_id\_age\_group\_dw

* Ostali atributi: counter

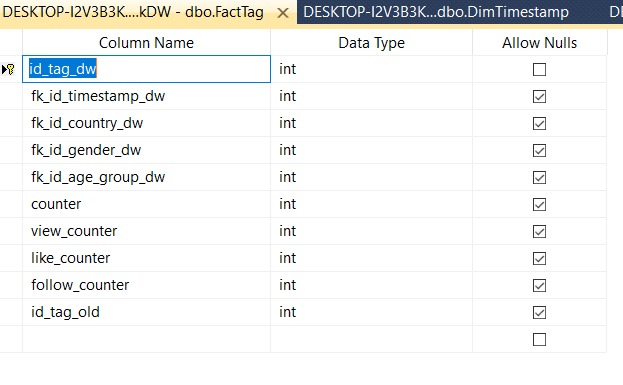
view\_counter

like\_counter

follow\_counter

id\_tag\_old

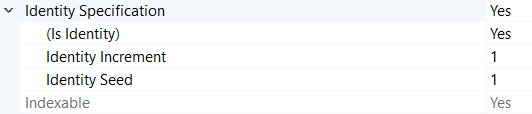
Tablica sadrži strane ključeve kojima je povezana za dimenzije, brojače koji se koriste za tik prije opisane podatke i na kraju id\_tag\_old, koji osigurava pohranu starog tag id-a, u slučaju mijenjanja trenutnog, što povezuje podatke trenutne i stare verzije.



Slika 15 Fact tablica

## Dimenzije

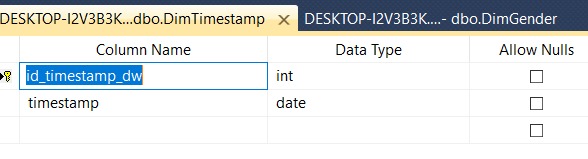
Dimenzijske tablice sadrže opis dimenzija, te sadrže dimenzijske ključeve, vrijednosti i atribute. Ukratko one nam daju opis proizvoda, vremena, a ne nose nikakvu brojčanu oznaku, tj. brojač. Dimenzijske tablice su tipično kratke, te variraju od nekoliko do nekoliko tisuća redaka. Podjela po dimenzijama omogućava izradu izvještaja po određenoj strukturi, odnosno zahtjevu. Svaka dimenzijska tablica sadrži primarni ključ kojim je vezana za transakcijsku bazu. Za sve dimenzije postavljamo svojstvo primarnog ključa, inkrement za 1 prilikom popune skladišta podataka. Svojstvo prikazano na Slika 16.



Slika 16 Dimenzijski inkrement

### Dimenzija DimTimeStamp

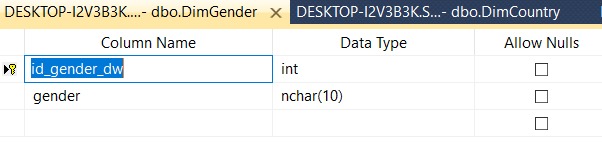
Dimenzija DimTimeStamp sadrži ID id\_timestamp\_dw, te timestamp polje, kao oznaku vremenskkog perioda u kojemu radimo mjerenje preko Fact tablice.



Slika 17 Tablica dimenzije DimTimeStamp

### Dimenzija DimGender

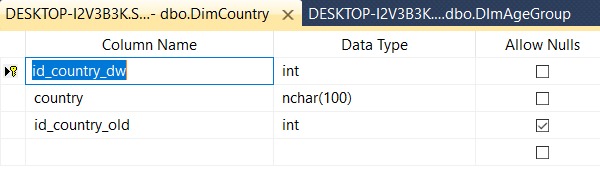
Dimenzija DimGender sadrži ID id\_gender\_dw, te gender polje, koje nam služi za pregled po rodu korisnika prilikom pretrage/analize.



Slika 18 Tablica dimenzije DimGender

### Dimenzija DimCountry

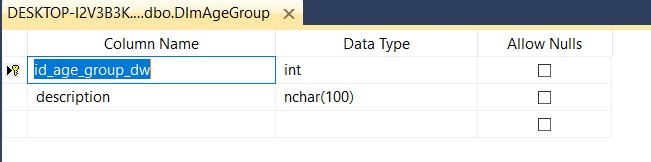
Dimenzija DimCountry sadrži ID id\_country\_dw, te polje country, koje služi za analiziranje podataka prema rodu. Također uz to u ovoj tablici se nalazi id\_country\_old, koji osigurava poveznicu sa starim podatcima, ako dođe do promjene državljanstva, što je moguće.



Slika 19 Tablica dimenzije DimCountry

### Dimenzija DimAgeGroup

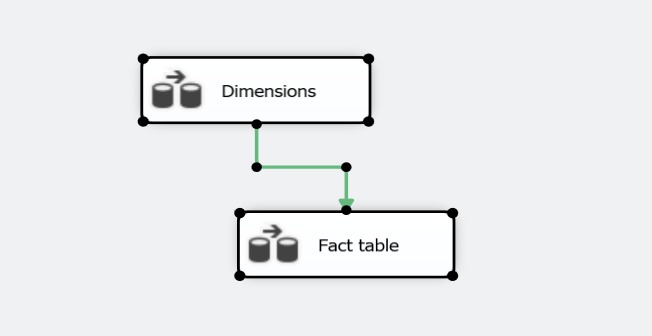
Posljednja dimenzija je DimAgeGroup, koja sadrži ID id\_group\_dw, te polje description, u kojemu se nalazi opis dobne skupine prema kojoj možemo napraviti pretragu.



Slika 20 Tablica DimAgeGroup

# Poveznica OLTP-a i Skladišta Podataka

Nakon što smo izgradili transakcijsku bazu podataka, popunili je, te na osnovu toga izgradili skladište podataka s Fact tablicom i dimenzijama, sljedeći zadatak je prebaciti podatke iz transakcijske ( OLTP ) baze podataka u naše skladište podataka ( DW ). Za to koristimo SSIS[[15]](#footnote-15). SSIS je integracijski servis s kojim možemo definirati kojim tokom će podatci teći.  
  
Kao prvi korak obavljamo kopiranje podataka iz OLTP-a u DW, prema određenim pravilima.  
Potrebno je u Control Flow-u[[16]](#footnote-16) namjestiti Data Flow Task[[17]](#footnote-17) za naše dimenzije i Fact tablicu. DFT u svom dosegu sadrži mehaniku koja nam omogućava transformaciju, pročišćavanje i izmjenu podataka prilikom premještanja. Također DFTEngine ima za zadatak pratiti red izvršavanja kopiranja podataka ako imamo više izvora za jedan zadatak.



Slika 21 Control Flow

Nakon toga definira se Data Flow Control, koji je potreban jer se Fact tablica popunjava podatcima iz dimenzija. Tu se definiraju pravila za dimenzije.

Data Flow Control definira sljedeće parametre:

* Izvor podataka (OLTP)
* Pravila za transfer
* Odredište (DW)

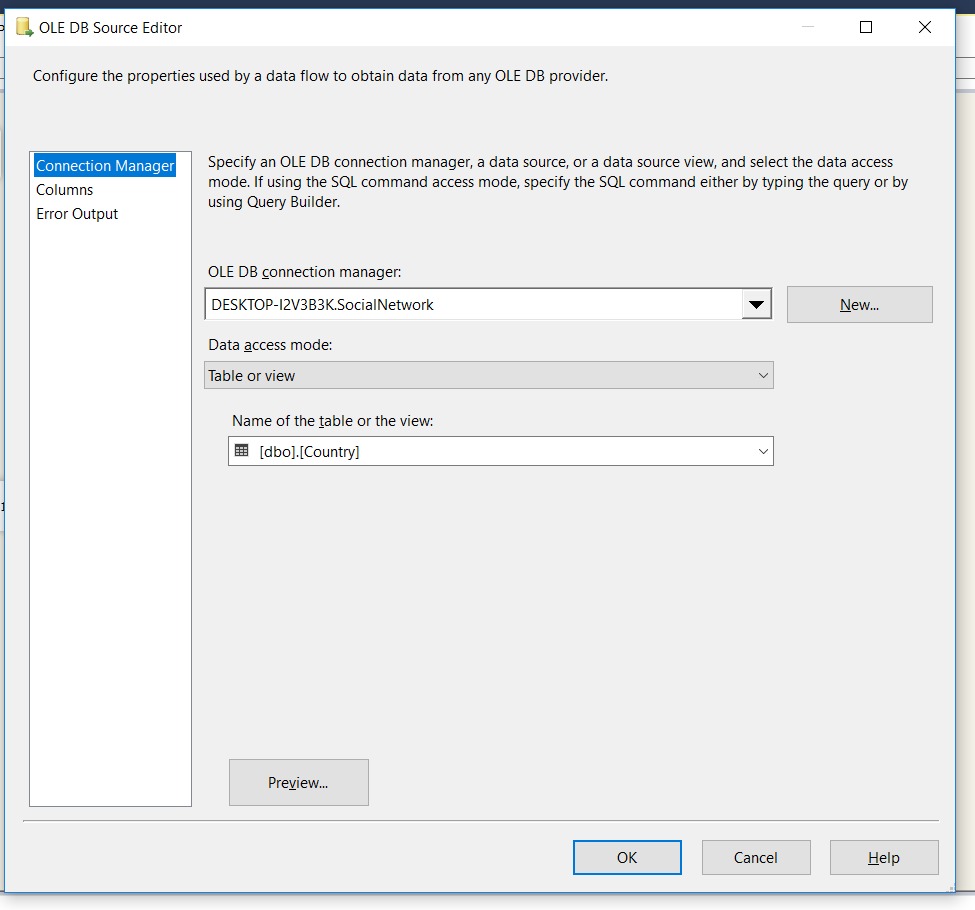
Da bi postavili pravila slijedimo korake:

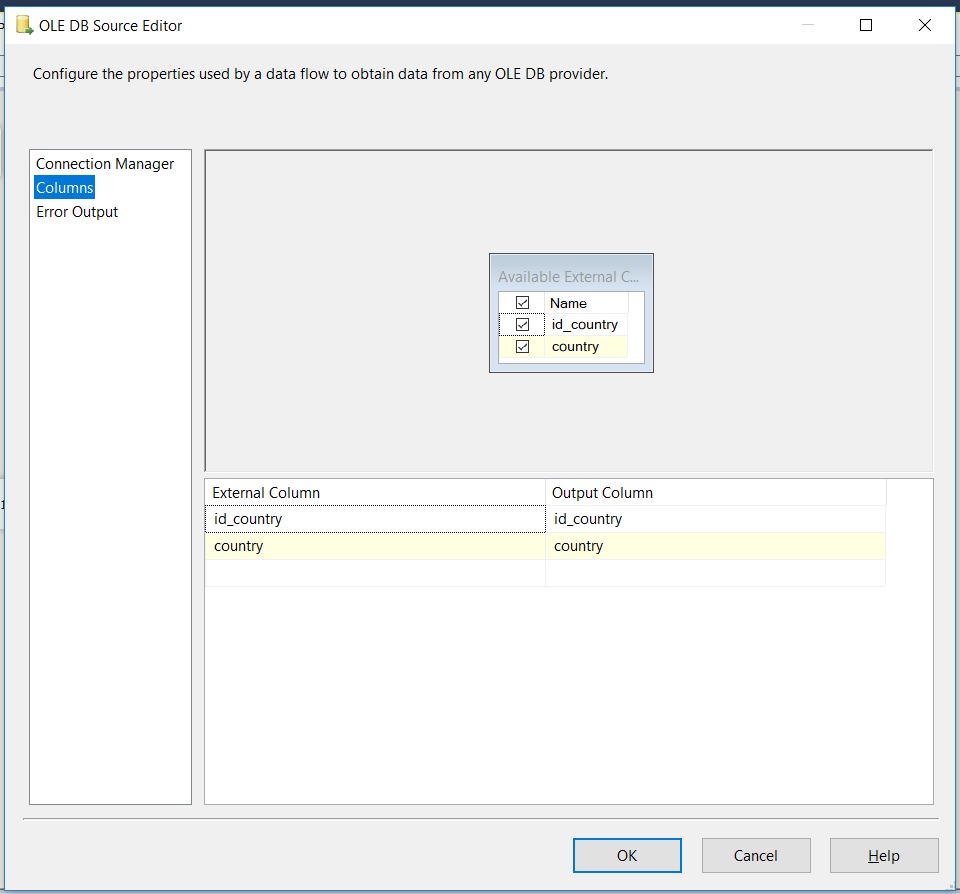
1. OLEDB[[18]](#footnote-18) Source – određujemo koja je baza izvor podataka, te koje su pripadne tablice.
2. LookUp – povezuje primarni ključ OLTP-a s id\_old dimenzijom
3. Conditional Split – koristimo za provjeru da li podatci postoje. To se obavlja pomoću ISNULL funkcije. Ako postoji id\_old, odnosno nije rezultat funkcije NULL, želimo spriječiti kopiranje podatak, jer su već postojeći.
4. Data Conversion – obavlja pretvorbu podataka (string), u jedinstveni format da bi se izbjegle greške u danjim koracima. Ovo je opcionalan korak, ali ga je dobro obaviti za sigurnost podatak DW. ( U našem slučaju nije napravljeno ).
5. OLEDB Destination – definira se spremanje podataka u dimenzijsku tablicu DW-a.

## Pravila za dimenzije Country

Obzirom na već opisani proces, ovdje ćemo kroz primjer objasniti korake.

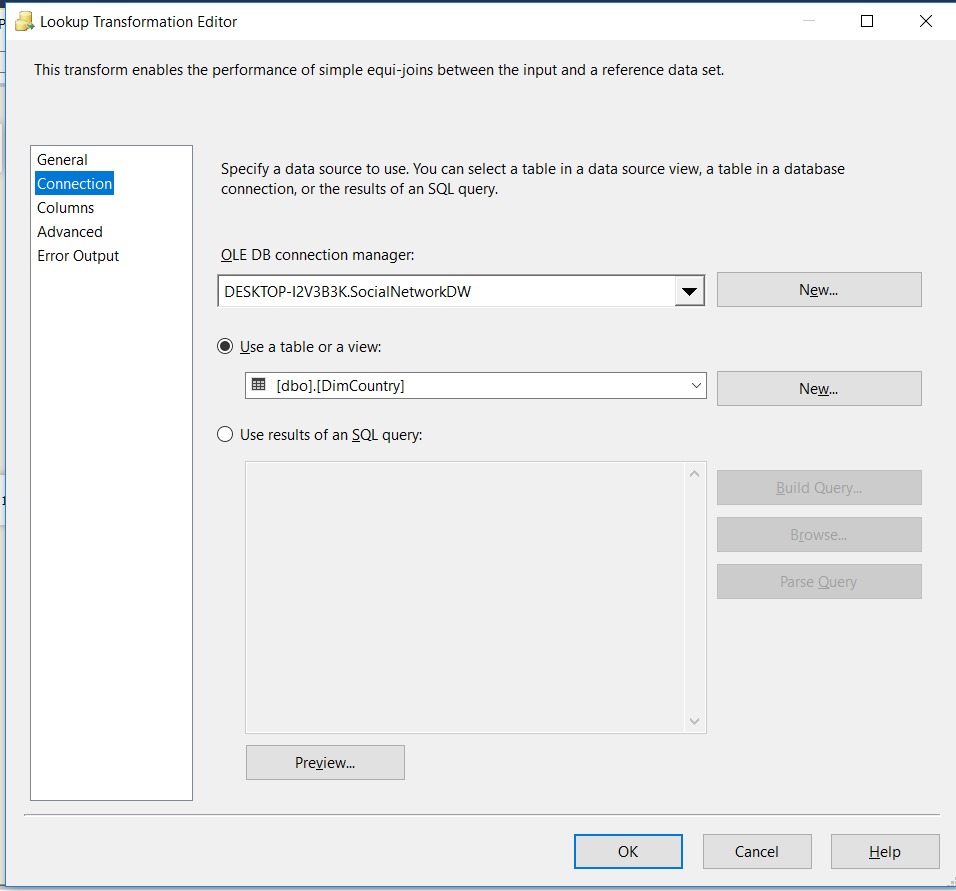
### OLEDB Source

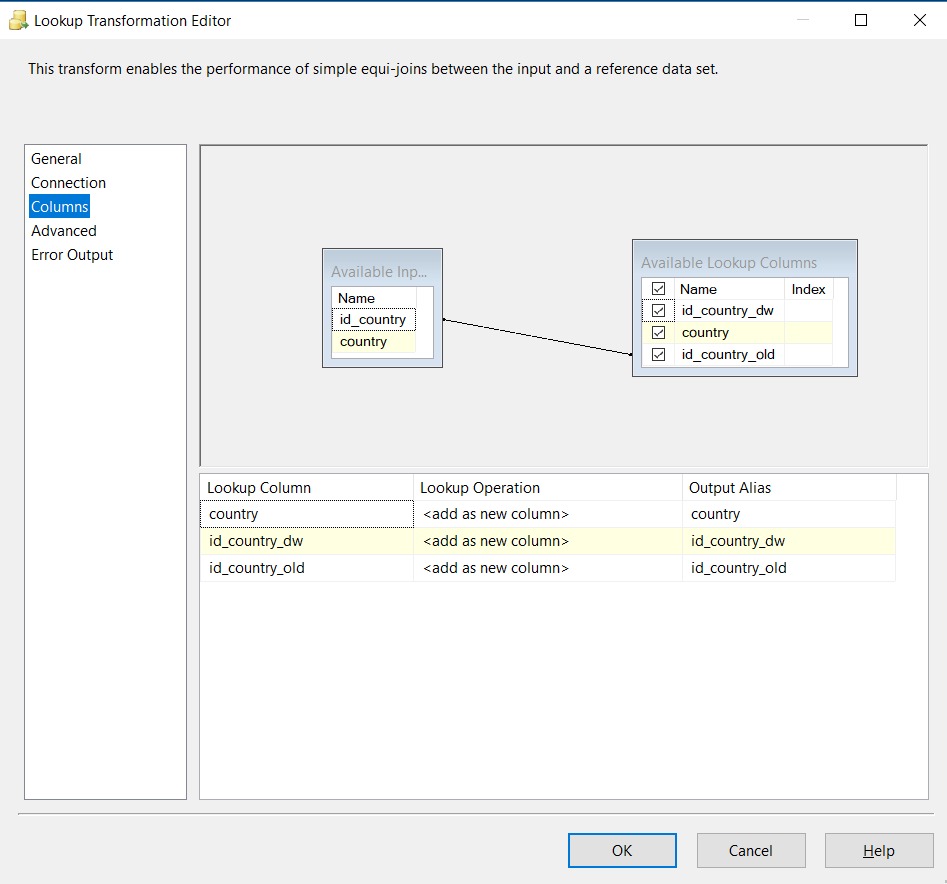




Slika 22 OLEDB Source za dimenziju County

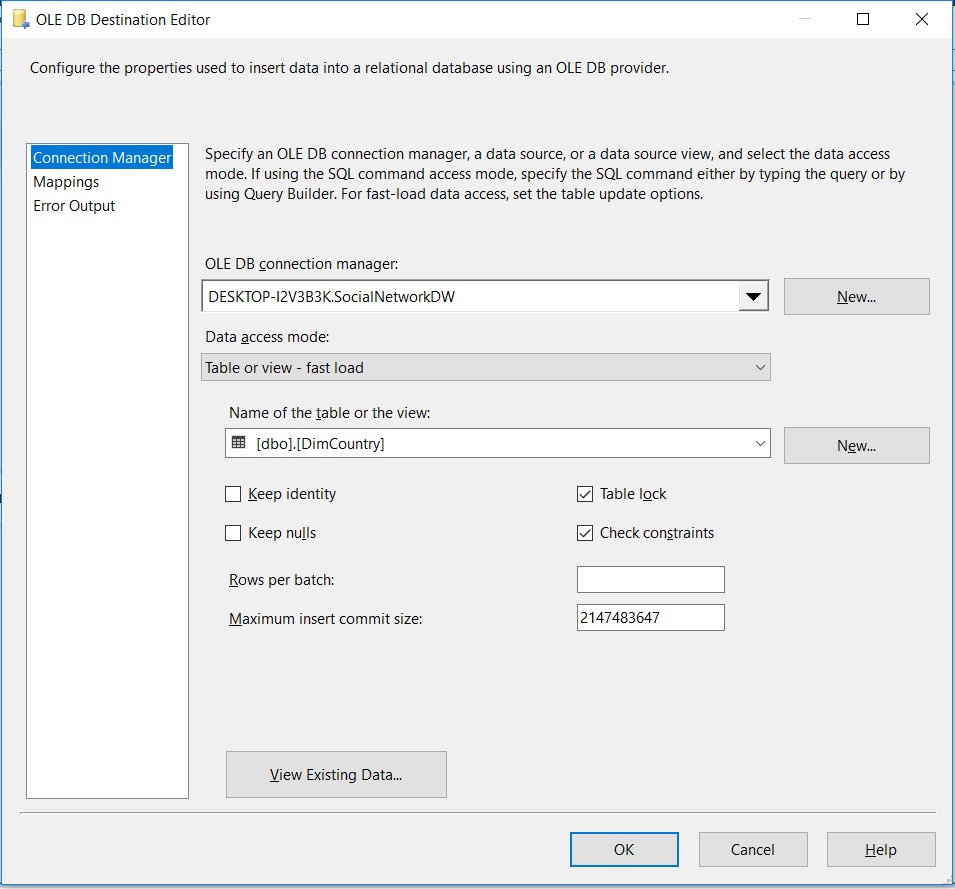
### Lookup



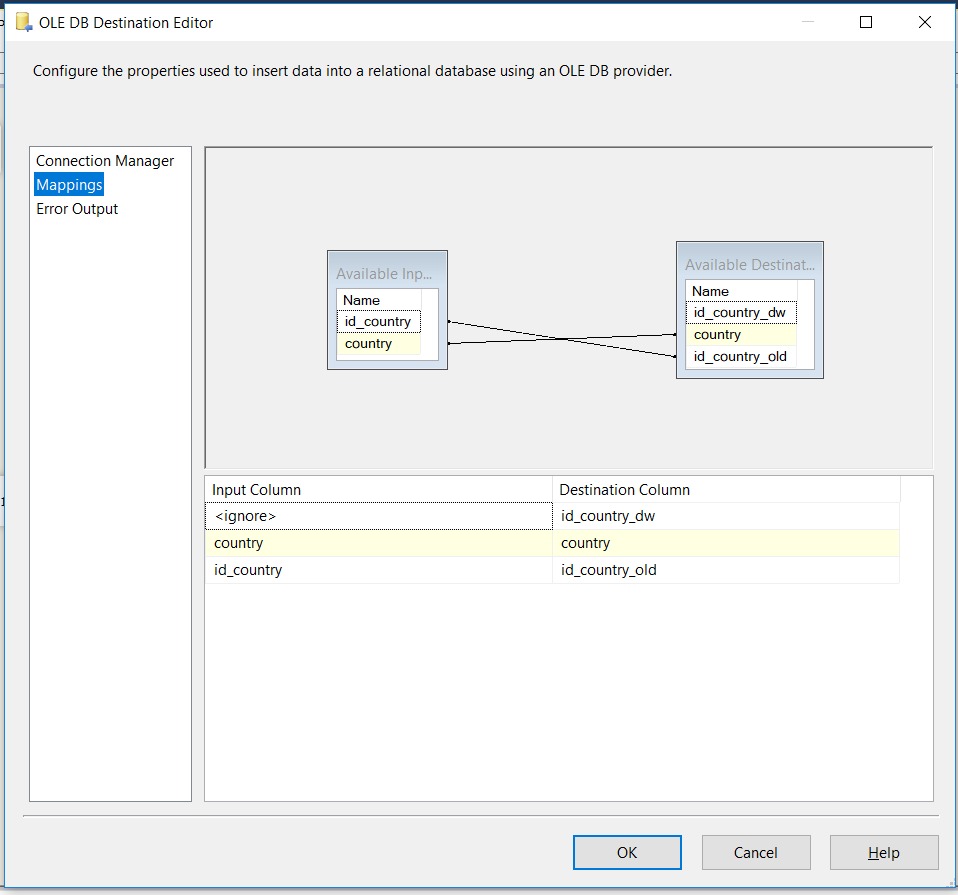


Slika 23 Postavke LookUp-a za dimenziju Country

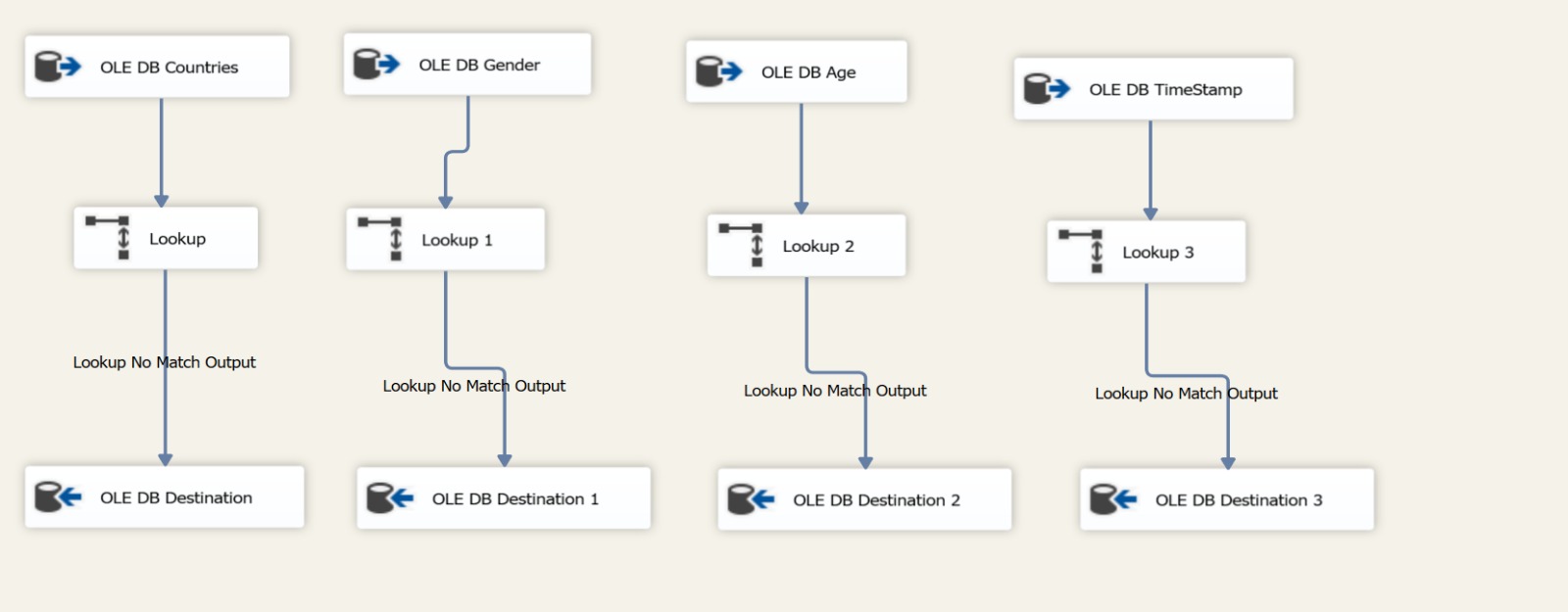
### OLEDB Destination



Slika 24 Postavke odabira DW i dimenzije



Slika 25 Postavke OLEDB Destination za dim\_Country



Slika 26 Pravila za dimenzije

# Definiranje pravila Fact tablice

Nakon što smo odredili pravila za dimenzije, krećemo s popunjavanjem Fact tablice. Fact tablica je brojčana tablica po kojoj pratimo ukupan iznos. Fact tablica u ovom slučaju je FactTag, preko koje pratimo iznos.

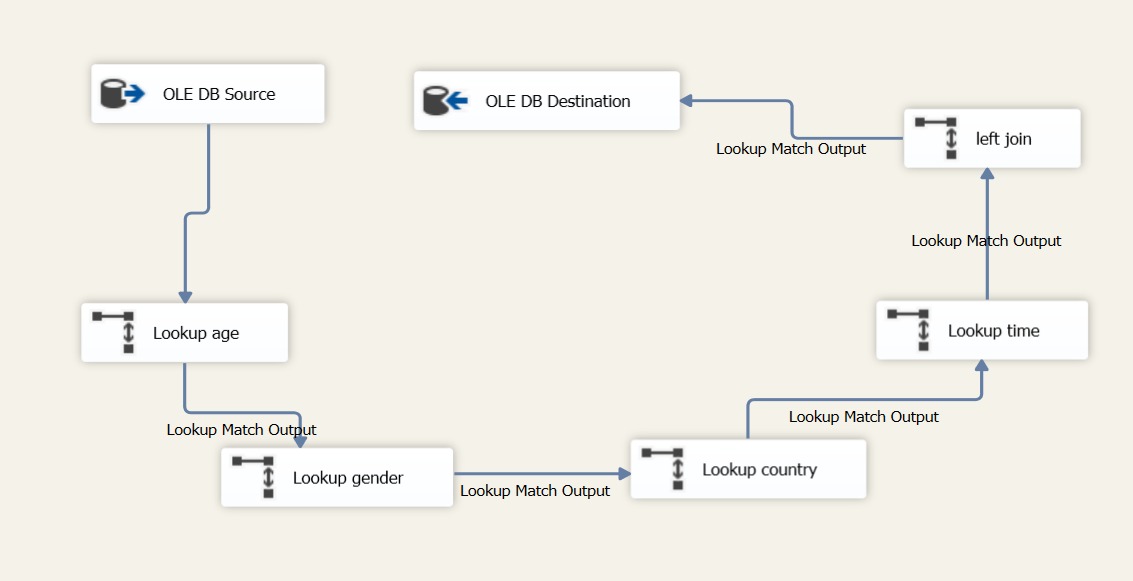
Opet je potrebno namjestiti Data Flow, ali za Fact tablicu. Na isti način, postavljamo izvor podataka, pravila za transfer i odredište podataka, uz neke preinake.

Koraci za postavljanje pravila:

1. OLEDB Source – definiramo izvornu bazu za Fact tablicu podataka, a nakon koristimo BuildQuery, u kojem biramo sve tablice i stupce koji su nam potrebni za Fact tablicu.
2. LookUp – povezujemo primarni ključ OLTP-a sa old\_id atributom dimenzije. Dobiva se novi primarni ključ dimenzije na koji dolazi prefiks „*lkp*“.
3. Left Join – lookup u kojem se odabire Fact tablica kojom se povezu svi „*lkp*“ primarni ključevi s njihovim pozivima u Fact tablici. Ostali podatci se propuštaju uz naznaku prefiksa „*test“.*
4. Conditional Split – isto kao i prošli put, provjera da li su podatci uneseni. Funkcija je ISNULL. Sada provjeravamo podatke s prefiksom „test“. Npr. ISNULL(test\_country).
5. OLEDB Destination – pristigli podatci se mapiraju u Fact tablicu DW baze. LKP PrimaryKey se poveze s ID poljem, a ostali podatci s orginalnim poljima(ne test).

## Pravila Fact tablice

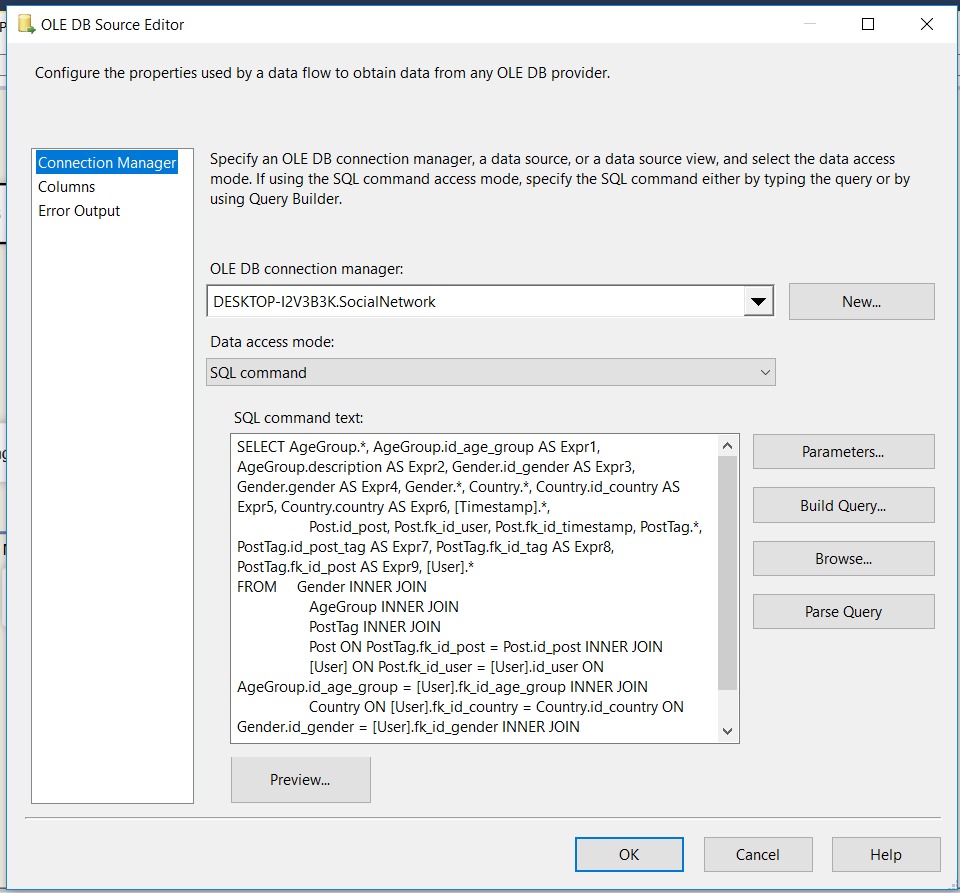
U ovom poglavlju opisana su pravila za fact tablicu pomoću, Slika 22. Na njoj je opisan proces od OLEDB izvora, prema LookUp-u za dob, rod, državu i vrijeme. Nakon čega je napravljen Left-Join, te zadnji LookUp za podatke s „*lkp*“ oznakom, dok se oni s „*test*“ propuštaju. Te na kraju OLEDB Destination gdje se podatci mapiraju u DW baze, te se povezuju podatci po ID-u ili originalnim poljima.



Slika 27 Pravila za Fact Tablicu

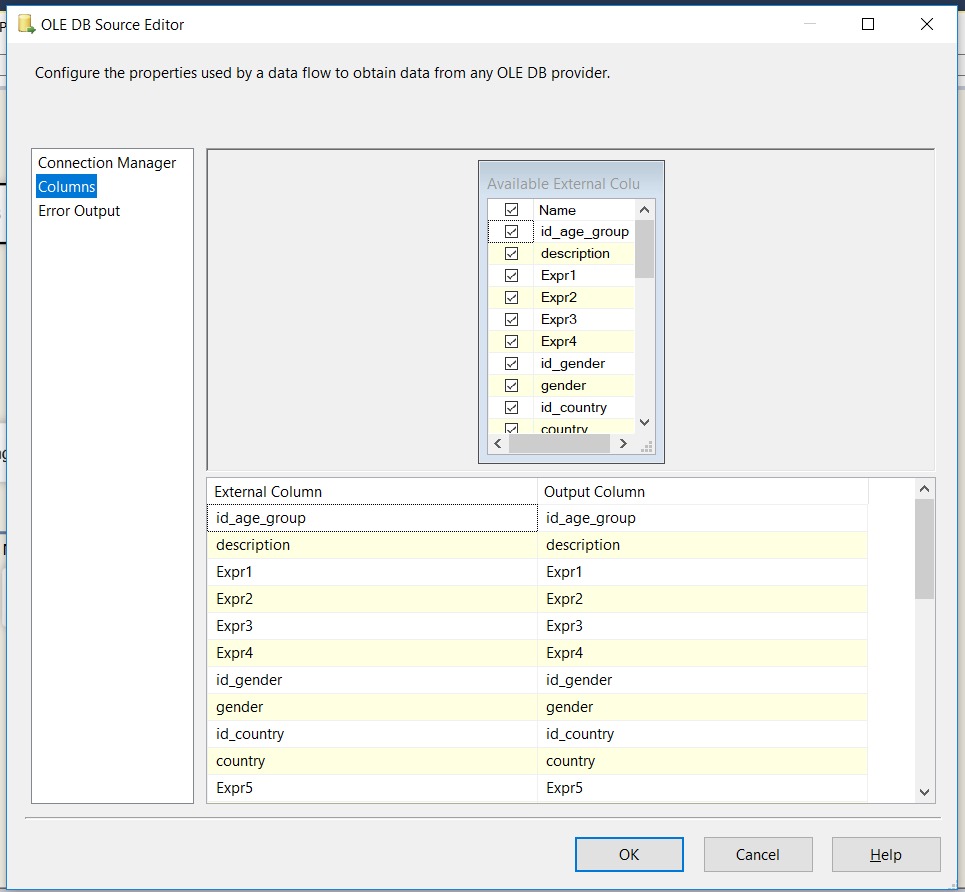
### Postavke za OLEDB Source

S obzirom na objašnjenje koje smo dali na početku poglavlja, Definiranje pravila Fact tablice, nema potrebe objašnjavati proces. Koristimo Build Query, te odredimo koje parametre koristimo, a sam SQL komanda je automatski generirana. Moguć je i ručni unos komande.



Slika 28 Postavljanje OLEDB Source koraka

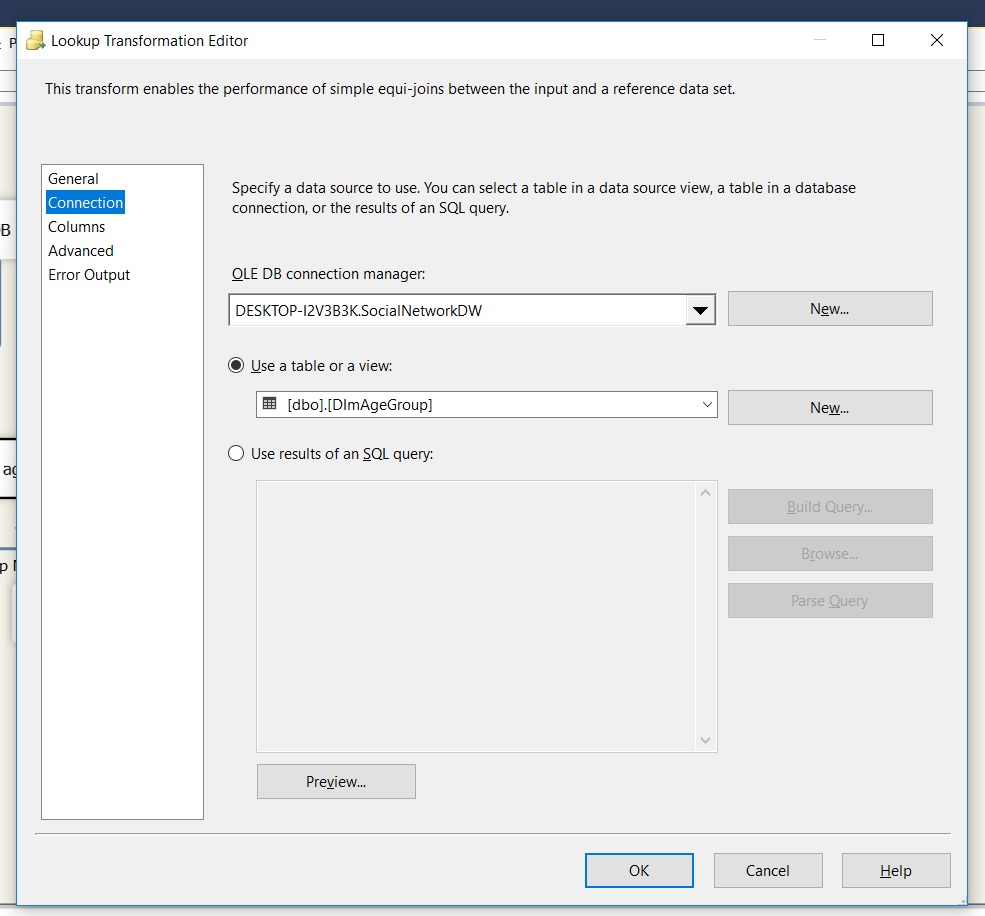
Nakon toga dobijemo u „Columns“ sekciji atribute kojima Fact tablica raspolaže. Tu možemo obaviti promjene ako nam je potrebno. Ovo je ujedno izgled OLEDB Source Fact tablice koja nastaje prethodnim korakom.



Slika 29 OLEDB Source Fact tablica

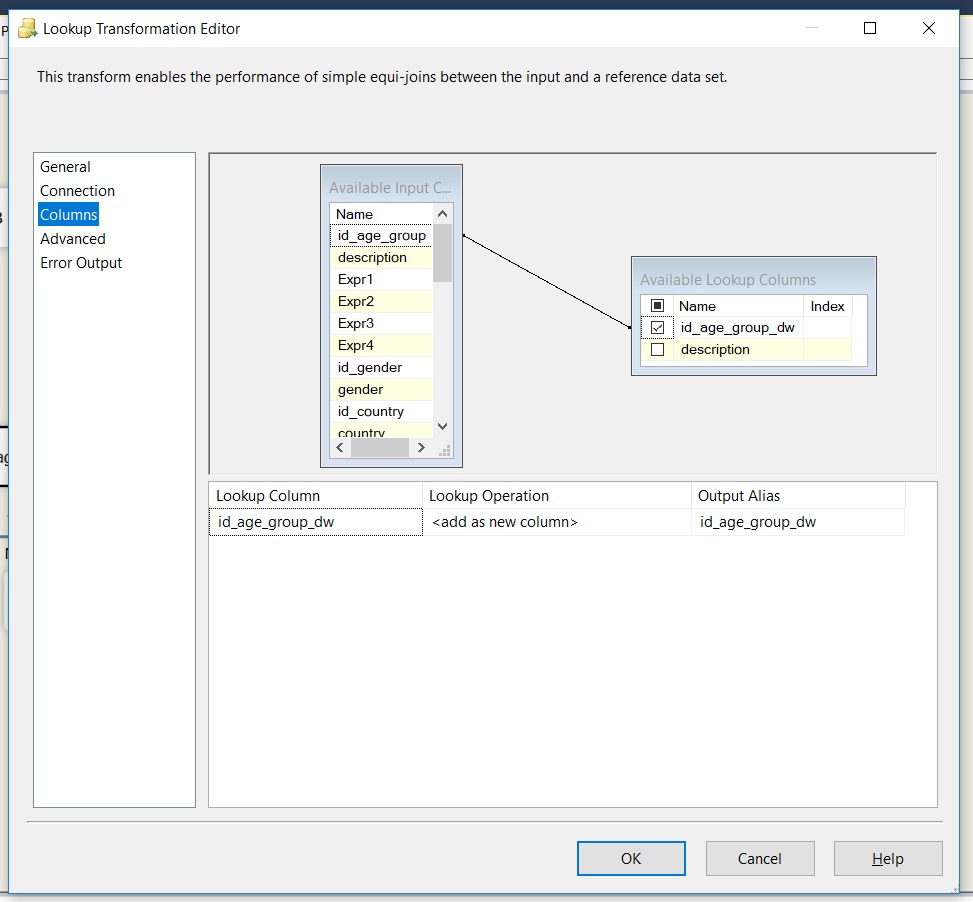
### Definiranje LookUp-a

Nakon što smo odredili Source, potrebno je napraviti LookUp-e. U „Connection“ polju pronalazimo OLEDB connection manager, gdje biramo naš DW, a ispod njega biramo tablicu koju želimo povezati.



Slika 30 LookUp Connection postavke

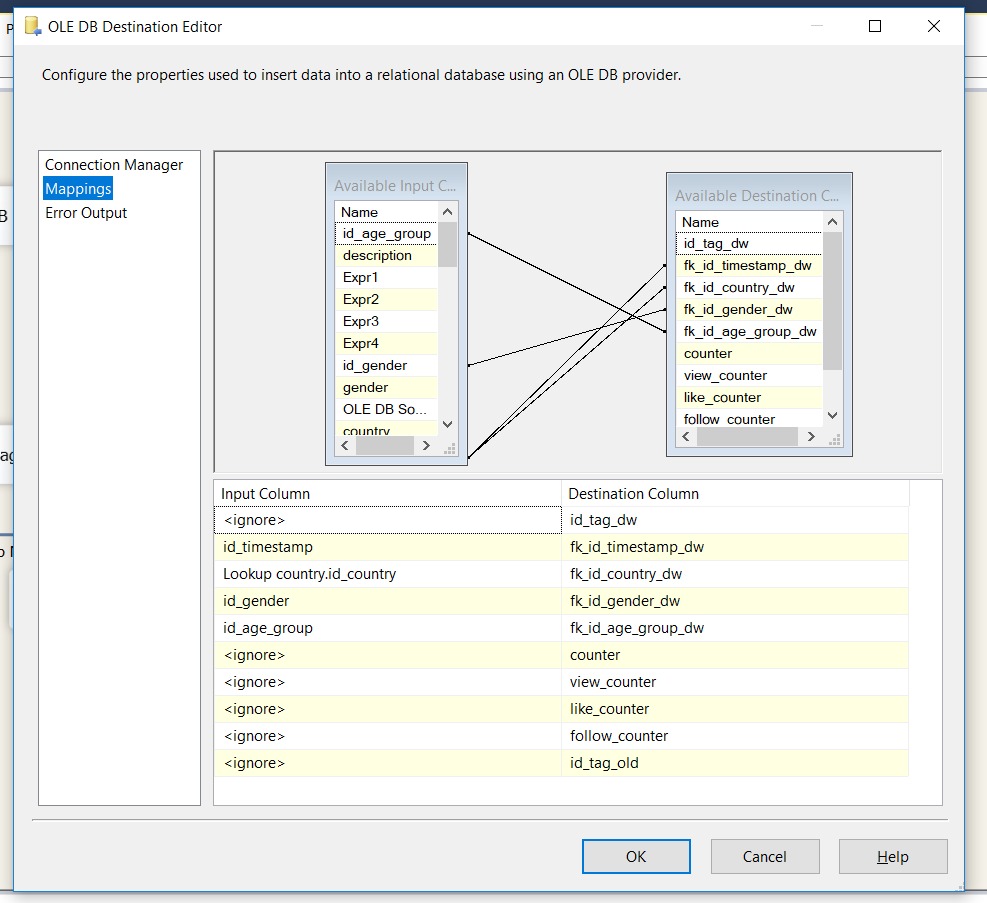
Nakon što smo to odabrali, u polju ispod „Connection“ nalazi se polje „Columns“. U njemu odabiremo polja koja povezujemo preko LookUp operacije.



Slika 31 LookUp Fact tablice

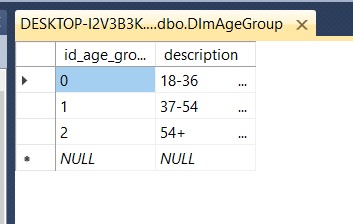
### OLEDB Destination

Ovdje je opisano mapiranje podataka u DW.

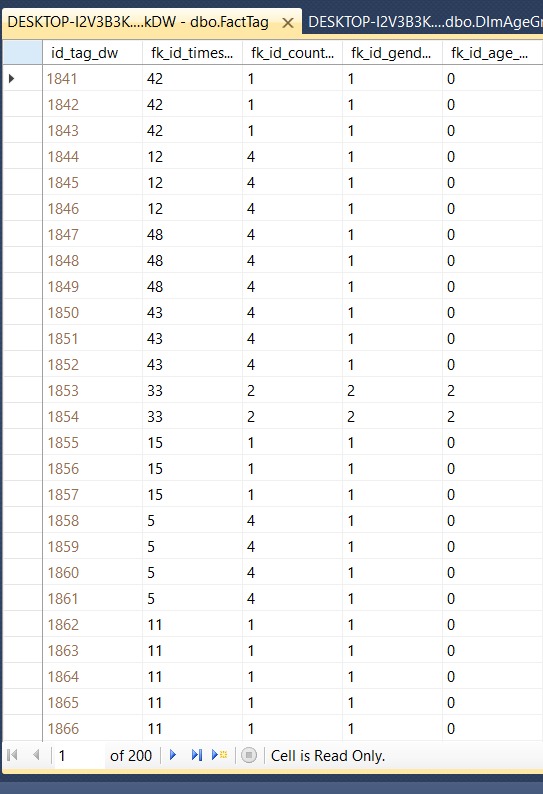


Slika 32 OLEDB Destination Fact Tablice

Temeljem tih pravila popunjavamo Fact tablicu:



Slika 33 Dimenzija AgeGroup za Fact tablicu



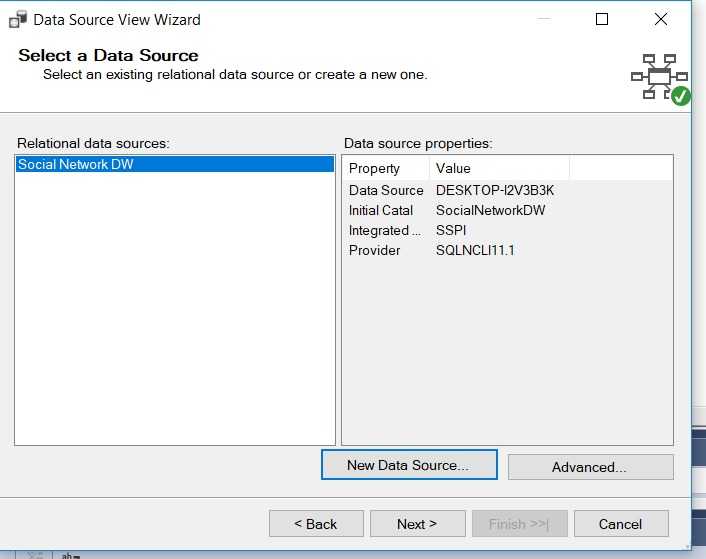
Slika 34 Napunjena Fact tablica

# SSAS kao povezinca DW i Kocke

Na temelju skladišta podataka potrebno je definirati OLAP kocku. Skladište je napunjeno podatcima kroz prethodne korake. Kocka je poseban zapis podataka, koji nam omogućava više dimenzionalno pretraživanje, te nam tako olakšava pristup podatcima koji se koriste u izradi izvještaja. OLAP kocka se izrađuje u Server Analysis Services.

## Izrada SQL Server Analysis Servis projekta (SSAS)

Da bi izgradili Kocku, moramo prvo kreirati SSAS projekt u SSDT-BI[[19]](#footnote-19). Nakon kreiranja dodajemo DW za izvor podataka iz koje će se generirati kocka.  
Odabiremo server na kojemu se nalazi skladište podataka, a za bazu se odabire naš data warehouse, SocialNetworkDW.



Slika 35 Data Source Wizard

## Definiranje pogleda

Da bi odabrali dimenzije iz izvora podataka ( SocialNetworkDW ) za izgradnju kocke, moramo definirati pogled, tj. Data Source view(DSV).

DSV sadrži:

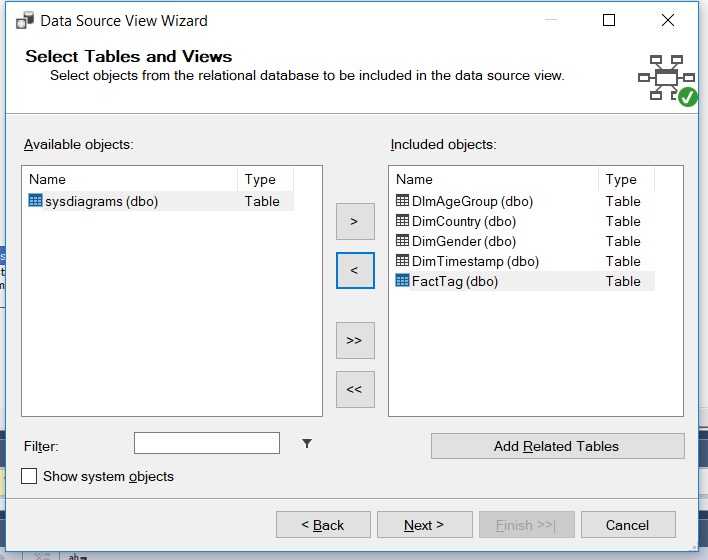
1. Naziv i opis
2. Definiciju bilo kojeg podskupa sheme preuzetih iz jednog ili više izvora podataka, uključujući cijelu shemu, uključujući sljedeće:

* Nazivi tablica
* Imena stupaca
* Tipovi podataka
* Null ability
* Duljine stupaca
* Primarni ključevi
* Primarni ključ – odnosi s vanjskim ključevima

1. Komentari na shemu iz temeljnih izvora podataka, uključujući sljedeće:

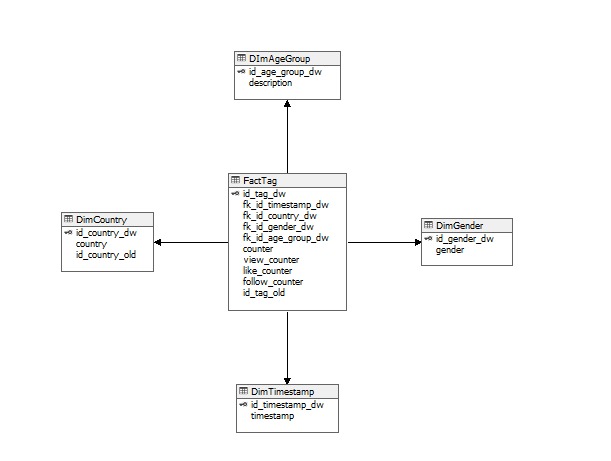
* Prijateljski nazivi za tablice, prikaze i stupce.
* Imenovani upiti koji vraćaju stupce iz jednog ili više izvora podataka (koji se prikazuju kao tablice u shemi).
* Imenovani izračuni koji vraćaju stupce iz izvora podataka (koji se prikazuju kao stupci u tablicama ili pogledima).
* Logički primarni ključevi (potrebni ako primarni ključ ne definira u temeljnoj tablici ili nije uključen u prikaz ili imenovani upit).
* Logički primarni ključ - odnosi stranih ključa između tablica, pogleda i imenovanih upita.

Nakon što smo kreirali DSV, moramo odabrati Fact tablicu i dimenzije. Prvo se radi odabir Fact, a onda dimenzijskih tablica.



Slika 36 Odabir Fact tablica i dimenzija

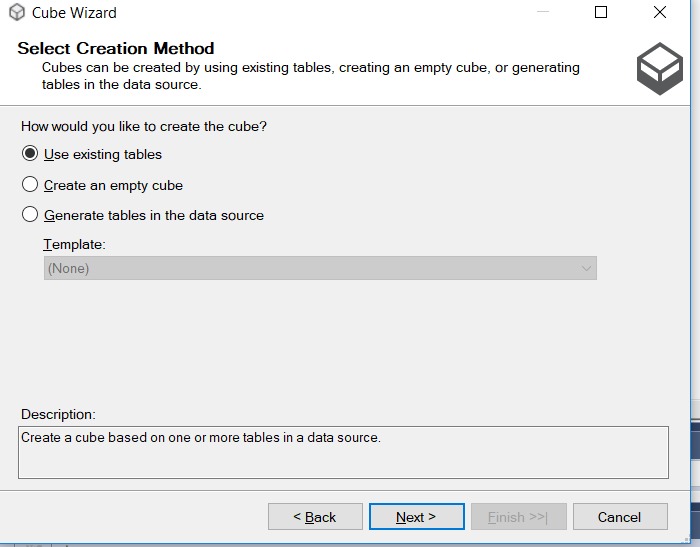
Na kraju dobijemo naš view – pogled.



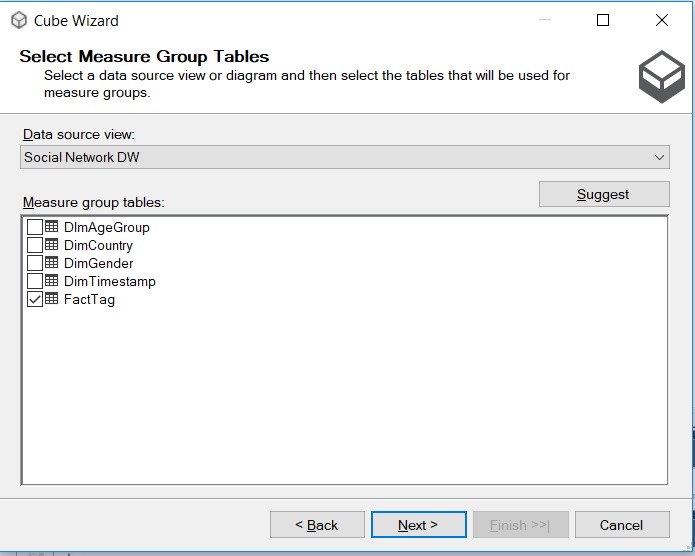
Slika 37 Data Source view

## Stvaranje kocke

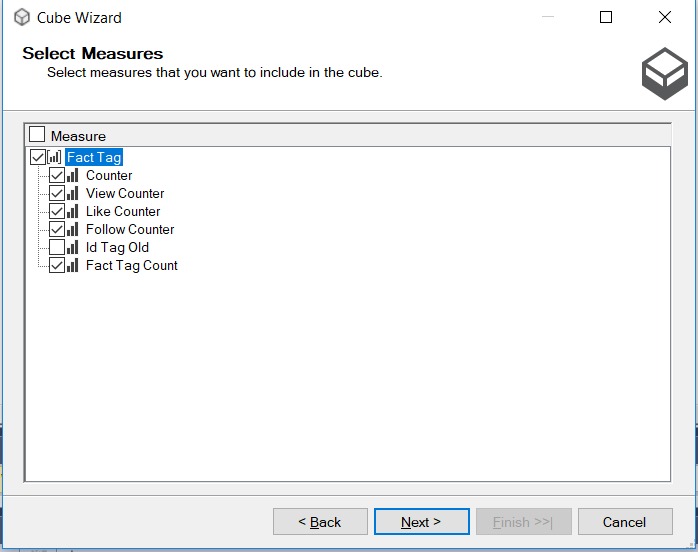
U ovom poglavlju opisano je kreiranje kocke na osnovu pogleda (view) iz prethodnog koraka. Koristimo Cube – Wizard za obradu Fact tablice, dimenzije i mjere. Opisano kao slika-korak.  
  
Odabiremo Use Existing Tables,



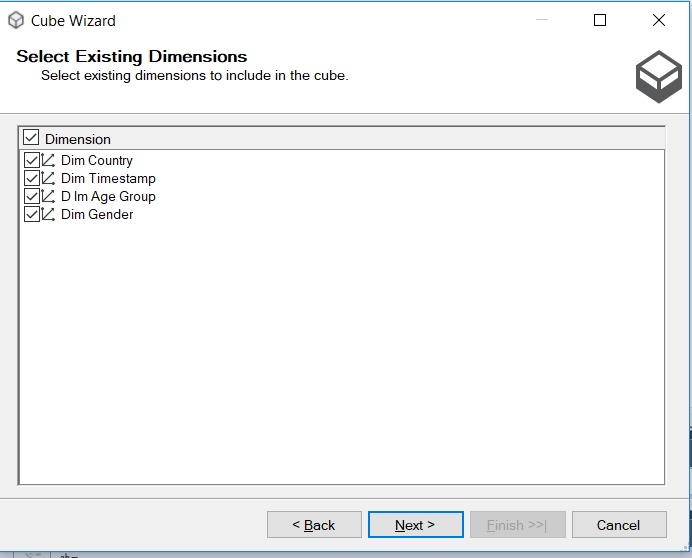
Odabiremo Fact tablicu,



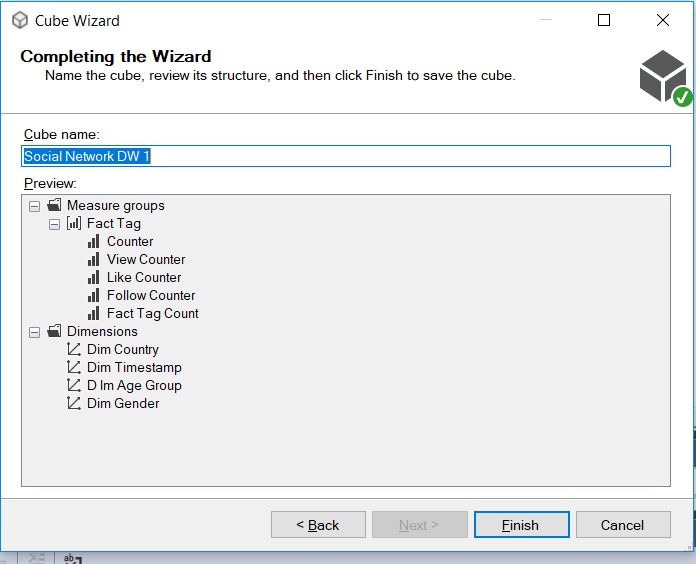
Odabiremo mjere,



Odabiremo dimenzije,

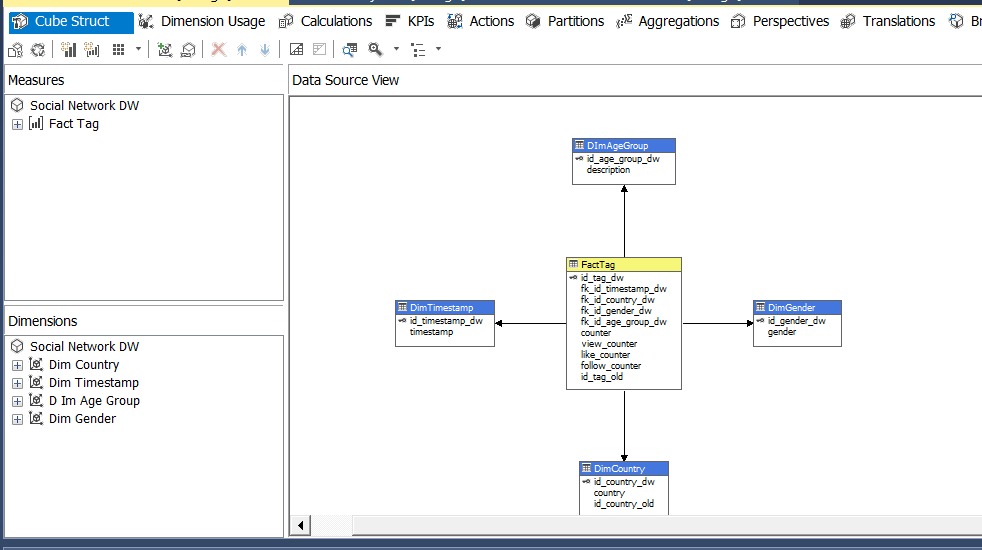


Završavamo proces izrade kocke, imenujemo ju i dobijamo na uvid strukturu.



Slika 38 Ukupan proces izrade kocke

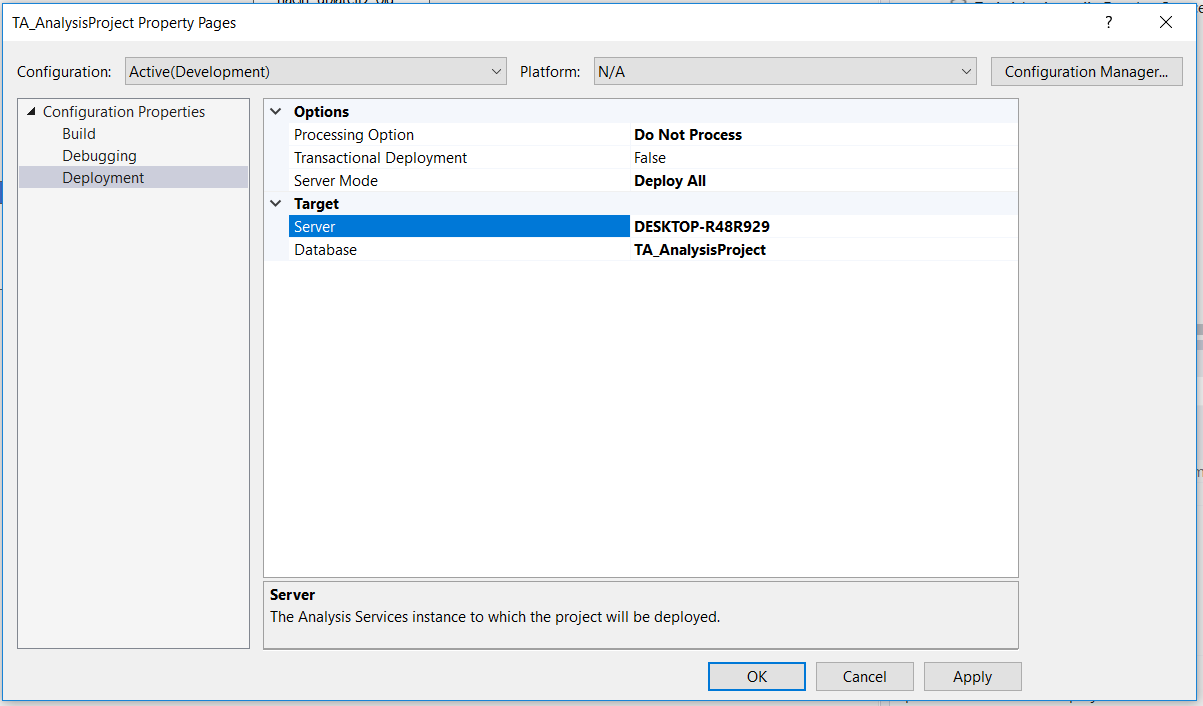
Na kraju dobijemo strukturu kocke oblika:



Slika 39 Struktura Kocke

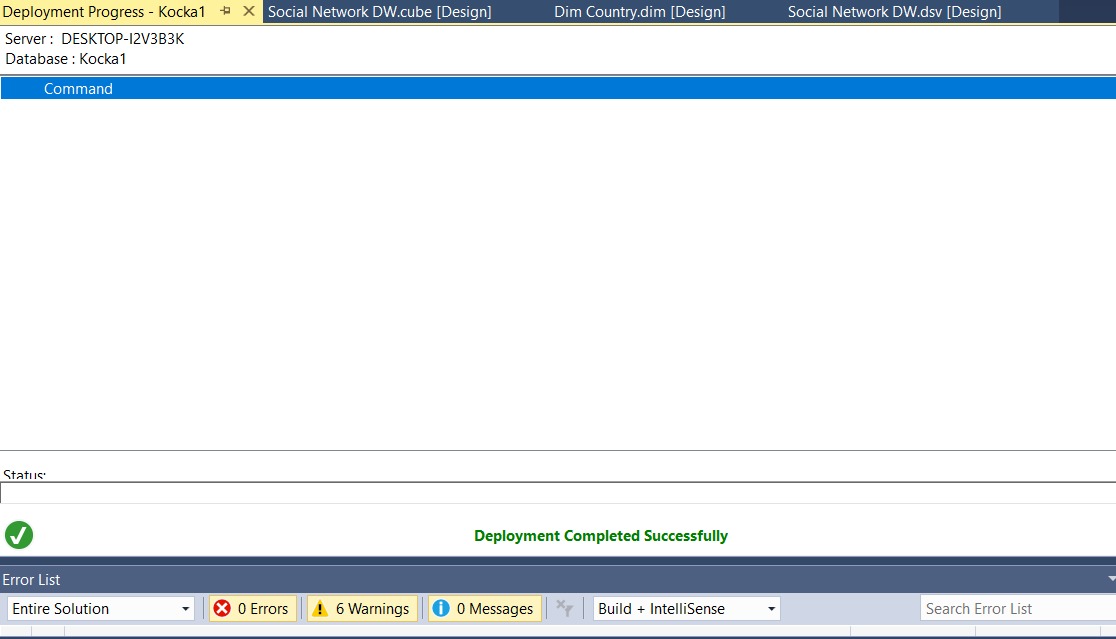
## Deploy Kocke

Kocku moramo razviti preko SQL Server Analysis Services Server-a, da bi pomoću alata CubePlayer mogli spojiti na Kocku i generirati izvještaj.



Slika 40 Postavke za spajanje na server

Da bi deploy-ali kocku, koristimo debugging, (prečica F5). To započinje procesiranje podataka.



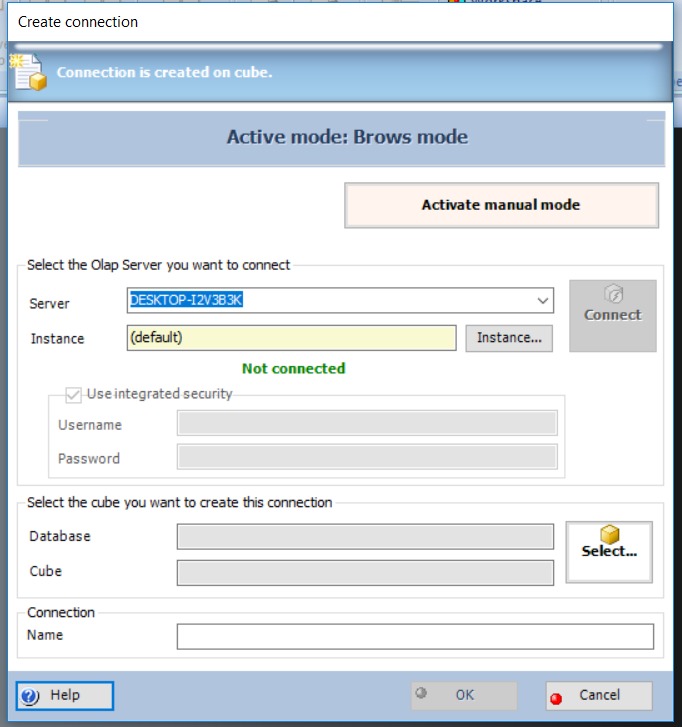
Slika 41 Uspješan deploy

# Analiza podataka kocke

Nakon što smo napokon stvorili kocku, sada se spajamo na server i kreiramo izvještaj.

## SSAS

Prvi korak je spajanje na SSAS.



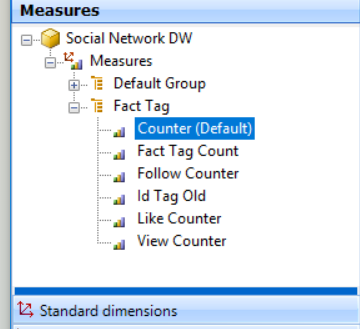
Slika 42 Spajanje na SSAS



Slika 43 Pokrenuti meni SSAS

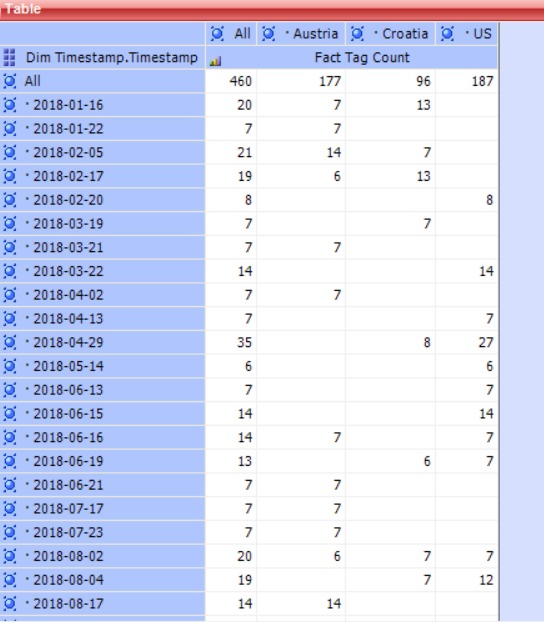
## Izrada izvješća

Da bi izradili izvještaj, definiramo mjere koje nas zanimaju. U našem slučaju biramo mjeru „Counter“ internu mjeru, a vrijeme pratimo prema dimenziji „Timestamp“.



Slika 44 Interna mjera Counter

Izvještaj nam javlja koliko se pojedinih „tag-ova“ koristilo, u danom vremenskom periodu.



Slika 45 Izvještaj

# Zaključak

Jedan od najvažnijih procesa poslovnih sustava je upravo praćenje statistike ( izvještaj ).  
Zato je kroz ovaj seminarski rad obrađen razvoj i analiza DW pomoću MSQL Server tehnologije. Za izradu izvještaja se koristio CubePlayer alat.

Bitno je paziti na pravilan unos podataka u skladište, odnosno imati mjeru zaštite ako ulaze neispravni ( dirty ) podatci u skladište, da se promjene po vrsti podatka ili se odbace.  
Ako imamo loš unos podataka sam izvještaj postaje irelevantan, jer mala greška može rezultirati katastrofalnim pomakom u statistici. U najboljem slučaju ćemo dobiti nešto što je očito krivo te moći odbaciti logički, što nije uvijek slučaj.

Podatke koje pratimo kroz vrijeme grupiramo u dimenzije. Mjera nam određuje koliko određenog sadržaja imamo po uvjetu dimenzije. Skupa dimenzije i mjera čine Fact tablicu.

Pomoću Fact tablice izrađujemo izvještaj, koji je uveliko brži način obrade podataka od staromodnog transakcijskog sustava. Zbog tog i velikog niza drugih prednosti danas je upravo u korištenju ovaj sustav.

1. TimeStamp – vremenski žig(oznaka) [↑](#footnote-ref-1)
2. Post – objava [↑](#footnote-ref-2)
3. User – korisnik [↑](#footnote-ref-3)
4. Tag – oznaka [↑](#footnote-ref-4)
5. ActionTaken – poduzeta akcija [↑](#footnote-ref-5)
6. View – pogled [↑](#footnote-ref-6)
7. Like – sviđati [↑](#footnote-ref-7)
8. Follow – slijediti [↑](#footnote-ref-8)
9. Country – država [↑](#footnote-ref-9)
10. Gender – rod [↑](#footnote-ref-10)
11. AgeGroup – dobna skupina [↑](#footnote-ref-11)
12. DW – Data Warehouse [↑](#footnote-ref-12)
13. Dimension – dimenzija [↑](#footnote-ref-13)
14. Fact – činjenica [↑](#footnote-ref-14)
15. SSIS – SQL server integration services [↑](#footnote-ref-15)
16. Control Flow – kontrola toka(podataka) [↑](#footnote-ref-16)
17. Data Flow Task(DFT) – zadatak toka podataka [↑](#footnote-ref-17)
18. Object Linking and Embedding, Database [↑](#footnote-ref-18)
19. SSDT-BI – SQL Server Data Tools - Business Intelligence [↑](#footnote-ref-19)