



Seminarski rad iz predmeta Skladišta podataka
Tema: Skladišta podataka za maloprodaju kompjuterske opreme

Studenti: Harun Hadžić (20-ST) i Bakir Karović (6-ST)

Sarajevo, maj 2019. godine

1. Uvod.....	2
2. UML reprezentacija relacijske baze podataka	2
2.1. ER dijagram	3
2.2. Use case dijagram.....	4
2.3. Dijagram aktivnosti	5
2.4. Dijagram klasa.....	7
3. Implementacija relacijske baze podataka u MS SQL-u	8
4. Implementacija DWH	8
5. ETL proces.....	10
5.1. Ekstrakcija/izdvajanje	11
5.2. Transformacija	11
5.3. Učitavanje	11
5.4. Opis procesa.....	11
6. Kreiranje kocke	16
7. Implementacija izvještaja.....	22
8. Zaključak	28
9. Reference.....	28

1. Uvod

DWH predstavlja bazu podataka koja omogućuje brzo i jednostavno izvođenje pretraga i upita nad velikim količinama podataka, kao i skup podataka na kojem se bazira sistem podrške u odlučivanju. Namijenjena je menadžerima, ali i svima koji u svom poslu obavljaju analitičke zadatke. Sadrži ogromne količine podataka, koji se koriste u svrhu poslovnih analiza i postizanja što boljih tržišnih rezultata. Očevima DWH se smatraju Bill Inmon i Ralph Kimball.

U ovom radu bit će prikazana kako relacijska baza podataka tako i DWH za maloprodaju kompjuterske opreme.

2. UML reprezentacija relacijske baze podataka

UML (Unified Modelling Language) služi za:

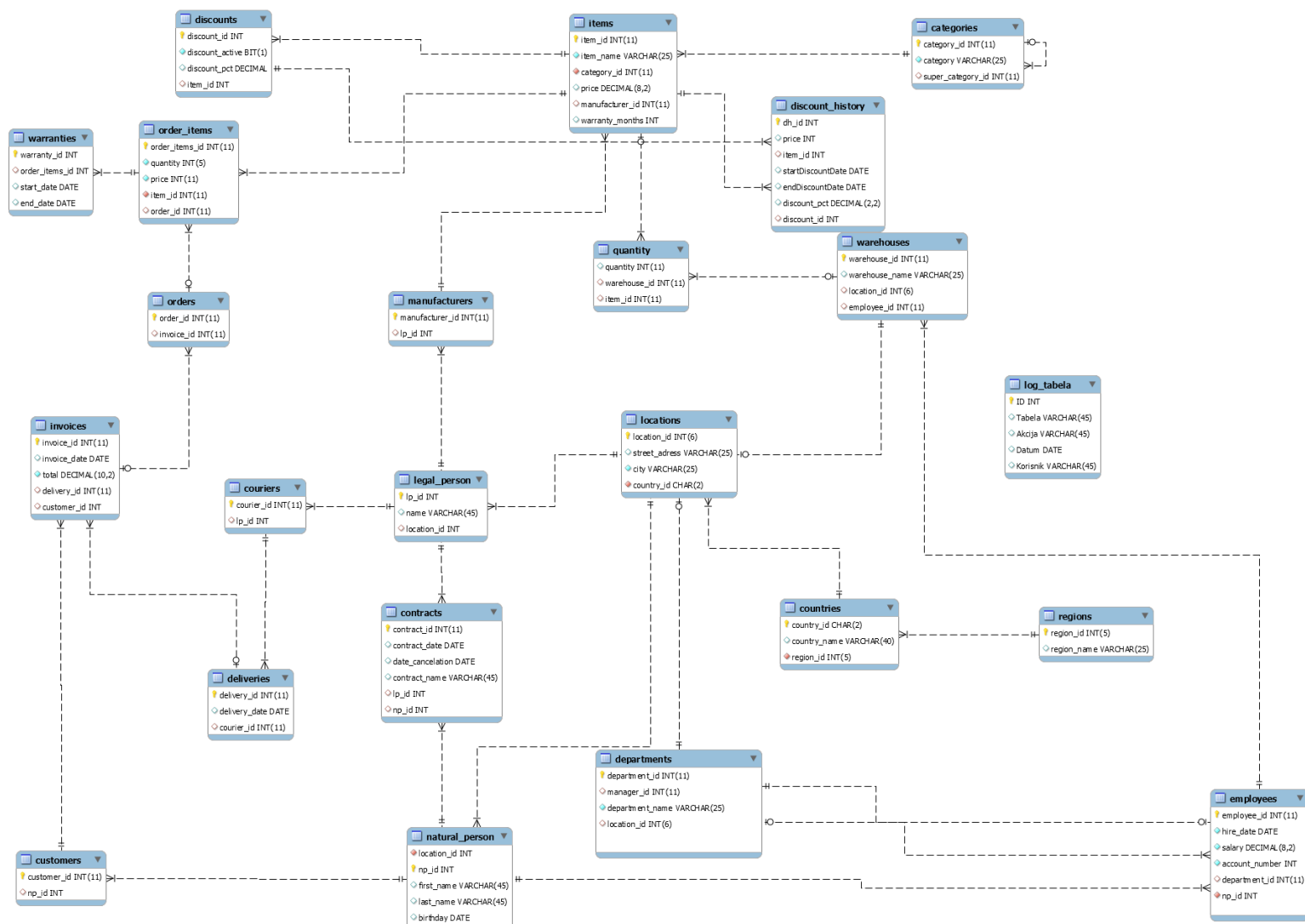
- Specifikaciju
- Vizuelizaciju
- Konstrukciju
- Dokumentaciju razvoja softvera

2.1. ER dijagram

Entity-Relationship (ER) model ili model objekti-veze (MOV) predstavlja detaljan logički prikaz podataka preko skupa entiteta, njihovih atributa i međusobnih veza.

Osnovni elementi modela su:

- entiteti (objekti)
- veze
- atributi

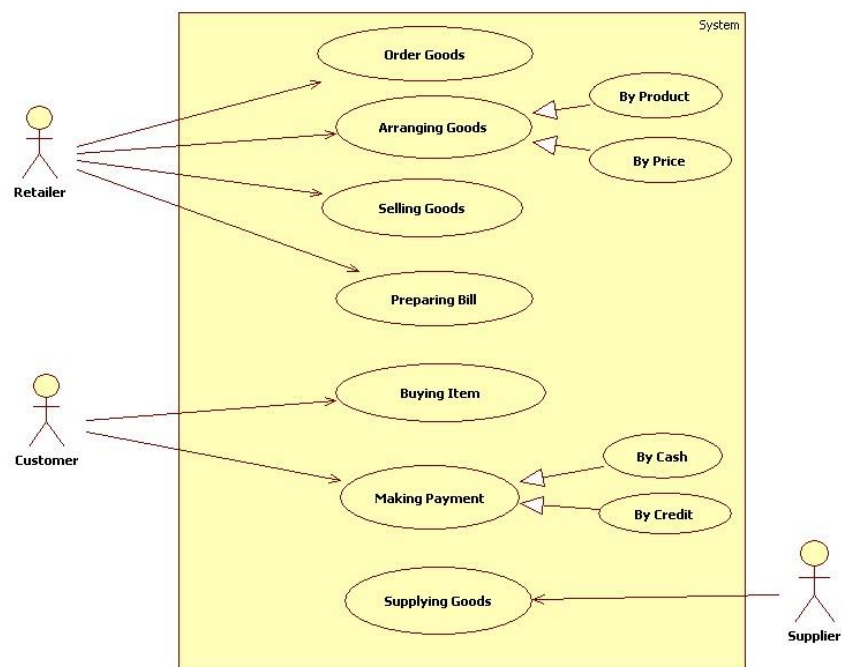


Slika 1 - ERD

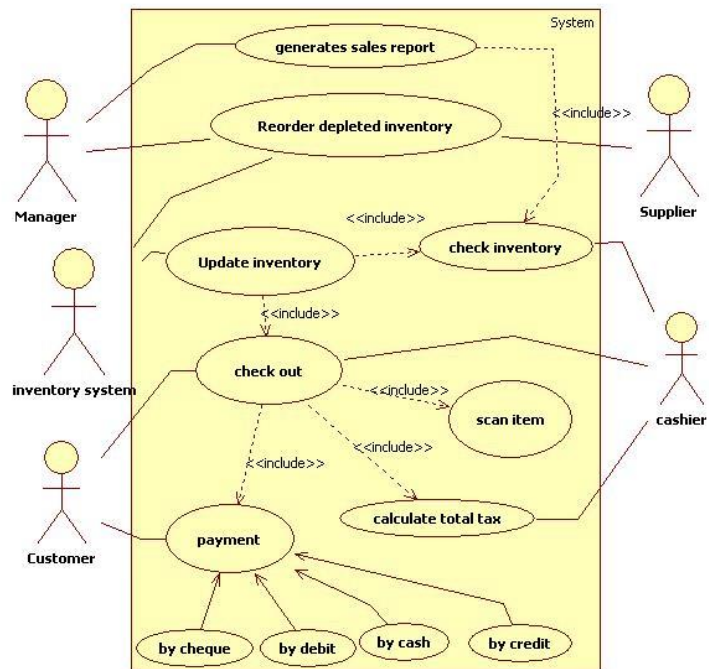
2.2. Use case diagram

Izlaz iz faze Analize korisničkih zahtjeva je dokument Specifikacije korisničkih zahtjeva u kojem su precizno opisana očekivanja korisnika od sistema.

Jedan od alata koji se tu često koriste su dijagrami slučajeva upotrebe (use-case diagram) - na slici desno. Na dijagramu su skicirani učesnici tj. aktori ("čiča gliše"), slučajevi upotrebe (ovalni oblici) i njihovi međusobni odnosi (različite vrste strelica), te granice sistema.



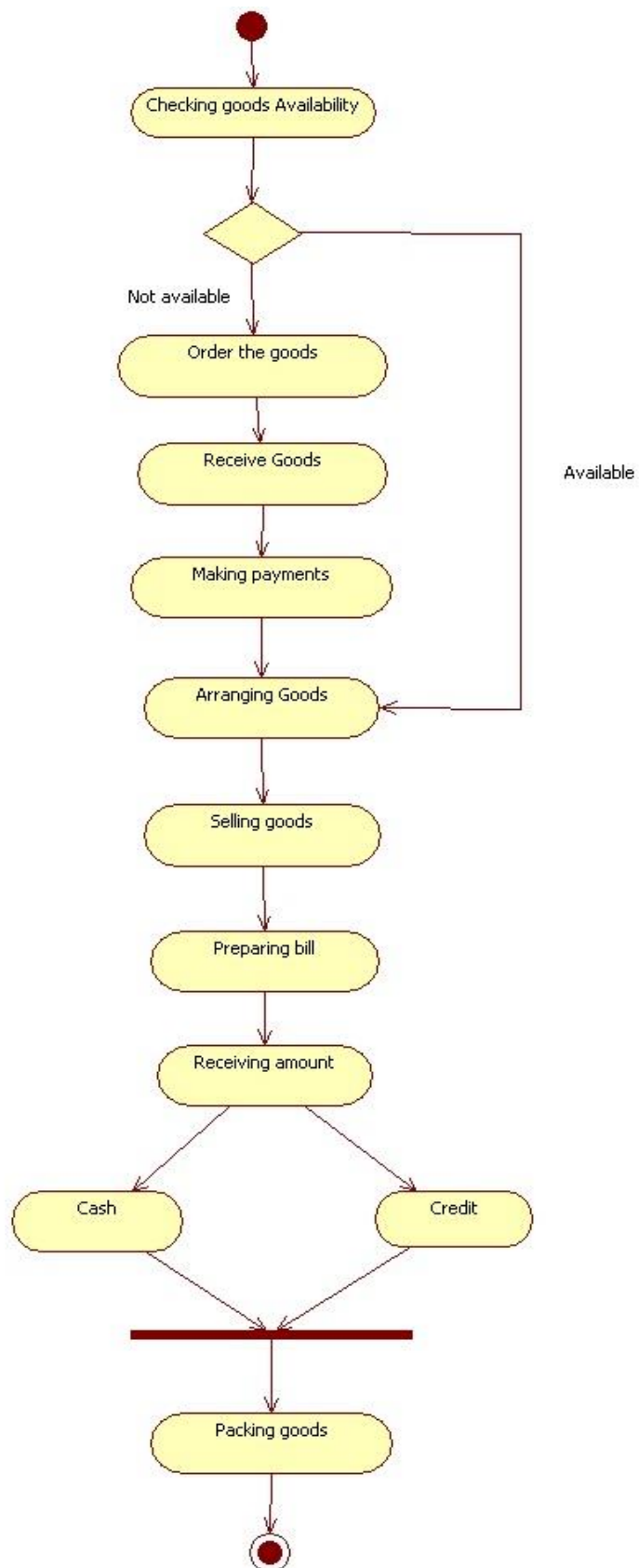
Slika 2 - Use Case dijagram za proces kupovine



Slika 3 - Use Case dijagram za cjelokupni sistem

2.3. Dijagram aktivnosti

Dijagram aktivnosti (eng. activity diagram) dozvoljava nam da specificiramo kako sistem radi. Dijagram aktivnosti je posebno dobar za opisivanje poslovnih procesa i poslovnih tokova. Ovaj dijagram je jedini dijagram koji pripada procesnom pogledu na sistem.



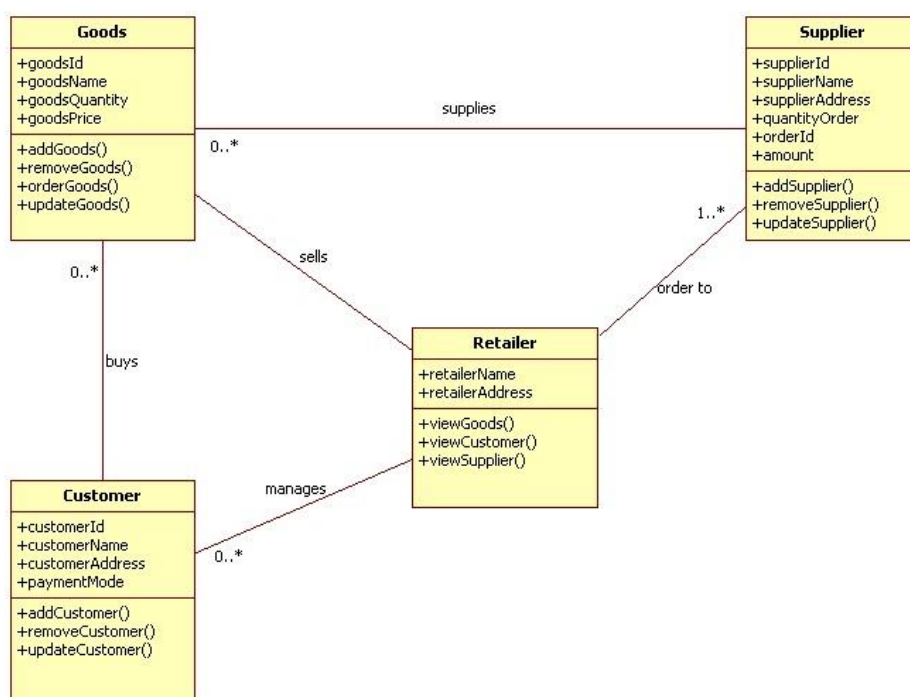
Slika 4 - Dijagram aktivnosti

2.4. Dijagram klasa

Osnovna gradivna jedinica dijagrama klasa je *klasa*. Klasa se predstavlja pravougaonikom podijeljenim po vertikali na tri dijela: naziv klase, atributi i metode.

Oznaka **atributa** sastoji se od: anotacije vidljivosti (+ public, - private, # protected, ~ package), naziv atributa : tip atributa, te opcionalno: oznaka kardinalnosti ([m..n] gdje m označava minimalan broj vrijednosti, a n maksimalan), default vrijednost itd.

Za **metode** se također navodi vidljivost, naziv : tip povratne vrijednosti, te parametri (naziv : tip). Posebne metode su <<create>> (konstruktor) i <<destroy>> (destruktor).



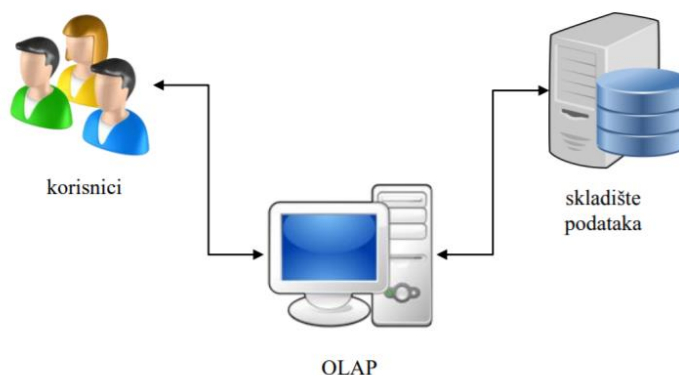
Slika 5 - Dijagram klasa

3. Implementacija relacijske baze podataka u MS SQL-u

Sav kod potreban za kreiranje relacijske baze podataka nalazi se u priloženom dokumentu. Također, analizom ERD-a (slika 1) možete zaključiti funkcionalnosti ovog sistema kao i sve što ova baza nudi. Ova baza će nam služiti kao izvorna baza podataka za naše skladište. Iskoristit ćemo priliku da spomenemo OLTP i OLAP sisteme koji su neophodni kako za relacijske tako i za DWH.

Baze podataka koriste OLTP (Online Transaction Processing) sisteme da trenutno obrađuju podatke nastale u poslovnim transakcijama i bazirani su na relacionoj tehnologiji

OLAP baze podataka omogućavaju upite poslovne inteligencije (Business Intelligence – BI). OLAP je tehnologija baze podataka koja je optimizovana za upite i pravljenje izvještaja, a ne za obradu transakcija. Izvorni podaci za OLAP su OLTP (Online Transactional Processing) baze podataka, koje se obično nalaze u skladištima za podatke.



Slika 6 - OLAP sistem

4. Implementacija DWH

Implementacija skladišta podataka se sastoji od sljedećih koraka:

- Proces izgradnje započinje definiranjem korisničkih zahtjeva.
- Analiziraju se zahtjevi i postojeći podaci u operacijskoj bazi.
- Izgrađuje se logički model skladišta podataka.
- Podaci se logički transformiraju prilikom izvlačenja iz transakcijske baze i učitavanja u skladište podataka.
- Podaci u skladištu podataka se modeliraju u dimenzijski model podataka pogodan za izvođenje miliona upita istovremeno.
- Podaci se zatim fizički dimenzioniraju (pročišćavaju i pojednostavljaju) što osigurava razumijevanje pojedinih podataka.
- Osigurava se učitavanje novih podataka u dogovorenim/potrebnim vremenskim intervalima.

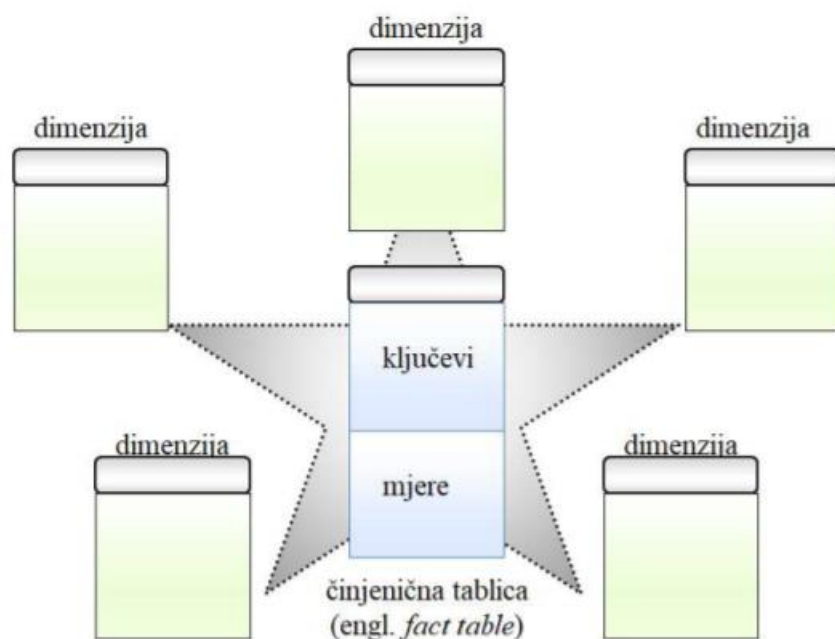
- Izgradnja skladišta je gotova, ali je skladište potrebno nadgledati i upravljati njime, te ako se javi potreba i rekonstruisati/modificirati (djelimično ili potpuno).

Naše skladište podataka će koristiti zvjezdastu shemu. Zvjezdasta shema podrazumijeva da je svaka dimenzija za analizu u potpunosti denormalizirana, dok pahuljičasta shema nije u potpunosti denormalizirana. Denormalizacija se radi kako bi broj tablica koje se dohvaćaju u korisničkom upitu bio što manji, te kako bi upiti radili brže. Dimenzijsko oblikovanje jest tehnika logičkog oblikovanja koja ima za cilj predstaviti podatke u jednostavnom, intuitivnom obliku koji dopušta vrlo učinkovit pristup (pregled). Dakle, dimenzijski model pokušava riješiti nedostatak relacijskog modela (u analitičkom smislu) i pružiti krajnjim korisnicima brz i jednostavan pregled podataka.

Dimenzijska tablica odgovara subjektu/objektu koji sudjeluje u procesu koji se prati u skladištu podataka. Dimenzijske tablice daju kontekst, odnosno objašnjavaju i opisuju činjenice pohranjene u činjeničnoj tablici. Dimenzijske tablice nisu normalizirane.

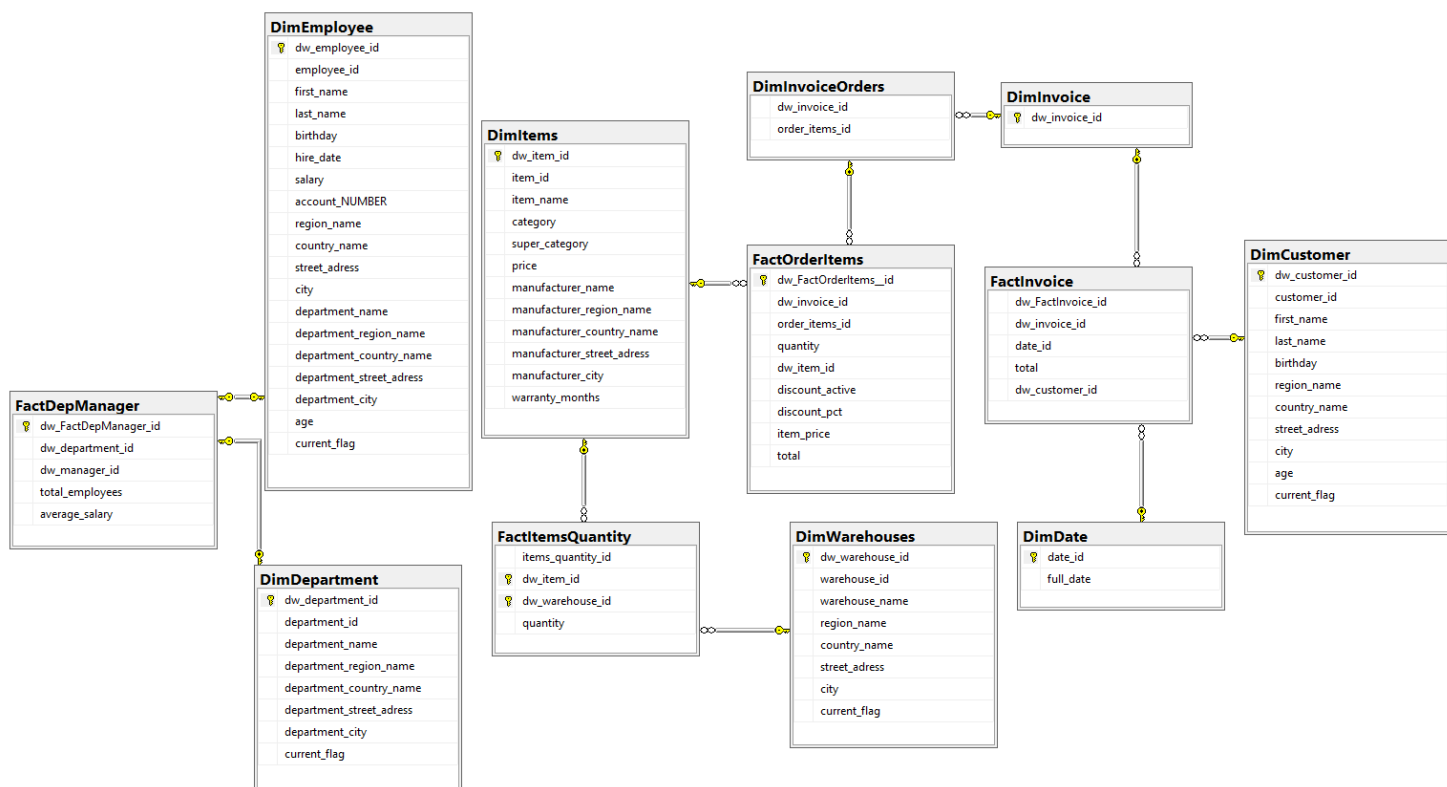
U središtu zvjezdastog modela je činjenična tablica koja odgovara procesu koji se prati u skladištu podataka. U nekom skladištu podataka možemo naravno imati N zvjezdastih spojeva i N činjeničnih tablica. Činjenična tablica sadrži dvije skupine numeričkih atributa:

- Ključevi dimenzijskih tablica
- Mjere (eng. measures)



Slika 7 - Zvjezdasta shema

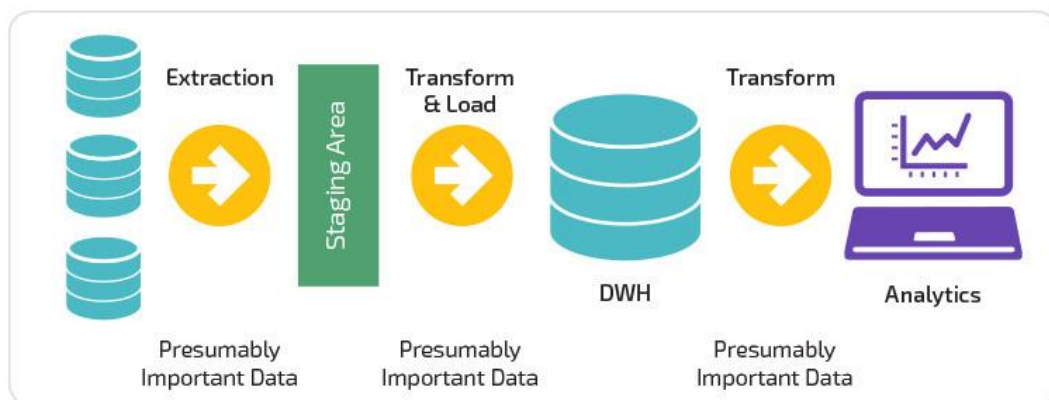
Naša DWH se sastoji o 4 star sheme koje će biti predstavljene u nastavku kao i krajnje skladište podataka. Atribute primarne, surogatne ključeve te njihove veze možete veoma lahko analizirati sa dijagramima.



Slika 8 - ER dijagram DW-a

5. ETL proces

ETL predstavlja skraćenicu za ekstrakciju/izdvajanje podataka (iz različitih OLAPa), transformaciju (svođenje podataka iz različitih OLAP-a na isti kontekst) i učitavanje podataka (u DWH).



Slika 9 – ETL proces

5.1. Ekstrakcija/izdvajanje

Podaci se trebaju uzeti iz izvora podataka, kako bi se mogli staviti u DW. Postoje potpuna i inkrementalna izdvajanja podataka.

Inicijalno izdvajanje podataka podrazumijeva prvo izvlačenje podataka. Također zahtjeva pripremu logičke mape.

Za naredna (inkrementalna) izvlačenja podataka podrazumijevaju da će se samo promjene učitavati nakon prvog/inicijalnog učitavanja.

5.2. Transformacija

Proces promjene strukture, sadržaja ili drugih karakteristika podataka, kako bi se podaci iskoristili za odgovarajuću upotrebu naziva je transformacija.

Sastoji se od dva glavna koraka:

- Čišćenje podataka
- Integracija podataka

5.3. Učitavanje

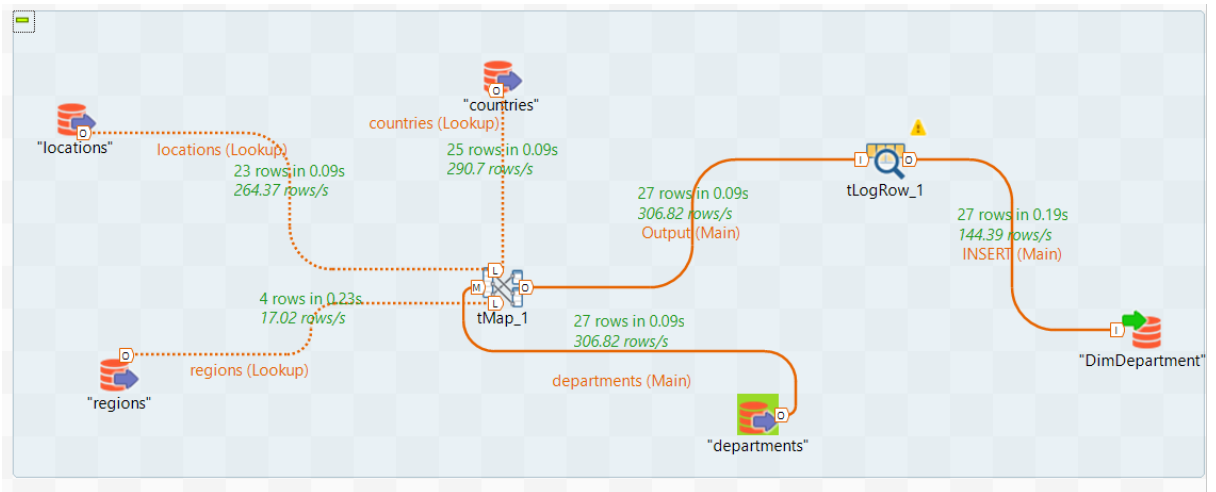
Proces učitavanja podataka može se kao i kod izdvajanja podijeliti na:

- Proces inicijalnog učitavanja
- Naredna učitavanja (učitavanja tokom vremena)

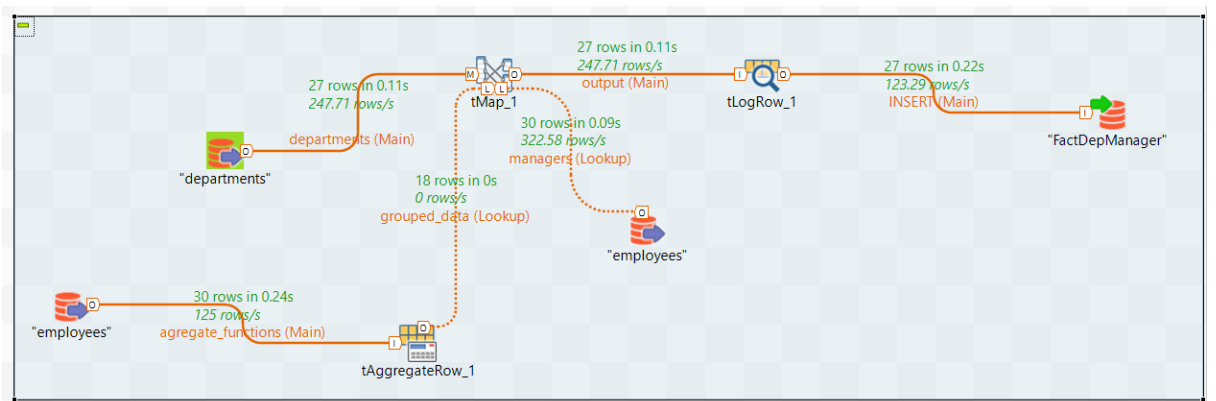
5.4. Opis procesa

Za ETL proces je korišten besplatni open source alat Talend Open Studio. Proces je podijeljen na 11 različitih poslova koji popunjavaju 12 tabela unutar dw-a. Šematski prikaz procesa je dat u nastavku. Cijeli projekat je moguće preuzeti sa linka u mail-u. Prvo je potrebno popuniti tj. pokrenuti poslove koji popunjavaju dimenzije pa onda tabele činjenica.

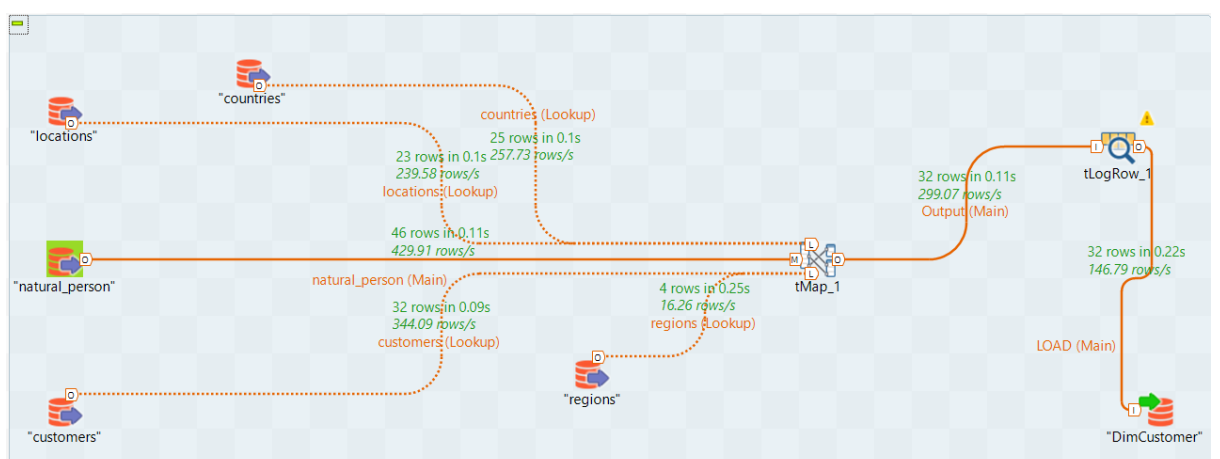
Pri opisu ETL procesa, potrebno je specificirati tabele iz relacione baze koje se dobiju kopiranjem same šeme pri konekciji na bazu. Nakon toga je te podatke potrebno filtrirati ili grupisati te ih spojiti (JOIN) preko tMape. Sljedeći korak je prikaz istih na ekran radi ručne inspekcije te dodavanje output tabele dw-a u koju će se isti podaci zapisati.



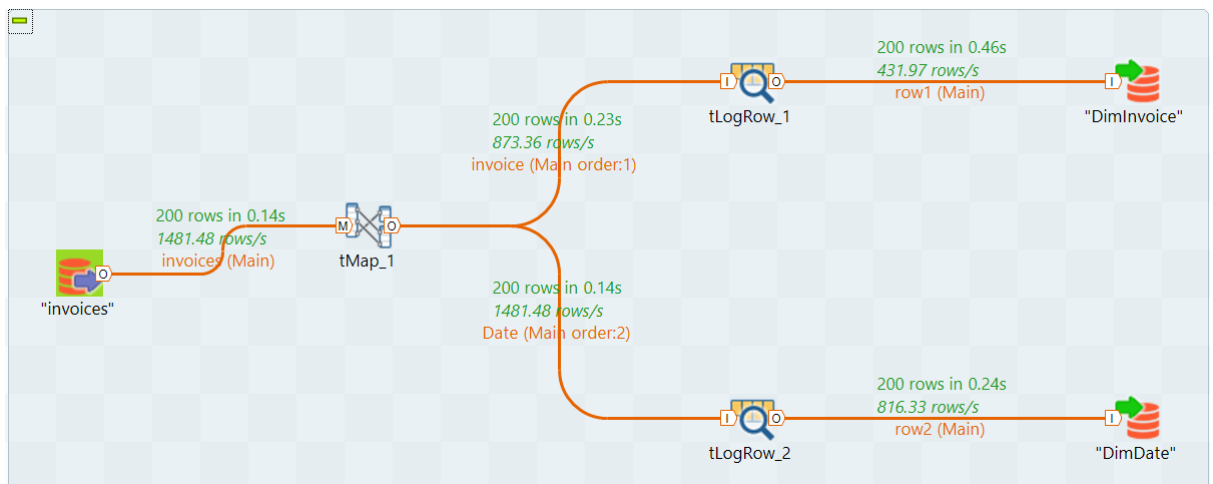
Slika 10 - ETL za DimDepartment



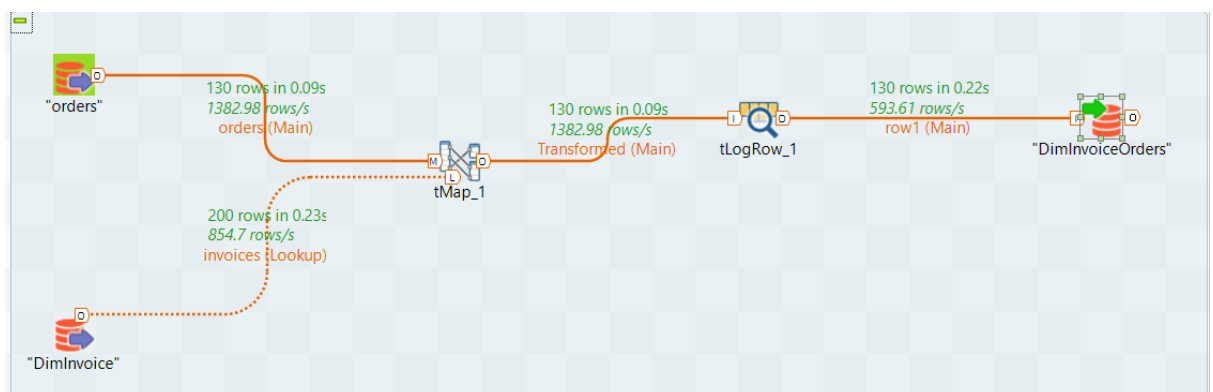
Slika 11 - ETL za FactDepManager



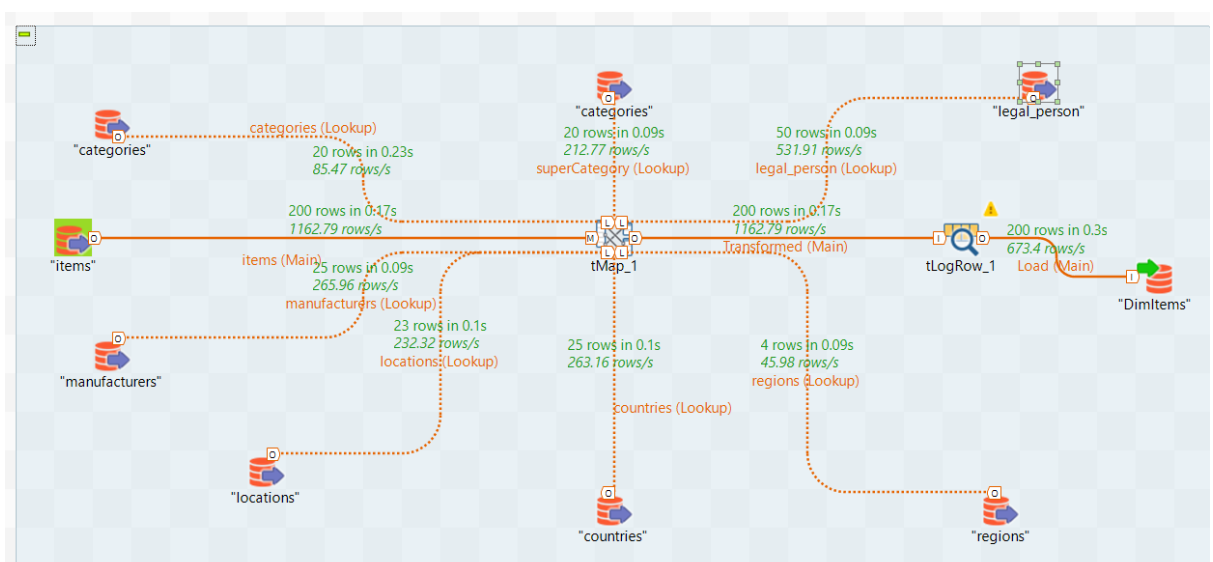
Slika 12 - ETL za DimCustomer



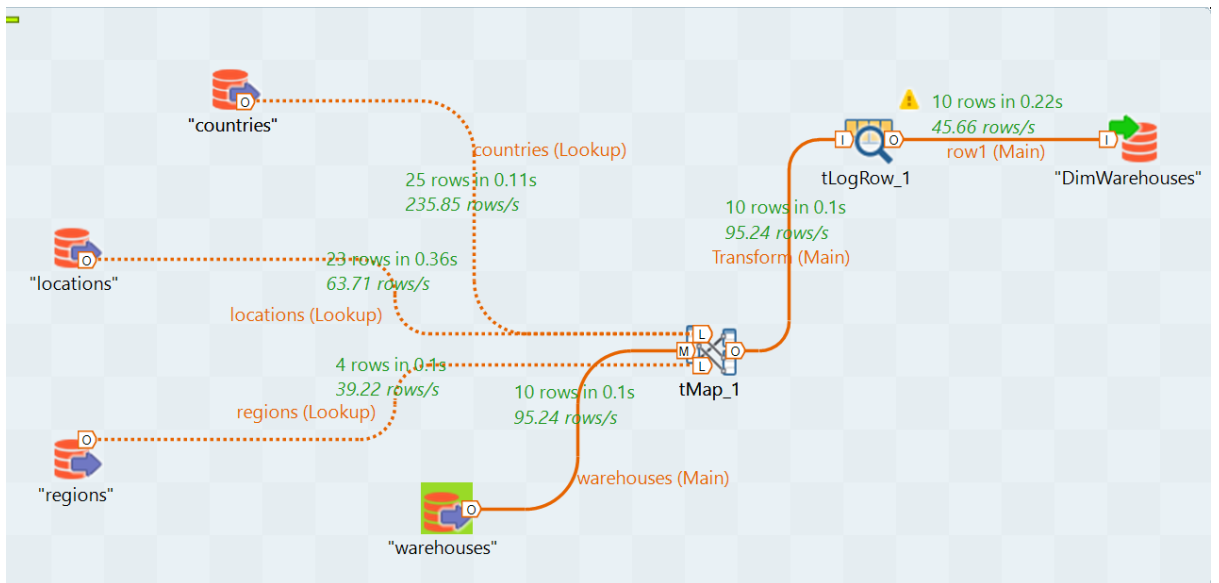
Slika 13 - ETL za DimInvoice i DimDate



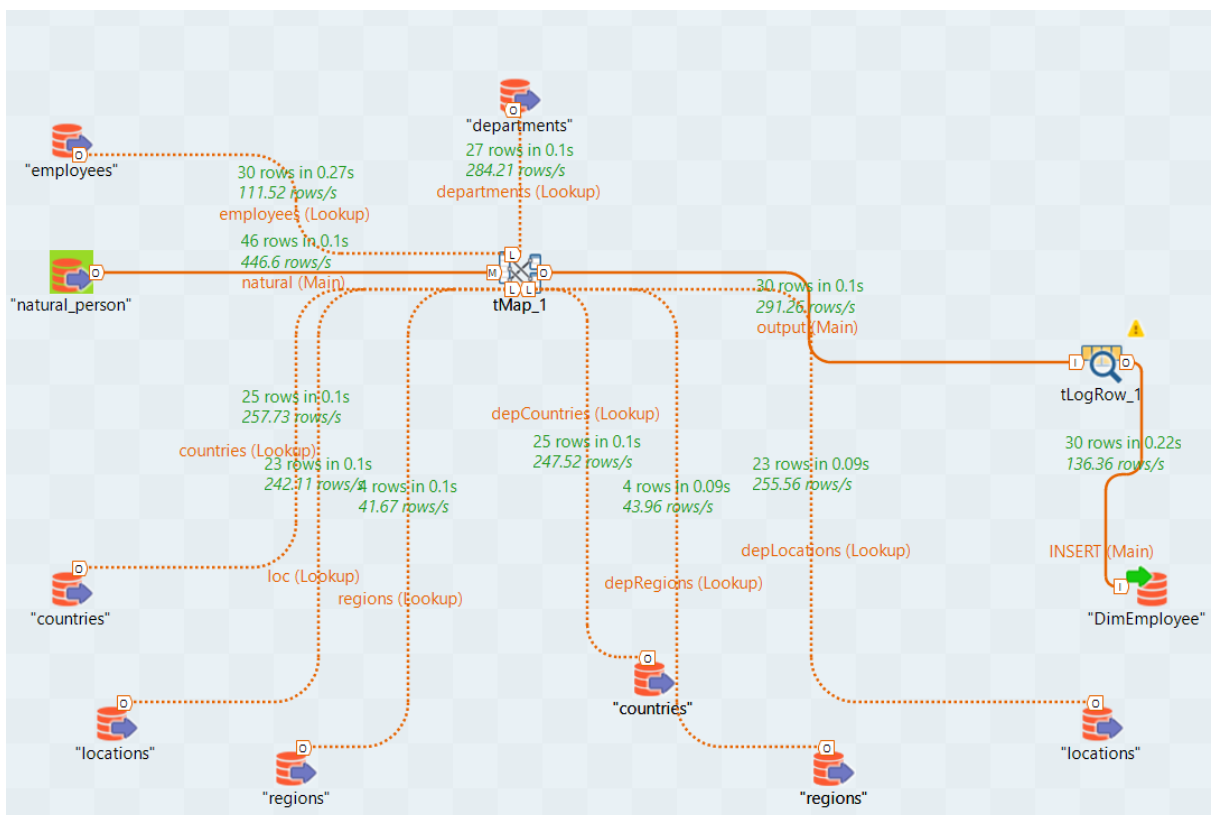
Slika 14 - ETL za DimInvoiceOrders



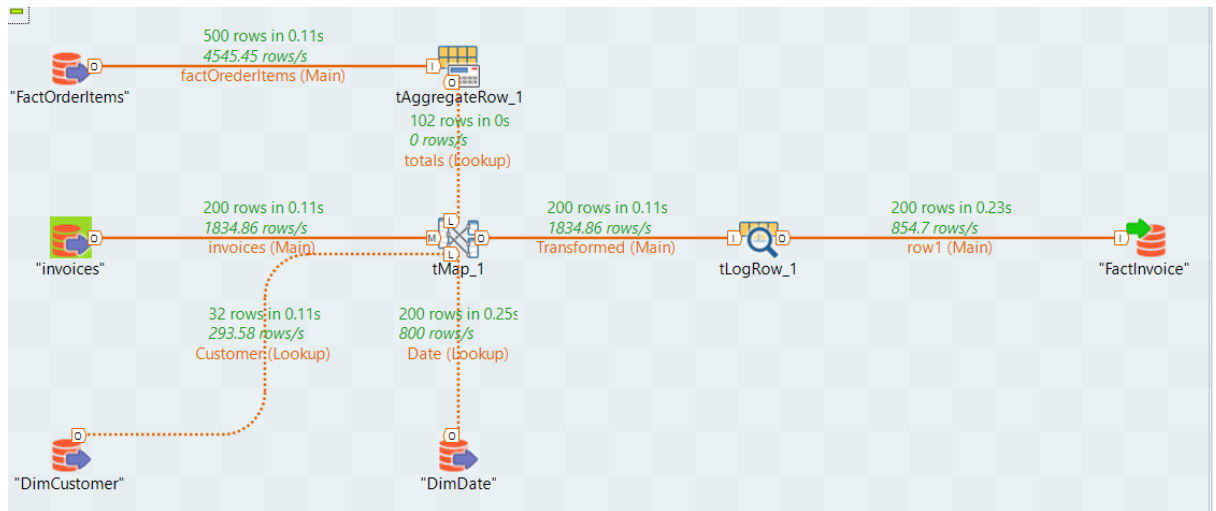
Slika 15 - ETL za DimItems



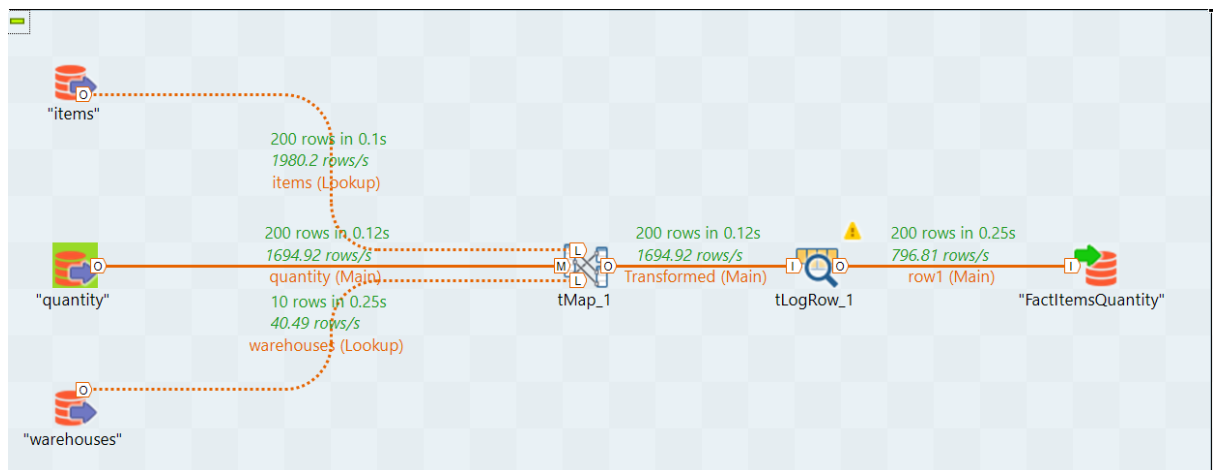
Slika 16 - ETL za DimWarehouses



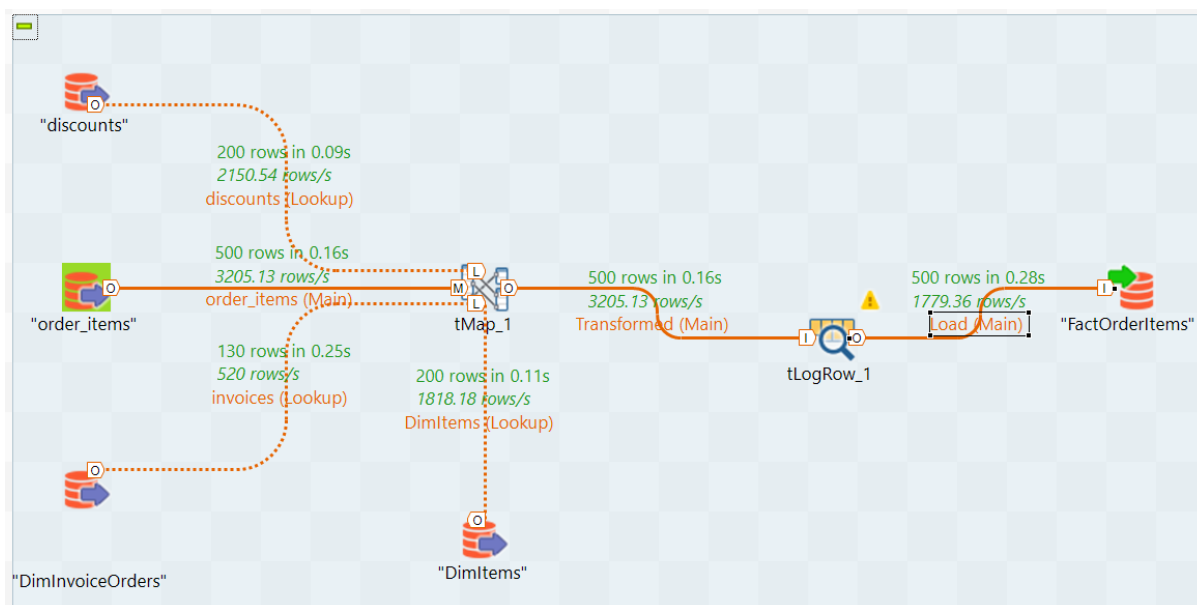
Slika 17 - ETL za DimEmployee



Slika 18 - ETL za FactInvoice



Slika 19 - ETL za FactItemsQuantity



Slika 20 - ETL za FactOrderItems

6. Kreiranje kocke

OLAP je pristup analizi i izvještavanju koji omogućuje korisniku da lako i selektivno izdvaja i pregledava podatke s različitih stajališta temeljeno na multidimenzijskoj strukturi podataka zvanj kocka (engl. cube). Kocka tipično odgovara jednom zvjezdastom spoju i tipično se unutar OLAP sistema definira uzevši postojeći zvjezdasti spoj kao izvor podataka.



Slika 21 – primjer kocke

OLAP kocka predstavlja multidimenzionalni niz podataka, odnosno skupinu ćelija podataka raspoređenih po dimenzijama. Ako govorimo o OLAP kocki podrazumijeva se da

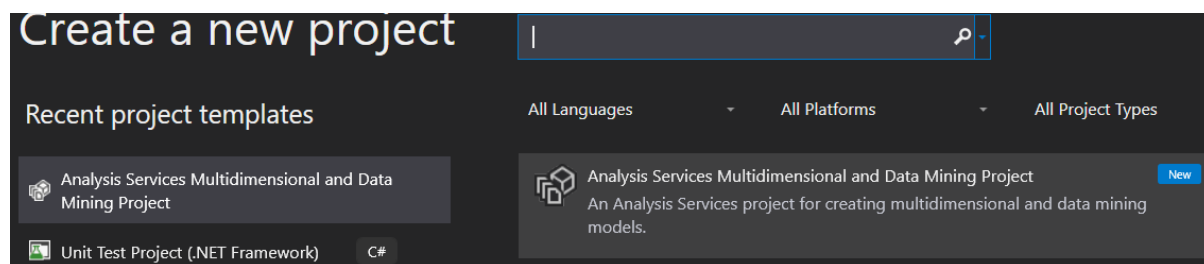
postoje tri dimenzije kojima se definiraju podaci. Multidimenzionalni niz podataka može imati i više od tri dimenzije, ali takav niz je teško predočiti i naziva se hiper-kocka.

Operacije koje se mogu izvoditi nad kockom su sljedeće:

- Roll-upp / Drill up – operacija koja se sastoji od agregiranja odnosno sakupljanja podataka u kocki čime ovu operaciju možemo nazvati agregiranjem.
- Roll-down / Drill down – operacija koja omogućuje korisniku analizu podataka prema većoj razini detalja u hijerarhiji koncepata.
- Slice and Dice – operacije koje pokreću korisnici, a uključuje interaktivnu navigaciju kroz pogled stranice, kroz specifikaciju kriški (slices) 5 pomoću rotacija i agregiranja i detaljiziranja. Postoji mogućnost horizontalnog i vertikalnog analiziranja podataka.
- Pivot – operacija rotacije koja omogućava rotiranje osi, mijenjajući određene dimenzije kako bi se stekao drugi pogled na multidimenzionalnu kocku.

Prije nego što se krene sa obrazloženjem kako se definiranju dimenzije u alatu, autor će objasniti postupak kreiranja novog projekta te definiranje izvora podataka koje će se koristiti za implementaciju OLAP kocke. Svi koraci detaljno su objašnjeni na stranicama Microsofta u Analysis Services Tutorials biblioteci pod cjelinom „Multidimensional Modeling (Adventure Works Tutorial)“.

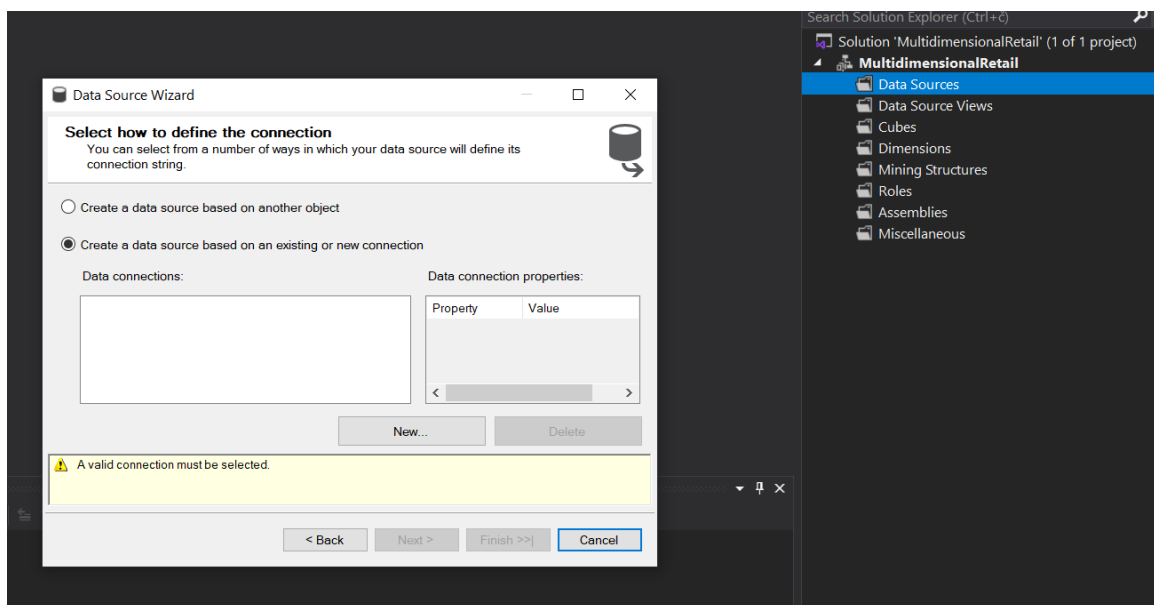
Pokrećemo SQL Server Data Tools koji će pokrenuti radno okruženje Microsoft Visual Studio, kreiramo novi projekt, ali biramo Business Intelligence instalirani predložak i Analysis Services Multidimensional Data Mining Project.



Slika 22 – Kreiranje novog projekta

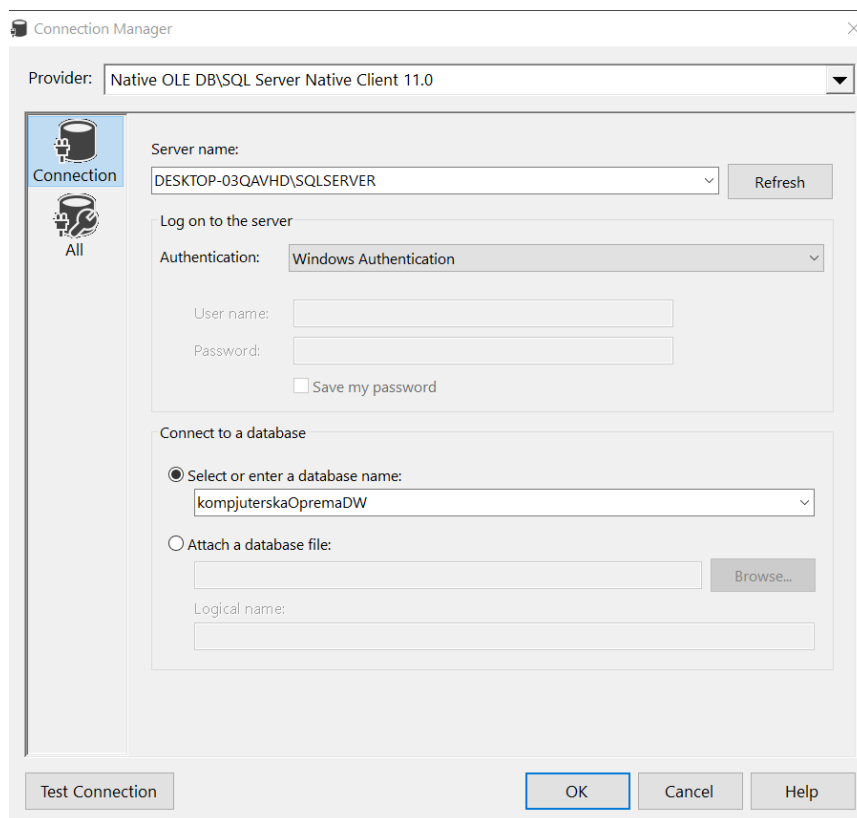
Nakon toga krećemo sa implementacijom naše OLAP kocke. Prvi korak do toga jeste definiranje izvora podataka. Način na koji ćemo to napraviti je:

1. u Solution Explorer-u kliknemo desnim klikom miša na Data Sources te odaberemo opciju New Data Source te nam se pojavljuje wizard za odabiranje izvora podataka (Data Source Wizard)



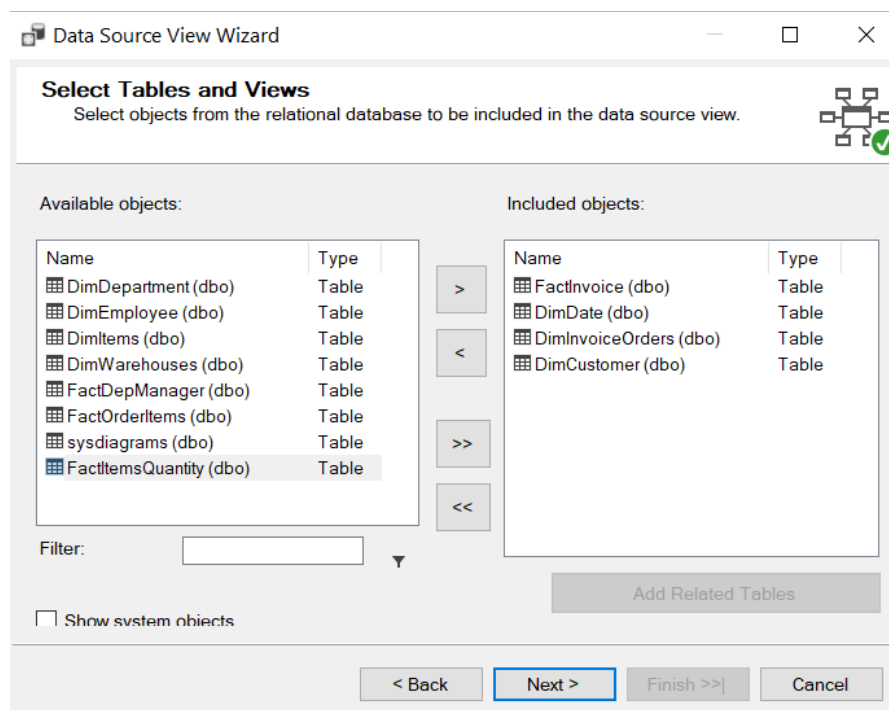
Slika 23 – Definisanje izvora podataka

2. Potom odabiremo novi izvor podataka te definiramo izvor podataka odnosno bazu podataka ovisno o tome gdje nam se nalazi. Primjer takvih postavki se može vidjeti na slici 24.



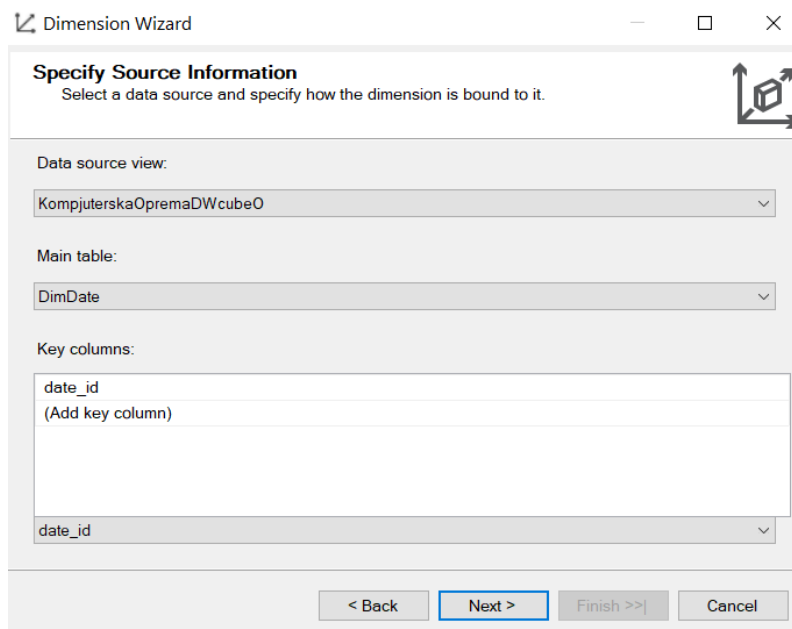
Slika 24 – Spajanje na bazu

3. Nakon što se kreira odgovarajući izvor podataka preostaje nam kreirati pogled na izvor podataka odnosno Data Source View , to radimo desnim klikom miša na Dana Source Views (isto tako u Solution Explorer-u) te odabirom mogućnosti New Dana Source View. Na slici 25 možete vidjeti proces odabira tabela.



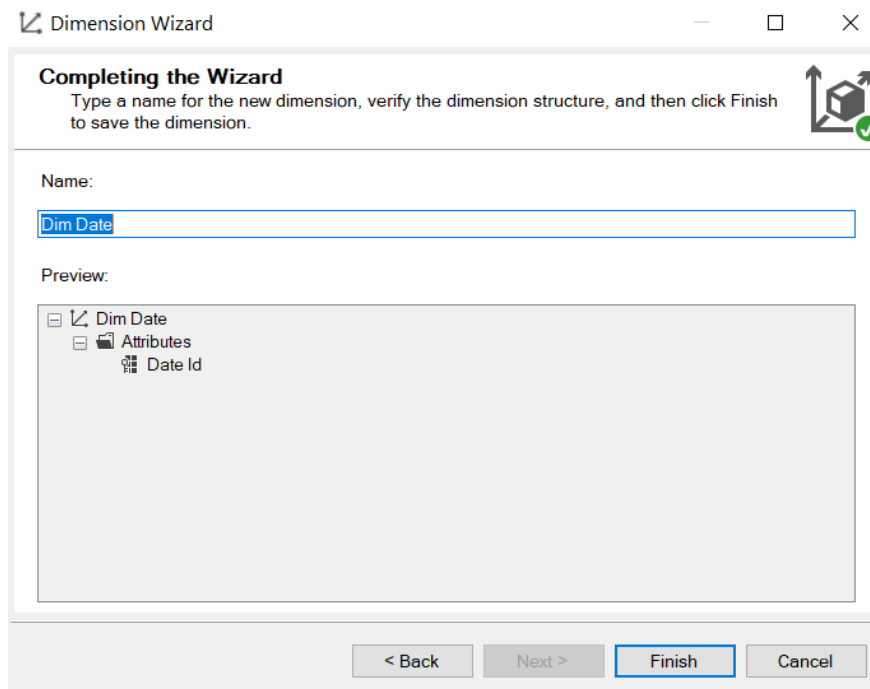
Slika 25 – Odabir tabela

4. Preostaje nam definiranje dimenzija naš kocke, a to radimo desnim klikom na Dimensions te odabirom opcije New dimension, čime nam se pojavljuje čarobnjak za kreiranje nove dimenzije. Potom odabiremo izvor podataka koji smo definirali ranije.



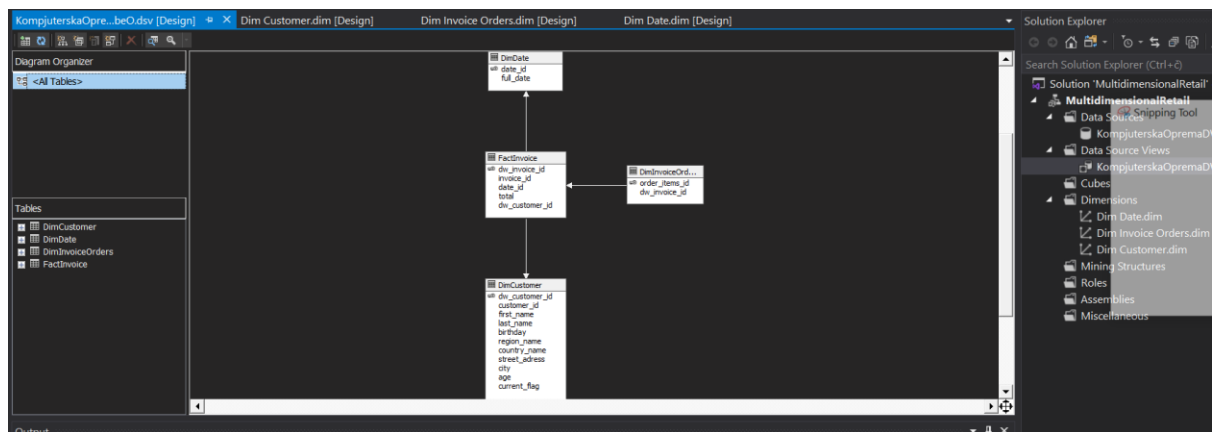
Slika 26 – Odabir dimenzijske tabele

Preostaje nam još samo nazvati dimenziju što možemo vidjeti na slici 27.



Slika 27 – Odabir naziva dimenzije

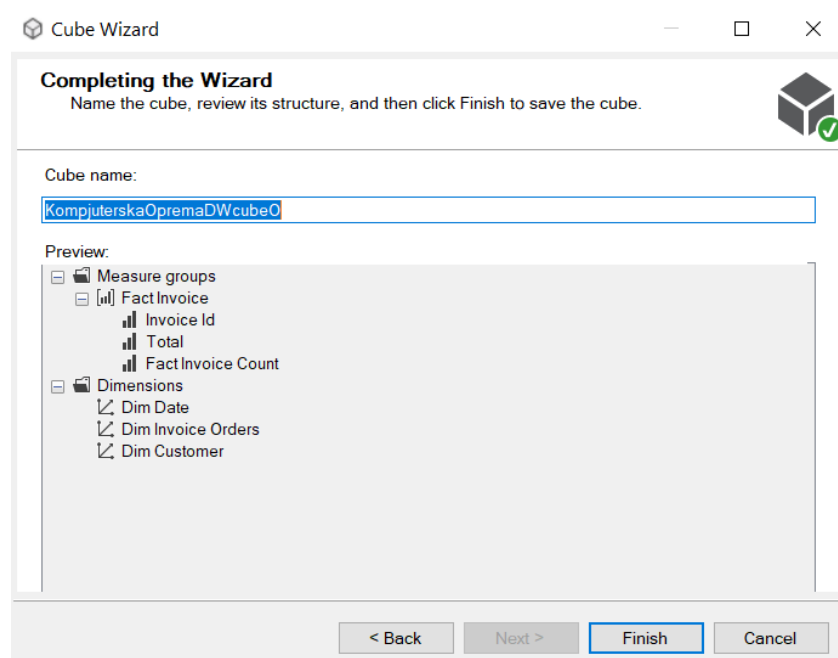
Ukoliko ove korake ponovimo za sve tablice koje su predviđene da budu dimenzijske tablice. U Solution Explorer-u moći ćemo vidjeti prikaz svih dimenzija. Prikaz svih kreiranih dimenzijskih tablica možemo vidjeti na slici 28.



Slika 28 – Prikaz Solution Explorera

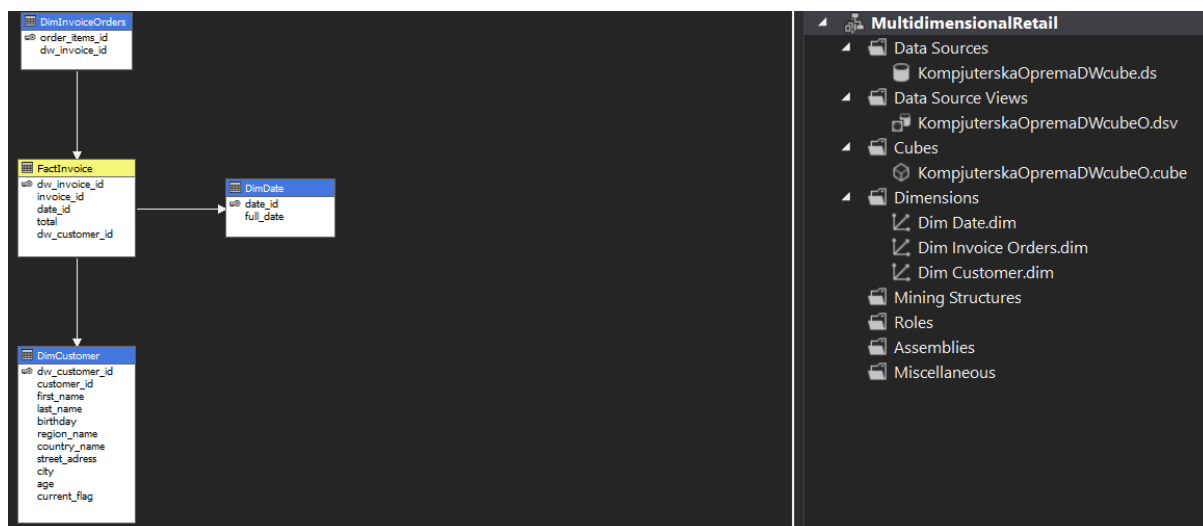
- Kreiranje OLAP kocke se izvodi desnim klikom na Cubes te odabirom New Cube.

Dimenzije će sam prepoznati jer smo ih kreirali prije samog kreiranja kocke, a kako izgleda zadnji korak implementacije kocke možemo vidjeti na slici 29, koja nam prikazuje mjerljive vrijednosti i dimenzije novo kreirane OLAP kocke.



Slika 29 – OLAP kocka

OLAP kocka je kreirana, a pogled na kocku i primjer zvjezdaste sheme smo prikazali na slici 30.



Slika 30 – OLAP kocka – pokazatelj maloprodaje

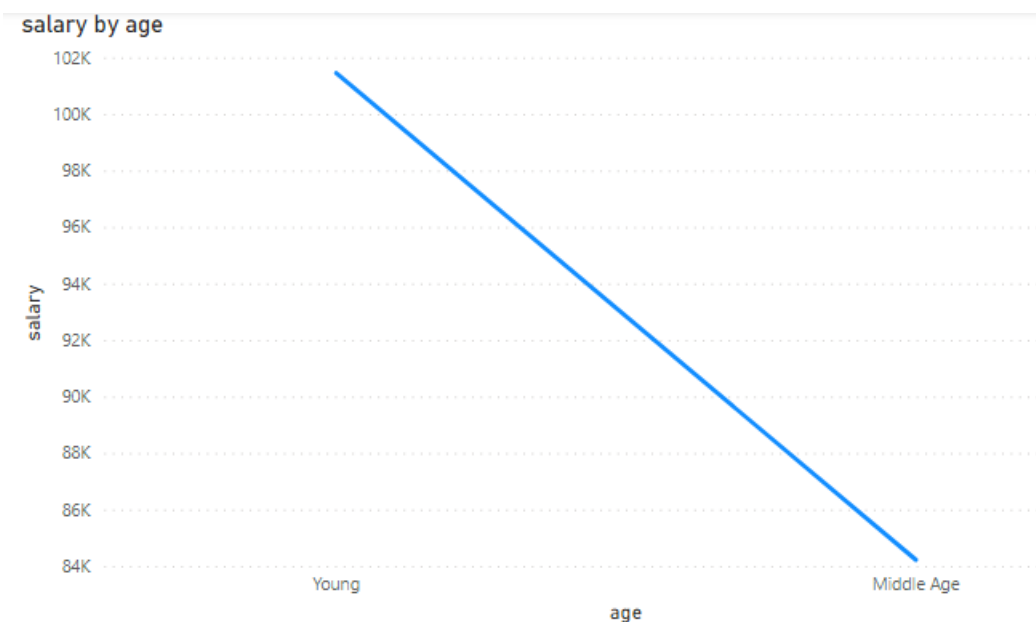
7. Implementacija izvještaja

Prvi i prilično standardan zahtjev za korištenje informacija je kreiranje statičnih izvještaja i ad-hoc upita na podatke u.

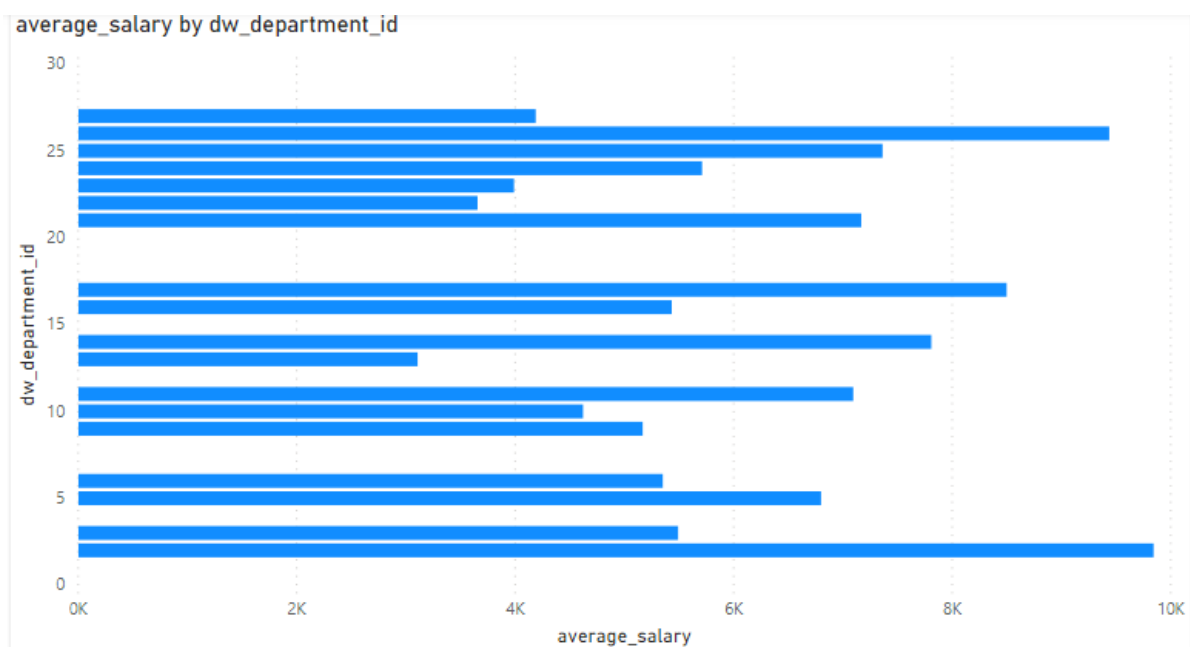
Često se izvještavanje dovodi u vezu sa „Business Intelligence“. To je termin koji se kod nas često prevodi kao "poslovna inteligencija" - što je zbunjujuće, jer najčešće ne uključuje nikakvu vještačku inteligenciju. Umjesto toga preferirani prijevod je "poslovno izvještavanje". Ovaj termin obuhvata prikupljanje i obradu podataka vezanih za poslovanje firme, njihovo prezentovanje i vizualizaciju sa ciljem donošenja bitnih poslovnih odluka.

Često korišteni pristupi su: Data mining, Data warehousing, OLAP. Naglasak je na vizualizaciji podataka i interaktivnosti.

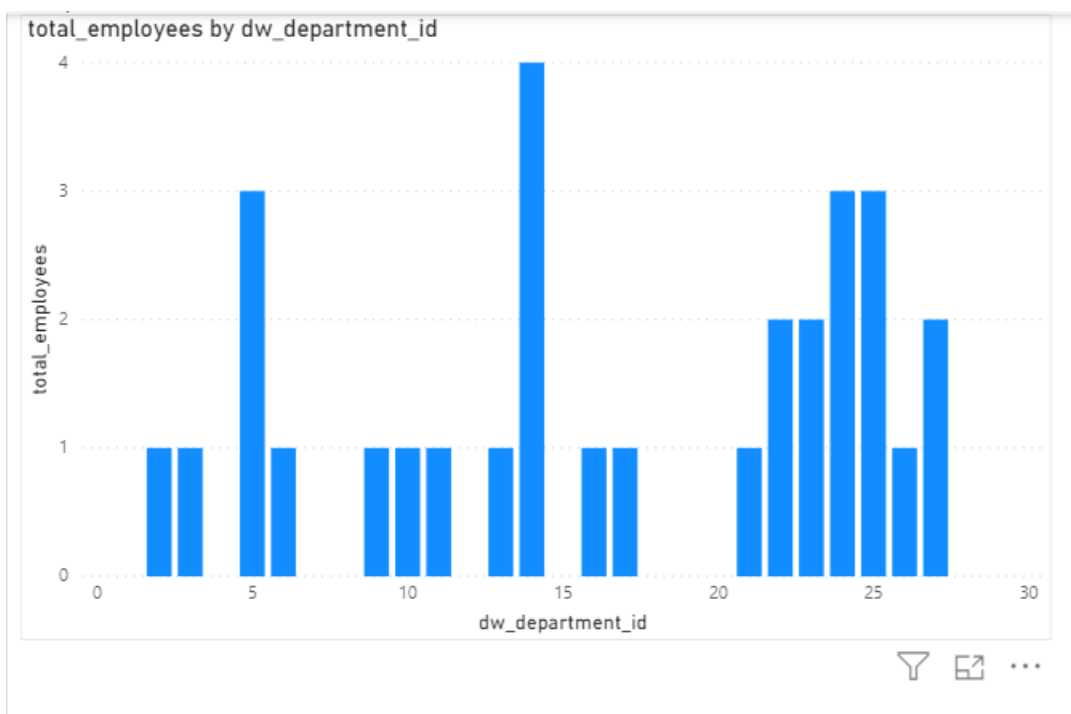
Za kreiranje izvještaja koristili smo alat Power BI. To je alat koji nam omogućava veoma lahko i efikasno prikazivanje raznih tipova izvještaja kao i vizualizaciju istih. Jednostavno prilikom otvaranja alata potrebno je najprije odabrati podatke nad kojima ćemo kreirati izvještaje. U našem slučaju biramo opciju SQL Server, usnosimo naziv servera i naravno tabele koje nas zanimaju. Dalje jednostavno odaberemo kolone te tip dijagrama za prikaz podataka, te Power BI kreira izvještaj.



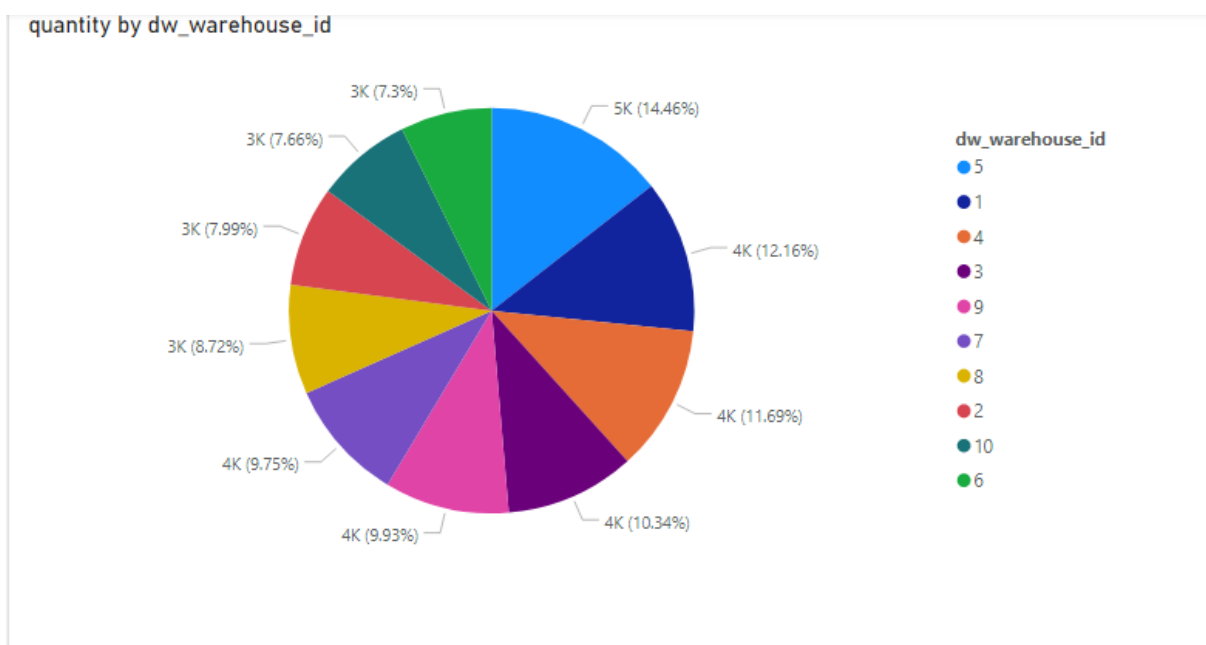
Slika 31 – plata po unaprijed izračunatoj dobi uposlenika



Slika 32 - prosječna plata po odjelu

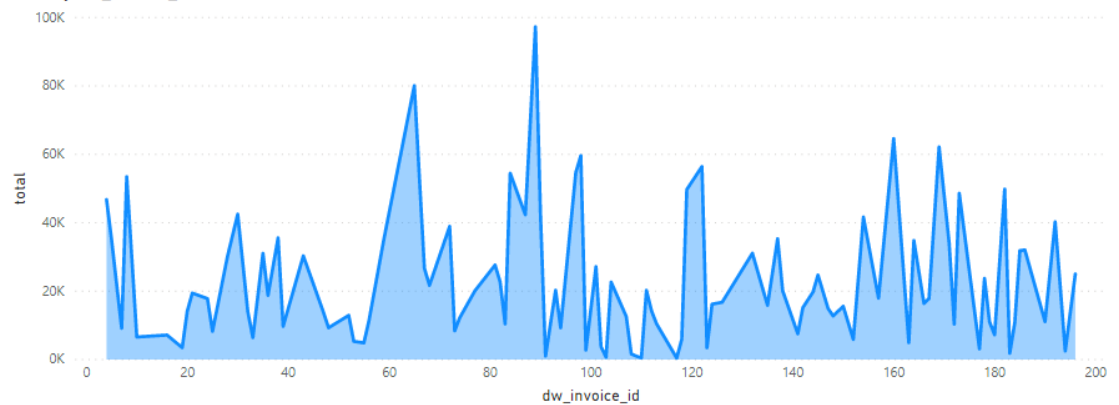


Slika 33 - ukupni broj uposlenika po odjelu



Slika 9 - ukupna količina artikala po skladištu

total by dw_invoice_id

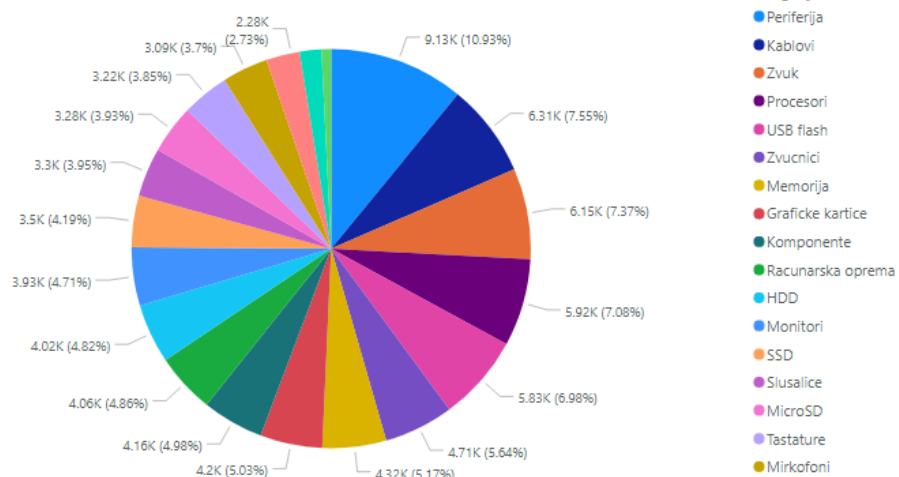


Slika 35 - unaprijed izračunata vrijednost fakture po pojedinačnoj fakturi

item_name	category	manufacturer_city	manufacturer_country_name
Mikrofon	Graficke kartice	Beijing	China
SSD	Graficke kartice	Beijing	China
Camera	Graficke kartice	Bern	Switzerland
Punjac	Graficke kartice	Bombay	India
Samsung s8	Graficke kartice	Hiroshima	Japan
Napojna jedinica	Graficke kartice	London	United Kingdom
Mobitel	Graficke kartice	Roma	Italy
Procesor	Graficke kartice	Sydney	Australia
Camera	Graficke kartice	Toronto	Canada
Zvucna kartica	HDD	Beijing	China
DVD	HDD	Hiroshima	Japan
Testatura	HDD	London	United Kingdom
Bass	HDD	South San Francisco	United States of America
Punjac	HDD	South San Francisco	United States of America
Zvucna kartica	HDD	South San Francisco	United States of America
Mikrofon blue	HDD	Sydney	Australia
Maticna	Kablovi	Beijing	China
Mikrofon	Kablovi	Beijing	China
Mikrofon blue	Kablovi	Beijing	China
Mobitel	Kablovi	Bern	Switzerland
iPhone X	Kablovi	Hiroshima	Japan
SSD	Kablovi	London	United Kingdom
Bass	Kablovi	Oxford	United Kingdom
RAM memorija	Kablovi	South San Francisco	United States of America
Ventilator	Kablovi	South San Francisco	United States of America
Procesor	Kablovi	Sydney	Australia
Samsung s8	Kablovi	Toronto	Canada
Graficka kartica	Kablovi	Venice	Italy
SSD	Kablovi	Whitehorse	Canada
DVD	Komponente	Beijing	China
Graficka kartica	Komponente	Beijing	China
Nokia 7 plus	Komponente	Beijing	China
Printer	Komponente	Beijing	China
Graficka kartica	Komponente	Bern	Switzerland
Kabl	Komponente	Hiroshima	Japan
Volan	Komponente	Hiroshima	Japan
Zvucnik	Komponente	Hiroshima	Japan
Mis	Komponente	London	United Kingdom
iPhone 12	Komponente	Oxford	United Kingdom
Zvucna kartica	Komponente	South San Francisco	United States of America
Printer	Komponente	Whitehorse	Canada

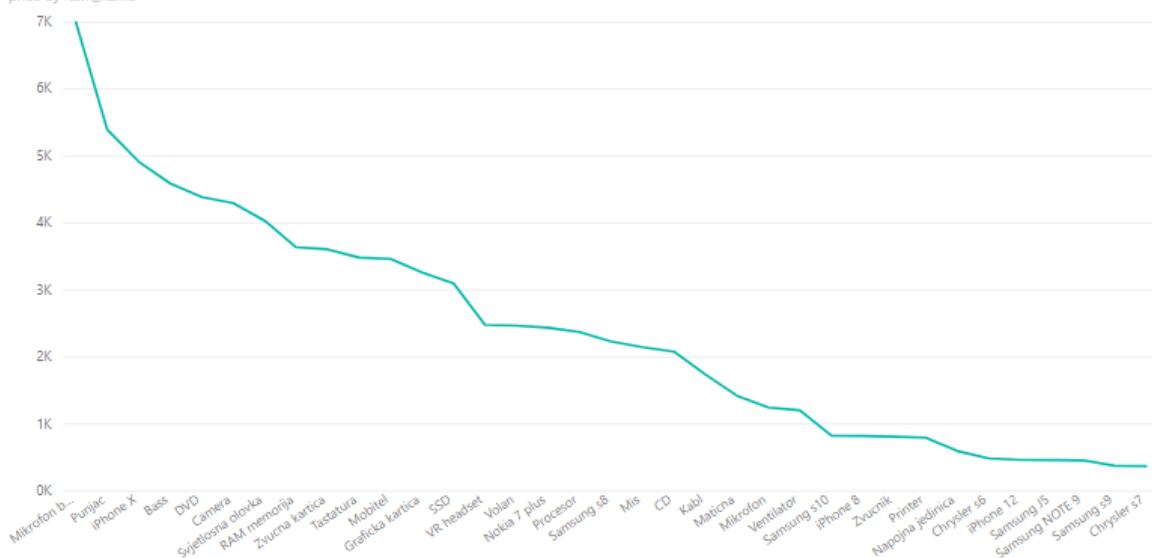
Slika 36 – prikaz artikala, kategorija i sjedišta njihovih proizvođača

price by category



Slika 37 - ukupna vrijednost artikala po kategoriji

price by item_name



Slika 38 - cijena po pojedinačnom artiklu



Slika 39 – prikaz lokacija sjedišta proizvođača na mapi



Slika 40 – prikaz lokacija skladišta na mapi

8. Zaključak

U sve složenijim uvjetima poslovanja vođenje posla „po osjećaju“ (ili instinktu) već odavno nije dovoljno za stabilno poslovanje i rast. Poznavanje i vladanje osnovama statističke analize omogućava menadžeru bolji uvid i razumijevanje stanja i procesa na tržištima na kojima djeluje, kao i u poslovanja vlastitog ili ostalih poduzeća koja djeluju u istoj okolini.

Unatoč poplavi dostupnih podataka, donošenje odluka u poslovanju nije lakše nego prije. Za razumijevanje i komuniciranje onoga što podaci govore potrebna je osnovna statistička pismenost, a često i nešto više od toga. Iz tog razloga, potrebno je omogućiti što bolji, brži i efikasniji sistem upravljanja skladištem podataka kao i sam DWH, kako bi menadžeri i svi koji učestvuju u poslovnom odlučivanju bili što produktivniji. Sve ovo skupa može dovesti do boljeg poslovanja cijele kompanije, održanja kvalitete proizvoda i zadovoljstva klijenata što i jeste glavni i osnovni cilj.

9. Reference

- 1) Howson C (2008) Successful Business Intelligence. The McGraw-Hill Companies.
- 2) Vercellis C (2009) Business Intelligence: Data Mining and Optimization for Decision Making. John Wiley & Sons Ltd.
- 3) R. Kimball, The Data Warehouse Toolkit, New York: John Wiley, 1996.
- 4) R. Kimball, “A Dimensional Manifesto,” 1997.
- 5) W.H. Inmon (2005) Building the Data Warehouse (fourth edition). Indianapolis: Wiley Publishing
- 6) Microsoft Developer Network.
[http://msdn.microsoft.com/enus/library/ms170208\(v=sql.110\).aspx](http://msdn.microsoft.com/enus/library/ms170208(v=sql.110).aspx)

Korisni linkovi:

- <https://www.slideshare.net/siddharthchaudhary39/data-warehouse-project-on-retail-store-86821154>
- <https://www.sqlshack.com/olap-cubes-in-sql-server>