

Seminarski rad iz predmeta Skladište podataka

Skladište podataka prodavnice kompjuterske opeme

Studenti: Profesor:

Emir Feratović Doc. Dr. Damir Omerašević

Ensar Kapur

Harun Ćemalović

Emir Skalonjić

Meho Aliefendić

Sarajevo Decembar 2020

Sadržaj

[1.Uvod 3](#_Toc60241140)

[2. UML reperezntacija relacijske baze podataka 4](#_Toc60241141)

[2.1 ER dijagram 4](#_Toc60241142)

[2.2. Use case dijagram 5](#_Toc60241143)

[2.3. Dijagram aktivnosti 6](#_Toc60241144)

[2.5 Dijagram klasa 7](#_Toc60241145)

[3. Kreiranje baze podataka u MS SQL-u 8](#_Toc60241146)

[4. Implementacija DWH 9](#_Toc60241147)

[4.1. Shema WDH datog projekta 11](#_Toc60241148)

[5.ETL proces 12](#_Toc60241149)

[5.1. Ekstrakcija/izdvajanje 12](#_Toc60241150)

[5.2. Transformacija i Integracija 12](#_Toc60241151)

[6. Kreiranje kocke 14](#_Toc60241152)

[6.1. Kreiranje kocke 15](#_Toc60241153)

[7. Generiranje izvještaj 17](#_Toc60241154)

[8. Zaključak 19](#_Toc60241155)

[9. Reference 20](#_Toc60241156)

# 1.Uvod

 Skladište podataka je baza podataka koja se koristi za pohranu podataka. To je središnje spremište podataka u koje se pohranjuju podaci iz različitih izvora. Skladište podataka se također koriti za izvještavanje i analizu podataka.

Nadalje DWH se koristi kao baza 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.

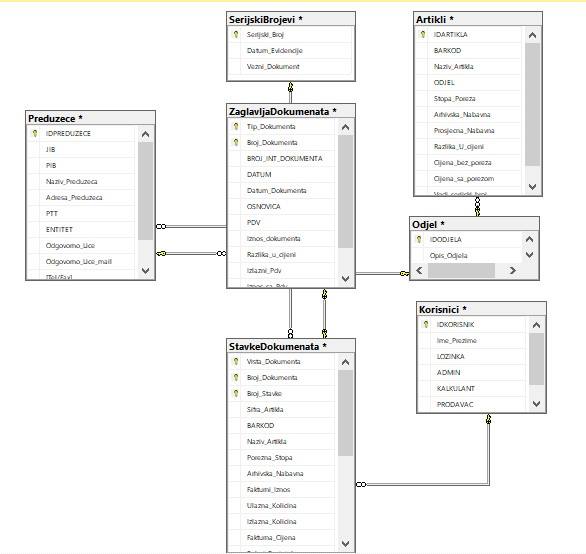
Baza podataka, s druge strane, je skup međusobno povezanih podataka, pohranjenih u vanjskoj memoriji računara. Podaci su istovremeno dostupni raznim korisnicima i aplikacijskim programima. Ubacivanje, promjena, brisanje i čitanje podataka obavlja se posredstvom zajedničkog softvera. Korisnici i aplikacije pritom ne moraju poznavati detalje fizičkog prikaza podataka, već se referenciraju na logičku strukturu baze.

# 2. UML reperezntacija relacijske baze podataka

UML (Unified Modeling Language) je standardizovani grafički jezik za projektovanje, specifikaciju, vizuelizaciju i dokumentovanje softverskih sistema u različitim faza životnog ciklusa. Prvi razlog za njegovu široku upotrebu jeste bogata notacija za modelovanje koja je nezavisna od procesa modelovanja. Drugi razlog proističe iz otvorenosti koncepta, jer UML može da se proširuje, odnosno specijalizuje za određenu oblast.

## 2.1 ER dijagram

Dijagram vezanja ER entiteta je dijagram toka koji ilustrira kako se “entiteti” (ljudi, objekti ili koncepti) međusobno odnose u sistemu. ER dijagram je model koji se najčešće koristi za razvoj ili uklanjanje relacijskih baza podataka u području softvera, poslovnih informacijskih sistema i istraživanja. Ona koristi skup geometrijskih simbola, kao što su pravokutnik, romb, oval i linije, za prikaz odnosa objekata, odnosa i njihovih atributa. ER dijagram projekta je prokazan na slici 2.



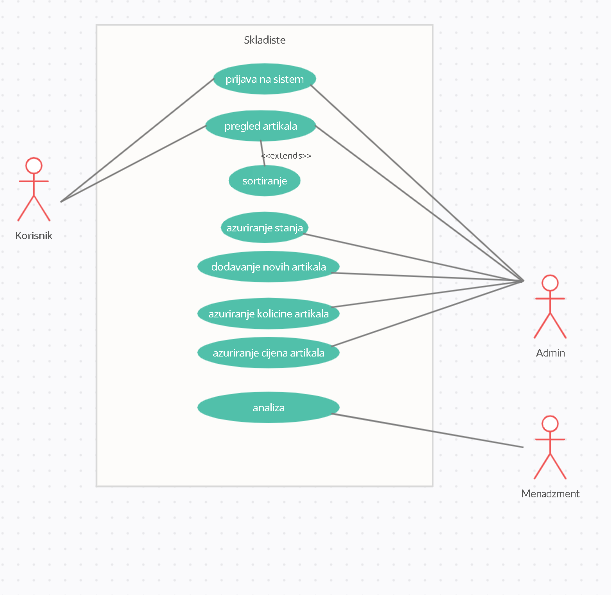
Slika 2.1. ER dijagram projekta

## 2.2. Use case dijagram

Use case dijagram prikazuju samo ono šta aplikacija treba raditi (funkcionalne zahtjeve), zatim otkriva bitne funkcionalne zahtjeve na samom početku projekta a sve to radi ušteda u vremena i novca kasnije. Također se prikazuje ponašanje sistema kako ga vidi korisnik.

Use case dijagram ovog projekta (slika 2.1.) će dati na nekoliko pitanja:

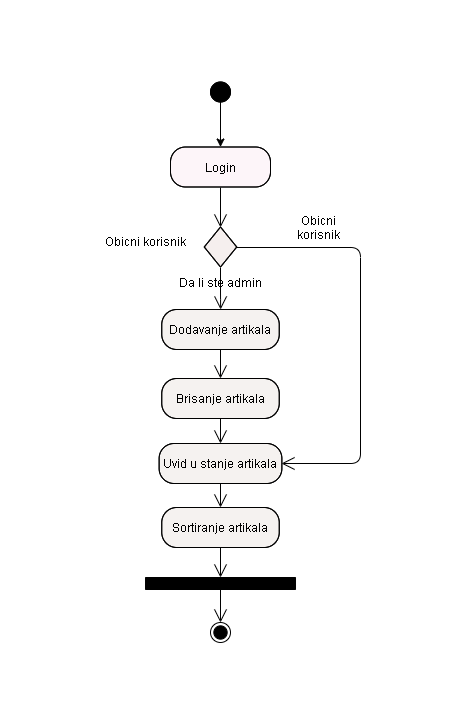
* Što se opisuje? (aplikacija)
* Ko ima interakciju sa sustavom? (sudionici)
* Šta korisnici mogu činiti? (slučajevi korištenja)



Slika 2.2. Use case dijagram datog projekta

## 2.3. Dijagram aktivnosti

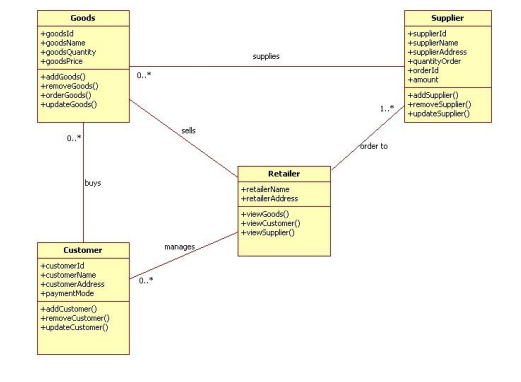
Dijagram aktivnosti je UML dijagram koji spada u grupu dijagrama posašanja, zajedno sa dijagramom stanja, dijagramom slučajeva upotrebe i dijagramima interakcije. On predstavlja tok u sistemu iz jedne aktivnosti u drugu. Slično kao ostali dijagrami ponašanja, dijagram aktivnosti se koristi za reprezentovanje dinamičkog ponašanja sistema. Njime se predstavljaju poslovni procesi višeg nivoa, tokovi podataka i složeni logički elementi sistema. Dijagram aktivnosti (eng. activity diagram) dozvoljava nam da specificiramo kako sistem radi. Dijagram aktivnosti ovog projekta je prikazan na slici 2.3.



Slika 2.3. Dijagram aktivnosti datog projekta

## 2.5 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 takoĎer navodi vidljivost, naziv : tip povratne vrijednosti, te parametri (naziv : tip). Posebne metode su <<create>> (konstruktor) i <<destroy>> (destruktor).



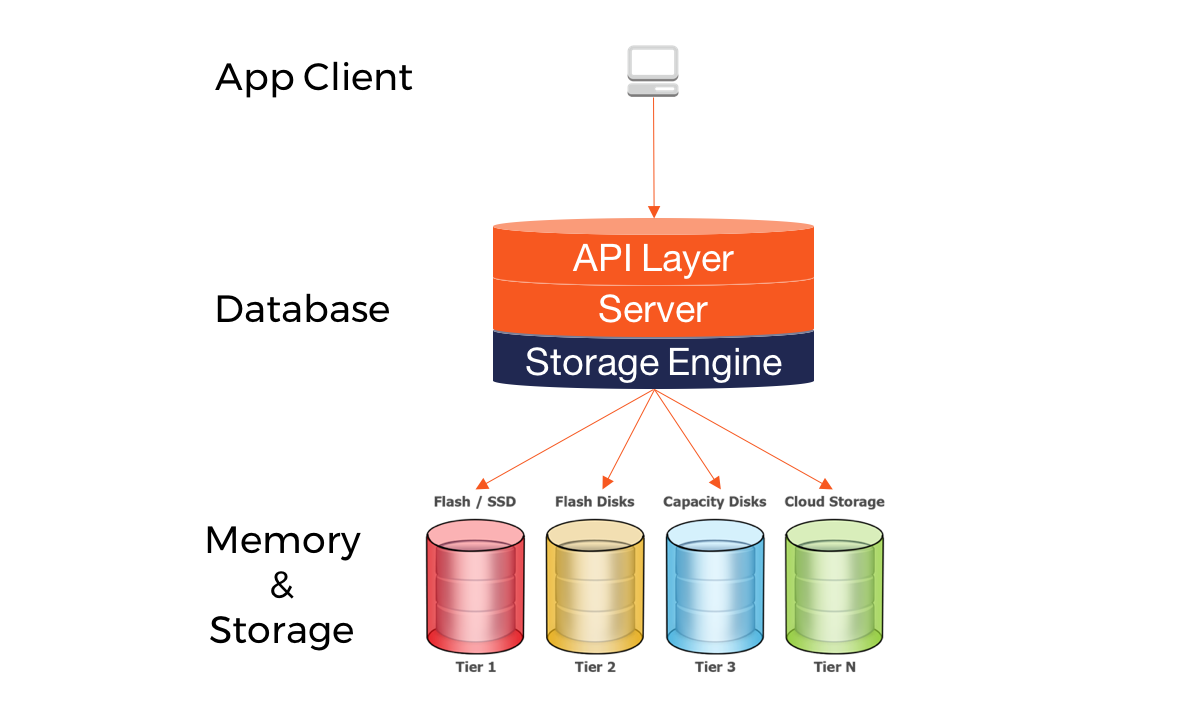
Slika 2.4 Dijagram klasa

# 3. Kreiranje baze podataka u MS SQL-u

Baza podatak je kreirana na osnovu prethodno nacrtanog ER dijagrama (Slika 2.1.) dok se sav kod potreban za kreiranje relacijske baze podataka nalazi u priloženom dokumentu. Sve što nudi baza na, kao i funkcionalnosti ovog sistema se mogu zaključiti detalnom analizom ERD-a. 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 3.1. - OLAP sistem

# 4. Implementacija DWH

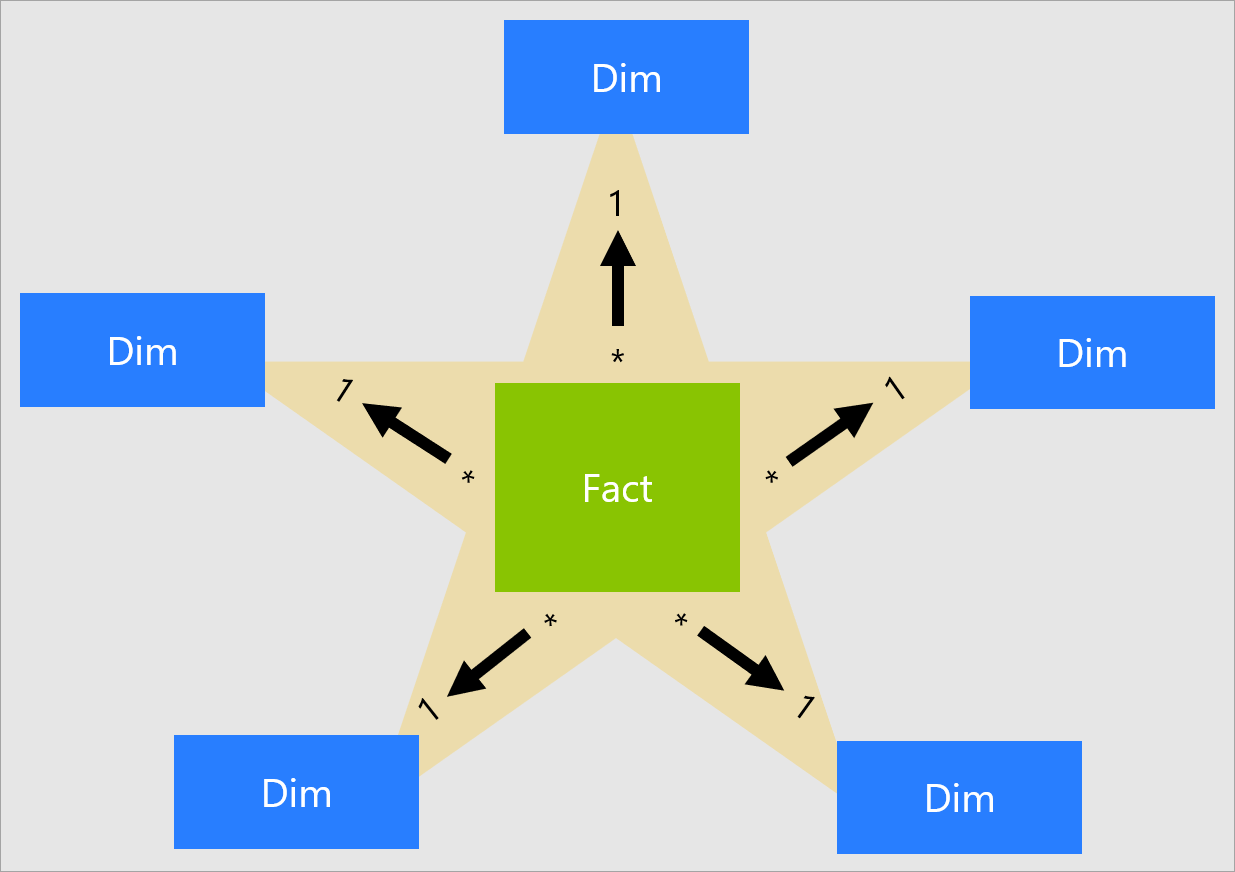
Implementacija skladišta podataka se sastoji od slijedećih koraka:

* 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 pojednostavljuju) što osigurava razumijevanje pojedinih podataka.
* Osigurava se učitavanje novih podataka u dogovorenim/potrebnim vremenskim intervalima.
* Izgradnja skladišta je gotova, ali je skladiše 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. Denoramalizacija 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 22 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:

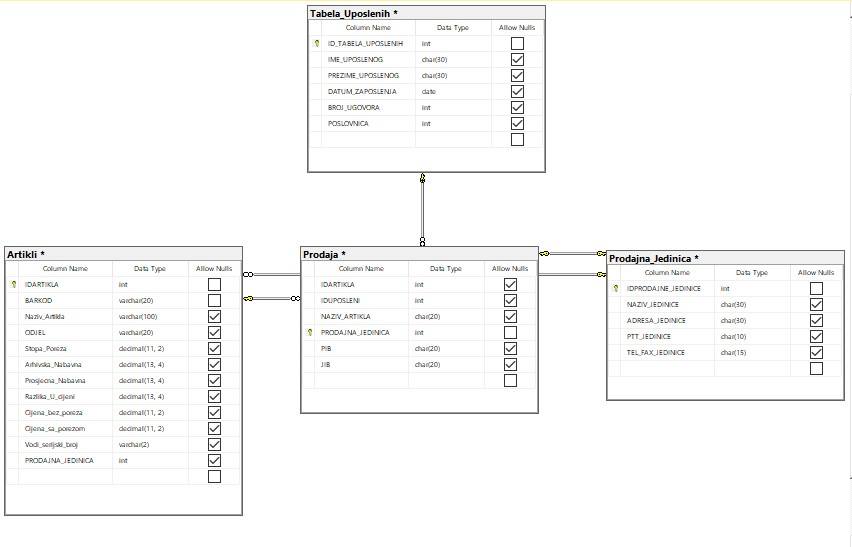
* Ključevi dimenzijskih tablica
* Mjere (eng. measures)



Slika 4.1. Zvjezdasta šema

## 4.1. Shema WDH datog projekta

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 4.1.).



Slika 4.2 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).

## 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 podrzumijevaju da će se samo promjene učitavati nakon prvog/inicijalnog učitavanja.

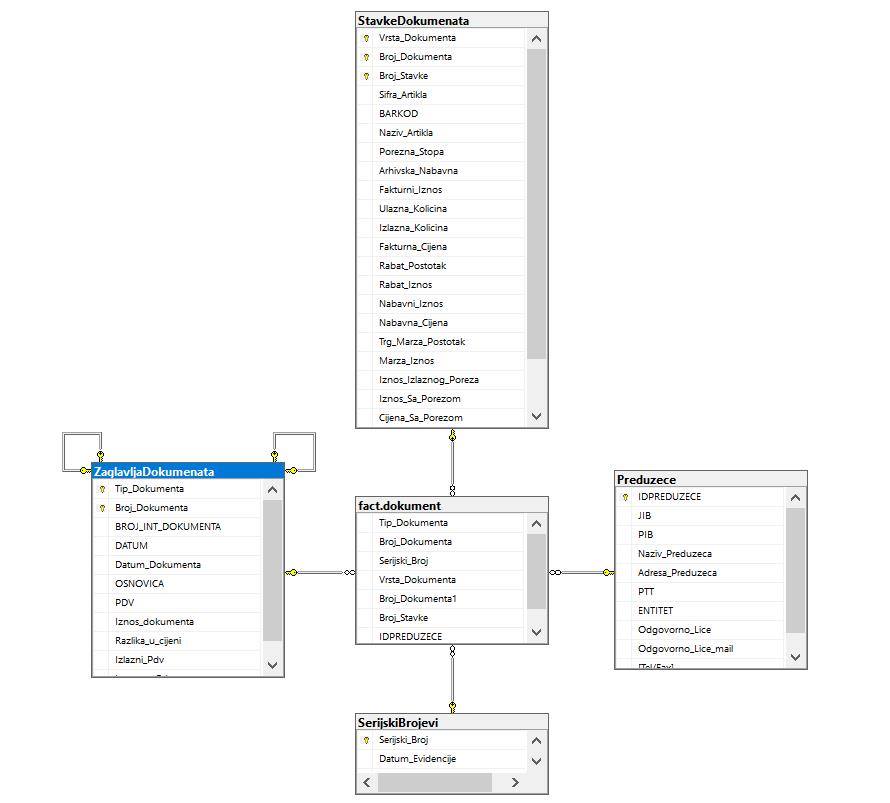
## 5.2. Transformacija i Integracija

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

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

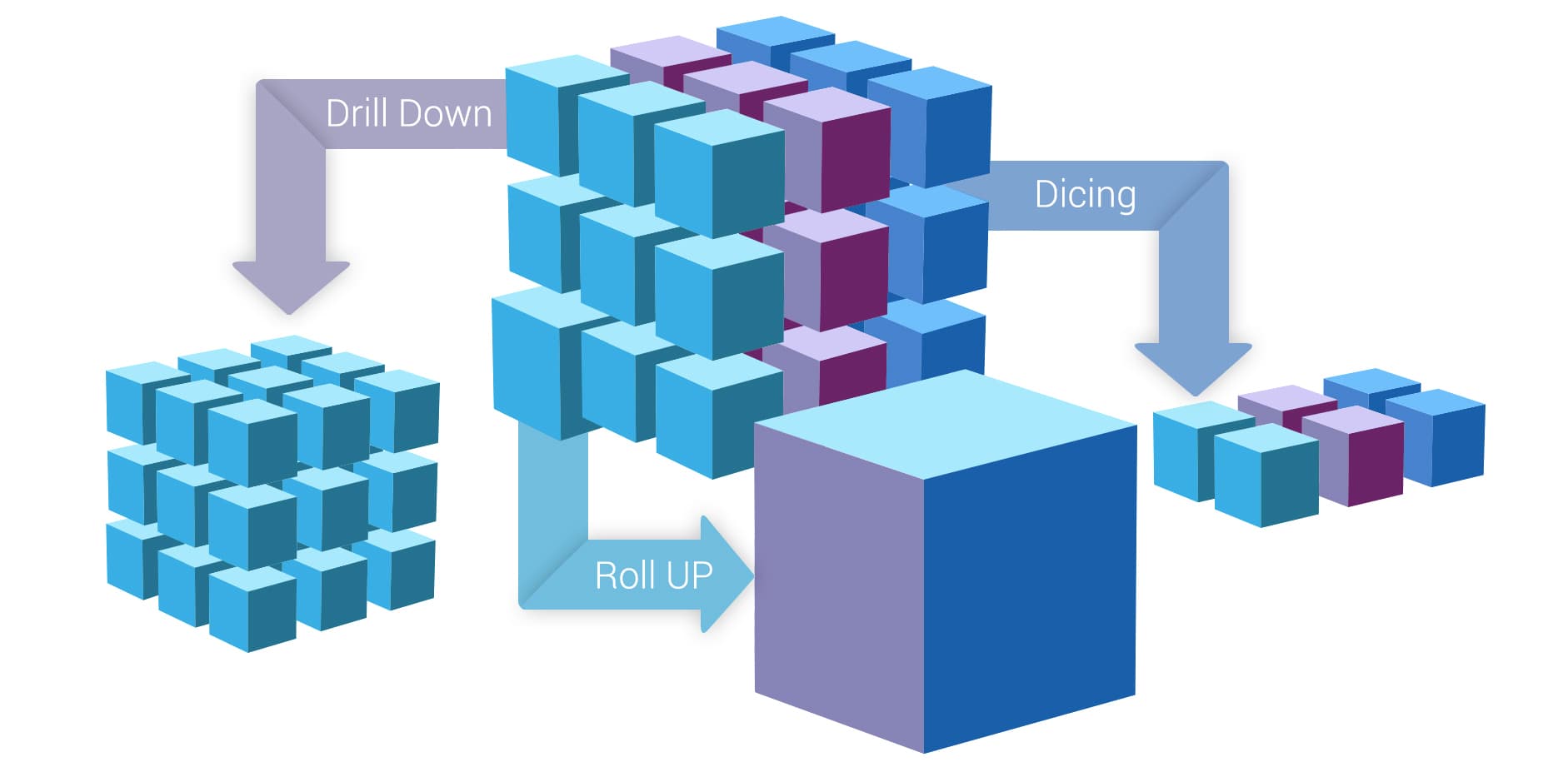
* Proces inicijalnog učitavanja
* Naredna učitavanja (učitavanja tokom vremena)



Slika 4.3 Dijagram koji se dobije nakon primjene svih etapa ETL procesa dokumentacije artikla

# 6. Kreiranje kocke

Struktura podataka koja objedinuje mjere prema razinama i hijerarhijama svake od dimenzija koje se analiziriraju. Kocke kombiniraju nekoliko dimenzija, kao što su vrijeme, zemljopis i crte proizvoda, s sažetim podacima, kao što su brojke o prodaji i zalihama. Kockice (slika 6.1.) nisu "kocke" u strogo matematičkom smislu jer ne moraju nužno imati jednake strane.



Slika 6.1 OLAP kocka

Kocka (eng. cube) omoguićava korisniku analizu podataka kao i izvještavanje, tako da selektivno izdvajanja i pregledava s različitih stajališta temeljeno na multidimenzionalnoj strukturi . Kocka tipično odgovara jednom zvjezdastom spoju i tipično se unutar OLAP sistema definira uzevši postojeći zvjezdasti spoj kao izvor podataka. 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
* Roll-down / Drill down
* Slice and Dice
* Pivot

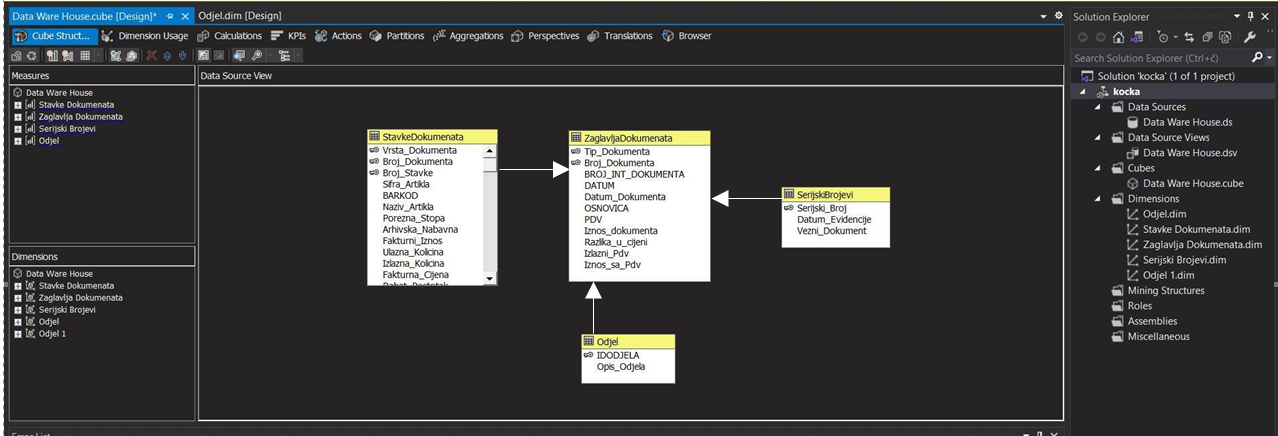
## 6.1. Kreiranje kocke

Prije implementacije OLAP kocke na sistemu koji razvijamo, potrebno je pokrenuti SQL Server Data Tools,a koji će pokrenuti radno okružje Microsoft Visual Studio.

U nastavku biće kratko objašnjen način kreiranja kocke. Pri kreiranju projekta potrebno je odabrati karticu Business Intelligence. Za pojavljivanje i funkcionisanje Business Intelligence potrebno je instalirati predložak i Analysis Services Multidimensional Dana Mining Project.

Zatim se vrši kreiranje Date Source uz pomoć Data Source Wizard, potom odabiremo novi izvor podataka te definiramo izvor podataka odnosno bazu podataka ovisno o tome gdje nam se nalazi. Na ovakav način se kreira odgovarajući izvor podataka. Sljedeći korak je kreirati pogled na izvor podataka odnosno Data Source View . Klikom na „next“ u kartici Data Source View otvara se kartica Dimension Wisard koja služi za definiranje dimenzija naš kocke. Nakon odabira dumenzije u u kartici New dimension, čime nam se pojavljuje čarobnjak za kreiranje nove dimenzije, gdje moramo odabrati izvor podataka koji smo definirali ranije.

Nakon odabira imena dimenzije, klikom na dugme „Finish“ otvara nam se prozor kao na slici 6.2.



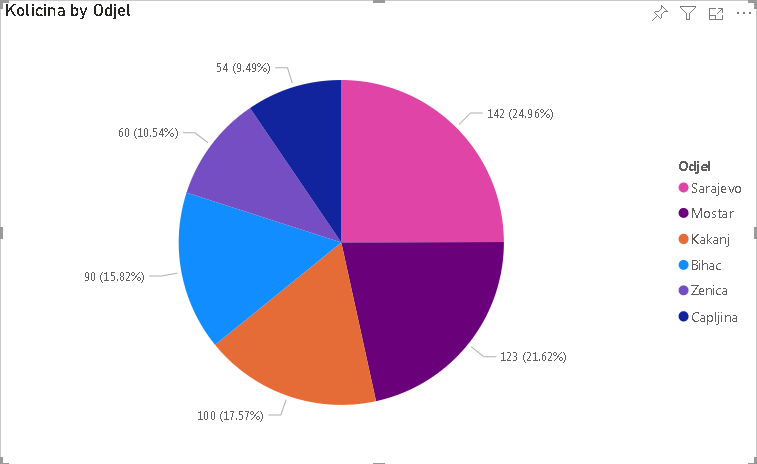
Slika 6.2 O LAP kocka Prikaz dokumentacije artikla

# 7. Generiranje izvještaj

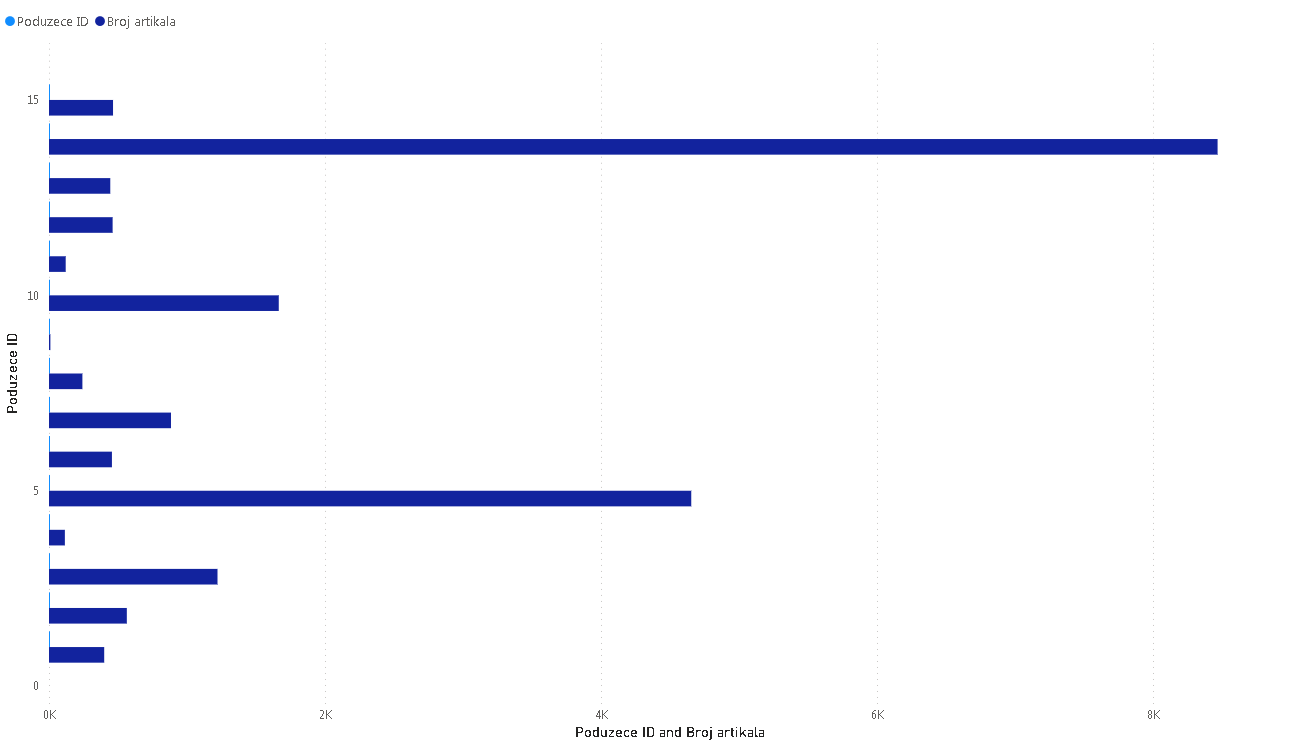
Generiranje izvještaja obuhvata prikupljanje i obradu podataka vezanih za poslovanje firme, njihovo prezentovanje i vizualizaciju sa ciljem donošenja bitnih poslovnih odluka. Često korišteni pristupi su: Data mining, Data warehousing, OLAP.

Naglasak je na vizualizaciji podataka i interaktivnosti. Za kreiranje izvještaja najčesće se koristi alat Power BI. Power BI 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.

Da bi prezentovali podatke iz SQL-a moram izabrati 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. Neki od dijagrama biće prikazan u nastavku.



7.1 Ukupna količina artikla po gradovima



7.2 Ukupna količina proizvoda po odjelu u jednom gradu

# 8. Zaključak

OLAP je jedna od komponenti poslovne inteligencije, a označava mrežnu analitičku obradu podataka. Kroz ovaj seminarski rad pokušano prikazati kako se kreiraju ER dijagram, Use case dijagram, Dijagram klasa, Dijagram aktivnosti, kao i kako se provodi analiza određene količine podataka, a za to su se koristili sljedeći alati i dodaci:

* Microsoft SQL Server 2020:
* Database Engine
* Analysis Services
* SQL Server Dana Tools for Visual studio 2020
* Power BI

Kao što se moglo vidjeti kroz sistem koji su predstavljeni u radu, postoje različite mogućnosti analize i obrade podataka, a gore navedeni alati nam uveliko pomažu u tome.

# 9. Reference

1. Vercellis C (2009) Business Intelligence: Dana Mining and Optimization for Decision Making. John Wiley & Sons Ltd.
2. R. Kimball, The Data Warehouse Toolkit, New York: John Wiley, 1996.
3. W.H. Inmon (2005) Building the Dana Warehouse (fourth edition). Indianapolis: Wiley Publishing
4. Microsoft Developer Network.
5. https://www.slideshare.net/siddharthchaudhary39/data-warehouse-project-on-retail-store86821154