



Univerzitet u Zenici

Politehnički fakultet

Odsjek: Softversko inženjerstvo

TEMA

Integracija poslovne inteligencije unutar e-commerce

Student: Ahmet Durmić, 53

Kolegij: Elektronsko poslovanje

Profesor, asistent: Doc.dr. Esad Kadušić

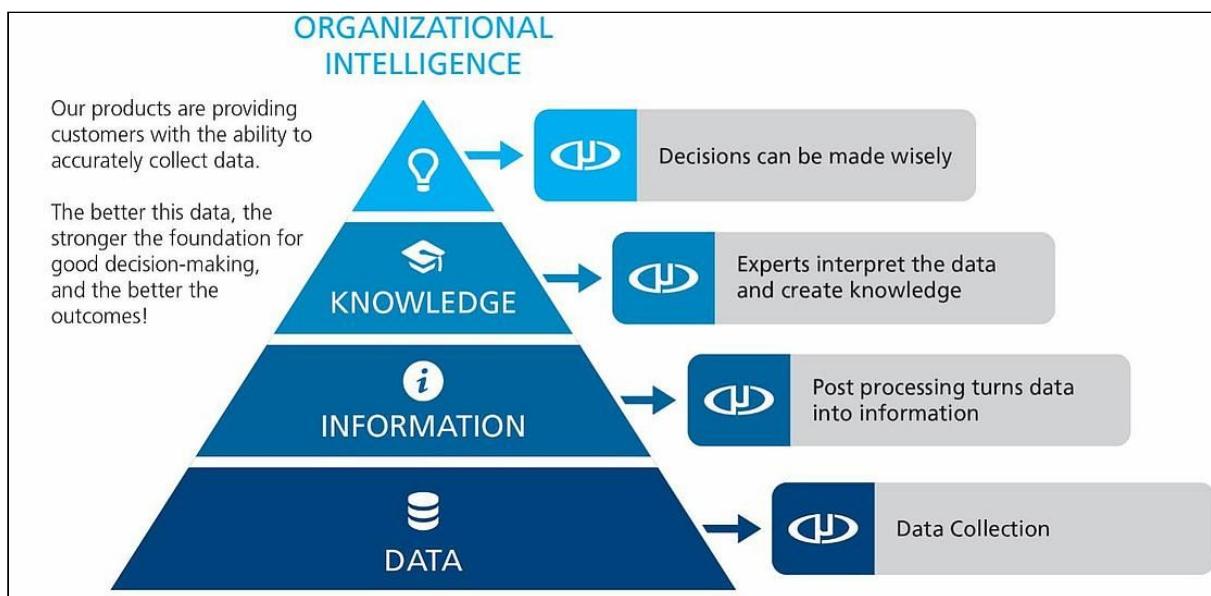
Akadembska godina: 2020/2021

Sadržaj

1. Uvod	2
2. Tehno-ekonomski značaj poslovne inteligencije	3
3. Skladište podataka	5
3.1 Dizajniranje skladišta podataka	8
3.2 Dimenzijske tabele	11
3.3 Dizajniranje sporo promjenljivih dimenzija	14
3.3.1 Sporomijenjajuća dimenzija tipa 1	14
3.3.2 Sporomijenjajuća dimenzija tipa 2	15
3.3.3 Sporomijenjajuća dimenzija tipa 3	15
3.4 Činjenične tabele	15
3.4.1 Mjere	16
4. ETL proces (Extract transform & load)	17
4.1 Planiranje extrakcije podataka (Extract)	18
4.2 Planiranje transformacije podataka (Transform)	19
4.3 Planiranje procesa punjenja (Load)	20
4.4 Platforme za ETL proces	20
5. Kocka podataka (Cube)	21
6. Tabularni model	24
7. Izvještavanje	26
8. Implementacija	27
8.1 Kreiranje skladišta i import podataka	27
8.1.1 Kreiranje i import podataka u dimenzijsku tabelu SalesTerritory	27
8.1.2 Kreiranje i import podataka u dimenzijsku tabelu Product	28
8.1.3 Kreiranje i import podataka u dimenzijsku tabelu Date	29
8.1.5 Kreiranje i import podataka u činjeničnu tabelu SalesFreight	31
8.1.6 ER Dijagram skladišta podataka	32
8.2 Analiza zarade i troškova prodaje	33
8.3 Analiza zarade i troškova po kategoriji prodavnih proizvoda	34
8.4 Prijedlog za poboljšanje poslovanja	36
9. Literatura	38

1. Uvod

Poslovna inteligencija (BI) je tehnološki postupak za analizu podataka i pružanje djelotvornih informacija koje pomažu rukovoditeljima, menadžerima i radnicima u donošenju utemeljenih poslovnih odluka. Kao dio BI procesa, organizacije prikupljaju podatke iz kako internih IT sistema tako i vanjskih izvora, pripremaju ih za analizu, pokreću upite prema podacima i izrađuju vizualizacije podataka, te uz pomoć BI alata kreiraju izvještaje (*engl. reports*) kako bi analitički rezultati postali dostupni poslovnim korisnicima za operativne odluke. Konačni cilj poslovne inteligencije je donošenje boljih poslovnih odluka koje organizacijama omogućuje povećanje prihoda, poboljšanje operativne učinkovitosti i stjecanje konkurenčkih prednosti u odnosu na poslovne rivale. Da bi se postigao taj cilj, BI uključuje kombinaciju analitike, alata za upravljanje podacima i izvještavanja, plus razne metodologije za upravljanje i analizu podataka.



Slika br.1: Proces transformacije podataka u odluke

Postavlja se pitanje, kako ustvari funkcioniра cjelokupni proces poslovne inteligencije. Arhitektura poslovne inteligencije uključuje više od samog softvera. Podaci poslovne inteligencije obično se pohranjuju u skladišta podataka(*engl. Data warehouses*) koja sadrže podskupove poslovnih podataka za pojedine odjele i poslovne jedinice. Podaci mogu uključivati historijske informacije i podatke u stvarnom vremenu prikupljene iz izvornih sistema, omogućujući BI alatima procese donošenja odluka. Prije nego što se koriste u BI aplikacijama, sirovi podaci iz različitih izvornih sistema obično se moraju integrirati, objediniti i očistiti pomoću alata za integraciju podataka i upravljanje kvalitetom podataka kako bi se osiguralo da BI timovi i poslovni korisnici analiziraju tačne i dosljedne informacije. Također, kako je bitno naznačiti da skladište podataka polazna osnova svakog BI rješenja, te u konačnici zadatak BI jeste upravo da se osigura upotreba istog.

Proces poslovne inteligencije možemo podijeliti u nekoliko ključnih koraka a to su:

1. Priprema podataka, u kojoj se skupovi podataka organiziraju i modeliraju za analizu
2. Analitičko ispitivanje pripremljenih podataka
3. Distribucija ključnih pokazatelja uspješnosti (*engl. Key performance indicator*) i drugih informacija koje su od koristi poslovnim korisnicima
4. Korištenje informacija prilikom donošenja poslovnih odluka

2. Tehno-ekonomski značaj poslovne inteligencije

Živimo u dobu tehnološkog napretka te sam digitalni napredak potpuno je revolucionirao našu svakodnevnicu, a svakako da je poslovni svijet osjetio najveći uticaj istog. Kompanije sada imaju pristup alatima i strategijama koji im omogućuju da saznaju o svojim kupcima, poslovnim procesima, prodaji i sebi više nego ikad prije, ali ne koriste ih svi. Možemo reći da je poslovna inteligencija presudna za uspjeh neke organizacije. Prije nego što krenemo u duboko razmatranje procesa sprovodenja poslovne inteligencije, prvo moramo razumjeti šta je to poslovna inteligencija i kako se odnosi na strateške inicijative neke kompanije. Pojam poslovne inteligencije (BI) odnosi se na tehnologije, aplikacije, strategije i prakse koje se koriste za prikupljanje, analizu, integraciju i predstavljanje relevantnih poslovnih informacija. Cjelokupna svrha poslovne inteligencije je podržati i olakšati bolje poslovne odluke. Sada kada smo objasnili šta je to ustvari poslovna inteligencija, postavlja se pitanje: Zašto je poslovna inteligencija toliko važna za moderne organizacije? Glavni razlozi za ulaganje u BI strategije su:

1) Pronalaženje i targetiranje novih kupaca

Jedan od glavnih razloga zbog kojih kompanije ulažu svoje vrijeme, novac i napore u poslovnu inteligenciju jest zato što im daje veću sposobnost promatranja i analize trenutnih trendova kupovine koju obavljaju kupci. Kada se poslovna inteligencija upotrijebi da bi se došlo do podataka šta potrošači kupuju i koji je njihov motiv kupovine, isti se mogu koristiti za stvaranje proizvoda i poboljšanja proizvoda kako bi udovoljili njihovim očekivanjima i potrebama, a sve u cilju poboljšanja samog posovanja i donošenja pravih poslovnih odluka.

2) Prikupljanje ključnih informacija i podataka

Učinkovit sistem poslovne inteligencije služi kao sredstvo za prepoznavanje ključnih organizacijskih poslovnih procesa i trendova. BI sistem također omogućuje razumijevanje implikacija različitih organizacijskih procesa i promjena, omogućujući vam donošenje utemeljenih odluka i djelovanje u skladu s njima. Bez ključnih informacija i podataka niti jedna od prethodno navedenih akcija ne bi bila moguća.

3) Poboljšanje organizacijske učinkovitosti

BI sistemu pomaže u poboljšanju organizacijske učinkovitosti što posljedično povećava produktivnost i potencijalno može povećati prihod. Sistemi poslovne inteligencije omogućuju kompanijama da s lakoćom dijele vitalne informacije među odjelima, štедеći vrijeme na izdvajanju i interpretaciji podataka. Olakšavanje razmjene informacija lakšim i učinkovitijim omogućuje organizacijama uklanjanje

suvišnih uloga i dužnosti, omogućujući zaposlenicima da se fokusiraju na svoj posao, umjesto na obradu podataka.

4) Uvid u prodaju

Prodajni i marketinški timovi prate svoje kupce preko CRM-a (*engl. Customer relationship management*). CRM-ovi su dizajnirani da obrađuju sve interakcije s kupcima. Budući da sadrže sve komunikacije i interakcije s kupcima, postoji mnoštvo podataka i informacija koji se mogu protumačiti i koristiti za strateške inicijative. BI sistemi pomažu organizacijama u svemu, od identificiranja novih kupaca, praćenja i zadržavanja postojećih te pružanja usluga nakon prodaje.

5) Prikaz podataka u stvarnome vremenu (*engl. Real-Time Data*)

BI sistemi pružaju korisnicima pristup podacima u stvarnom vremenu na različite načine, uključujući proračunske tabele, vizualne nadzorne ploče itd. Velike količine podataka mogu se brzo i precizno asimilirati, protumačiti i distribuirati koristeći alate za poslovnu inteligenciju.

6) Konkurentska prednost

Uz sve prethodno navedene stavke, poslovna inteligencija pomaže da se stekne uvid u ono što rade vaši konkurenti, omogućujući vašoj organizaciji donošenje obrazovanih odluka i planiranje budućih poslovnih poteza.

Ukratko, BI omogućuje kombiniranje podataka iz više izvora, analizu informacija u probavljeni format i zatim širenje informacija relevantnim dionicima. To kompanijama omogućuje uvid u širu sliku i donošenje pametnih poslovnih odluka. Uvijek postoje svojstveni rizici kada je riječ o donošenju bilo koje poslovne odluke, ali ti rizici nisu toliko istaknuti ili zabrinjavajući kada se implementira učinkovito i pouzdano BI rješenje. Organizacije poslovne inteligencije mogu se kretati naprijed u klimi koja se sve više temelji na podacima s povjerenjem, znajući da su spremne za svaki izazov koji se pojavi.



Slika br.2: Top 5 pokazatelja, zašto svaka ozbiljna organizacija treba da sprovede proces poslovne inteligencije

3. Skladište podataka

Skladište podataka (*engl. Data warehouse*) je centralizovano mjesto svih poslovnih podataka u cilju izvještavanja i vršenja različitih poslovnih analiza. Skladište podataka polazna je osnova svakog BI sistema. Spoj je tehnologija i komponenata koji pomažu u strateškoj upotrebi podataka. To je baza koja je optimizovana za elektroničko pohranjivanje velike količine podataka koje je dizajnirano za upite i analizu umjesto za obradu transakcija. To je proces pretvaranja podataka u informacije i pravodobnog stavljanja istih na raspolaganje a sve u cilju donošenja ispravnih poslovnih odluka. Skladišta podataka pružaju generalizirane i objedinjene podatke u višedimenzionalnom prikazu. Zajedno s uopćenim i objedinjenim prikazom podataka, skladišta podataka pružaju nam i alate za mrežnu analitičku obradu (OLAP). Ovi nam alati pomažu u interaktivnoj i učinkovitoj analizi podataka u višedimenzionalnom prostoru. Ova analiza rezultira generalizacijom i rudarenjem podataka. Prije nego što uđemo u detalje dimenzijskog modeliranja, potrebno je definisati temeljne ciljeve skladištenja podataka i poslovne inteligencije[1]. Hodajući hodnicima bilo koje organizacije iste teme se ponavljaju već više od tri desetljeća[1]:

- "Prikupljamo tone podataka, ali im ne možemo pristupiti."
- "Poslovni ljudi moraju lako doći do podataka."
- "Samo mi pokaži što je važno."
- „Provodimo čitave sastanke raspravljujući se oko toga ko ima prave brojke umjesto donošenja poslovnih odluka. "
- „Želimo da ljudi koriste informacije kako bi podržali odluku zasnovanu na činjenicama"

Na temelju našeg iskustva, ove su brige još uvijek toliko univerzalne da su iste temeljni zahtjevi prilikom implementacije DW / BI sistema[1].

S obzirom na prethodno referencirane stavke DW/BI sistem mora da zadovolji sljedeće aspekte kako bi cijelokupan proces poslovne inteligencije bio uspješan.

- DW / BI sistem mora učiniti podatke lako dostupnima[1]. Sadržaji DW / BI sistema mora biti razumljiv[1]. Podaci moraju biti intuitivni poslovnom korisniku, a ne samo programeru[1]. Strukture podataka a označe bi trebale oponašati misaoni proces i rječnik poslovnih korisnika[1]. Poslovni korisnici žele odvojiti i kombinirati analitičke podatke u bezbroj kombinacija[1]. Alati i aplikacije za poslovnu inteligenciju kojima pristupaju podaci moraju biti jednostavni i laki za upotrebu[1]. Također moraju vratiti rezultate upita korisniku s minimalnim vremenom čekanja[1].
- DW / BI sistem mora dosljedno prikazivati informacije[1]. Podaci u DW / BI sistemu moraju biti vjerodostojni i dosljedni[1]. Dosljednost također podrazumijeva uobičajene označe i definicije za sadržaj DW / BI sistema koji se koriste u izvorima podataka[1].
- DW / BI sistem mora se prilagoditi promjenama[1]. Potrebe korisnika, uvjeti poslovanja, podaci i tehnologija podložni su promjenama[1]. DW / BI sistem mora biti dizajniran za graciozno rješavanje ove neizbjegne promjene kako ne bi onesposobio postojeće podatke ili aplikacije[1]. Postojeći podaci i aplikacije ne bi smjeli biti promijenjeni ili poremećeni kad poslovna zajednica postavlja nova pitanja ili se novi podaci dodaju u skladište[1].

- DW / BI sistem mora pravovremeno predstaviti informacije[1]. Kako DW / BI sistem se intenzivnije koristi za operativne odluke, sirovi podaci se trebaju pretvoriti u djelotvorne informacije u roku od nekoliko sati, minuta, ili čak sekundi[1].
- DW / BI sistem mora biti siguran[1]. Informativni dragulji organizacije pohranjeni su u skladištima podataka[1]. DW / BI sistem mora učinkovito kontrolirati pristup na povjerljive podatke organizacije[1].
- DW / BI sistem mora poslužiti kao mjerodavan i pouzdan temelj za poboljšanje donošenja odluka[1]. Skladište podataka mora imati tačne podatke koji podržavaju donošenje odluka[1]. Najvažniji rezultati iz DW / BI sistema su odluke koje se donose na temelju analitičkih dokaza[1].
- Poslovna zajednica mora prihvati DW / BI sistem da bi ga ocijenila uspješnim[1]. Nije važno što ste izgradili elegantno rješenje koristeći proizvode najbolje vrste i platforme[1]. Ako poslovna zajednica ne prihvati DW / BI okruženje i aktivno ga ne koristi, pali ste na testu prihvaćanja[1]. Za razliku od implementacije operativnog sistema gdje poslovni korisnici nemaju izbora nego koristiti novi sistem, upotreba DW / BI ponekad nije obavezna[1]. Poslovni će korisnici prihvatiti DW / BI sistem ako je "jednostavan i brz" izvor informacija koje se mogu iskoristiti[1].

Iako je svaki zahtjev s ovog popisa važan, konačna dva su najviše kritična, i nažalost, često najviše previđena[1]. Uspješno skladištenje podataka i poslovna inteligencija zahtijeva više od zvjezdanih arhitekata, tehničara, modelar ili administrator baze podataka[1].

Svojstva tipičnog skladišta podataka su

1. Velike količine historijskih zapisa
2. Optimizovana za upite a ne dodavanje i ažuriranje podataka
3. Dopunjava se inkrementalno sa novim podacima u predefinisanim vremenskim intervalima
4. Osnova svakog BI rješenja

Skladište podataka u konačnici nije ništa drugo nego baza podataka koja je dizajnirana i implementirana na osnovu jasno definisanih pravila. Bitno je napomenuti da se dizajn i implementacija skladišta podataka dosta razlikuje od dizajna i implementacije standardne baze podataka koja je optimizovana za OLTP procese. Skladišta podataka čuvaju se odvojeno od relacionih baza podataka iz sljedećih razloga.

Sr.No.	Data Warehouse (OLAP)	Operational Database(OLTP)
1	It involves historical processing of information.	It involves day-to-day processing.
2	OLAP systems are used by knowledge workers such as executives, managers, and analysts.	OLTP systems are used by clerks, DBAs, or database professionals.
3	It is used to analyze the business.	It is used to run the business.
4	It focuses on Information out.	It focuses on Data in.
5	It is based on Star Schema, Snowflake Schema, and Fact Constellation Schema.	It is based on Entity Relationship Model.
6	It focuses on Information out.	It is application oriented.
7	It contains historical data.	It contains current data.
8	It provides summarized and consolidated data.	It provides primitive and highly detailed data.
9	It provides summarized and multidimensional view of data.	It provides detailed and flat relational view of data.
10	The number of users is in hundreds.	The number of users is in thousands.
11	The number of records accessed is in millions.	The number of records accessed is in tens.
12	The database size is from 100GB to 100 TB.	The database size is from 100 MB to 100 GB.
13	These are highly flexible.	It provides high performance.

Slika br.3: Usporedba relacione i data warehouse baze podataka

Kao što je već spomenuto, skladište podataka pomaže poslovnim rukovoditeljima da organiziraju, analiziraju i koriste svoje podatke za donošenje poslovnih odluka. Skladište podataka služi kao jedini dio sistema povratnih informacija plana-izvršenja-procjene. Upravo iz svih navedenih razloga, skladišta podataka su svoje mjesto u primjeni našla sljedećim poljima

- Financijske usluge
- Bankarske usluge
- Potrošačka roba
- Maloprodajni sektori
- Kontrolirana proizvodnja

Nakon što smo se upoznali sa skladištima podataka u sljedećem dijelu ćemo detaljno objasniti kako se vrši planiranje i implementacija istog.

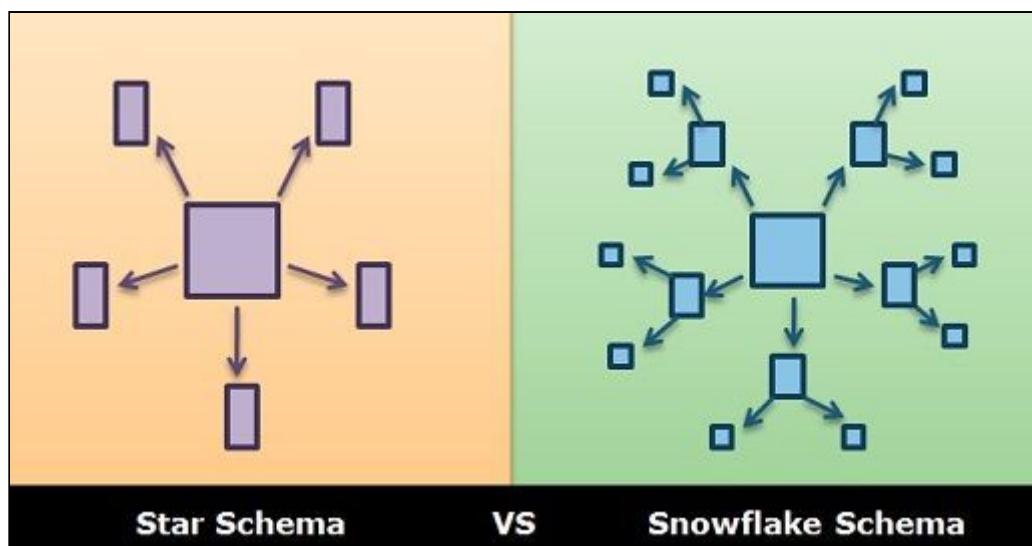
3.1 Dizajniranje skladišta podataka

Sad kad razumijemo ciljeve DW / BI sistema, razmotrimo osnove dimenzionalnog modeliranja. Dimenzionalno modeliranje je široko prihvaćeno kao preferirana tehnika za prezentiranje analitičkih podataka jer su istovremeno zadovoljena 2 zahtjeva:

- Dostaviti podatke koji su poslovnim korisnicima razumljivi
- Omogućite brzu izvedbu upita

Dimenzionalno modeliranje je tehnika kreiranja strukture podataka koja je optimizirana za pohranu podataka u skladištu podataka. Svrha dimenzionalnog modeliranja je optimizacija baze podataka za brži pristup podacima. Koncept dimenzionalnog modeliranja razvio je Ralph Kimball, a sastoji se od tablica „činjenica“ (engl. *Fact tables*) i „dimenzija“ (engl. *Dimension tables*). Dimenzionalni model u skladištu podataka dizajniran je za čitanje, sažimanje, analizu numeričkih podataka poput vrijednosti, stanja, brojanja itd. Suprotno tome, relacijski modeli optimizirani su za dodavanje, ažuriranje i brisanje podataka u sistemu mrežnih transakcija u realnom vremenu. Ovi dimenzionalni i relacijski modeli imaju svoj jedinstveni način pohrane podataka koji ima specifične prednosti. Na primjer, u relacijskom načinu rada, normalizacija i ER modeli smanjuju redundanciju nad podacima. Suprotno tome, dimenzionalni model u skladištu podataka raspoređuje podatke na takav način da je lakše doći do podataka i generisati izvještaje. Također, bitno je napomenuti da dimenzionalni model sadrži iste informacije kao i normalizirani model, ali su podaci u formatu koji omogućava razumljivost te optimizirano slanje upita koji su od krucijalnog značaja prilikom donošenja bitnih poslovnih odluka. Prilikom dimenzionalnog modeliranja postoje dva tipa schema a to su

- Star schema
- Snowflake schema



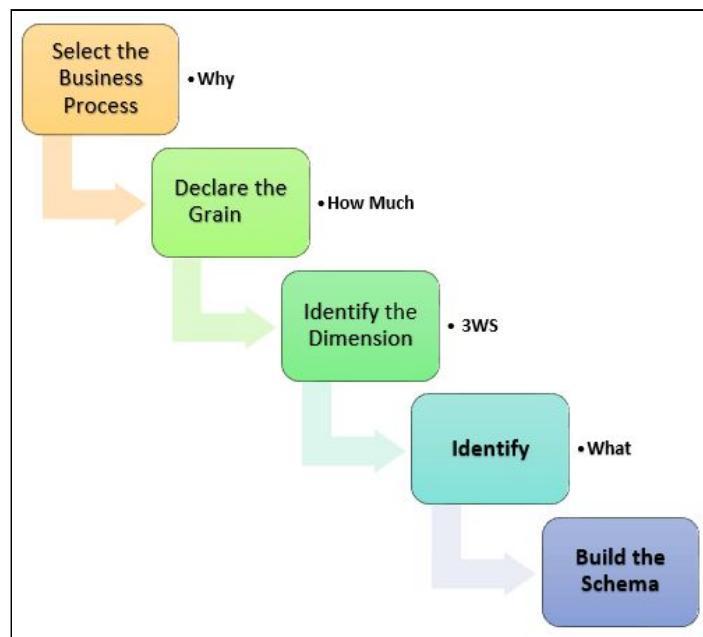
Slika br.4: Prikaz star i snowflake scheme

Dakle, ključne razlike između star i snowflake scheme su:

- Star schema sadrži samo jednu dimenzijsku tabelu, dok kod snowflake schema mogu imati tabele dimenzija koje se referenciraju na druge poddimenzije

- Normalizacija se koristi u snowflake schemi koja eliminira redundanciju nad podacima. Suprotno tome, normalizacija se ne izvodi u star schemi što rezultira redundancijom
- Star schema je jednostavna, lako razumljiva i uključuje manje složene upite. Naprotiv, snowflake shemu teško je razumjeti i uključuje složene upite.
- Pristup podatkovnom modelu u star shemi je down top, dok kod snowflake sheme se koristi top down pristup.
- Star schema koristi manji broj join operatora. S druge strane, snowflake shema koristi veliki broj join operatora.
- Prostor koji zauzima star schema veći je u usporedbi sa snowflake shemom.
- Vrijeme potrebno za izvršavanje upita u star shemi je manje. Suprotno tome, snowflake schema troši više vremena zbog pretjerane upotrebe join operatora.

Star i snowflake scheme koriste se za projektovanje skladišta podataka. Obe scheme imaju određene prednosti i nedostatke. Međutim, uporaba snowflake scheme smanjuje redundanciju, ali nije popularna kao star schema koja se u većini slučajeva koristi prilikom dizajna skladišta podataka. Prije nego krenemo sa modeliranjem dimenzionalnih i fact tabela potrebno je identifikovati poslovne procese te ispitati izvore podataka identifikovanih poslovnih procesa. Prilikom modeliranja skladišta podataka, proces identificiranja poslovnih procesa i izvora podataka ima krucijalnu važnost, jer u odnosu na provedenu analizu će biti modeliranje kako dimenzione tako i činjenične tabele. Možemo sa pravom reći da u ovome dijelu mjeseta za greške nema. Naš model bi trebao opisivati zašto, koliko, kada, gdje, ko i koji poslovni proces se analizira.



Slika br.5: Koraci prilikom kreiranja scheme skladišta podataka

Referencirajući se na *sliku br.4* pogledajmo ključne stvari na koje trebamo obratiti pažnju prilikom kreiranja scheme skladišta podataka.

1) Identifikacija poslovnih procesa

Podrazumijeva identificiranje stvarnog poslovog procesa koji bi skladište podataka trebalo obuhvatiti. To može biti marketing, prodaja, HR itd. Odabir poslovog procesa također ovisi o kvaliteti podataka dostupnih za taj proces. To je najvažniji korak u procesu modeliranja podataka. Ovdje bi neuspjeh imao kaskadne i nepopravljive nedostatke. Prilikom opisivanja poslovog procesa, možemo koristiti običan tekst ili "osnovnu notaciju modeliranja poslovnih procesa" (*engl. BPMN*) ili "jedinstveni jezik za modeliranje" (*engl. UML*).

2) Prepoznavanje granularnosti

Granularnost opisuje razine detalja za poslovni problem / rješenje. To je postupak identificiranja najniže razine informacija za bilo koju tabelu u skladištu podataka. Ako tabela sadrži podatke o prodaji za svaki dan, to bi trebala biti dnevna granulacija. Ako tabela sadrži ukupne podatke o prodaji za svaki mjesec, ona ima mjesecnu detaljinost.

3) Prepoznavanje dimenzija

Dimenzije su imenice poput datuma, proizvoda, uposlenika itd. U tim dimenzijama trebaju se pohraniti svi podaci. Na primjer, dimenzija datuma može sadržavati podatke poput godine, mjeseca i dana u sedmici.

4) Prepoznavanje činjenica

Ovaj je korak povezan s poslovnim korisnicima sistema jer tu dobivaju pristup podacima pohranjenim u skladištu podataka. Pored referentnih ključeva ka dimenzionim tabelama, sadržaj redova tabele činjenica su aditivne i poluaditivne numeričke vrijednosti.

5) Kreiranje scheme skladišta podataka

U ovom koraku implementiramo prethodno isplanirani dimenzioni model.

3.2 Dimenzijske tabele

Dimenzijske tabele sadrže tekstualni kontekst povezan s događajem mjerjenja poslovnog procesa. Dimenzijske tabele odgovaraju na pitanja

- tko
- šta
- gdje
- kada
- kako
- zašto

Kao što je prikazano *na slici br.5*, dimenzijske tabele često imaju mnogo atributa. Nerijetko je da dimenzijska tabela ima 50 do 100 atributa, iako neke dimenzijske tabele prirodno imaju samo nekoliko atributa. Svaka je dimenzija definirana jednim primarnim ključem , koji služi kao osnova za referencijalni integritet s bilo kojom fact tabelom na koju se referencira.

Product Dimension
Product Key (PK)
SKU Number (Natural Key)
Product Description
Brand Name
Category Name
Department Name
Package Type
Package Size
Abrasive Indicator
Weight
Weight Unit of Measure
Storage Type
Shelf Life Type
Shelf Width
Shelf Height
Shelf Depth
...

Slika br.6: Primjer dimenzijske tabele

Atributi dimenzija služe kao primarni izvor ograničenja upita i grupiranja prilikom kreiranja izvještaja. Atributi dimenzijske tabele igraju vitalnu ulogu u DW / BI sistemu upravo jer su oni izvor gotovo svih ograničenja prilikom kreiranja bilo kakvih izvještaja. Dimenzijski atributi su u konačnici osnova za kreiranje upotrebljivog i razumljivog DW / BI sistema. Atributi bi trebali sastoje se od stvarnih deskriptivnih riječi, a ne od kriptičnih skraćenica. Na mnogo je načina skladište podataka jednako dobro kao i atributi dimenzija koje se nalaze u istom. Analitička snaga DW / BI sistema izravno je proporcionalna kvalitetu i dubini atributa dimenzija. Prije nego objasnimo najbolje prakse za modeliranje dimenzijskih tabela pogledajmo tzv. Enterprise Bus matrix koju je izmislio Ralph Kimball. Matrica sabirnice(engl. *Enterprise bus matrix*) nadovezuje se na koncept uskladištenih dimenzija koja omogućava stvaranje strukture

zajedničkih dimenzija koju idealno mogu koristiti sva preduzeća u svim poslovnim procesima koji su povezani sa sistemom. Koncept sabirnice dobro je poznat u jeziku informacionih tehnologija i on je ono što odražava usklađeni koncept dimenzija u skladištu podataka, stvarajući osnovnu strukturu u kojoj se svi dijelovi sistema povezuju, osiguravajući interoperabilnost i dosljednost podataka.

BUSINESS PROCESSES	COMMON DIMENSIONS						
	Date	Product	Warehouse	Store	Promotion	Customer	Employee
Issue Purchase Orders	X	X	X				
Receive Warehouse Deliveries	X	X	X				X
Warehouse Inventory	X	X	X				
Receive Store Deliveries	X	X	X	X			X
Store Inventory	X	X		X			
Retail Sales	X	X		X	X	X	X
Retail Sales Forecast	X	X		X			
Retail Promotion Tracking	X	X		X	X		
Customer Returns	X	X		X	X	X	X
Returns to Vendor	X	X		X			X
Frequent Shopper Sign-Ups	X			X		X	X

Slika br.7: Primjer enterprise bus matrice

Sada, kada smo objasnili šta su to ustvari dimenzijske tabele, pogledajmo neke od najboljih praksi prilikom modeliranja istih.

1) Struktura dimenzionih tabela

Svaka dimenziona tabela ima jedan i samo jedan primarni ključ. Ovaj je primarni ključ ugrađen kao strani ključ u bilo kojoj pridruženoj tabeli činjenica. Atributi dimenzijske tabele su primarni cilj ograničenja i grupiranje specifikacija iz upita i BI aplikacija.

2) Dimenzioni surogatni ključevi

Dimenzijske tabele dizajnirane su tako da imaju jednu i samo jednu kolonu koja služi kao jedinstveni primarni ključ. Ovaj primarni ključ ne može biti prirodni ključ operativnog sistema (primarni ključ koji se koristi u relacionoj bazi podataka), jer će se vremenom naše skladište podataka puniti podacima koji se referenciraju za taj prirodni ključ kad se promjene prate tijekom vremena. DW / BI sistem također zatražuje kontrolu nad primarnim ključevima svih dimenzija, podaci se mogu importovati sa više različitih lokacija i upravo zbog tog razloga se koristi surogatni primarni ključ.

3) Prirodni ključevi

Prirodni ključevi stvoreni u izvornim operativnim sistemima podliježu poslovnim pravilima izvan DW / BI sistema. Na primjer, broj zaposlenika (primarni ključ u

relacionoj bazi podataka) može se promijeniti ako zaposlenik podnese ostavku, a zatim bude ponovno zaposlen. Kada skladište podataka želi imati jedan ključ za tog zaposlenika, mora se stvoriti novi trajni ključ koji je doslijedan (*engl. Persistent*) i ne mijenja se u ovoj situaciji. Taj se ključ ponekad naziva i trajni superprirodni ključ (*engl. Durable supernatural key*). Najbolji ključevi imaju format koji je neovisan o izvornom poslovnom procesu i stoga bi trebali biti jednostavni cijeli brojevi dodijeljeni u nizu počevši od 1.

4) Drilling down

Analiza podataka najvažniji je način na koji poslovni korisnici analiziraju podatke. Drilling down jednostavno znači dodavanje reda u postojeći upit. Novi red je atribut dimenzije dodan izrazu GROUP BY u SQL upitu. Atribut može doći iz bilo koje dimenzije pridružene tablici činjenica u upitu. Drilling down down ne zahtijeva definiciju unaprijed zadanih hijerarhija.

5) Degenerativne dimenzije

Ponekad je potrebno definisati dimenziju koja nema sadržaj osim svog primarnog ključa. Degenerativne dimenzije se najčešće pojavljuju kod činjeničnih tabela koje pohranjuju informacije o transakcijama i akomuliraju trenutačno stanje.

6) Višestruke hijerarhije u dimenzijama

Mnoge dimenzijske tabele sadrže više od jedne prirodne hijerarhije. Na primjer, datumska dimenzija može sadržavati hijerarhiju razdoblja od dana do sedmice, kao i hijerarhiju iz dana u mjesec u godinu. Dimenzije lokacija mogu biti višestruke geografske hijerarhije. U svim tim slučajevima zasebne hijerarhije koegzistiraju u istoj dimenzionoj tabeli.

7) Null atributi u dimenzijama

Atributi dimenzija koji imaju null vrijednost rezultiraju kada zadani red dimenzije nije bio potpuno popunjeno ili kada postoje atributi koji nisu primjenjivi na sve redove dimenzije. U oba slučaja preporučuje se zamjenu opisnog niza, kao što je nepoznato (*engl. Unknown*) umjesto null vrijednosti. Null attribute treba izbjegavati jer različite baze podataka handleaju grupiranje i ograničavanje null vrijednosti na drugačiji način.

8) Tabela za vremenske dimenzije

Tabela za vremenske dimenzije se referenciraju na gotovo svaku tabeli činjenica kako bi se omogućila navigacija tabele činjenica kroz poznate datume, mjesecе, fiksna razdoblja i posebne dane. Tabela za vremenske dimenzije tipično ima mnogo atributa koji opisuju karakteristike kao što su broj sedmice, naziv mjeseca, fiksno razdoblje. Da bi se olakšalo razdvajanje, primarni ključ za vremenske dimenzije se uglavnom stavlja u formatu GGGGMMDD, umjesto sekvenčialno dodijeljenog surogat ključa. Preporuke kojih bi se trebali držati prilikom kreiranja vremenskih dimenzija su:

- Surogatni ključ u formatu GGGGMMDD
- Granularnost
- Opseg
- Višestruki kalendari

DateKey	DateAltKey	MonthDay	WeekDay	Day	MonthNo	Month	Year
00000000	01-01-1753	NULL	NULL	NULL	NULL	NULL	NULL
20130101	01-01-2016	1	3	Tue	01	Jan	2016
20130102	01-02-2016	2	4	Wed	01	Jan	2016
20130103	01-03-2016	3	5	Thu	01	Jan	2016
20130104	01-04-2016	4	6	Fri	01	Jan	2016

Slika br.8: Tabela vremenske dimenzije

9) Junk dimenzije

Junk dimenzija je tablica dimenzija koja se sastoji od atributa koji ne pripadaju tablici činjenica ili bilo kojoj postojećoj tablici dimenzija. Priroda ovih atributa obično je tekst ili razne zastave, npr. generički komentari ili samo jednostavni indikatori da / ne ili istina / laž. Ove vrste atributa obično ostaju kada su identificirane sve očite dimenzije u poslovnom procesu i stoga se dizajner suočava s izazovom gdje treba staviti te atributi koji ne pripadaju drugim dimenzijama.

3.3 Dizajniranje sporo promjenljivih dimenzija

Sporo promjenjive dimenzije (Slowly Changing Dimensions, SCD) su dimenzijske relacije čiji se atributi mijenjaju sporo odnosno niskom frekvencijom[2]. Autor koncepta je Ralph Kimball (1994. godine), a termin i akronim pojavili su se naknadno nakon što je koncept široko prihvaćen kako u istraživačkoj zajednici tako i u pratećoj industriji [1]. Ralph Kimball je izvorno predložio tri strategije – Tip 1, Tip 2 i Tip 3[2]. Iako su ove tehnike izdržale test vremena pojavile su se i neke dodatne te njihove kombinacije tako da se danas spominje čak osam vrsta dimenzija, ali ovdje nećemo navesti sve[2]. Kada se atribut koji nam je zanimljiv mijenja moguće su tri strategije (te njihove razne kombinacije): 1. prebrisati vrijednost 2. dodati novi zapis 3. dodati novi atribut Prije nastavka, valja uočiti da se koristi termin sporomijenjajuća dimenzija, ali zapravo se strategija postupanja definira na razini dimenzijskog atributa, te dimenzija u teoriji može imati više tipova[2]. U praksi, ako dimenzija ima barem jedan dimenzijski atribut čiju povijest pratimo, nazivamo ju sporomijenjajućom dimenzijom[2].

3.3.1 Sporamijenjajuća dimenzija tipa 1

Ova se strategija sastoji u prepisivanju stare vrijednosti atributa novom[2]. Pristup je prikladan za ispravljanje pogrešaka u podacima te za održavanje aktualne vrijednosti dimenzijskih atributa uz uvjet da nije potrebno pamtitи prethodne vrijednosti (ispravne ili pogrešne)[2]. Očiti nedostatak pristupa je nemogućnost rekonstruiranja povijesnih podataka jer se prati samo najnovija, trenutna vrijednost atributa[2]. Također, postojeće agregirane tablice treba preračunati[2].

Dimenzijske tipove 1 su pogodne kad [2]:

- izmjena vrijednosti je rezultat popravljanja pogreške u izvorišnom sustavu [2]
- praćenje povijesne vrijednosti atributa nije zanimljivo odnosno vrijedno [2]

3.3.2 Sporomijenjajuća dimenzija tipa 2

Kada se govori o sporomijenjajućim dimenzijama, onda se tipično misli na tip 2 [2]. U ovom pristupu povijesni podaci se prate dodavanjem nove n-torce u dimenzijsku relaciju svaki put kad se promijeni vrijednost bilo kojeg atributa tipa 2 [2]. Ova strategija podrazumijeva da je u shemu dimenzijske relacije uključen prirodni ključ izvorne relacije [2]. Ako je prirodni ključ kompozitni, svi atributi ključa moraju biti sadržani u shemi dimenzijske relacije [2]. Prirodni ključ preuzet iz izvornog sustava omogućuje direktno povezivanje n-torce u dimenzijskom s izvornom n-torkom u relacijskom modelu [2]. Primarni ključ dimenzijske relacije je uvijek surogatni - serijski broj bez značenja koji se automatski generira [2]. Ovim pristupom se postiže da je za isti entitet (s istim prirodnim ključem) u dimenzijskoj relaciji pohranjeno više n-torki [2].

3.3.3 Sporomijenjajuća dimenzija tipa 3

Sporomijenjajuća dimenzije tipa 3 koriste se u situacijama u kojima je potrebno nove činjenične podatke promatrati u kontekstu starih vrijednosti dimenzijskih podataka i obratno [2]. Za razliku od sporo promjenjivih dimenzija tipa 2 koje omogućuju neograničeno praćenje povijesnih podataka, ovim pristupom neće biti moguće zapamtiti sve povijesne podatke ako se vrijednost sporo promjenjivog atributa promijeni više od jedan put [2]. Dok strategija Tip 2 može uzrokovati ubrzani porast broja zapisa u dimenzijskoj relaciji i značajan porast prostora potrebnog za pohranu, kod ove strategije to nije slučaj [2]. Iako je inicijalna shema relacije proširena dodatnim atributima kako broj n-torki ne raste s promjenama ne očekuje se nagli porast potrebe za dodatnim prostorom za pohranu [2]. Sporomijenjajuća dimenzije tipa 3 se rijetko koriste [2].

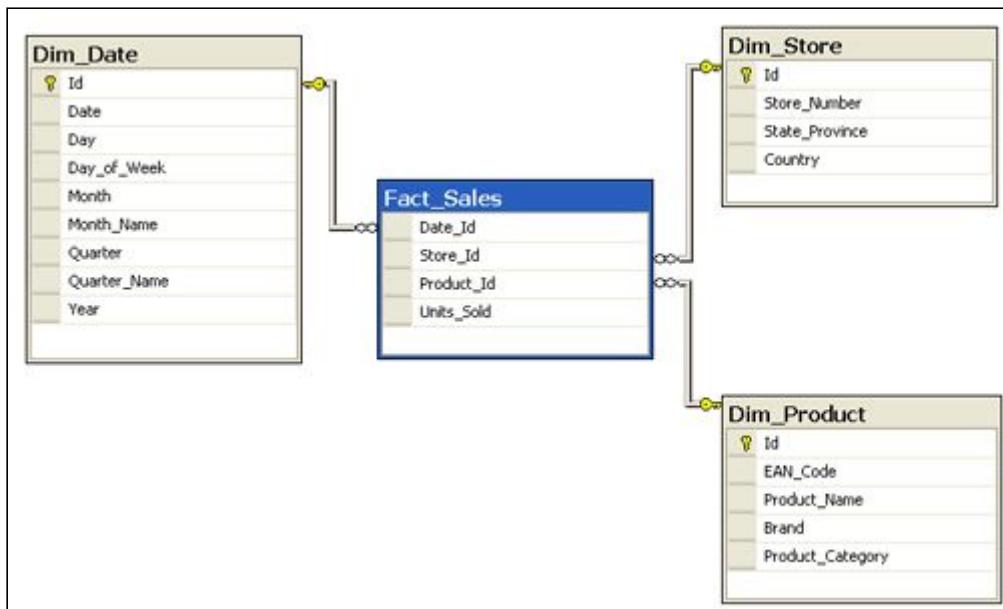
3.4 Činjenične tabele

U skladištu podataka tablica činjenica sastoji se od mjernih podataka ili činjenica poslovnog procesa. Smješteno je u središtu star ili snowflake scheme. Tabela činjenica obično ima dvije vrste atributa:

- one koje sadrže činjenice
- one koje su strani ključ prema dimenzijskim tabelama.

Primarni ključ tablice činjenica može biti surogatni ključ ili složeni ključ (kompozitni) koji se sastoji od svih njenih stranih ključeva. Tabele činjenica sadrže sadržaj skladišta podataka i pohranjuju različite vrste mjera poput aditivnih, neaditivnih i poluaditivnih mjera. Unutar činjeničnih tabela se pohranjuju uglavnom aditivne vrijednosti koje djeluju kao neovisne varijable pomoću kojih se analiziraju dimenzijski atributi. Tabele činjenica često se definiraju

prema granularnosti. Granularnost činjeničnih tabela predstavlja najatomičniju razinu pomoću koje se činjenice mogu definirati. Činjenične tablice tipično imaju velik broj zapisa [2]. Upravo zato se činjenične tablice i drže normaliziranim, pri čemu sadrže samo numeričke atributе (koji troše malo prostora na disku) [2]. Drugim riječima, jedan redak činjenične tablice zauzima malo prostora na disku, čime se pokušavamo nositi s potencijalno velikim brojem zapisa [2].



Slika br.9: Primjer činjenične tabele na koju se refenciraju dimenzijske tabele

3.4.1 Mjere

Tabela činjenica može sadržavati ili činjenice na razini detalja ili činjenice koje su agregirane.

Tablica činjenica može pohraniti različite vrste mjera a to su

- Aditivne mjere su mjere koje se mogu agregirati po svim dimenzijama. Mjera je aditivna samo ako rezultat aggregatne funkcije SUM ima smisla i odgovara istini bez obzira na dimenzijske pojmove koje sudjeluju u upitu, odnosno bez obzira kako formulirali upit.
- Poluaditivne mjere su mjere koje se mogu agregirati samo po nekim dimenzijama. Mjera je semiaditivna ako je aditivna s obzirom na barem jednu dimenzijsku pojmove i nije aditivna s obzirom na barem jednu drugu dimenzijsku pojmove.
- Neaditivne mjere su mjere koje se ne mogu agregirati ni po jednoj dimenzijskoj pojmove, ali se iste nalaze u činjeničnoj tabeli u smislu nekog filtera.

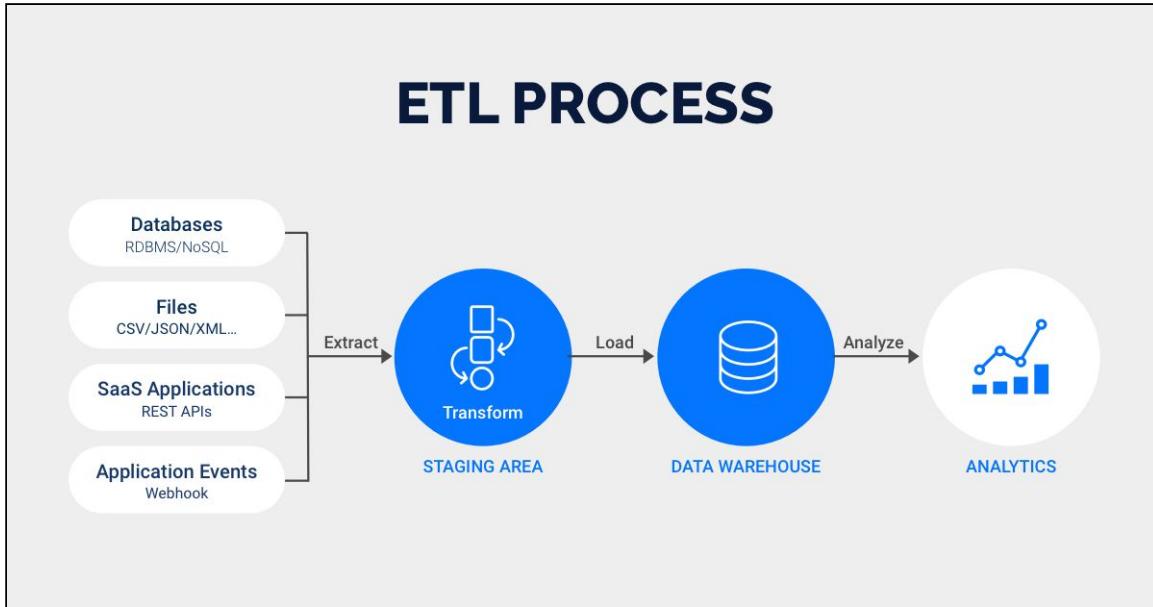
4. ETL proces (Extract transform & load)

Extract transform and load je postupak kopiranja podataka iz jednog ili više izvora u skladište podataka. Izdvajanje podataka uključuje izdvajanje podataka iz homogenih ili heterogenih izvora. Transformacija podataka obrađuje podatke čišćenjem podataka i pretvaranjem istih u odgovarajući format za potrebe upita i analiza. Ispravno dizajnirani ETL sistem izvlači podatke iz izvornih sistema, provodi standarde kvalitete i konzistentnosti podataka, usklađuje podatke tako da se odvojeni izvori mogu koristiti zajedno i na kraju isporučuje podatke u formatu spremnom za prezentaciju tako da manageri mogu donositi bitne poslovne odluke. ETL sistemi obično integriraju podatke iz više aplikacija, koje obično razvijaju i podržavaju različiti dobavljači ili su hostirani na zasebnom računarskom hardveru. Odvojenim sistemima koji sadrže izvorne podatke često upravljaju različiti zaposlenici. Kompanije su se dugi niz godina oslanjale na ETL postupak kako bi stekle uvid u podatke koji pokreću bolje poslovne odluke. Danas je ova metoda integriranja podataka iz više sistema i izvora i dalje ključna komponenta alata za integraciju podataka neke organizacije.

Pogledajmo neke od stavki zašto je ETL proces ključan u procesu poslovne inteligencije:

- 1) ETL pruža duboki historijski kontekst za poslovanje.
- 2) Pružajući objedinjeni prikaz, ETL olakšava poslovnim korisnicima analizu i izvještavanje o podacima relevantnim za njihove inicijative.
- 3) ETL može poboljšati produktivnost stručnjaka za obradu podataka jer koristi procese koji premještaju podatke bez potrebe za tehničkim vještinama pisanja koda ili skripti
- 4) ETL se s vremenom razvio kako bi podržao nove integracijske zahtjeve
- 5) ETL proces omogućuje objedinjavanje podataka, a sve u cilju održanja tačnost i revizije koja je obično potrebna za skladištenje podataka, izvještavanje i analitiku.
- 6) Kako se izvori podataka mijenjaju, skladište podataka se automatski ažurira.
- 7) ETL proces omogućuje poređenje podataka između izvora i ciljnog sistema.
- 8) ETL je unaprijed definirani postupak za pristup i manipulaciju izvornim podacima u ciljnau bazu podataka.

ETL je provjerena metoda na koju se mnoge organizacije oslanjaju svakodnevno - poput trgovaca koji moraju redovito pregledavati podatke o prodaji. ETL može kombinirati i prikazati podatke o transakcijama iz skladišta ili nekog drugog izvora podataka, tako da su isti poslovnim ljudima spremni za pregled u formatu koji razumiju. ETL se također koristi za migraciju podataka iz naslijедenih sistema u moderne sisteme s različitim formatima podataka. Pristup širokom opsegu podataka kompanijama može pružiti konkurentnu prednost. Danas kompanije trebaju pristup svim vrstama podataka - od videozapisa, društvenih medija, Interneta stvari (IoT-a), prostornih podataka itd. ETL dobavljači (*engl. ETL vendors*) često dodaju nove transformacije svojim alatima kako bi podržali nove zahtjeve i nove izvore podataka. ETL i drugi softverski alati za integraciju podataka, koji se koriste za u ETL procesu osiguravaju pouzdanost podataka. ETL alati integriraju se s alatima za kvalitetu podataka, a ETL dobavljači u svoja rješenja uključuju srodne alate, poput onih koji se koriste za mapiranje i poređenje podataka. Također, jako je bitno napomenuti da je ETL proces ponavljajuća aktivnost (dnevno, sedmično, mjesечно) te istii mora biti agilan, automatizovan i dobro dokumentovan. Bilo koja analiza i proces poslovne inteligencije je nezamisliv bez kvalitetnih podataka koji su konzistentni i ujednačeni.



Slika br.10: Slikovni prikaz ETL procesa

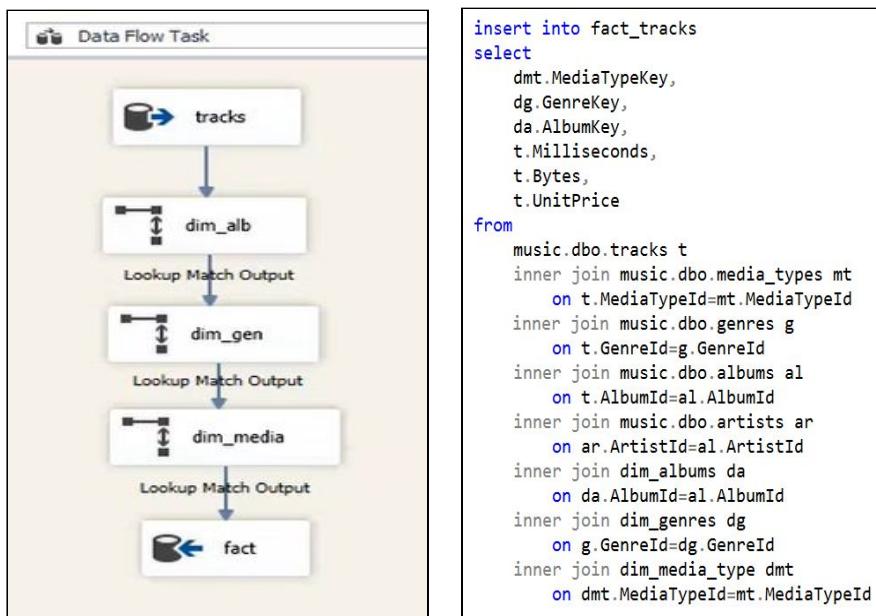
Jedan od nedostataka ETL procesa jeste što se isti ne može uvijek raditi tj. Ne možemo operetiti izvornu bazu podataka kada to mi želimo odnosno kada mi mislimo da je za to najbolji trenutak. Povlaci kada se povlače je tema detaljne analize svakog sistema ponaosob. Tu se također javljaju i problemi prilikom prenosa podataka. Zamislimo otkazani transfer velike količine podataka. Zdrav ritam kompletнnog procesa je jako važan.

4.1 Planiranje extrakcije podataka (Extract)

Prije samo ekstrakcije podataka javo je bitno izvršiti profiliranje izvora podataka jer na osnovu toga se vrši planiranje extrakcije podataka i validacije istih, od kojih zavisi sljedeći korak a to je transformacija. U ovom koraku ETL procesa, podaci se iz izvornog sistema izdvajaju u tzv međuprostor (*engl. staging area*) radi bolje kontrole. Nakon izdvajanja podataka u međuprostor vršimo identifikovanje novih i modifikovanih zapisa putem vremenskih atributa u tabelama. Nije poenta da se data warehouse napuni samo jednom. Data warehouse se puni konstantno, međutim čime se isti puni je dobro pitanje. Data warehouse se inkrementalno puni sa novim zapisima i sa onim koji su modifikovani kako bi poslovna analiza obuhvatila novo i stanje koje je promijenjeno u odnosu na izvorne podatke. Pomoću audit mehanizama, imamo uvid ko i šta mijenja na izvornom sistemu. U zavisnosti od proizvođača DBMS sistema, postoje Change data capture mehanizmi koji prate bilo kakve izmjene nad podacima. Također, ne treba zaboraviti i na namjenske metode koje ćemo sami implementirati a sve u cilju detekcije izmjena nad podacima. Vremenski period u kojem se naš data warehouse puni je svakako faktor koji ne smijemo izostaviti. Za određivanje inkrementalnog vremenskog perioda punjenja skladišta podataka, potrebno je poznavati izvorni sistem. Koliko se često isti puni podacima, koliko se podaci zadržavaju u sistemu (*engl. Retention period*), koliko se podataka povlači iz izvornog sistema, koja je frekfentna visina transakcija tog sistema su samo neka od pitanja na koje moramo imati jasne odgovore. Vidimo da mnogo faktora igra ulogu tokom ekstrakcije podataka, te isti utiču na trajanje samog ETL procesa. Povlačenje podataka iz izvornog sistema treba biti u vremenskom periodu kada je isti najmanje frekfentan.

4.2 Planiranje transformacije podataka (Transform)

Podaci povučeni iz izvornog OLTP sistema su sirovi i ne mogu se koristiti u izvornom obliku. Stoga iste treba očistiti, mapirati i transformirati. Zapravo je ovo ključni korak u kojem ETL postupak dodaje vrijednost i mijenja podatke tako da se mogu generisati pronicljive poslovne analize. U koraku transformacije se mogu izvoditi prilagođene operacije nad istima. Bitno je naglasiti da proces planiranja transformacije podataka priprema iste za analizu. Fazu transformacije možemo uraditi na više mesta. Počevši od izvorišta podataka, u međuprostoru ili pak unutar samoga skladišta podataka. Danas postoje alati koji su konkretno namijenjeni za ETL fazu i koji poprilično elegantno rade ovaj proces, međutim ako dođe do nekog erora prilikom ovakvog načina transformacije podataka, poznavanje SQL-a je neophodno. Kada pričamo o transformaciji podataka, međutim iste su u većini slučajeva kombinacija SQL-a i BI alata.



Slike br.11 i 12: Prikaz povlačenja podataka korištenjem data flow alata i sql koda

Elegantnost data flow alata vidi se kada imamo kompleksne transformacije upravo zato što imamo vizuelni prikaz iste. Međutim, ako dođe do bilo kakvih erora, poznavanje SQL-a je mandatorno za debagiranje i rješavanje istog. Prilikom transformacije potrebno je također voditi računa o samoj optimizaciji upita kojeg smo pripremili. Prilikom optimizacije trebamo imati neke od ključnih stavki na našem umu, a to su:

- 1) Optimizacija upita: odabrati samo one kolone i zapise koji su nam potrebni
- 2) Izbjegavati nepotrebna sortiranja: koristiti već sprtorane podatke ako je to moguće
- 3) Provjetiri tempdb postavke
- 4) Koristit skripte ako govorimo o SQL pristupu a ne front-end komponente

Prilikom transformacija trebamo voditi računa i o samim greškama. Nažalost, nismo iste u mogućnosti predvidjeti, međutim kada se iste pojave, procesiranje istih ima ključnu ulogu. Minimizacijom rollback operacija, logiranjem zapisa nad kojim je došlo do same greške, audit tabela korisnika su efikasne metode koje nam omogučavaju uspješno procesiranje grešaka.

4.3 Planiranje procesa punjenja (Load)

Faza učitavanja učitava podatke u krajnji cilj, što može biti bilo koja pohrana podataka, uključujući jednostavni import podataka u skladište. Bitno je napomenuti da izvor iz kojeg se podaci importuju u data warehouse ne mora isključivo biti neki RDBMS sistem. To mogu biti excel, csv, plain text, xml, json, pdf, fajlovi ili pak neki cloud provider. Manje više, kroz poglavlja 4.1 i 4.2 smo objasniji i posljednju fazu, te u ovome dijelu čemu se referencirati na dijelove koje smo u tim poglavljima obradili svrsishodno. Dakle kao što smo naglasili, nije poenta da se data warehouse napuni samo jednom. Data warehouse se puni konstantno, međutim čime se isti puni je dobro pitanje. Data warehouse se inkrementalno puni sa novim zapisima i sa onim koji su modifikovani kako bi poslovna analiza obuhvatila novo i stanje koje je promijenjeno u odnosu na izvorne podatke. Također, vremenski period u kojem se naš data warehouse puni je svakako faktor koji ne smijemo izostaviti. Za određivanje inkrementalnog vremenskog perioda punjenja skladišta podataka, potrebno je poznavati izvorni sistem. Povlačenje podataka iz izvornog sistema treba biti u vremenskom periodu kada je isti najmanje frekfentan.

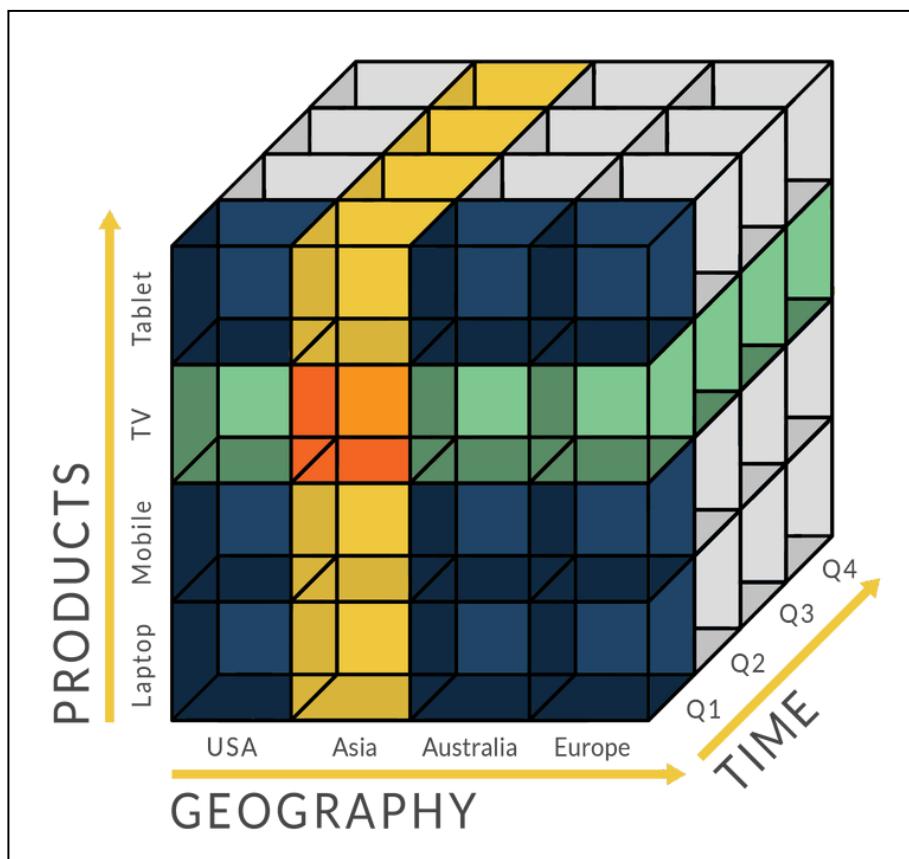
4.4 Platforme za ETL proces

Na samome kraju, elaboracija ETL procesa bi bila nepotpuna bez spomena alata koji su automatizirali i olakšali cjelokupan proces ekstrakcije, transformacije i punjena skladišta podataka. ETL alati mogu prikupljati, čitati i migrirati podatke iz višestrukih struktura podataka i na različitim platformama. ETL alati uključuju operacije spremne za upotrebu poput filtriranja, transformacije, sortiranja, spajanja i agregiranja. Uz to, ETL alati također podržavaju raspoređivanje transformacija, kontrolu verzija, nadzor i objedinjeno upravljanje metapodacima, dok su neki od alata integrirani s BI alatima. U nastavku pogledajmo neke od alata koji nam olakšavaju ETL proces.

- 1) Informatica PowerCenter
- 2) Business Objects Data Integrator
- 3) IBM InfoSphere DataStage
- 4) Microsoft SQL Server Integration Services
- 5) Oracle Warehouse Builder / Data Integrator
- 6) Pentaho Data Integration (Open Source)
- 7) Jasper ETL (Open Source)

5. Kocka podataka (Cube)

Prije nego objasnimo kocku podataka, podsjetimo se šta je to OLAP. OLAP je proces kreiranja analiza i izvještavanja koji omogućuje korisniku da lako i selektivno izdvaja i pregledava podatke s različitih stajališta temeljeno na multidimenzijskoj strukturi podataka zvanoj kocka (*engl. cube*). Operacije nad podacima i analize obično se izvode pomoću jednostavne proračunske tabele, gdje su vrijednosti podataka poredane u formatu redova i kolona. Međutim, OLAP kocka sadrži višedimenzijske podatke, a podaci se obično dobivaju iz drugog i nepovezanog izvora. Snaga kocke podataka ogleda se u tome jer ista može pohraniti i analizirati višedimenzijske podatke na logičan i uredan način. Kocka podataka je višedimenzijski ("n-D") niz vrijednosti. Količina podataka unutar kocki podataka se mjeri u terabajtima i petabajtima, te optimizovani lookup nad tolikom količinom podataka je mandatoran. Kocka podataka koristi se za predstavljanje podataka uz neku mjeru od interesa. Kocka podataka općenito je višedimenzijski koncept koji može biti jednodimenzijski, dvodimenzijski, trodimenijski ili višedimenzijski. Svaka dimenzija predstavlja neki atribut u bazi podataka, a čeliće u kocki podataka predstavljaju mjeru od interesa. Na primjer, mogu sadržavati broj koliko puta se kombinacija atributa dogodi u bazi podataka, ili minimalnu, maksimalnu, zbir ili prosječnu vrijednost nekog atributa.



Slika br.13: Kocka podataka

OLAP sistemi podržavaju četiri vrste osnovnih analitičkih operacija nad kockom podataka a to su:

1) Roll-up / Drill down

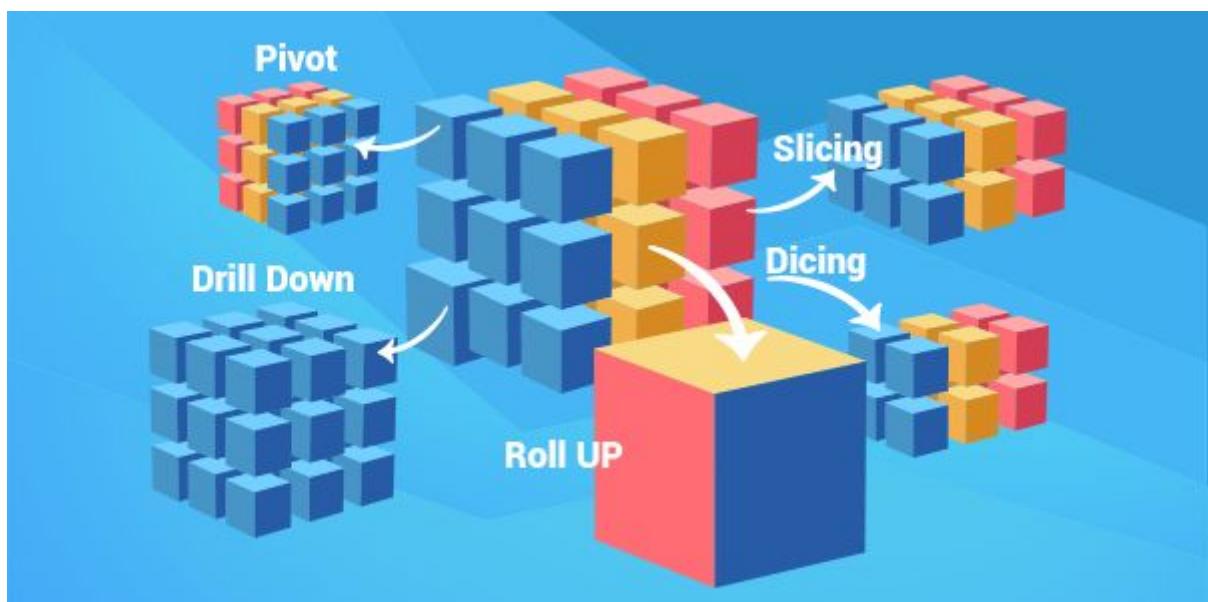
Operacija koja se sastoji od agregiranja odnosno sakupljanja podataka u kocki čime ovu operaciju možemo nazvati agregiranjem. Isto kao i drill down to je jedna od analitičkih tehnika gdje korisnik ima mogućnost navigacije razina podataka od najsajetijih (drill up) do najdetaljnijih (drill down)[3].

2) 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. Pogled stranice je pojam koji označava trenutnu orientaciju odnosno pogled na multidimenzionalnu krišku. Horizontalne dimenzije idu preko pogleda definirajući dimenziju stupca, dok vertikalne dimenzije idu niz pogled stranice definirajući tako sadržaj reda dimenzija . [3]

3) Pivot tabela

Operacija rotacije koja omogućava rotiranje osi, mijenjajući određene dimenzije kako bi se stekao drugi pogled na multidimenzionalnu kocku. Operacija se može sastojati od mijenjanja redova i stupaca, pomicanja jednog reda dimenzije u dimenziju stupca, itd. Ovisno o tome koliko kocka ima dimenzija, povećava se broj mogućnosti pivotiranja. [3]



Slika br:14: Grafički prikaz analitičkih operacija unutar OLAP sistema

Kada pričamo o dimenzijama unutar skladišta podataka, iste osiguravaju kontekst za činjenice. Imamo tri tipa opcija za pohranu dimenzija a to su:

1) MOLAP

Multidimenzionalni OLAP, skraćeno MOLAP predstavlja alate kod kojih su podaci spremljeni u multidimenzionalne kocke, kako se ne mogu koristiti relacijske baze podataka većinom su ti podaci spremljeni u vlastite optimizirane baze podataka višedimenzionalne matrične strukture. Panian i Klepac (2003) navode kako je najveća prednost ovakve OLAP arhitekture zapravo brzina, a Nagabhushana (2008) navodi kako takvi alati jednostavno moraju pružiti visoke performanse, sofisticirane multidimenzionalne kalkulacije i funkcionalnosti. No baš zbog toga, mana ovih alata se vidi u zauzeću ogromnog prostora u slučaju postojanja većeg broja dimenzija. [3]

2) ROLAP

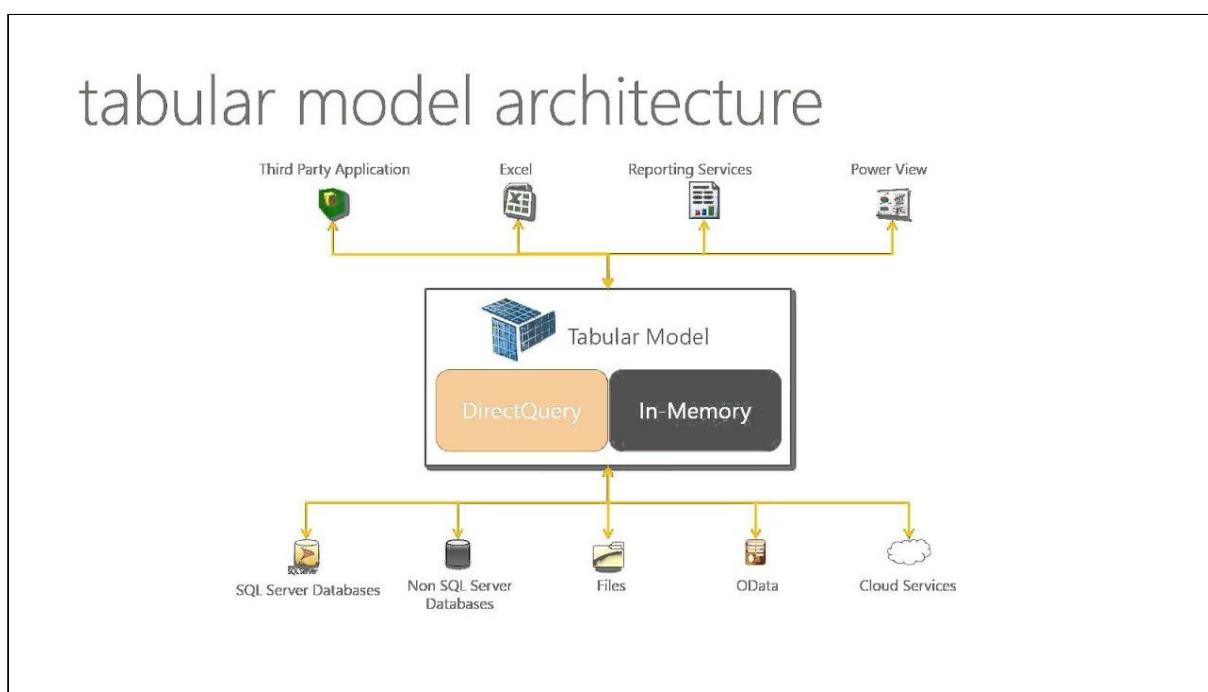
Relacijski OLAP ili ROLAP alati sve podatke koje dobiju spremaju u standardne sustave relacijskih baza podataka, ne spremivši tako ništa u eksterne repozitorije. Sposobni su raditi s vrlo velikim skupom podataka, ali su kompleksni što dovodi do toga da su vrlo skupi za implementirati, a isto tako imaju spore performanse izvođenja upita te nisu sposobni izvoditi složenije finansijske kalkulacije. ROLAP alati ne spremaju nužno podatke za analizu odvojeno od ostalih izvornih podataka. Panian i Klepac (2003) isto tako navode kako u MOLAP ROLAP HOLAP DOLAP performase kapacitet funkcionalnost razmjestivost 16 ROLAP alati imaju otežanu primjenu standardnih upitnih jezika (npr. SQL-a), na koje su korisnici zapravo navikli pa neke druge oblike interakcije s bazama podataka ne smatraju zapravo prednošću nego nedostatkom. [3]

3) HOLAP

Hibridni OLAP koristi kombinaciju tehnika spremanja podataka. Predstavljaju kombinaciju ROLAP i MOLAP arhitekture jer će određene podatke spremati u multidimenzionalne baze podataka dok će drugi dio podataka spremiti u relacijske baze podataka te tako koristi neke pogodnosti obaju tehnika. Panian i Klepac (2003) su opisali kako HOLAP pravne podatke drži u relacijskoj bazi podataka i pristupa njima pomoću drillthrough procedura, a sažetak drži u multidimenzionalnoj kocki, čime se postiže velika brzina pristupa i relativno malo zauzimanje prostora. [3]

6. Tabularni model

Tabularni modeli su baze podataka koje se izvode u memoriji ili u DirectQuery načinu rada, povezujući se s podacima iz pozadinskih relacijskih izvora podataka. Korištenjem najsvremenijih algoritama za kompresiju i procesiranje upita u višeprocesorskom načinu rada, Analytics Engine Analysis Services Vertipaq omogućuje brzi pristup objektima i podacima tabularnog modela izvještavanjem klijentskih aplikacija poput Power BI-a i Excela. U "In-Query" načinu rada tabularni modeli su baza podataka koja uključuje sve podatke u memoriji. Tabularni model postiže zavidne performanse uz pomoć xVelocity engine-a. xVelocity omogućuje da tabučarni model učita objekte i podatke u memoriju s najmodernejim algoritmima kompresije i pruža mu funkciju izvršavanja upita s više niti. "Direct Query" djelomično radi s načinom "In-Query". Iako su zadani zadani modeli u memoriji, DirectQuery je alternativni način upita za modele koji su ili preveliči da bi se uklopili u memoriju ili kada velika količina podataka isključuje razumnu strategiju obrade. DirectQuery postiže jednakost s modelima u memoriji kroz podršku za širok raspon izvora podataka, sposobnost obrade izračunatih tabela i kolona u DirectQuery modelu. Tabularni modeli kreiraju se u programu Microsoft Visual Studio uz pomoć Analysis Services projektnog templatea. Proširenje instalira dizajner tabličnog modela, koji pruža površinu dizajna za stvaranje objekata semantičkog modela poput tabela, particija, odnosa, hijerarhija, mjera i KPI-a. Tabularni modeli mogu se implementirati u Power BI Premium-u, Azure Analysis Services ili uz pomoć instance SQL Server Analysis Services konfiguriranih za način tabelarnog posluživanja [4].



Slika br.15: Arhitektura tabularnog modela

Nakon što smo objasnili šta je to tabularni model referencirajući se na microsoftovu dokumentaciju, objasniti ćemo šta su to zrazi za analizu podataka(*engl. Data Analysis Expression*). Izrazi za analizu podataka (DAX) je jezik koji se koristi za stvaranje prilagođenih izračuna u uslugama Analysis Services, Power BI i Power Pivot u programu Excel. Formule DAX uključuju funkcije, operatore i vrijednosti za izvođenje naprednih izračuna nad podacima. Prilikom korištenja tabularnog modela, DAX formule se koriste u izračunatim kako kolonama tako i tabelama, mjerama i filterima.

1) Izračunate kolone

Izračunate kolone su kolone koje se dodaju u postojeću tabelu, a uz pomoć DAX-a se definiše formula koja definira vrijednosti kolone. Bitno je napomenuti da izračunate kolone nisu podržane za modele koji su importovani iz relacijskog izvora u DirectQuery načinu rada. Kada izračunata kolona sadrži valjanu DAX formulu, vrijednosti se izračunavaju za red čim se unese formula. Vrijednosti se zatim spremaju u bazu podataka. Rezultat za svaki red u izračunatoj koloni izračunava se odmah i prikazuje se. Vrijednosti se preračunavaju samo ako se podaci ponovno obrade.

2) Izračunate tabele

Izračunate tabele su objekti, koji se kreiraju na temelju DAX upita ili izraza, izvedeni iz svih ili dijela drugih tabela u istom modelu.

3) Mjere

Mjere su dinamičke formule u kojima se rezultati mijenjaju ovisno o kontekstu. Mjere se koriste u formatima izvještavanja koji podržavaju kombiniranje i filtriranje podataka ovisno o modelu, korištenjem više atributa kao što su Power BI izvješće ili Excel PivotTable ili PivotChart. Mjere definira autor modela pomoću mreže mjera u dizajneru modu. Formula u mjeri može koristiti standardne funkcije agregiranja automatski stvorene pomoću opcija Autosum, COUNT ili SUM. Za razliku od izračunatih kolona i filtera redova, sintaksa mjere uključuje naziv mjerne koja prethodi formuli.

4) Filteri

Filteri redova definiraju koji su redovi u tabeli vidljivi članovima određene uloge. Filteri redova kreiraju se za određenu ulogu pomoću SSMS-a. Kada se definira filter redova pomoću DAX formule, stvara se dopušteni skup redova. Ovo ne uskraćuje pristup ostalim redovima nego se jednostavno ne vraćaju kao dio dopuštenog skupa.. Ako je korisnik član druge uloge, a filteri redova te uloge omogućuju pristup određenom skupu redova, korisnik može pregledati te podatke . Redni filteri primjenjuju se na navedene redove kao i na povezane redove. Kada tabela ima više odnosa, filteri primjenjuju sigurnost na odnos koji je aktivan. Redni filteri presijecat će se s ostalim filterima koji su definisani za povezane tabele.

7. Izvještavanje

Veliki podaci igraju presudnu ulogu u analizi podataka, poslovnim informacijama i inteligentnom izvještavanju. Kompanije se moraju prilagoditi dvosmislenosti podataka i ponašati se u skladu s njima. Proračunske tabele više ne pružaju adekvatna rješenja za ozbiljne kompanije koje žele precizno analizirati i koristiti sve prikupljene poslovne podatke. Tu izvještavaju stupaju na snagu i zaista se pokazuju presudnim u osnaživanju organizacija da učinkovito prikupe podatke i naprave kvalitetnu poslovnu analizu. Kao rezultat toga, poslovna inteligencija može imati koristi od cijelokupnog razvoja, kao i od profitabilnosti kompanije, bez obzira na nišu ili industriju. Kreiranje izvještaja (*engl. Reports*) je proces prikupljanja podataka korištenjem različitih softvera i alata za dobivanje relevantnih uvida. U konačnici, izvještavanje nam daje prijedloge i zapažanja o poslovnim trendovima. Na osnovu izvještaja, možemo razviti strateški pristup i prikupiti uvide koji bi inače ostali zanemareni. Pogledajmo detaljnije kakve su koristi od takvih vrsta izvještavanja i kako kompanije, bile one male ili velike, mogu stvoriti profitabilne rezultate. Pogledajmo koje su ključne prednosti:

- 1) **Povećanje produktivnosti**
Budući da ljudi obrađuju vizualne informacije 60.000 puta brže od teksta, ritam rada može se značajno povećati korištenjem umjetne inteligencije u obliku interaktivnih vizualnih podataka u stvarnom vremenu. Sve se informacije mogu prikupiti u jedinstvenu nadzornu ploču, koja će u konačnici osigurati brz, jasan, jednostavan i učinkovit tok rada. Ova vrsta izvještaja postat će vizualna, lako joj se pristupa i postojano u prikupljanju uvida.
- 2) **Primjena u bilo kojoj industriji**
Stvaranje sveobuhvatnog izvještaja može biti zastrašujući zadatak za bilo koji odjel, zaposlenika ili menadžera. Ciljevi pisanja uspješnih, pametnih izvješća uključuju smanjenje troškova i poboljšanje učinkovitosti.
- 3) **Korištenje podataka u stvarnom vremenu**
Uz tradicionalna sredstva izvještavanja teško je iskoristiti i shvatiti golemu količinu prikupljenih podataka. Stvaranje jednostavne prezentacije od obimnih informacija može biti problem i najiskusnijim menadžerima. Izvješće napisano kao word dokument neće pružiti istu količinu informacija i koristi kao analiza podataka u stvarnom vremenu, s implementiranim alarmima koji mogu upozoriti na bilo koju poslovnu anomaliju, a takva vrsta softvera za podršku posljedično će povećati poslovnu učinkovitost i smanjiti troškove.
- 4) **Operativna optimizacija i predviđanje**
Svaka ozbiljna kompanija koristi ključne pokazatelje uspješnosti za mjerjenje procjenu uspjeha. Samo pravilno praćenje i analize mogu donijeti profitabilne rezultate. Poslovna inteligencija i izvještavanje nisu fokusirani samo na dio praćenja, već uključuju predviđanje temeljeno na prediktivnoj analitici i umjetnoj inteligenciji koje lako mogu izbjegći donošenje skupe i dugotrajne poslovne odluke. Izvještavanje u poslovnoj inteligenciji stoga je istaknuto iz više uglova koji mogu pružiti uvide koji se inače mogu previdjeti.

Prikupljanje podataka u današnjem svijetu koji je digitalno vođen važno je, ali analiza istih do još je presudnija ako komšanija želi uživati u održivom uspjehu suočena s neprestanim promjenama. Izvještavanje i poslovna inteligencija igraju presudnu ulogu u dobivanju osnovnih podataka za objašnjavanje odluka i prezentiranje podataka na način koji nudi izravne koristi za poslovanje. Kao što smo ranije spomenuli, ne postoji industrija na koju trenutno ne utiče važnost podataka i analiza. U ovom smjelom novom svijetu podatkovne inteligencije, kompanije svih veličina mogu koristiti BI alate kako bi transformirali uvid u bi postali lideri u svom polju. Uočavanje poslovnih problema, uz BI rješenja koja pružaju detaljna izvješća o poslovnoj inteligenciji, mogu samo stvoriti prostor za budući razvoj, smanjenje troškova i sveobuhvatnu analizu strateškog i operativnog stanja tvrtke.

8. Implementacija

U procesu implementacije jednog BI rješenja simuliran je proces u kojem je kompanija dostavila podatke na osnovu kojih je potrebno sprovesti poslovnu analizu.

Zahtjev klijenta: Potrebno je kreirati skladište podataka i izvršiti import u isti. Nakon kreiranja skladišta i importa podataka potrebno je analizirati poslovni proces prodaje, te na osnovu kreirane analize donijeti prijedlog poboljšanja poslovanja. Svaka poslovna odluka treba imati odgovarajuće obrazloženje.

Prilikom analize relevantnih tabela, kreirano je sljedeće skladište podataka. U nastavku se nalazi SQL kod uz pomoću kojeg je kreirano skladište podataka te import istih.

8.1 Kreiranje skladišta i import podataka

```
create database AdwentureDWSeminarski  
go
```

```
use AdwentureDWSeminarski  
Go
```

8.1.1 Kreiranje i import podataka u dimenzijsku tabelu SalesTerritory

```
create table DIM_SalesTerritory  
(  
    SalesTerritoryKey int not null identity(1,1),  
    SalesTerritoryId int not null,  
    TerritoryName nvarchar(50) not null,  
    CountryRegionCode nvarchar(3) not null,  
    TerritoryGroup nvarchar(50) not null,  
    constraint PK_DIMSalesTerritory primary key(SalesTerritoryKey)
```

```

)
go

insert into DIM_SalesTerritory
select
    TerritoryID,
    Name,
    CountryRegionCode,
    [Group]
from
    AdventureWorks2014.Sales.SalesTerritory
Go

```

8.1.2 Kreiranje i import podataka u dimenzijsku tabelu Product

```

create table DIM_Product
(
    ProductKey int not null identity(1,1),
    ProductId int not null,
    ProductName nvarchar(50) not null,
    Color nvarchar(15) not null,
    ProductSubCategoryName nvarchar(50) not null,
    ProductCategoryName nvarchar(50) not null,
    ProductModelName nvarchar(50) not null,
    constraint PK_DimProduct primary key(ProductKey)

)
go

insert into DIM_Product
select
    p.ProductID,
    p.Name,
    isnull(p.Color,'N/A'),
    isnull(psc.Name,'N/A'),
    isnull(pc.Name,'N/A'),
    isnull(pm.Name,'N/A')
from
    AdventureWorks2014.Production.Product          p      left      join
AdventureWorks2014.Production.ProductSubcategory psc
        on p.ProductSubcategoryId=psc.ProductSubcategoryId
    left join AdventureWorks2014.Production.ProductCategory pc
        on pc.ProductCategoryID=psc.ProductCategoryID
    left join AdventureWorks2014.Production.ProductModel pm
        on p.ProductModelID=pm.ProductModelID
go

```

8.1.3 Kreiranje i import podataka u dimenzijsku tabelu Date

```
create table DIM_Date
(
    DateKey int not null ,
    Date datetime not null,
    Year int not null,
    Quarter int not null,
    Month int not null,
    Day int not null,
    constraint PK_DimDate primary key(DateKey)
)
go

insert into DIM_Date
select distinct
    (YEAR(OrderDate) * 10000 + MONTH(OrderDate) * 100 + DAY(OrderDate))
DateKey,
    OrderDate,
    YEAR(OrderDate),
    CASE
        when MONTH(OrderDate) <=3 then 1
        when MONTH(OrderDate) >3 and MONTH(OrderDate)<=6 then 2
        when MONTH(OrderDate) >6 and MONTH(OrderDate)<=9 then 3
        when MONTH(OrderDate) >9 and MONTH(OrderDate)<=12 then 4
    END,
    MONTH(OrderDate),
    DAY(OrderDate)
from AdventureWorks2014.Sales.SalesOrderHeader
go
```

8.1.4 Kreiranje i import podataka u činjeničnu tabelu Sales

```
create table FACT_Sales
(
    SalesKey int not null identity(1,1),
    SaleId int not null,
    DimSalesTerritoryKey int not null,
    DimProductKey int not null,
    DimDateKey int not null,
    OrderQty int not null,
    UnitPrice money not null,
    constraint PK_Sales primary key(SalesKey),
```

```

        constraint FK_Sales_SalesTerritory foreign key(DimSalesTerritoryKey) references
DIM_SalesTerritory(SalesTerritoryKey),
        constraint FK_Sales_Product foreign key(DimProductKey) references
DIM_Product(ProductKey),
        constraint FK_Sales_Date foreign key(DimDateKey) references DIM_Date(DateKey)
)
go

```

```

insert into FACT_Sales
select
    soh.SalesOrderID,
    dst.SalesTerritoryKey,
    dp.ProductKey,
    dt.DateKey,
    sod.OrderQty,
    sod.UnitPrice
from
    AdventureWorks2014.Sales.SalesOrderHeader soh
        inner join AdventureWorks2014.Sales.SalesOrderDetail
            sod
        on sod.SalesOrderID=soh.SalesOrderID
        inner join AdventureWorks2014.Sales.SalesTerritory st
        on soh.TerritoryID=st.TerritoryID
        inner join AdventureWorks2014.Production.Product p
        on sod.ProductID=p.ProductID
        left join AdventureWorks2014.Production.ProductSubcategory psc
        on p.ProductSubcategoryID=psc.ProductSubcategoryID
        left join AdventureWorks2014.Production.ProductCategory pc on
        psc.ProductCategoryID=pc.ProductCategoryID
        inner join DIM_SalesTerritory dst
        on st.TerritoryID=dst.SalesTerritoryId
        inner join DIM_Product dp
            on p.ProductID=dp.ProductId
        inner join DIM_Date dt
            on soh.OrderDate=dt.Date
order by sod.SalesOrderID
go

```

8.1.5 Kreiranje i import podataka u činjeničnu tabelu SalesFreight

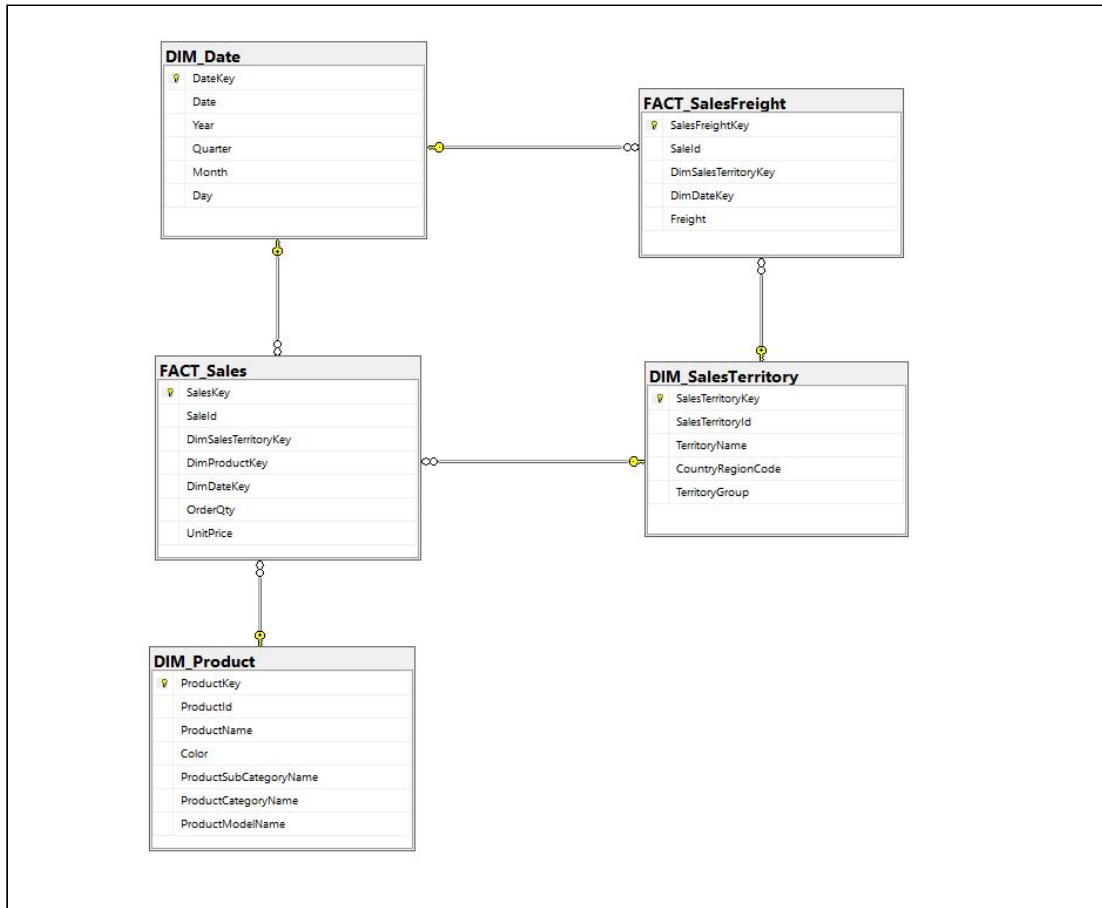
```
create table FACT_SalesFreight
(
    SalesFreightKey int not null identity(1,1),
    SaleId int not null,
    DimSalesTerritoryKey int not null,
    DimDateKey int not null,
    Freight money not null,
    constraint PK_SalesFreight primary key(SalesFreightKey),
    constraint FK_SalesFreight_SalesTerritory foreign key(DimSalesTerritoryKey)
references
    DIM_SalesTerritory(SalesTerritoryKey),
    constraint FK_SalesFreight_DimDate foreign key(DimDateKey) references
    DIM_Date(DateKey)
)
go

insert into FACT_SalesFreight
select distinct
    soh.SalesOrderID,
    dst.SalesTerritoryKey,
    dt.DateKey,
    soh.Freight
from
    AdventureWorks2014.Sales.SalesOrderHeader soh
        inner join AdventureWorks2014.Sales.SalesOrderDetail
            sod
        on sod.SalesOrderID=soh.SalesOrderID
        inner join AdventureWorks2014.Sales.SalesTerritory st
        on soh.TerritoryID=st.TerritoryID
        inner join DIM_SalesTerritory dst
        on st.TerritoryID=dst.SalesTerritoryId
        inner join DIM_Date dt
            on soh.OrderDate=dt.Date
order by soh.SalesOrderID
go

select * from FACT_SalesFreight
go
```

8.1.6 ER Dijagram skladišta podataka

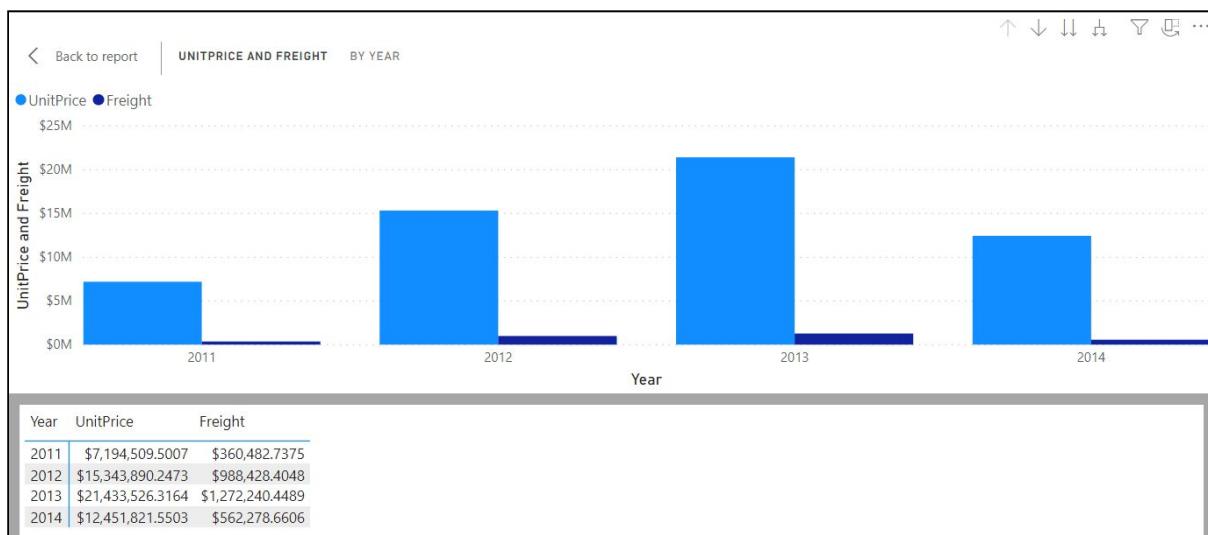
Pogledajmo ER dijagram kreiranog skladišta podataka. Možemo vidjeti da imamo 2 činjenične tabele a to su FACT_Sales i FACT_SalesFreight. ER dijagram skladišta podataka smo mogli kreirati iz 2 odvojena dijela. Prva schema bi obuhvatila činjeničnu tabelu FACT_Sales i dimenzijske tabele Date,Product,SalesTerritory, dok bi druga schema obuhvatila činjeničnu tabelu FACT_SalesFreight te Date i SalesTerritory dimenzije. Radi preglednosti, sve tabele su obuhvaćene jednim dijagramom.



Slika br.16: ER Dijagram kreiranog skladišta podataka

8.2 Analiza zarade i troškova prodaje

Prilikom analize kreirana je kocka podataka, te dijagram koji slikovito prikazuje zaradu i troškove na godišnjem nivou.



Slika br.17: Slikovni prikaz zarade i troškova na godišnjem nivou

Iz prikazanog dijagrama možemo vidjeti da je odnos zarade i troškova na godišnjem nivou na poprilično zavidnom nivou. Prikažimo odnos između troška i zarade procentualno.

Year	UnitPrice	Freight	Freight divided by UnitPrice
2011	\$7,194,509.5007	\$360,482.7375	\$0.050105255607061
2012	\$15,343,890.2473	\$988,428.4048	\$0.06441837036562
2013	\$21,433,526.3164	\$1,272,240.4489	\$0.059357495827765
2014	\$12,451,821.5503	\$562,278.6606	\$0.045156337836086

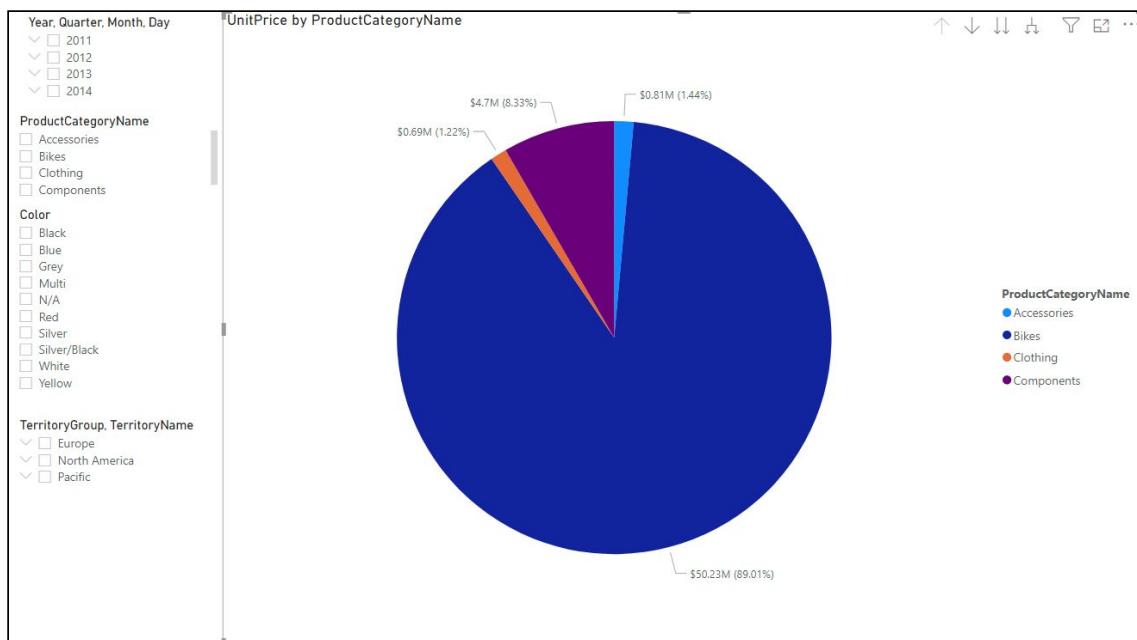
Slika br.18: Odnos troška na osnovu zarade na godišnjem nivou

S obzirom da je tip podatka monetaran zanemarimo u trećoj koloni monetarnu jedinicu. Dakle procenat(zaokružen na 2 decimale) troška u odnosu na zaradu po godinama iznosi sa :

- 2011 godina: 0.05 % je omjer troška na osnovu zarade
- 2012 godina: 0.06% je omjer troška na osnovu zarade
- 2013 godina: 0.05% je omjer troška na osnovu zarade
- 2014 godina: 0.04% je omjer troška na osnovu zarade

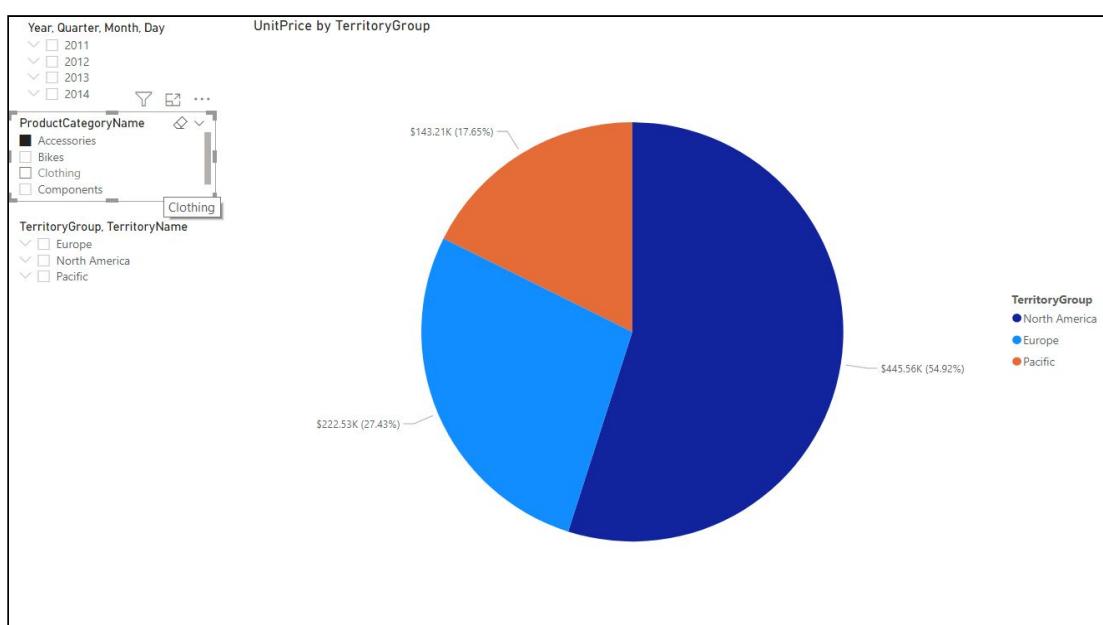
8.3 Analiza zarade i troškova po kategoriji prodavnih proizvoda

Pogledajmo pie chart koji je kreiran.

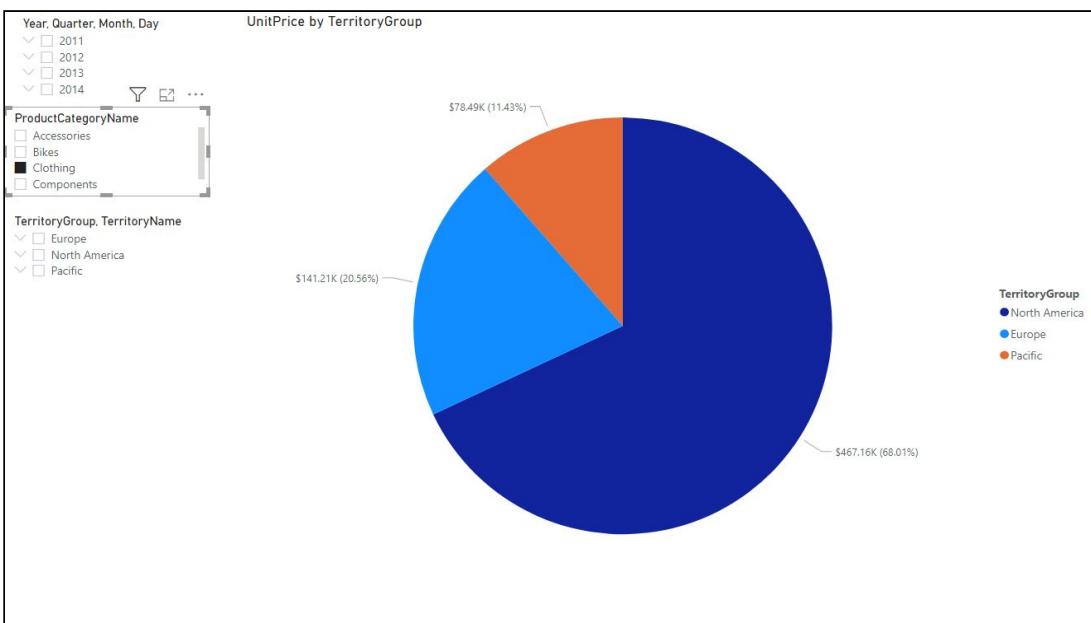


Slika br.17: Slikovni prikaz zarade po kategorijama prodavnih proizvoda

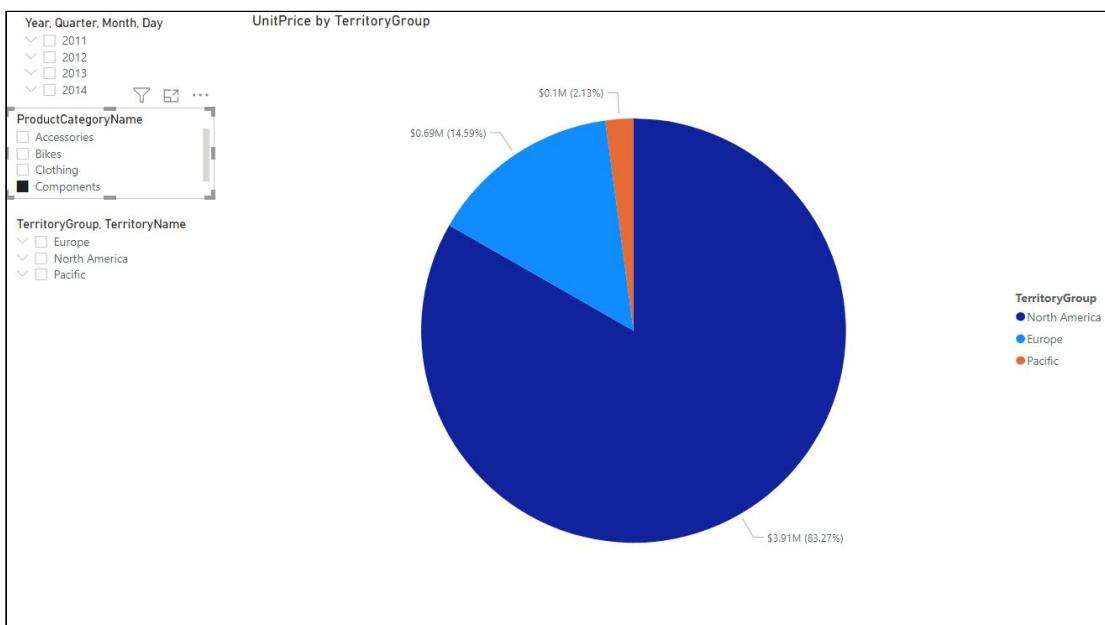
Možemo vidjeti da 89.01% ukupnih prihoda je ostvareno na Bikes kategoriji proizvoda. Drugo mjesto je zauzela kategorija Components sa 8.33% ukupne zarade. Treće utješno mjesto pripalo je kategoriji Accessories sa 1.44% prihoda od ukupne zarade i na kraju imamo kategoriju Clothing sa kojom se ostvarilo 1.22% ukupne zarade. Sa strane vidimo slicere na kreiranom reportu te ako uključimo bilo koji od istih vidimo da se situacija drastično ne mijenja. Očigledno je da se kategorije Components, Accessories i Clothing slabije prodaju. Pogledajmo sljedeća 3 pie char dijagrama:



Slika br.18: Slikovni prikaz zarade na kategoriji Accessories po teritorijalnim grupama

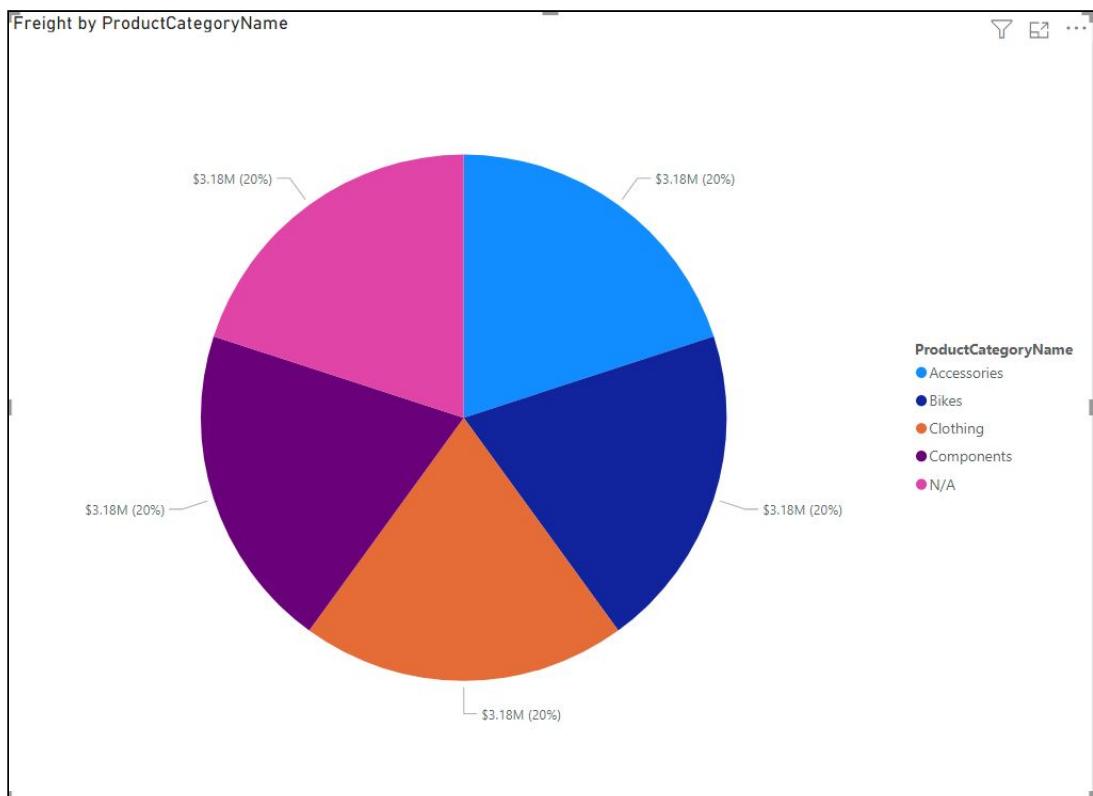


Slika br.19: Slikovni prikaz zarade na kategoriji Clothing po teritorijalnim grupama



Slika br.20: Slikovni prikaz zarade na kategoriji Components po teritorijalnim grupama

Kako bi izvršili precizniju analizu, kreirali smo pie chart dijagram gdje su prikazane teritorijalne grupe, a naknadno uz pomoć ProductCategoryName slicera prikazali procentualnu zaradu po teritorijama u odnosu na izabranu kategoriju. Iako slabija prodaja navedenih kategorija ne utiče na cjelokupan proces prodaje koji smo prikazali u analizi 8.2, možemo vidjeti da na sva 3 reporta, prodaja kategorija koje se slabije prodaju dominira na području Sjeverne Amerike, dok s druge strane vidimo da na području Pacifika i Evrope prodaja istih je dosta slabija. Analizom troškova po kategorijama proizvoda, dolazimo do zanimljive analize. Pogledajmo sljedeći pie char dijagram.

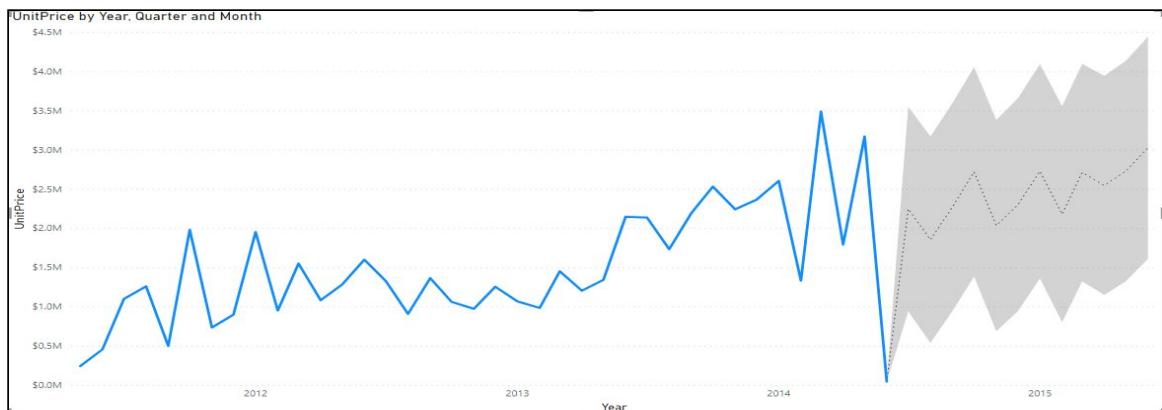


Slika br.21: Slikovni prikaz troškova u odnosu na kategoriju proizvoda

Ako uporedimo dijagram sa slike br.17 vidimo da u toku 2011,2012,2013 i 2014 godine na proizvodima koji nemaju definisanu kategoriju nije ostvaren nikakav prihod, dok na slici br.21 vidimo da je trošak za iste nas koštao zapetih 3.18 miliona dolara.

8.4 Prijedlog za poboljšanje poslovanja

Prije nego što dadnemo prijedlog za poboljšanje poslovanja, pogledajmo kakve su predikcije procesa prodaje za narednih 12 mjeseci sa 95% intervalom vjerovatnosti korištenjem Power BI-ovog forecasting featura, koji isti generiše u odnosu na prethodno dostupni set podataka.



Slika br.22: Kriva prodaje i predikcije iste za narednih 12 mjeseci

Plavom linijom vidimo krivu koja nam pokazuje oscilacije prodaje na godišnjem nivou. Isprekidana kriva nam pokazuje predviđene oscilacije prodaje u odnosu na podatke koji su dostupni. Područje označeno sivom bojom nam govori u kojem rasponu su moguće oscilacije predviđene krive. Bitno je naglasiti da je predikcija napravljena samo na osnovu postojećeg seta podataka, te shodno uvažavanju sljedećih prijedloga, rast isprekidane krive bi mogao biti i veći u odnosu na situaciju koja je prikazana na *slici br.22*.

Nakon izvršenih analiza i prikazanih reporta u prilogu se nalazi prijedlog za poboljšanje poslovanja.

- 1) Kontaktirati dobavljače proizvoda koji nemaju definisanu kategoriju i prekinuti nabavku istih
- 2) Na području Sjeverne Amerike, Evrope i Pacifika predlaže se uvođenje totalne rasprodaje i akcijske cijene za proizvode koji nemaju definisanu kategoriju a koji su $\geq 50\%$ u odnosu na inicijalnu cijenu proizvoda u vremenskom periodu do isteka zaliha
- 3) Na području Evrope i Pacifika predlaže se uvođenje popusta i akcijskih cijena prve sedmice u mjesecu za kategorije Accessories, Clothing koji su u rasponu od 5-15% u odnosu na inicijalnu cijenu proizvoda
- 4) Na području Pacifika predlaže se uvođenje popusta i akcijskih cijena prve sedmice u mjesecu za kategoriju Components koji su u rasponu 15-30% u odnosu na inicijalnu cijenu proizvoda

9. Literatura

- 1) The Data Warehouse Toolkit, 3rd Edition - Ralph Kimball, Margy Ross
- 2) https://www.fer.unizg.hr/_download/repository/SKRIPTA_-_Skladista_podataka_i_poslovna_inteligencija.pdf [2]
- 3) https://bib.irb.hr/datoteka/718474.1-Anto_Toma_-_Poslovna_inteligencija_oblikovane_OOLAP_kocke_na_primjeru.pdf [3]
- 4) <https://docs.microsoft.com/en-us/analysis-services/tabular-models/tabular-models-as?view=asallproducts-allversions> [4]
- 5) Slika br. 1:
https://upload.wikimedia.org/wikipedia/commons/thumb/0/06/DIKW_Pyramid.svg/1200px-DIKW_Pyramid.svg.png
- 6) Slika br 2:
<https://lh3.googleusercontent.com/proxy/NsKEufpi4dsqwh5ez0qJp5BilnIBjQzuqVbk-VHtXtcSUP5vBLHWlfXyrsDzLX0w1yOj5pu6cxOe1ja9q4SFJuFWn6R0hBp0GNIJz6nYERJS0Yn0sAr7qdLwKI0Rfvn5Ch8nLwCdr8SAFxrn>
- 7) Slika br:3: https://www.tutorialspoint.com/dwh/dwh_overview.htm
- 8) Slika br:4:
<https://techdifferences.com/difference-between-star-and-snowflake-schema.html>
- 9) Slika br:5: <https://www.guru99.com/dimensional-model-data-warehouse.html>
- 10) Slika br.6:The Data Warehouse Toolkit, 3rd Edition - Ralph Kimball, Margy Ross 50 str
- 11) Slika br.7:
<https://www.kimballgroup.com/wp-content/uploads/2013/08/Data-Warehouse-Bus-Matrix1.png>
- 12) Slika br.8:
https://www.youtube.com/watch?v=eWMR6QY8sHk&list=PLy-wPT_Y-K8zj7VqofdEcWI2eMZzgB42s&index=3 25:19
- 13) Slika br.9: <https://zentut.com/wp-content/uploads/2012/10/fact-table-example1.png>

14) Slika br.10 :

<https://rivery.io/wp-content/uploads/2020/05/ETL-Process-for-linkedin3.png>

15) Slika br:13:

<https://images.maritimeprofessional.com/images/maritime/w1200h1200/photothe-tide-tech-generate-metocean-different-87456.jpg>

16) Slika br.14:

https://galaktika-soft.com/wp-content/uploads/2018/01/olap_operations2.png

17) Slika br.15: <https://i.ytimg.com/vi/AAX1KdWJUfw/maxresdefault.jpg>