

Skladište podataka i poslovna inteligencija

Lesar, Alen

Undergraduate thesis / Završni rad

2017

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Pula / Sveučilište Jurja Dobrile u Puli**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:137:648449>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-29**



Repository / Repozitorij:

[Digital Repository Juraj Dobrila University of Pula](#)



Sveučilište Jurja Dobrile u Puli

Odjel za informacijsko-komunikacijske tehnologije

ALEN LESAR

SKLADIŠTA PODATAKA I POSLOVNA INTELIGENCIJA

Završni rad

Pula, rujan 2017.

Sveučilište Jurja Dobrile u Puli
Odjel za informacijsko-komunikacijske tehnologije

ALEN LESAR

SKLADIŠTA PODATAKA I POSLOVNA INTELIGENCIJA
Završni rad

JMBAG: 0303055721, redovni student

Studijski smjer: Sveučilišni preddiplomski studij Informatika

Predmet: Baze podataka

Znanstveno područje: Društvene znanosti

Znanstveno polje: Informacijske i komunikacijske znanosti

Znanstvena grana: Informacijski sustavi i informatologija

Mentor: doc.dr.sc. Tihomir Orehovački

Pula, rujan 2017.



IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisani Alen Lesar, kandidat za prvostupnika informatike ovime izjavljujem da je ovaj Završni rad rezultat isključivo mogega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio Završnog rada nije napisan na nedozvoljen način, odnosno da je prepisan iz kojega necitiranog rada, te da ikoji dio rada krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

Student

U Puli, _____, _____ godine



IZJAVA

o korištenju autorskog djela

Ja, Alen Lesar dajem odobrenje Sveučilištu Jurja Dobrile u Puli, kao nositelju prava iskorištavanja, da moj završni rad pod nazivom Skladišta podataka i poslovna inteligencija koristi na način da gore navedeno autorsko djelo, kao cjeloviti tekst trajno objavi u javnoj internetskoj bazi Sveučilišne knjižnice Sveučilišta Jurja Dobrile u Puli te kopira u javnu internetsku bazu završnih radova Nacionalne i sveučilišne knjižnice (stavljanje na raspolaganje javnosti), sve u skladu s Zakonom o autorskom pravu i drugim srodnim pravima i dobrom akademskom praksom, a radi promicanja otvorenoga, slobodnoga pristupa znanstvenim informacijama.

Za korištenje autorskog djela na gore navedeni način ne potražujem naknadu.

U Puli, _____(datum)

Potpis

Skladišta podataka i poslovna inteligencija

Alen Lesar

Sažetak: U ovom radu su objašnjeni pojmovi skladišta podataka i poslovne inteligencije. Navedene su tri najčešće arhitekture skladišta podataka kao i vrlo bitan ETL proces te dimenzijsko modeliranje. Također su navedeni neki od stilova poslovne inteligencije kao i prednosti njezina uvođenja te neki od uobičajenih problema koji se javljaju u poslovnoj inteligenciji.

Ključne riječi: skladište podataka, poslovna inteligencija, ETL proces, dimenzijsko modeliranje

Data warehouses and business intelligence

Abstract: This bachelor thesis explains concepts of data warehousing and business intelligence. Three most commonly used data warehouse architectures, as well as a very important ETL process and dimensional modeling have been mentioned. Also, some of the business intelligence styles as well as the advantages of its introduction and some of the common problems that arise in business intelligence are listed.

Keywords: data warehouse, business intelligence, ETL process, dimensional modeling

Sadržaj

1. Uvod	1
2. Ciljevi skladišta podataka i poslovne inteligencije.....	3
3. Skladište podataka	5
3.1 Definicija skladišta podataka	5
3.2 Arhitektura skladišta podataka	6
3.2.1 Pristup odozgo-dolje	6
3.2.2 Pristup odozdo-gore	8
3.2.3 Hibridni pristup.....	10
3.3 ETL proces.....	11
3.3.1 Dohvaćanje podataka	12
3.3.2 Transformacija podataka	13
3.3.3 Učitavanje podataka	15
3.4 Dimenzijsko modeliranje	16
3.4.1 Zvezdana shema	17
3.4.2 Činjenične tablice	18
3.4.3 Dimenzijske tablice.....	20
3.4.4 Pahuljasta shema	23
4. Poslovna inteligencija	24
4.1 Definicija poslovne inteligencije	24
4.1.2 Osnovna svojstva poslovne inteligencije	25
4.2 Stilovi poslovne inteligencije	26
4.2.1 Poslovno izvještavanje	27
4.2.2 Ad hoc upiti i analize.....	28
4.2.3 Rudarenje podataka	30
4.2.4. Analize pomoću OLAP kocke	31
4.3 Prednosti uvođenja poslovne inteligencije u poslovanje	33
4.4 Najčešći problemi poslovne inteligencije.....	34
5. Zaključak	37
6. Literatura	38

Popis slika

Slika 1. Odozgo-dolje pristup	7
Slika 2. Pristup odozdo-gore	9
Slika 3. Hibridni pristup	11
Slika 4. Grafički prikaz ETL procesa (Faizaan, 2015.)	12
Slika 5. Primjer transformacije podataka	15
Slika 6. Prikaz zvjezdaste sheme (Datawarehouse4u, 2008-2009)	17
Slika 7. Skeniranje proizvoda i prijevod u činjenične tablice (Kimball i Ross, 2013.) .	18
Slika 8. Primjer dimenzijske tablice (Kimball i Ross, 2013.)	20
Slika 9. Primjer redaka iz tablice dimenzije s denormaliziranim hijerarhijama	22
Slika 10. Primjer pahuljaste sheme	23
Slika 11. Komponente modela poslovne inteligencije (Panian i Klepac, 2003)	26
Slika 12. OLAP kocka (Houston, 2011.)	31
Slika 13. Različite analize pomoću OLAP kocke (Panian et al., 2007)	32

1. Uvod

Donošenje ispravnih poslovnih odluka je vrlo značajno kod praćenja i zadovoljavanja sve zahtjevnijeg tržišta. Da bi se donosile ispravne poslovne odluke potrebno je imati znanje o samome poduzeću i naravno o okruženju u kojemu to poduzeće posluje. Za stjecanje znanja je potrebno imati odgovarajuće informacije koje se temelje na različitim izvorima. Prikupljanje podataka te dobivanje potrebnih informacija iz tih podataka mora biti što kraće, jer je podataka sve više a vremena za njihovo prikupljanje i analiziranje sve manje. Potrebno je obratiti pozornost na kvalitetu podataka jer se samo na temelju kvalitetnih podataka mogu dobiti pouzdane informacije na temelju kojih se mogu donositi ispravne poslovne odluke.

Kako bi se dobile kvalitetne informacije iz raznih podataka koristi se skladište podataka. Skladištenje podatka služi za prikupljanje potrebnih podataka kako bi se dobile kvalitetne informacije te steklo potrebno poslovno znanje pomoću kojega se mogu donositi ispravne poslovne odluke. Skladištenje podataka odnosi se na potrebnu podatkovnu infrastrukturu koja uključuje sveobuhvatne, istinite i pravodobne integrirane podatke, transformirane u obliku koji je pogodan za dobivanje potrebnih informacija. Skladište podataka je konsolidacija podataka iz različitih izvora koje je osmišljeno kako bi podržalo strateško i taktičko odlučivanje. Glavna svrha skladišta podataka je pružiti koherentnu sliku poslovanja u određeno vrijeme.

Poslovna inteligencija donosi rješenja koja pomažu u dobivanju jasnije slike internih informacija, kupaca, opskrbnog lanca i financijske uspješnosti u tvrtkama. Također omogućava značajan povrat ulaganja, učinkovitije reagiranje na hitne slučajeve te brže reagiranje na nove mogućnosti. Rješenja poslovne inteligencije stvaraju organizacije za učenje tako što omogućuju tvrtkama da prate ciklus prikupljanja i analiziranja informacija, osmišljavanja i djelovanja na buduće planove. Kako bi se podržao taj ciklus i stekao uvid u mogućnosti poslovne inteligencije, organizacije moraju implementirati sustav poslovne inteligencije koji obuhvaća skladištenje podataka i analitička okruženja.

Dobro dizajnirani sustavi poslovne inteligencije su prilagodljivi, omogućuju kontinuirano mijenjanje kako bi odgovorili na nova i drugačija poslovna pitanja. Rješenja poslovne inteligencije se mogu nositi s velikom količinom strukturiranih a

ponekad i nestrukturiranih podataka kako bi se identificirale, razvile i na drugi način stvorile nove strateške poslovne mogućnosti. Poslovna inteligencija koristi tehnologije koje se usredotočuju na statistiku i poslovne ciljeve za poboljšanje uspješnosti poslovanja.

U ovom završnom radu će na početku biti objašnjeni osnovni ciljevi skladišta podataka i poslovne inteligencije. Nakon toga će biti pojašnjena definicija skladišta podataka, njegove tri različite arhitekture a to su pristup odozgo prema dolje, pristup odozdo prema gore te hibridni pristup. Zatim će detaljno biti pojašnjen ETL proces i njegovi potproces i nakon čega slijedi dimenzijsko modeliranje, neki od modela shema i dimenzijskog modeliranja.

U drugom dijelu bit će riječ o definiciji poslovne inteligencije te njezinim osnovnim svojstvima i nekim stilovima kao što su poslovno izvještavanje, ad hoc upiti i analize, rudarenje podataka i te analize pomoću OLAP kocke. Na samom kraju ovog rada bit će objašnjene prednosti uvođena poslovne inteligencije u poslovanje te neki njezini najčešći problemi.

2. Ciljevi skladišta podataka i poslovne inteligencije

U nastavku slijede neki od glavnih ciljeva skladišta podataka i poslovne inteligencije (Kimball i Ross, 2013:3*, vlastiti prijevod):

„Sustav mora učinite informacije lako dostupnima. - Sadržaj sustava skladišta podataka i poslovne inteligencije mora biti razumljiv. Podaci moraju biti intuitivni i očiti poslovnom korisniku, a ne samo razvojnom programeru. Strukture i oznake podataka trebale bi oponašati procese i vokabular poslovnih korisnika. Poslovni korisnici žele razdvojiti i kombinirati analitičke podatke u beskonačnim kombinacijama. Alati poslovne inteligencije i aplikacije koje pristupaju podacima moraju biti jednostavne i lake za korištenje. Također moraju vratiti rezultate upita korisniku s minimalnim vremenom čekanja. Možemo sažeti ovaj zahtjev jednostavno govoreći da moraju biti jednostavni i brzi.“

„Sustav mora dosljedno prezentirati podatke. - Podaci u sustavu moraju biti vjerodostojni. Oni moraju biti pažljivo sastavljeni iz raznih izvora, očišćeni, osigurani kvalitetom i pušteni samo kada su prikladni za korisničku potrošnju. Dosljednost također podrazumijeva zajedničke oznake i definicije za sadržaj sustava koji se koriste preko izvora podataka. Ako dvije mjere izvedbe imaju isti naziv, oni moraju značiti istu stvar. Suprotno, ako dvije mjere ne znače istu stvar, one bi trebale biti drugačije označene.“

„Sustav se mora prilagođavati promjenama. - Korisničke potrebe, uvjeti poslovanja, podaci i tehnologija podložni su promjenama. Sustav mora biti osmišljen tako da se tim neizbježnim promjenama nosi dostojanstveno kako ne bi poništio postojeće podatke ili aplikacije. Postojeći podaci i aplikacije se ne smiju mijenjati ili poremetiti kada poslovna zajednica postavlja nove zadatke ili se novi podaci dodaju u skladište. Ako se deskriptivni podaci u sustavu moraju mijenjati, treba na odgovarajući način računati na izmjene i učiniti ih jasnima za korisnicima.“

„Sustav mora pravodobno prikazati podatke. - Budući da se sustav intenzivnije koristi za operativne odluke, neobrađeni podaci možda će se morati pretvoriti u podatke koji se mogu aktivirati u roku od nekoliko sati, nekoliko minuta ili čak sekundi. Tim sustava poslovne inteligencije i skladišta podataka i poslovni korisnici moraju imati realna

očekivanja za ono što znači dostaviti podatke kada ima malo vremena za čišćenje ili potvrdu tih podataka.“

„Sustav mora biti siguran obrambeni sustav koji štiti imovinu informacija. - Najbitniji informacijski podaci organizacije se čuvaju u skladištu podataka. Skladište će barem minimalno sadržavati informacije o tome što organizacija prodaje, komu prodaje, po kojoj cijeni i to potencijalno predstavlja štetne pojedinosti u rukama pogrešnih ljudi. Sustav mora učinkovito kontrolirati pristup povjerljivim informacijama organizacije.“

„Sustav mora poslužiti kao mjerodavan i pouzdan temelj za poboljšanje odlučivanja. - Skladište podataka mora imati prave podatke za podršku odlučivanju. Najvažniji rezultati iz sustava su odluke koje se donose na temelju prikazanih analitičkih dokaza. Te odluke daju utjecaj poslovanja i vrijednost koja se može pripisati sustavu. Izvorna oznaka koja prethodi skladištu podataka i poslovnoj inteligenciji još uvijek je najbolji opis onoga što se dizajnira a to je sustav podrške odlučivanju.“

„Poslovna zajednica mora prihvatiti sustav kako bi se smatrao uspješnim. - Nije bitno izgraditi elegantno rješenje pomoću najboljih proizvoda i platformi. Ako poslovna zajednica ne prihvaća okruženje sustava i ne koristi ga aktivno, test prihvatanja je neuspješan. Za razliku od implementacije operativnog sustava u kojem poslovni korisnici nemaju izbora nego koristiti novi sustav, korištenje Sustav skladišta podataka i poslovne inteligencije ponekad je opcionalan. Poslovni korisnici će prihvatiti sustav ako on predstavlja jednostavan i brz izvor informacije u koje se mogu pouzdati.“

Iako je svaki zahtjev na ovom popisu važan, posljednja dva su najkritičnija, i često najnezavisnija. Uspješno skladištenje podataka i poslovna inteligencija zahtijevaju više od odličnog arhitekta, tehničara, modela ili administratora baze podataka. S inicijativom skladišta podataka i poslovne inteligencije, njihovi korisnici se jednim dijelom nalaze u svojoj komfornoj zoni informacijske tehnologije, a drugim dijelom se nalaze na nepoznatom području poslovnih korisnika (Kimball i Ross, 2013).

3. Skladište podataka

3.1 Definicija skladišta podataka

Pojam „Skladište podataka“ prvi je upotrijebio Bill Inmon 1990. godine. Prema Inmonu skladište podataka je usmjereno području, integrirano, vremenski obilježeno te nepromjenjivo/neizbrisivo (Ćurko i Španić, 2016):

- Usmjereno području (engl. *subject oriented*) - Pohranjeni podaci ciljaju određena područja. Primjer: može pohraniti podatke o ukupnoj prodaji, broju kupaca i slično a ne opće podatke o svakodnevnom poslovanju.
- Integrirano (engl. *integrated*) – Skladište podataka integrira podatke iz više izvora podataka. Integriranje raznorodnih podataka iz različitih izvora je važno kako bi skladište obuhvatilo i povezalosve potrebne podatke bez obzira na izvor i osiguralo podatke za cjelovit pogled na poslovanje. Na primjer, izvor A i izvor B mogu imati različite načine prepoznavanja proizvoda, ali u skladištu podataka postoji samo jedan način prepoznavanja proizvoda. Potrebno je voditi računa o standardizaciji, verifikaciji te raznim transformacijama prilikom implementacije skladišta podataka.
- Vremenski obilježeno (engl. *time-variant*) – Povijesni podaci se čuvaju u skladištu podataka. Na primjer, može dohvatiti podatke iz 3 mjeseca, 6 mjeseci, 12 mjeseci ili čak i starijih podataka iz skladišta podataka. To je u suprotnosti s transakcijskim sustavom, gdje se čuvaju samo najnoviji podaci. Na primjer, sustav transakcija može imati najnoviju adresu klijenta, gdje skladište podataka može sadržavati sve adrese povezane s kupcem.
- Nepromjenjivo/neizbrisivo (engl. *nonvolatile*) – Nepromjenjivost i neizbrisivost je naglašena s obzirom na to da podaci ulaze u statičnom obliku iz različitih izvora u skladište podataka. Nakon što se podaci nađu u skladištu podataka, neće se više izmjenjivati, stoga se tako mogu čuvati povijesni podaci kojih u izvorima podataka više nema.

Skladišta podataka pružaju opće i konsolidirane podatke u višedimenzionalnom pogledu. Uz generaliziranu i konsolidiranu sliku podataka, skladišta podataka pružaju

i alate za online analitičku obradu podataka (engl. *OLAP – Online analytical processing*¹). Ti alati pomažu nam u interaktivnoj i učinkovitoj analizi podataka u višedimenzionalnom prostoru (Ćurko i Španić, 2016).

Skladište podataka je baza podataka koja se čuva odvojeno od operativne baze podataka organizacije te u njemu nema čestog ažuriranja. Posjeduje konsolidirane povijesne podatke koji pomažu organizaciji da analizira svoje poslovanje. Ono pomaže rukovoditeljima da razumiju i koriste svoje podatke kako bi donijeli strateške odluke. Sustavi skladišta podataka olakšavaju integraciju različitih aplikacijskih sustava te pomaže u konsolidiranoj povijesnoj analizi podataka.

3.2 Arhitektura skladišta podataka

Svako skladište podataka jedinstveno je jer se mora prilagoditi potrebama poslovnih korisnika u različitim funkcionalnim područjima, čije se tvrtke suočavaju s različitim uvjetima poslovanja i konkurentskim pritiscima. Ipak, postoje 3 glavna pristupa izgradnji okruženja za skladištenje podataka. Ove arhitekture općenito se nazivaju: pristup odozgo-dolje (engl. *top-down*), pristup odozdo-gore (engl. *bottom-up*) i hibridni pristup.

Razlike između dva najutjecajnija pristupa izgradnje arhitekture su zagovarani od strane već prije spomenutog Billa Inmona i Ralpa Kimballa. Obojica su plodni autori i konzultanti u polju skladišta podataka. Inmon kojem se pripisuje osmišljenje pojma „skladišta podataka“, zagovara pristup odozgo-dolje a Ralph Kimball je zagovornik pristupa odozdo-gore.

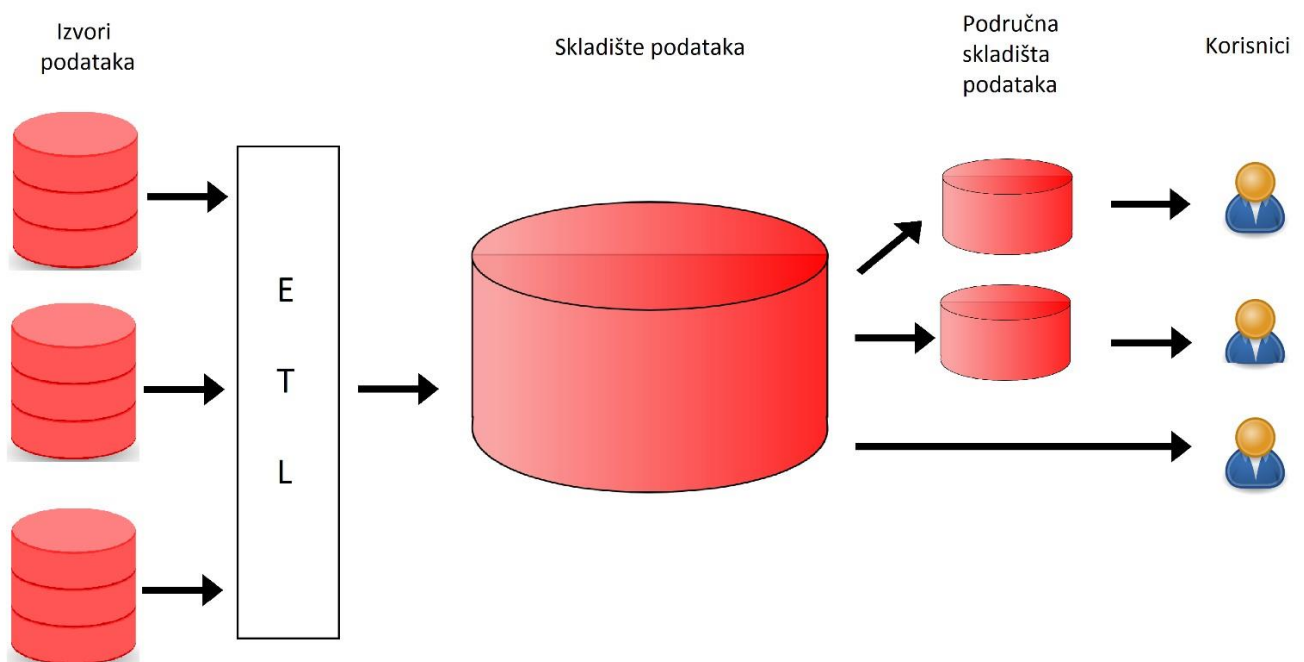
3.2.1 Pristup odozgo-dolje

Pristup odozgo-dolje gleda na skladište podataka kao ključni element cjelokupnog analitičkog okruženja. Prvo se iz izvora podataka dohvate i transformiraju svi podaci te se nakon toga spremaju u skladište podataka. Od tuda su podaci su sažeti,

¹ OLAP - računalna obrada koja korisniku omogućuje jednostavno i selektivno izdvajanje i pregled podataka s različitih gledišta.

dimenzionirani i distribuirani u jedan ili više ovisnih područnih skladišta podataka (engl. Data Marts). Područno skladište podataka je podskup ili agregacija podataka pohranjenih u primarno skladište podataka. Uključuje skup informacija koji su relevantni za određeno poslovno područje, korporacijski odjel ili kategoriju korisnika. Područna skladišta često se nazivaju nesamostalnim ili ovisnim jer izvode sve svoje podatke iz centraliziranog skladišta podataka. Iako područna skladišta podataka nisu potrebna, vrlo su korisna za sustave skladišta podataka u srednje velikim poduzećima jer (Seiner, 2007):

- Koriste se kao građevni blokovi dok se skladište podataka razvija inkrementalno.
- Označavaju informacije koje zahtijeva određena skupina korisnika kako bi obavili upite.
- Mogu pružiti bolje performanse jer su manji od primarnih skladišta podataka.



Slika 1. Odozgo-dolje pristup

Glavna prednost odozgo-dolje pristupa je da on pruža integriranu, fleksibilnu arhitekturu koja podržava nizvodne strukture analitičkih podataka. To znači da skladište podataka predstavlja polaznu točku za sve podatke, utječući na dosljednost

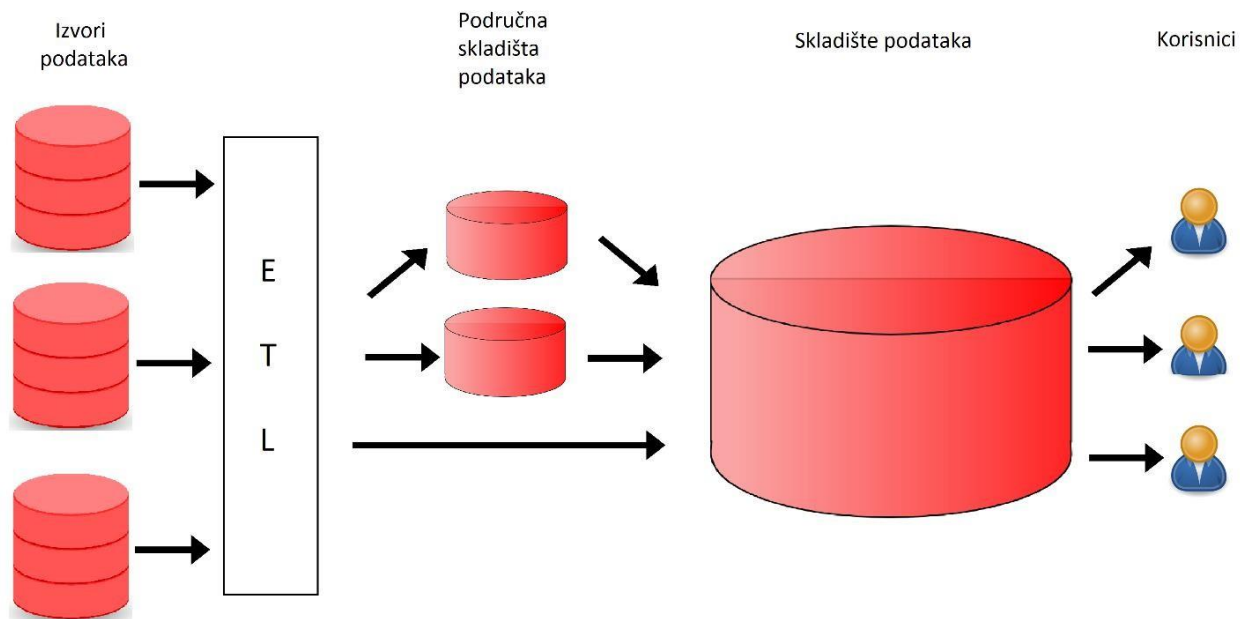
i standardizaciju kako bi organizacije postigle „jedinstvenu verziju istine (engl. *SVOT - Single version of the truth*)“. *SVOT* je tehnički koncept koji opisuje idealno skladištenje podataka koji ima jednu centraliziranu bazu podataka ili barem distribuiranu sinkroniziranu bazu podataka koja pohranjuje sve podatke organizacije u dosljedan i ne-redundantni oblik (Seiner, 2007).

Atomski podaci u skladištu omogućuju organizacijama da ponovno koriste te podatke na bilo koji broj načina kako bi zadovoljili nove i neočekivane poslovne potrebe. Na primjer, skladište podataka može se koristiti za stvaranje bogatih skupova podataka za statističare, isporuku operativnih izvješća i slično. Štoviše, korisnici mogu poslati upite za podatke u skladište podataka ako trebaju višefunkcionalne ili poslovne poglede na podatke.

S druge strane, pristupu odozgo-dolje može trebati duže vremena za implementaciju te implementacija može koštati više od ostalih pristupa, posebno u početnim koracima. To je zbog toga što organizacije moraju stvoriti razumno detaljan model podataka poduzeća, kao i fizičku infrastrukturu koja će obuhvaćati prostor i mjesto skladišta podataka prije razvijanja aplikacija ili izvješća (Seiner, 2007).

3.2.2 Pristup odozdo-gore

U pristupu odozdo-gore cilj je stvoriti skladište podataka koje mora biti brzo i razumljivo. Ovaj pristup dohvaća i transformira sve podatke, ali ih prvo sprema u područna skladišta pa tek onda u centralno skladište podataka. Za razliku od odozgo-dolje pristupa, područna skladišta podataka sadrže atomske i povijesne podatke koje korisnici mogu poželjeti ili trebaju sada ili u budućnosti. Smanjuje redundantnost podataka i olakšava proširivanje postojećih modela dimenzija za smještaj novih područja. Podaci su modelirani zvjezdastom shemom kako bi se optimizirala upotrebljivost i izvedba upita. Zvjezdasta shema će biti malo opširnije objašnjena u zasebnom poglavlju (Seiner, 2007).



Slika 2. Pristup odozdo-gore

Glavna prednost odozdo-gore pristupa jest da se usredotočuje na stvaranje jednostavnih, fleksibilnih struktura podataka pomoću dimenzijskih, zvjezdastih shema. Još jedna prednost jest da, budući da skladišta podataka sadrže i povijesne i atomske podatke, te korisnici ne moraju "kopati" iz područnog skladišta podataka u drugu strukturu kako bi dobili detaljan ili transakcijski podatak.

Jedan od problema ovog pristupa je da zahtjeva od organizacija da nametnu korištenje standardnih dimenzija i činjenica kako bi se osigurala integracija i *SVOT*. Kada su područna skladišta podataka logički raspoređena unutar fizičke baze podataka, lako se postigne integracija. No, u distribuiranoj, decentraliziranoj organizaciji, može biti previše tražiti da se odjeli i poslovne jedinice pridržavaju i ponovno koriste reference i pravila za izračunavanje činjenica. Tu organizacija može imati tendenciju stvaranja "neovisnih" ili neintegriranih skladišta podataka (Seiner, 2007).

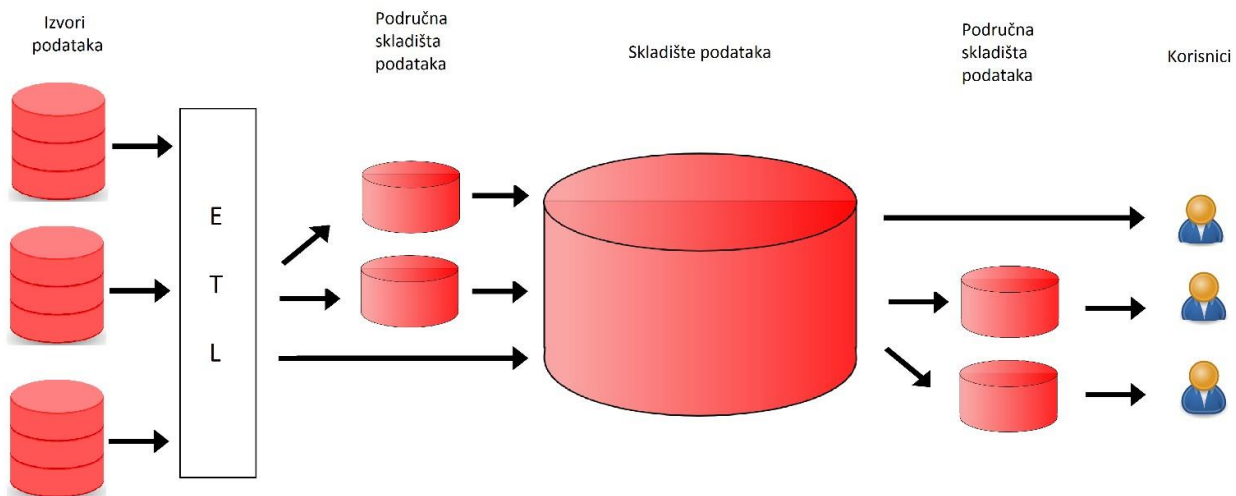
3.2.3 Hibridni pristup

Hibridni pristup pokušava spojiti najbolje od odozgo-dolje i odozdo-gore pristupa. Pokušava iskoristiti brzinu i korisničku orijentaciju odozdo-dolje pristupa bez žrtvovanja integracije koju primjenjuje skladište podataka s odozgo-dolje pristupom. Pieter Mimno, nezavisni konzultant koji predaje na TDWI (*Transforming Data with Intelligence*) konferencijama je trenutno najglasniji zagovornik ovog pristupa (Seiner, 2007).

Hibridni pristup preporučuje da se potroše oko dva tjedna u razvoju poslovnog modela (engl. *Enterprise model*²) u trećem normalnom obliku prije nego što se razvije prvi podatkovni program. Prvih nekoliko područnih skladišta podataka su također dizajnirana u trećem normalnom obliku, ali se koriste pomoću fizičkih modela zvjezdane sheme. Ovaj dvostruki modelni pristup objedinjuje poslovni model bez žrtvovanja upotrebljivosti i performansi izvedbe upita zvjezdane sheme (Seiner, 2007).

Hibridni pristup se oslanja na ETL alat (dohvat, transformiranje, učitavanje – engl. *Extract, Transform, Load*) za pohranu i upravljanje poslovnim i lokalnim modelima podataka u područnim skladištima podataka te sinkroniziranje razlika među njima. To omogućuje primjerice da lokalne skupine razviju vlastite definicije i pravila za elemente podataka koji su izvedeni iz poslovnog modela bez žrtvovanja dugoročne integracije. Nakon implementacije prvih nekoliko područnih skladišta podataka, popunjava se centralno skladište podataka. Zatim se atomski podaci prenose iz područnih skladišta podataka u centralno skladište podataka i konsolidiraju se suvišni podatkovni izvori čime se organizaciji štedi vrijeme, novac i resursi za obradu. Organizacije obično popunjavaju skladište podataka kada korisnici tvrtke zahtijevaju prikaz atomskih podataka u više područnih skladišta podataka (Seiner, 2007).

² Enterprise model - Najviša razina predstavljanja ciljeva organizacije te funkcija, procesa i tokova informacija. Koristi se kao referenca za izradu svih ostalih modela.



Slika 3. Hibridni pristup

Glavna prednost hibridnog pristupa jest da kombinira brzu tehniku razvoja unutar okvira arhitekture poduzeća. Međutim, nadopunjavanje skladišta podataka u ovom pristupu može biti vrlo razoran proces koji ne donosi nikakvu očitu vrijednost pa stoga se može dogoditi da nikada ne bude isplativ. Osim toga, samo nekolicina alata za upite može dinamički i inteligentno pretraživati atomske podatke u jednoj bazi podataka (tj. skladištu podataka) i sumarne podatke u drugoj bazi podataka (tj. skladištu podataka). Korisnici biti zbunjeni jer neće znati kada trebaju slati upite na koju bazu (Seiner, 2007).

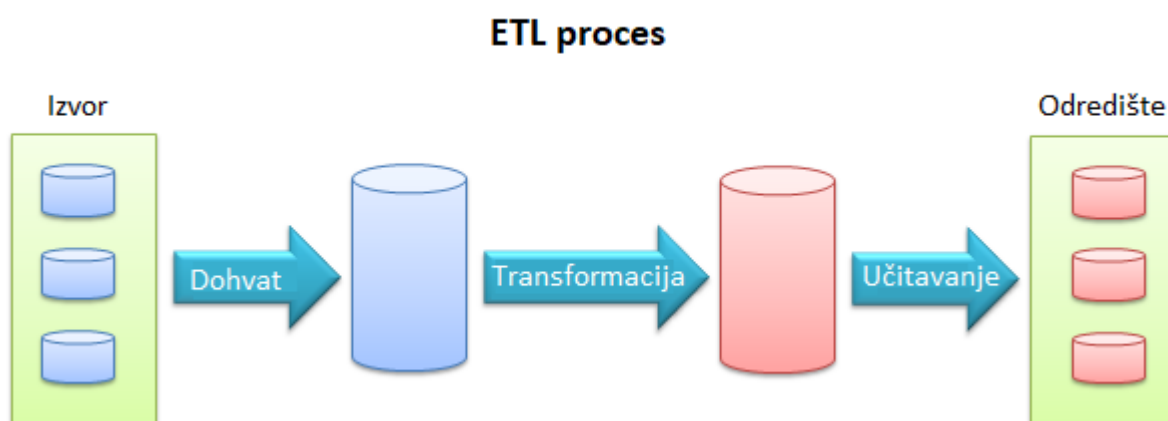
3.3 ETL proces

Procesi dohvata, transformacije i učitavanja podataka (engl. *skraćeno ETL*³) su odgovorni za operacije koje se odvijaju u pozadini arhitekture skladišta podataka. Detaljniji opis ETL procesa bi bio da se podaci dohvaćaju iz raznih izvora podataka koji mogu biti iz transakcijske obrade podataka (engl. *OLTP - Online Transaction Processing*⁴) ili baštinjenih sustava, datoteka u bilo kojem obliku, web stranica i različitih vrsta dokumenata kao što su proračunske tablice i tekstualni dokumenti.

³ ETL – Extract, Transform, Load

⁴ OLTP - klasa informacijskih sustava koji olakšavaju i upravljaju transakcijskim orijentacijskim aplikacijama, obično za unos podataka i obradu transakcija dohvata na sustavu upravljanja bazom podataka.

Obično se dohvaćaju samo podaci koje se razlikuju od prethodnog izvršenja ETL procesa, dakle novo umetnuti, ažurirani ili izbrisani podaci. Nakon ove faze, izvučeni podaci se prenose u prostor posebne namjene u skladištu podataka koji se naziva područje za obradu podataka (engl. *DSA – Data Staging Area*⁵), gdje se odvija njihova transformacija, homogenizacija i čišćenje. Potrebno je posebno voditi računa o potrebnim transformacijama, verifikaciji i čišćenju podataka jer upravo to najviše utječe na uspješno funkcioniranje skladišta podataka. Najčešće korištene transformacije uključuju filtre i provjere kako bi se osiguralo da podaci koji su preneseni u skladište podataka poštuju pravila i ograničenja integriteta, kao i sheme transformacije koje osiguravaju da se podaci uklapaju u shemu skladišta podataka ciljanih podataka. Na samom kraju, podaci se učitavaju u centralno skladište podataka i područna skladišta podataka (Vassiliadis i Simitsis, 2009).



Slika 4. Grafički prikaz ETL procesa (Faizaan, 2015.)

3.3.1 Dohvaćanje podataka

Dohvaćanje podataka je konceptualno najjednostavniji zadatak od svih, njegov cilj je utvrditi ispravni podskup izvornih podataka koji se moraju poslati na daljnju obradu. Ovaj zadatak ima i određenu težinu zbog sljedeća dva ograničenja (Vassiliadis i Simitsis, 2009):

- Izvor podataka tijekom dohвата mora pretrpjeti minimalne troškove budući da se u tom razdoblju obavljaju i druge administrativne aktivnosti.

⁵ DSA - privremeno skladište između izvora podataka i skladišta podataka.

- Zbog tehničkih razloga, administratori su vrlo neodlučni prihvatiti velike intervencije u svojoj konfiguraciji sustava, stoga mora postojati minimalno ometanje konfiguracije softvera na izvornoj strani.

Ovisno o tehnološkoj infrastrukturi i prirodi izvornog sustava (relacijskoj bazi podataka, proračunskoj tablici, web stranici i slično) kao i obujmu podataka koji se trebaju obraditi, mogu se usvojiti različita pravila za korak dohvata podatka. Najviše naivna mogućnost uključuje dohvat cijelog izvora podataka i njegovu obradu kao da se izvršava prvobitno učitavanje skladišta podataka. Bolja mogućnost uključuje dohvat snimke podataka koja se naknadno uspoređuje s prethodnom snimkom (bilo na izvoru ili na strani *DSA*), te se otkrivaju dodavanja, brisanja i ažuriranja. U ovom slučaju, nema potrebe daljnje obrade podataka koji ostaju isti. Druga mogućnost uključuje upotrebu okidača (engl. *trigger*⁶) kod izvora podataka koji se aktiviraju svaki put kad se odvija izmjena u izvornoj bazi podataka. Jasno, to se može učiniti samo ako je izvorna baza podataka relacijski sustav. Završna točka u koraku dohvata podataka uključuje nužnost šifriranja i komprimiranja podataka koji se prenose iz izvora u skladište zbog sigurnosti i izvedbenih razloga mreže (Vassiliadis i Simitsis, 2009).

3.3.2 Transformacija podataka

Ovisno o primjeni i korištenom alatu, ETL procesi mogu sadržavati puno transformacija. Općenito, poslovi transformacije i čišćenja se bave klasama sukoba i problemima koji se mogu podijeliti u dvije razine a to su shema i razina instance. U nastavku će biti prikazana šira klasifikacija problema koja uključuje i probleme na razini vrijednosti (Vassiliadis i Simitsis, 2009):

Problemi na razini shema – Glavni problemi s obzirom na razinu sheme su imenovanje sukoba, gdje se isto ime koristi za različite objekte (homonime) ili se za isti objekt upotrebljavaju različiti nazivi (sinonimi).

Problemi na razini zapisa - Najčešći problemi na razini zapisa odnose se na dvostruke ili kontradiktorne zapise. Nadalje, pojavljuju se problemi dosljednosti u pogledu preciznosti ili pravovremenosti podataka, budući da se dizajner suočava s problemom integracije skupova podataka s različitim razinama agregacije (npr., prodaja po danu

⁶ Trigger - Pohranjeni program, koji se automatski izvršava ili aktivira kada se dogode neki događaji

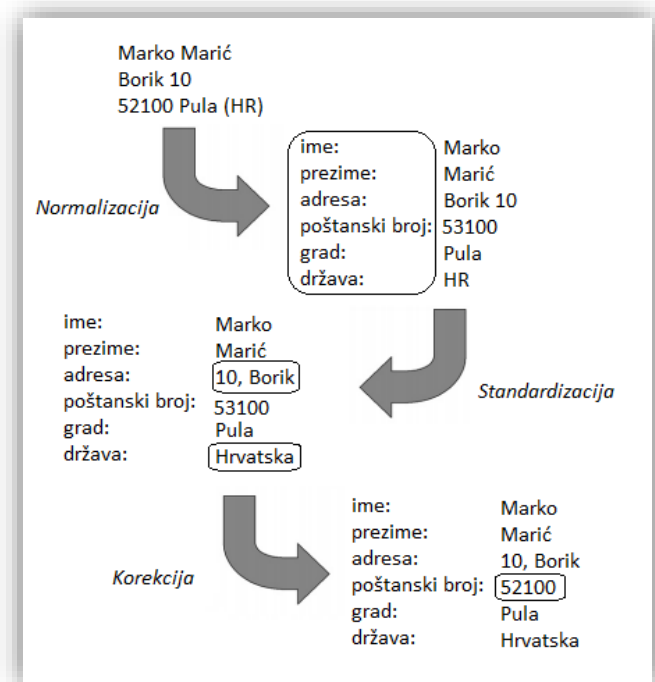
u odnosu na godišnju prodaju) ili upućivanjem na različite vremenske točke (npr., trenutna prodaja od jučer za određeni izvor u odnosu na prošli mjesec za neki drugi izvor).

Problemi na razini vrijednosti – Brojni tehnički problemi niske razine mogu se pojaviti u različitim ETL scenarijima. Primjerice, mogu postojati problemi u primjeni i formata podataka (npr., za spol: 'muški', 'M', '1'), ili različita tumačenja vrijednosti (npr., formati datuma: europski 'dd / mm / gggg' u odnosu na američki 'mm / dd / gggg'). Ostali problemi na razini vrijednosti uključuju dodjelu zamjenskog upravljanja ključem, zamjenom konstanti, postavljanjem vrijednosti na NULL ili DEUFALT na temelju uvjeta ili korištenjem čestih SQL operatora kao što su UPPER (vraća znakovni izraz s malim slovima, pretvorene u velika slova), TRUNC (vraća datum skraćen na određenu mjernu jedinicu) i SUBSTR (vraća dio znaka, binarnog, tekstualnog ili slikovnog izraza u SQL serveru).

Da bi se takvi problemi riješili, zadaci integracije i transformacije uključuju širok raspon funkcija poput normalizacije⁷, denormalizacije⁸, reformatiranja, ponovnog izračuna, spajanja podataka iz više izvora, modifikacije ključnih struktura, sažimanja, dodavanje elemenata vremena, utvrđivanja zadanih vrijednosti i davanje naredbi za odlučivanje između više izvora (Vassiliadis i Simitsis, 2009).

⁷ Normalizacija - proces organiziranja stupaca (atributa) i tablica (odnosa) relacijske baze podataka radi smanjenja redundancije podataka i poboljšanja integriteta podataka.

⁸ Denormalizacija – proces koji se koristi na prethodno normaliziranoj bazi podataka radi povećanja performansi.



Slika 5. Primjer transformacije podataka

3.3.3 Učitavanje podataka

Posljednji korak je učitavanje transformiranih podataka u odgovarajuće tablice u skladištu podataka. Mora se održavati dosljednost podataka jer se zapisi mogu ažurirati prilikom učitavanja. Učitavanje se može provesti na dva načina (Vassiliadis i Simitsis, 2009):

- Osvježavanje (engl. *Refresh*) – Podaci skladišta podataka se u potpunosti prepisuju. To znači da se stari podaci zamjenjuju. Osvježavanje se obično koristi u kombinaciji sa statičkom ekstrakcijom kako bi se skladište podataka popunilo u početku.
- Ažuriranje (engl. *Update*) – U skladištu podataka dodaju se samo one promjene koje su primijenjene na izvornim podacima. Ažuriranje se obično provodi bez brisanja ili mijenjanja postojećih podataka. Ova tehnika se koristi u kombinaciji s inkrementalnom ekstrakcijom kako bi se skladišta podataka redovito ažurirala.

3.4 Dimenzijsko modeliranje

Dimenzijsko modeliranje je široko prihvaćeno kao poželjna tehnika za prikazivanje analitičkih podataka jer se bavi s dva istovremena zahtjeva:

- Isporučuje podatke koji su razumljivi poslovnim korisnicima.
- Isporučuje brzu izvedbu upita.

Dimenzijsko modeliranje je dugogodišnja tehnika za izradu jednostavnih baza podataka. Sa svakim novim slučajem, već više od pet desetljeća, IT organizacije, konzultanti i poslovni korisnici prirodno teže ka jednostavnoj dimenzijskoj strukturi koja odgovara temeljnoj ljudskoj potrebi za jednostavnošću. Jednostavnost je kritična jer osigurava da korisnici mogu lako razumjeti podatke, kao i da softver omogućuje brzu i učinkovitu navigaciju i isporuku rezultata (Kimball i Ross, 2013).

Dimenzijski dizajneri pažljivo slušaju kada je naglasak na proizvodu, tržištu i vremenu. Većina ljudi smatra da je intuitivno razmišljati o takvom poslu kao kocki podataka, s rubovima označenim proizvodom, tržištem i vremenom gdje se kocka rastavlja na manje dijelove. Točke unutar kocke su mjesto gdje se pohranjuju mjere kao što su prodana količina i profit za određenu kombinaciju proizvoda, tržišta i vremena. Sposobnost da se zamisli nešto apstraktno kao što je skup podataka na konkretan i razumljiv način je tajna razumljivosti. Model podataka koji je u početku jednostavan ima veću vjerojatnost ostati jednostavan na kraju dizajna. Model koji je u početku kompliciran će zasigurno biti prekompliciran na kraju, što će rezultirati sporom izvedbom upita i odbijanjem poslovnih korisnika (Kimball i Ross, 2013).

Iako se dimenzijski modeli često instanciraju u sustavima za upravljanje relacijskim bazama podataka, oni su poprilično različiti od modela treće normalne forme (skraćeno 3NF⁹) koji nastoje ukloniti suvišne podatke. Baza podataka o prodajnim nalogima može započeti sa zapisom za svaku liniju narudžbe, ali se može pretvoriti u složeni mrežni web dijagram kao 3NF model, koji može sadržavati stotine normaliziranih tablica.

Industrija se ponekad odnosi na 3NF model kao entitet-veza, *ER model*¹⁰ (engl. *entity-relationship*). *ER* dijagrami su crteži koji pokazuju komunikaciju između tablica. Kako

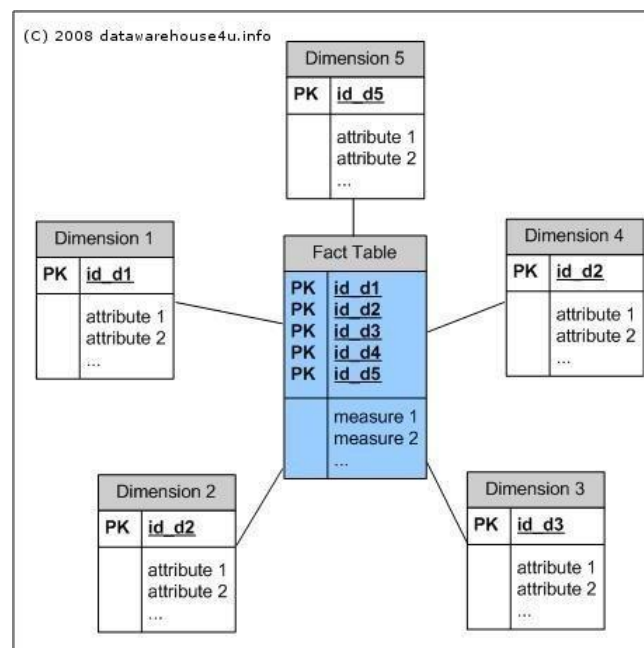
⁹ 3NF - normalna forma koji se koristi za normalizaciju dizajna baze podataka.

¹⁰ ER model - način grafičkog prikazivanja logičkih odnosa entiteta (ili objekata) kako bi se stvorila baza podataka.

3NF, isto tako i dimenzijski modeli mogu prikazati ER dijagrame jer se oboje sastoje od pridruženih relacijskih tablica. Ključna razlika 3NF i dimenzijskih modela je stupanj normalizacije. Normalizirane 3NF strukture su korisne u operativnoj obradi jer transakcije ažuriranja ili umetanja dotiču bazu podataka samo na jednom mjestu. Dimenzijski modeli sadrže iste podatke kao i normalizirani model, ali dimenzijski modeli pakiraju podatke u formatu koji daje korisnicima razumljivost, bolje performanse upita i otpornost na promjene (Kimball i Ross, 2013).

3.4.1 Zvezdana shema

Dimenzijski modeli implementirani u relacijskim sustavim za upravljanje bazama podataka nazivaju se zvezdanim shemama zbog njihove sličnosti sa strukturom zvijezde. Arhitektura sheme zvijezda najjednostavnija shema modela podataka. Zove se zvezdana shema jer dijagram sliči zvijezdi, s vrhovima koje izlaze iz središta. Središte zvijezde sastoji se od činjenične tablice i vrhova zvijezde koji su dimenzijske tablice. Obično se činjenice u tablici zvijezda nalaze u trećem normalnom obliku (engl. 3NF), dok se dimenzijske tablice denormaliziraju. Unatoč činjenici da je shema zvijezde najjednostavnija arhitektura, najčešće se koristi danas i preporučuje ju Oracle (Kimball i Ross, 2013).



Slika 6. Prikaz zvezdaste sheme (Datawarehouse4u, 2008-2009)

3.4.2 Činjenične tablice

Činjenična tablica u dimenzijskom modelu pohranjuje mjerenja performansi koja proizlaze iz događaja poslovnih procesa organizacije. Bitno je nastojati pohraniti niske razine mjernih podataka koji proizlaze iz poslovnog procesa u jednom dimenzijskom modelu. Budući da su mjerni podaci pretežno najveći skupovi podataka, oni ne bi trebali biti replicirani na više mjesta za više organizacijskih funkcija oko poduzeća. Omogućivanjem poslovnih korisnika iz više organizacija da pristupaju jednom centraliziranom spremištu za svaki skup mjernih podataka osigurava korištenje dosljednih podataka u cijelom poduzeću. Izraz činjenica (engl. *fact*) predstavlja poslovnu mjeru. Za sve proizvode koji se prodaju, zapisuju se jedinice količine i cijena u svakoj prodajnoj transakciji, onda se te mjere bilježe dok se proizvodi skeniraju kao što je prikazano na slici 6 (Kimball i Ross, 2013).



Slika 7. Skeniranje proizvoda i prijevod u činjenične tablice (Kimball i Ross, 2013.)

Svaki redak u činjeničnoj tablici odgovara mjernom događaju. Podaci u svakom retku nalaze se na određenoj razini detalja, odnosno zrnatosti detalja. Viša zrnatost znači manje detalja a manja zrnatost više detalja. Jedno od osnovnih načela dimenzijskog modeliranja jest da svi redovi s mjerama budu iste zrnatosti. Imajući mogućnost stvaranja činjeničnih tablica s jednom razinom detalja osigurava da se mjere ne bi nekim slučajem brojile dva puta. Ideja je da mjerni događaj u fizičkom svijetu ima vezu jedan naprema jedan s jednim redom u odgovarajućoj činjeničnoj tablici te je to temelj za dimenzijsko modeliranje. Sve ostalo se gradi iz tog temelja (Kimball i Ross, 2013).

Najkorisnije činjenice su numeričke i aditivne, kao što je iznos prihoda. Aditivnost je presudna jer aplikacije poslovne inteligencije rijetko dohvaćaju samo jedan redak

činjeničnih tablica. Umjesto toga, one dohvaćaju stotine, tisuće ili čak milijune činjeničnih redova u isto vrijeme, a najkorisnija stvar za napraviti s toliko puno redova je dodati ih. Činjenice su ponekad polu-aditivne ili čak neaditivne. Polu-aditivne činjenice, kao što su bilance računa, ne mogu biti sažete kroz dimenziju vremena. Neaditivne činjenice kao što su jedinične cijene, one nikad ne mogu biti dodane (Kimball i Ross, 2013).

Teoretski je moguće da mjerna činjenica bude tekstualna, međutim rijetko se pojavljuje. U većini slučajeva, tekstualne mjere su opis nečega i izvučene su iz nekog diskretnog popisa vrijednosti. Dizajner bi trebao učiniti sve kako bi tekstualne podatke stavio u dimenzije gdje se mogu učinkovitije korelirati s drugim tekstualnim dimenzijskim atributima te potrošiti što manje prostora. Treba izbjegavati pohranjivanje suvišnih tekstualnih podataka, osim ako je tekst jedinstven za svaki redak u činjeničnoj tablici i pripada dimenzijskoj tablici. Pravi tekstualna činjenica je rijetkost jer nepredvidljiv sadržaj teksta činjenice, poput komentara slobodnog teksta, gotovo je nemoguće analizirati (Kimball i Ross, 2013).

Ako ne postoji prodajna aktivnost za određeni proizvod, u tablicu se ne stavlja niti jedan redak. Važno je ne ispunjavati činjeničnu tablicu s nulama koje ne predstavljaju nikakvu aktivnost, jer bi te nule preopteretile većinu činjeničnih tablica. Uključivanjem samo istinskih aktivnosti, činjenične tablice nastoje biti poprilično oskudne. Unatoč njihovoj oskudnosti, činjenične tablice obično čine 90% ili više od ukupnog prostora koji troši dimenzijski model. Činjenične tablice nastoje da budu široke u smislu redaka, ali uske u odnosu na broj stupaca (Kimball i Ross, 2013).

Sve činjenične tablice imaju dva ili više stranih ključeva (oznaka FK¹¹ na slici 7.) koje se povezuju s primarnim ključevima dimenzijskih tablica. Na primjer, ključ nekog proizvoda u činjeničnoj tablici uvijek odgovara određenom ključu proizvoda u dimenzijskoj tablici proizvoda. Kada svi ključevi u činjeničnoj tablici ispravno odgovaraju primarnim ključevima u odgovarajućim dimenzijskim tablicama, tablice zadovoljavaju referencijalni integritet, prema kojemu veze tablica moraju uvijek biti konzistentne. Pristup činjeničnoj tablici se odvija preko dimenzijskih tablica koje su joj pridružene (Kimball i Ross, 2013).

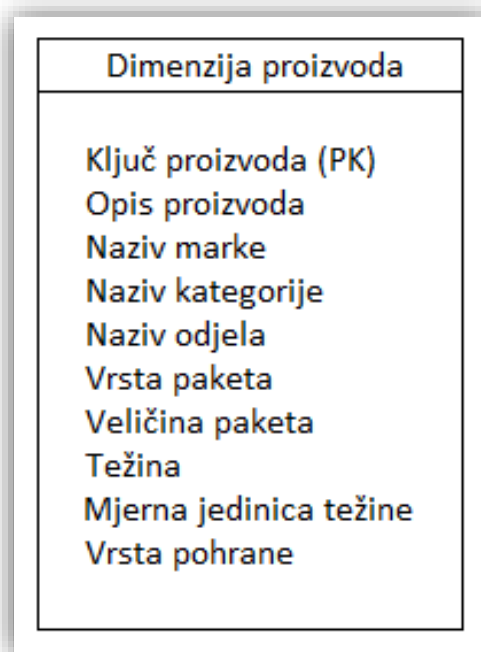
¹¹ FK – engl. Foreign Key – Strani ključ

Činjenična tablica obično ima svoj primarni ključ sastavljen od podskupina stranih ključeva. Taj se ključ često naziva kompozitni ključ te svaka tablica koja ima kompozitni ključ je činjenična tablica. One izražavaju veze više-prema-više.

Uobičajeno ima pregršt dimenzija koje jedinstveno identificiraju svaki redak činjeničnih tablica. Nakon što se identificira taj podskup ukupnog popisa dimenzija, ostatak dimenzija preuzima jednu vrijednost u kontekstu primarnog ključa retka činjenice (Kimball i Ross, 2013).

3.4.3 Dimenzijske tablice

Dimenzijske tablice sadrže tekstualni kontekst povezan s poslovnim procesom mjernog događaja. Opisuju „tko, što, gdje, kada, kako i zašto“ je povezano s događajem. Kao što je prikazano na slici 9., dimenzijske tablice često imaju puno atributa.



Dimenzija proizvoda
Ključ proizvoda (PK)
Opis proizvoda
Naziv marke
Naziv kategorije
Naziv odjela
Vrsta paketa
Veličina paketa
Težina
Mjerna jedinica težine
Vrsta pohrane

Slika 8. Primjer dimenzijske tablice (Kimball i Ross, 2013.)

Nije neuobičajeno da dimenzijska tablica ima 50 ili čak 100 atributa, iako neke po prirodi imaju samo nekoliko atributa. Dimenzijske tablice obično imaju manje redaka nego činjenične tablice, ali mogu biti široke s mnogim velikim tekstnim stupcima. Svaka

dimenzija je definirana jednim primarnim ključem (primjer na slici 8., oznaka PK¹²), koji služi kao osnova za referencijalni integritet s bilo kojom činjeničnom tablicom kojoj je pridružen (Kimball i Ross, 2013).

Dimenzijski atributi služe kao primarni izvor ograničenja upita, grupiranja i oznaka izvješća. U upitu ili zahtjevu za izvješće, atributi se identificiraju kao po riječima. Primjerice, kada korisnik želi vidjeti prihod po određenoj robnoj marki proizvoda, robna marka proizvoda mora biti dostupna kao dimenzijski atribut.

Atributi dimenzijskih tablica igraju vitalnu ulogu u sustavima skladišta podataka i poslovne inteligencije. Oni su bitni jer su izvor gotovo svih ograničenja i oznaka izvješća, te su kritični za izradu sustava skladišta podataka i poslovne inteligencije koji je korisnicima koristan i razumljiv. Atributi bi se trebali sastojati od stvarnih riječi a ne od kriptirane kratice. Treba težiti smanjivanju upotrebe kodova u dimenzijskim tablicama zamjenjujući ih s više opširnim tekstualnim atributima. Preporučuje se napraviti standardno dekodiranje za operativne kodove da budu dostupni kao dimenzijski atributi kako bi se osigurala konzistentna oznaka upita i izvješća. Vrijednosti dekodiranja nikada ne bi smjele biti zatrpane u aplikacijama za izvješćivanje gdje je nedosljednost neizbježna (Kimball i Ross, 2013).

Ponekad operativni kodovi ili identifikatori imaju legitimno poslovno značenje za korisnike ili su potrebni za povratnu komunikaciju u operativnom svijetu. U tim slučajevima, kodovi se trebaju pojaviti kao eksplicitni atributi dimenzija uz odgovarajuće tekstualne deskriptore koji su ugodni za korištenje. Operativni kodovi ponekad imaju ugrađeno značenje u sebi. Primjerice, prve dvije znamenke mogu identificirati poslovnu djelatnost, dok sljedeće dvije mogu identificirati globalnu regiju. Umjesto prisiljavanja korisnika na ispitivanje ili filtriranje podniza (engl. *substring*) unutar operativnih kodova, preporučuje se izvlačenje ugrađenih značenja i njihovo predstavljanje korisnicima kao zasebnih dimenzijskih atributa koji se jednostavno mogu filtrirati, grupirati ili izvijestiti (Kimball i Ross, 2013).

Na mnoge načine, skladište podataka je jednako dobro kao i dimenzijski atributi, analitička snaga okruženja skladišta podataka i poslovne inteligencije izravno je proporcionalna kvaliteti dimenzijskih atributa. Što je više vremena provedeno u

¹² PK – engl. Primary Key – Primarni ključ

pružanju atributa s opsežnom poslovnom terminologijom kao i popunjavanjem vrijednosti domene u stupcima atributa, to bolje.

Pri odabiru operativnih izvora podataka, ponekad je nejasno je li numerički podatkovni element atribut činjenice ili dimenzije. Ta odluka često se rješava tako da se provjeri da li je stupac (Kimball i Ross, 2013):

- mjera koja uzimo puno vrijednosti i sudjeluje u izračunima čineći ga tako činjenicom ili
- diskretno vrijedni opis koji je manje ili više konstantan i sudjeluje u ograničenjima i oznakama redova čineći ga dimenzijskim atributom.

Na primjer, standardna cijena proizvoda izgleda kao konstantan atribut proizvoda ali se može tako često mijenjati da ipak izgleda kao mjerna činjenica.

Dilema dizajnera o tome je li numerički podatkovni element činjenica ili atribut dimenzija, rijetko je problematična odluka. Kontinuirano vrijedna numerička promatranja gotovo su uvijek činjenice, dok diskretna numerička opažanja izvučena iz malog popisa gotovo su uvijek atributi dimenzija (Kimball i Ross, 2013).

Slika 10. prikazuje da dimenzijske tablice često predstavljaju hijerarhijske odnose.

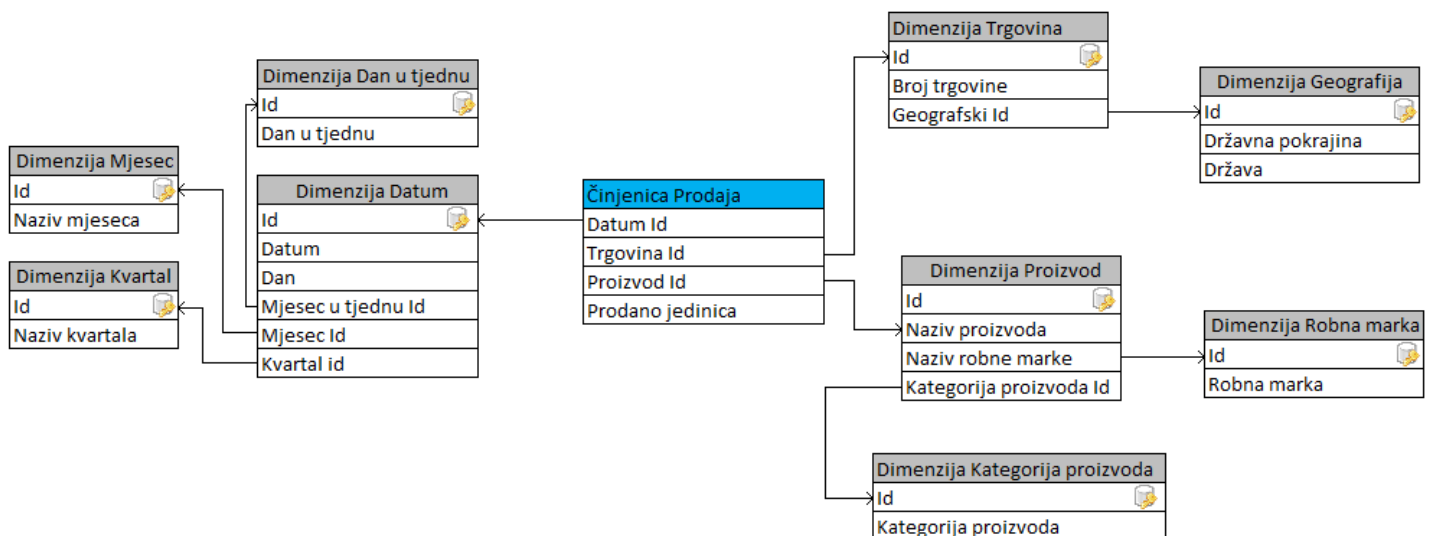
Ključ proizvoda	Naziv proizvoda	Robna marka	Naziv kategorije
1	Mission 24' Boy	X Fact	Brdski bicikl
2	Cross Sport	X Fact	Cross bicikl
3	Chronos LC	KTM	Cross bicikl
4	Peak 27,5'	KTM	Brdski bicikl
5	Benger Comfort	Benger	Treking bicikl
6	Dirt Rider 24	Cygnus	Brdski bicikl
7	Pic Nic Alu	Cygnus	Treking bicikl
8	Life Classic	Scirocco	Treking bicikl
9	Peak XT 29	KTM	Brdski bicikl
10	Jumper 16	Scirocco	Gradski bicikl
11	City Queen 20	Scirocco	Gradski bicikl

Slika 9. Primjer redaka iz tablice dimenzije s denormaliziranim hijerarhijama.

Primjerice, proizvode se uklapaju u robne marke, a zatim u kategorije. Za svaki redak u dimenziji proizvoda je potrebno pohraniti određenu robnu marku i opis kategorije. Hijerarhijske deskriptivne informacije pohranjuju se redundantno s ciljem lakšeg korištenja i boljih performansi upita (Kimball i Ross, 2013).

3.4.4 Pahuljasta shema

Normalizacija podataka pohranjivanjem samo oznake robne marke u dimenziju proizvoda i stvaranjem zasebne tablice za pretraživanje robne marke, a isto tako i za opis kategorije u zasebnoj tablici za pretraživanje kategorije naziva se pahuljiziranje (engl. *snowflaking*). Umjesto treće normalne forme, dimenzijske tablice su obično izrazito denormalizirane s izravnim više-naprema-jedan vezama unutar jedne dimenzijske tablice. Budući da su dimenzijske tablice obično geometrijski manje od činjeničnih tablica, poboljšanje učinkovitosti pohrane normalizacijom ili pahuljiziranjem gotovo da nema utjecaja na ukupnu veličinu baze podataka (Kimball i Ross, 2013).



Slika 10. Primjer pahuljaste sheme

4. Poslovna inteligencija

4.1 Definicija poslovne inteligencije

Poslovna inteligencija je način i metoda poboljšanja poslovnih performansi pružajući veliku pomoć izvršnim donositeljima odluka tako što im omogućuje da imaju djelotvorne informacije. Alati poslovne inteligencije se gledaju kao tehnologija koja omogućuje učinkovitost poslovanja pružajući povećanu vrijednost poduzetničkim informacijama, a time i način korištenja tih informacija. U nastavku slijedi nekoliko različitih definicija poslovne inteligencije (Panian i Klepac, 2003):

Larrisa T. Moss i Shaku Atre (2003.) - Poslovna inteligencija nije ni proizvod niti sustav, već arhitektura koju čini skup povezanih operativnih aplikacija i baza podataka potrebnih za poslovno odlučivanje te koja omogućuje poslovnim korisnicima jednostavan pristup do poslovnih podataka.

Steve Murfitt (2001.) – Poslovna inteligencija je način dostavljanja pravih informacija u pravom formatu u prave ruke u pravom trenutku. Dobar sustav poslovne inteligencije prikuplja informacije iz svih dijelova tvrtke, analizira ih, priprema potrebna izvješća i šalje ih ljudima koji ih trebaju.

Ravi Kalakota, Marcia Robinson (2001.) – Zadatak poslovne inteligencije je pretvaranje podataka u znanje. Poslovna inteligencija je skupina novih aplikacija oblikovanih tako da mogu organizirati i strukturirati podatke o poslovnim transakcijama na način koji omogućuje analizu korisnu u potpori odlučivanju i operativnim aktivnostima kompanije.

Vinod Badami (2003.) – Pojednostavljeno definirano, poslovna inteligencija je proces prikupljanja raspoloživih internih i relevantnih eksternih podataka, te njihove konverzije u korisne informacije koje mogu pomoći poslovnim korisnicima pri donošenju odluka.

Robert Stackowiak – Poslovna inteligencija je proces uzimanja velikih količina podataka, analiziranja tih podataka i prezentacije skupa izvješća visoke razine koji sažimaju suštinu tih podataka u osnove poslovnih aktivnosti, omogućujući menadžerima da donose temeljne dnevne poslovne odluke.

4.1.2 Osnovna svojstva poslovne inteligencije

Na temelju prethodno navedenih definicija mogao bi se steći dojam da poslovna inteligencija teži stvaranju što veće količine podataka i informacija o svim važnim aspektima djelovanja tvrtke za čije se potrebe stvara i razvija. No to nije točno jer s obzirom na količinu generiranih informacija, poslovna inteligencija se temelji na sljedećim temeljnim zamislima (Panian i Klepac, 2003):

- Cilj poslovne inteligencije nije stvaranje što veće količine podataka, već stvaranje kvalitetnijih i boljih informacija koje su potrebne kod donošenja poslovnih odluka.
- Upravo to je ono svojstvo poslovne inteligencije koje joj daje moć poticanja i stvaranja pozitivnih promjena u sredini u kojoj stvara i u kojoj se primjenjuje.
- Ona pruža korisnicima samo one informacije koje su im potrebne, te ih pruža pravovremeno i na način koji korisnicima najviše odgovara.
- Ako se poslovna inteligencija primjenjuje kako treba, njen koncept će smanjiti količinu podataka s kojom se susreću zaposlenici tvrtke te će istovremeno povećati kvalitetu tih podataka.

Osnovne značajke poslovne inteligencije mogu se ubrojiti u sljedeće a to je da se zasniva na personalizaciji, proaktivna je i nastaje iz operativnih podataka. Personalizacija znači primjenu tehnologije u svrhu proaktivnog zahvaćanja, organiziranja i dostavljanja informacija pojedincima. Poslovna inteligencija automatizira dostavu informacija primateljima, koristeći pritom posebne uvjete odstupanja od planiranih aktivnosti kao poticaje za takvu komunikaciju.

Tradicionalni sustavi za podršku odlučivanju (engl. *Decision Support System*¹³) ne personaliziraju informacije, te ih je stoga potrebno u svakoj novoj primjeni mijenjati i poboljšavati. Poslovna inteligencija funkcionira prema modelu koji se sastoji iz pet glavnih komponenata a to su informacije, analize i segmentacije, personalizacije, dostave informacija putem više kanala te akcije, interakcije i/ili transakcije. Tako

¹³ Decision Support System - računalni informacijski sustav koji podržava poslovne ili organizacijske aktivnosti odlučivanja.

ona utire put novim sustavima potpore odlučivanja, odnosno sustavima upravljanja znanjem (engl. *Knowledge Management Systems*¹⁴) (Panian i Klepac, 2003).



Slika 11. Komponente modela poslovne inteligencije (Panian i Klepac, 2003)

Kada su navedene komponente konfigurirane i usklađene, one čine snažnu, proširivu i prilagodljivu platformu poslovne inteligencije. Takve aplikacije pretvaraju tradicionalne sustave potpore odlučivanju (funkcioniraju prema na načelu pitanja i odgovora) u sustave upravljanja znanjem. Personalizacija zahtjeva prilagodbu korisničkog iskustva njegovim potrebama. Iako je došlo do promjene načina dolaženja do informacija i samog komunikacijskog medija, korisnik bi uvijek trebao dobivati podjednako kvalitetnu informaciju. Primjena poslovne inteligencije pretpostavlja da korisnici ne vole dobivati nepotrebne informacije te će zbog toga rado iznijeti svoja mišljenja o tipu informacija koje će primati, učestalosti primanja tih informacija te naravno komunikacijskog medija putem kojega dobivaju informacije (Panian i Klepac, 2003).

4.2 Stilovi poslovne inteligencije

Sama priroda poslovne inteligencije i njeni dometi još uvijek nisu u potpunosti razumljivi te stoga ne postoji ni dovoljno poznavanja i razumijevanja svih oblika u kojima se ona

¹⁴ Knowledge Management System - odnose se na bilo kakav informatički sustav koji pohranjuje i prima znanje, poboljšava suradnju, smješta izvore znanja, spremišta rudnika za skriveno znanje, bilježi i koristi znanje, ili na drugi način pojačava proces upravljanja znanjem.

danas pojavljuje. Funkcionalnosti alata za poslovnu inteligenciju posljednjih godina su evoluirale u nekoliko smjerova te zbog toga postoje razni stilovi.

4.2.1 Poslovno izvještavanje

Poslovno izvještavanje (engl. *Enterprise reporting*) se odnosi na široku lepezu izvještaja raznih formata koji su namijenjeni operativnom poslovnom izvještavanju te poredbenih tablica rezultata poslovanja i poslovnih kontrolnih ploča namijenjenih primjeni od strane korisnika informacija različitih profila.

Poslovno izvještavanje je popularni je model poslovne inteligencije koji proširuje mogućnosti izvještavanja i analize tako što poslovni izvještaji mogu biti namijenjeni svim tipovima korisnika, zaposlenicima tvrtke na svim razinama organizacijske strukture, na svim radnim mjestima, u svim poslovnim funkcijama i službama pa čak i partnerima u vrijednosnom lancu i klijentima tvrtke. Upravo zbog toga, ovo je najčešće primjenjivani stil poslovne inteligencije, koji obuhvaća mnoštvo operativnih izvještaja generiranih izravno iz unutarnjeg informacijskog sustava tvrtke ili njenog sustava za upravljanje odnosima s klijentima (Panian et al., 2007).

Inicijativa za poslovno izvještavanje o poduzeću je veliki pothvat jer različite korisničke skupine trebaju različite informacije u različitim formatima. Stoga platforma za poslovnu inteligenciju koja je odabrana kako bi se olakšalo izvještavanje o poduzećima, mora biti u mogućnosti savladati zahtjeve velikog broj korisnika bez da to utječe na performanse. Ona treba pružiti široku ponudu mogućnosti kao što su ad hoc izvještavanje, dinamična distribucija izvješća i drugih sadržaja poslovne inteligencije, financijsko izvješćivanje i fleksibilno izlazno oblikovanje kako bi podržala najširu paletu strateških, operativnih i analitičkih potreba za izvješćivanjem (Panian et al., 2007).

Prilikom odabira platforme za poslovno izvještavanje, organizacije trebaju potražiti sljedeće značajke i funkcije (Panian et al., 2007):

- Široki pristup podacima tako da se sve informacije u bilo kojem izvoru, uključujući real-time transakcijske sustave, relacijske skladišta podataka, kocke i sve aplikacije planiranja resursa poduzeća (engl. ERP - Enterprise resource planning) ili baštinjene aplikacije mogu iskoristiti za podršku aktivnostima odlučivanja.

- Veliku skalabilnost, tako da u okruženje se može uvesti na stotine, tisuće ili desetke tisuća korisnika.
- Maksimalnu upotrebljivost, tako da i ne-tehički poslovni korisnici mogu lako dohvatiti i komunicirati s podacima poduzeća, bez opsežne obuke.
- Fleksibilan pristup, tako da korisnici mogu pregledavati informacije putem svog uređaja po izboru – osobno računalo, laptop, smartphone ili tablet.

4.2.2 Ad hoc upiti i analize

Ovaj stil poslovne inteligencije predstavlja softver koji korisnicima omogućuje postavljanje praktički neograničenog broja i vrsta upita prema svim raspoloživim podacima i OLAP analizu cjelokupnih baza podataka, sve do razine detaljnih transakcijskih podataka ako je to neophodno. Namijenjen je naprednijim korisnicima kojima trebaju mogućnosti ispitivanja svake moguće kombinacije podataka. Osnovni način osiguranja podrške sustava ad hoc upitima i analizama jest pružiti korisnicima mogućnost kreiranja potpuno novih izvještaja koji će nastati bilo kakvom kombinacijom podataka u izvještaja te se tako otklanja potreba za prethodnim oblikovanjem svih mogućih kombinacija izvještaja (Panian et al., 2007).

Ad hoc upiti i analize se mogu koristiti u bilo kojem području poslovanja i to zahvaljujući sljedećim obilježjima (Panian et al., 2007):

- *Parametarski vođeno izvještavanje uz usmjeravanje analiza.* Ovo obilježje omogućuje korisnicima stvaranje radikalno različitih izvještaja tako što jednostavno odgovaraju na pitanja ili popunjavaju zahtjeve neposredno prije generiranja izvještaja. Sadržaj i izgled izvješća korisnici mogu prilagođavati pomoću određenih parametara. Parametarski vođeno izvještavanje je jako korisno jer omogućuje brze iteracije uz primjenu niza parametara i pohranjivanja rezultata takvih iterativnih obrada za neku buduću uporabu. Kako bi poboljšali svoje analize, korisnici mogu koristiti različite parametre za generiranje istih izvještaja.

- „Svrđlanje (engl. *drill*¹⁵)“ bilo gdje. Korištenjem OLAP funkcionalnosti, korisnici su u mogućnosti „skakati“ po bazi podataka. Mogu „svrdlati prema gore“ (engl. *drill-up*) i „svrdlati prema dolje“ (engl. *drill-down*) po hijerarhijski strukturiranim podacima. Također mogu obavljati raznovrsne unakrsne analize. Svakom novom akcijom „svrdlanja“ može se dinamički stvarati novi izvještaj.
- *OLAP analiza na razini cjelokupne baze podataka*. Korisnici bez ikakvih ograničenja mogu manipulirati izvještajima. Primjer tih manipulacija bi bio da mogu filtrirati podatke koji će biti prikazani u izvještaju koristeći granične vrijednosti ili intervale dopuštenih vrijednosti. Mogu razvrstavati i sortirati vrijednosti u izvještaju prema različitim kriterijima te prelamati (engl. *break*) izvještaje u pojedinačne stranice prema različitim atributima. Omogućeno im je izvoziti podatke iz izvještaja u dokumente nekog drugog tipa.
- *Sofisticirano filtriranje podataka*. Korisnici imaju mogućnost podjele podataka prema različitim poslovnim kriterijima s ciljem dotjerivanja i detaljnijeg prikazivanja određenog skupa podataka.
- *Grupiranje podataka prema definicijama koje daju korisnici*. Korisnicima se omogućuje dotjerivanje poslovnog modela bez potrebe za mijenjanjem ustroja baze podataka ili općeg poslovnog modela. Skladišta podataka se obično organiziraju u strukture koje vjerno odražavaju poslovni model cjelokupnog poduzeća, no ne moraju uvijek dobro odražavati i poslovne potrebe pojedinih odjela, radnih timova i pojedinačnih zaposlenika koji moraju donositi neke specifične odluke. Zbog toga dolazi do poteškoća na koje se nailazi u pokušajima spajanja podataka koji će pogađati takve specifične potrebe u jednom izvještaju. Ovaj stil poslovne inteligencije nudi rješenje tako što omogućuje grupiranje podataka prema samostalnim, neovisnim definicijama korisnika koje po njihovom mišljenju imaju smisla jer im omogućuju provođenje nekih specifičnih analiza.

¹⁵ Drill - odnosi se na bilo koju od različitih operacija i transformacija na tabličnim, relacijskim i višedimenzionalnim podacima.

4.2.3 Rudarenje podataka

Rudarenje podataka je definirano kao dohvaćanje podataka iz ogromnih skupova podataka. Drugim riječima, možemo reći da je rudarenje podataka postupak rudarenja znanja iz podataka. Ovaj stil poslovne inteligencije namijenjen je naprednijim korisnicima koji redovito provode korelaciju i analizu trendova te proizvode specifične projekcije budućih trendova i događaja. Uvid u poslovanje stvoren ovim stilom poslovne inteligencije je ključanje za svako poduzeće. Međutim problem stvaraju pojedini specijalizirani alati za rudarenje podataka zbog teškoće njihovog korištenja, stoga se proizvođači trude učiniti ih što jednostavnijima (Panian et al., 2007).

Informacije ili znanje izvađene ovim stilom se tako mogu koristiti za bilo koju od sljedećih primjena (Tutorialspoint, 2014):

- Analiza tržišta
- Otkrivanje prijevара
- Zadržavanje korisnika
- Kontrola proizvodnje
- Istraživanje znanosti

Rudarenje podataka vrlo je korisno u sljedećim područjima (Tutorialspoint, 2014):

- Analiza tržišta i upravljanje – Koristi se za profiliranje korisnika, odnosno pomaže u određivanju koja vrsta ljudi kupuje koju vrstu proizvoda. Pomaže u prepoznavanju najboljih proizvoda za različite korisnike tako što koristi predviđanja za pronalaženje čimbenika koji mogu privući nove kupce. Također pomaže pronaći skupine korisnika koji dijele iste karakteristike kao što su interesi, navike potrošnje, prihod i slično. Koristi se za određivanje uzorka kupnje kupca, odnosno tipičnog načina kupovine te pruža razna višedimenzionalna sažeta izvješća.
- Korporativna analiza i upravljanje rizicima – Koristi se za planiranje financiranja i procjena imovine, što uključuje analizu i predviđanje novčanog toka. Također se koristi za planiranje resursa što uključuje sažimanje i uspoređivanje resursa

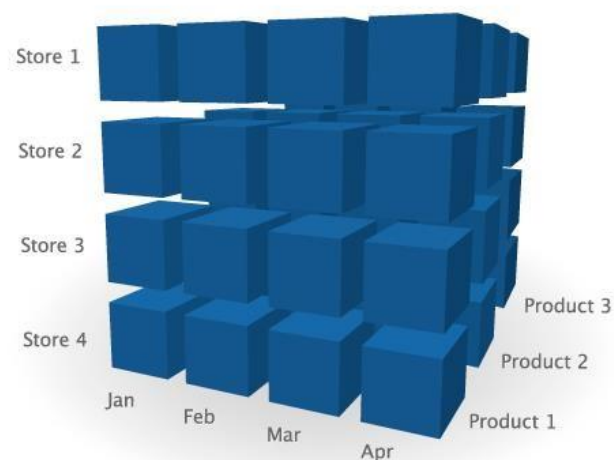
i potrošnje te analizu konkurencije što uključuje praćenje konkurenata i tržišnih smjernica.

- Otkrivanje prijevara – Koristi se u području usluga kreditnih kartica i telekomunikacija za otkrivanje prijevara. U telefonskim prijevarama pomaže pronaći odredište poziva, trajanje poziva, doba dana ili tjedna i slično. Također analizira obrasce koji odstupaju od očekivanih normi.

4.2.4. Analize pomoću OLAP kocke

Online analitička obrada (engl. *OLAP - Online Analytical Processing*¹⁶) se koristi za pripremu podataka za analizu. Sustav izravnih analitičkih obrada upravlja velikim količinama povijesnih podataka, uključuje sažimanja, spremanja i upravljanja informacijama i predstavlja najčešći način analiziranja podataka iz skladišta podataka (Mailvaganam, 2007).

Analize pomoću *OLAP* kocke su primjerene za predviđanje i anticipiranje nekih procesa koji bi se mogli dogoditi u bližoj budućnosti. Primjer primjene ovog stila poslovne inteligencije bi bile analize prodaje po određenim proizvodima ili zemljopisnim područjima za određeno vremensko razdoblje. Te analize uvelike mogu pomoći regionalnim voditeljima prodaje ako ih zanimaju neki određeni dijelovi rada njihovih zaposlenika i/ili službe.

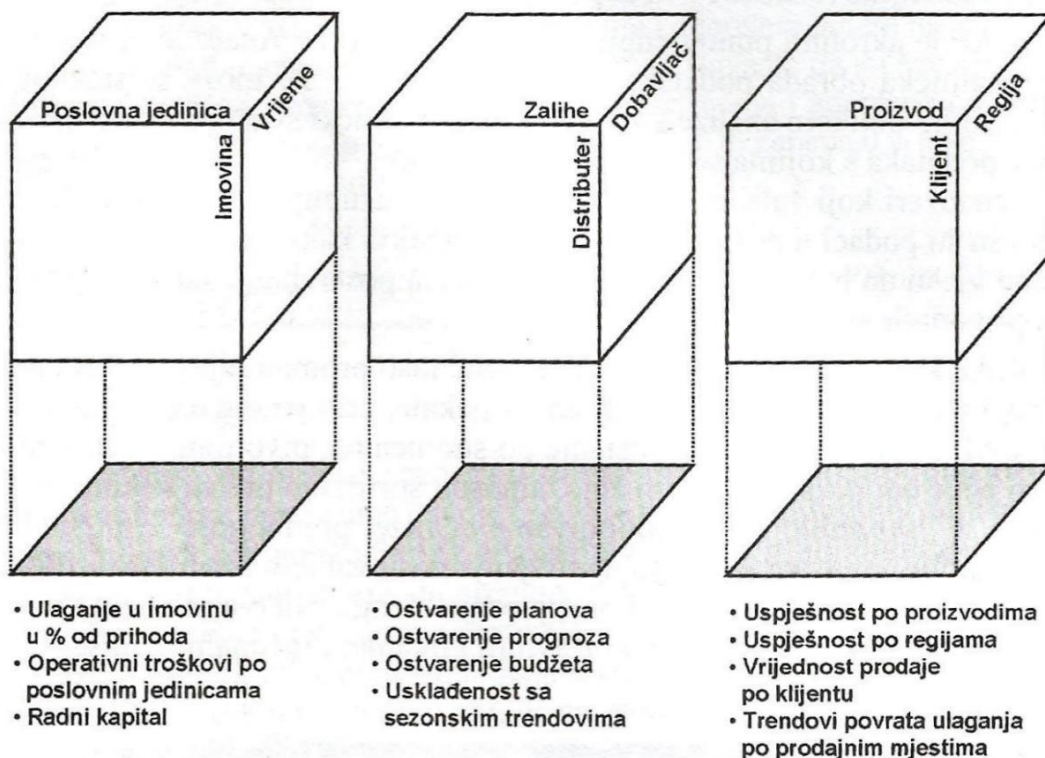


Slika 12. OLAP kocka (Houston, 2011.)

¹⁶OLAP – Online analitička obrada

OLAP kocke se smatraju najjednostavnijim oblikom analize koji omogućuje skoro pa svakom zaposleniku analiziranje podataka s kojima se susreće u svom radu, no njima se najčešće koriste menadžeri koji žele bolje razumjeti osnovne uzorke zašto su podaci u poslovnim izvještajima baš takvi kakvi jesu. Omogućuju korisnicima da promatraju podatke koji su sadržani u poslovnim izvještajima iz različitih aspekata. To je moguće zbog sljedećih *OLAP* obilježja: listanje po stranicama, rotacija oko neke osi kocke, razvrstavanje odnosno sortiranje prema određenom kriteriju, filtriranje i produblivanje analize prema gore ili dole već spomenutim „svrdlanjem“. Te funkcije omogućuju korisnicima „rezanje i presijecanje“ (engl. *Slice and Dice*) kocke podataka (Panian et al., 2007).

Kocke se osnovne komponente *OLAP* sustava te se kocka odnosi na podskup međusobno povezanih logičkih podataka koji su unaprijed organizirani tako da korisnik može kombinirati bilo koje attribute obuhvaćene kockom (primjer proizvodi, prodavaonice, dobavljači, skladišta i slično) s pokazateljima uspješnosti poslovanja koji su također obuhvaćeni kockom (primjer prodaja, dobit, starost i slično). Tako korisnik kreira različite dvodimenzionalne isječke koji se mogu prikazati na zaslonu računala (Panian et al., 2007).



Slika 13. Različite analize pomoću *OLAP* kocke (Panian et al., 2007)

4.3 Prednosti uvođenja poslovne inteligencije u poslovanje

Poslovna inteligencija pruža mnoge pogodnosti tvrtkama koje ju koriste. Može eliminirati mnoga nagađanja unutar organizacije, poboljšati komunikaciju između odjela tijekom koordinacije aktivnosti i omogućiti tvrtkama da brzo reagiraju na promjene u financijskim uvjetima, preferencijama kupaca i operacijama lanca opskrbe. Poslovna inteligencija poboljšava ukupnu učinkovitost tvrtke koja ga koristi.

Najvažnija imovina tvrtke su njezini ljudi, a Informacije se često smatraju drugim najvažnijim resursom tvrtke. Dakle, kada tvrtka može donijeti odluke temeljem pravovremenih i točnih informacija, tvrtka može poboljšati svoju izvedbu. Poslovna inteligencija također ubrzava donošenje odluka i poboljšava korisničko iskustvo, tako što omogućava pravovremeni i odgovarajući odgovor na probleme kupaca.

U nastavku su navedeni neki općeniti razlozi zbog čega je poslovna inteligencija važna (Ranjan, 2005-2009):

- S alatima poslovne inteligencije zaposlenici mogu jednostavno pretvoriti svoje poslovno znanje putem analitičke inteligencije kako bi riješili razne poslovne probleme kao što su povećanje stope odgovora od izravne pošte, e-pošte i internetskih marketinških kampanja.
- S poslovnom inteligencijom, tvrtke mogu identificirati svoje najprofitabilnije kupce i temeljne razloge lojalnosti tih klijenata, kao i identificirati buduće kupce.
- Brzo prepoznaje prijavljene probleme u vezi jamstva kako bi se smanjili učinci nedostatka dizajna proizvoda.
- Omogućuje analizu podataka klikova (engl. *clickstream data*) za poboljšanje e-trgovinske strategije.
- Otkriva kriminalne aktivnosti pranja novca.
- Analizira profitabilnost potencijalnog rasta potrošača i smanjuje izloženost rizicima pomoću točnijih financijskih bodovanja svojih klijenata.
- Pomaže u određivanju koje će se kombinacije proizvoda i usluga korisnici vrlo vjerojatno kupiti i kada.

- Smanjuje prekide rada opreme tako što omogućuje prediktivno održavanje.
- Omogućuje odrediti analizom odstupanja i uklanjanja zbog čega određeni kupci napuštaju konkurente i/ili postanu kupci.
- Otkriva i sprječava neovlašteno ponašanje kao što je korištenje ukradene kreditne ili telefonske kartice.

Klijenti su najkritičniji aspekt uspjeha tvrtke. Bez klijenata tvrtke ne mogu postojati stoga je vrlo važno imati informacije o korisničkim željama. Tvrtke se moraju brzo prilagoditi njihovim promjenjivim zahtjevima. Poslovna inteligencija omogućuje tvrtkama prikupljanje informacija o trendovima na tržištu i pronalaženje inovativnih proizvoda ili usluga u očekivanju promjenjivih zahtjeva kupaca (Ranjan, 2005.-2009.).

Konkurenti mogu biti velika prepreka na putu do uspjeha tvrtke. Njihovi ciljevi kao i ciljevi svih drugih tvrtki je povećanje profita i zadovoljstva kupaca. Kako bi bile uspješne, tvrtke moraju ostati jedan korak ispred konkurenata kako ne bi izgubili dragocjeni tržišni udio. Poslovna inteligencija pokazuje koje akcije poduzimaju konkurenti te tako omogućuje donošenje boljih poslovnih odluka (Ranjan, 2005-2009).

4.4 Najčešći problemi poslovne inteligencije

Za mnoge tvrtke, poslovna inteligencija ne daje očekivane rezultate. Ne daje im jasan pogled na poslovanje te ne nudi trenutne odgovore koje tvrtke očekuju. Nekim tvrtkama se to dešava jer jednostavno imaju prevelika očekivanja i pretpostavljaju da njihova rješenja poslovne inteligencije mogu pružiti rezultate koji su zapravo van njenih mogućnosti.

Jedan od problema bi mogao biti da poslovna inteligencija ne odgovara poslovanju. Sustav poslovne inteligencije je nepotpun i tvrtke pokušavaju raditi s rješenjima koja ne odgovaraju u potpunosti njihovom poslovanju. Nesporazum je jedan od čestih krivaca za to. Često se gube neki detalji u tumačenju ili su zanemareni tijekom faze prikupljanja zahtjeva. Kada IT odjel pokušava implementirati rješenje na temelju nepotpunih zahtjeva, to rješenje nikad neće odgovarati tvrtki. Izgradnja sustava poslovne inteligencije treba biti suradnički projekt između IT-a i poslovanja tvrtke.

Često IT još uvijek izvodi projekte metodom vodopada, koja ne funkcionira u kontinuiranom poslovnom okruženju današnjice (Stangarone, 2015).

Drugi problem poslovne inteligencije bi bio da podaci nisu čisti odnosno nisu kvalitetni. Tvrtka implementira najbolje rješenje za poslovnu inteligenciju i očekuje da će riješiti svoje probleme s podacima, no zanemaruju njihovu kvalitetu podataka. Sustavi poslovne inteligencije ovise o dobroj kvaliteti podataka, bez ikakvog načina da automatski prikupljaju te podatke (Stangarone, 2015).

Sljedeći problem su slijepe točke u podacima. U svakom projektu poslovne inteligencije, treba obratiti pozornost na dvije važne razlike u podacima a to su vrsta podataka koja se prikuplja i količina prikupljenih podataka. Vrsta podataka koja se prikupi razlikuje se ovisno o tome gdje je pohranjena. Ponekad se podaci nalaze u različitim sustavima, ili u oblacima podataka te se često pohranjuju u različitim formatima. Ovo obično nije velik problem jer sustav poslovne inteligencije treba omogućiti ETL sposobnosti za povezivanje tih podataka iz više izvora. Međutim, i dalje treba paziti na razlike u količini podataka koji se prikupljaju u različitim odjelima. Kao što je objašnjeno u nastavku, te razlike mogu stvoriti slijepe točke, što dovodi do neispravne analize. Postoje dijelovi tvrtke gdje se lako prikupljaju velike količine preciznih informacija, primjerice prodajnih mjesta ili računovodstva, dok podaci iz drugih dijelova tvrtke mogu biti teži za određivanje količine, primjerice ljudski resursi. Posljedica toga je da svaka analiza ovih podataka pati od pristranosti odabira podataka, odnosno dolazi do pogrešaka koje uzrokuju da se jedna skupina podataka odabire češće od druge skupine (Stangarone, 2015).

Još jedan od problema bi bio da korisnici ne koriste novi sustav. Korisničko prihvaćanje jedan je od najvećih izazova s kojima se poslovna inteligencija danas suočava. Uobičajeno su sljedeći problemi kod prihvaćanja (Stangarone, 2015):

- Korisnicima se ne sviđa rješenje – Ponekad se sustav poslovne inteligencije odabire bez prisustva korisnika. Kada je sustav konačno implementiran, korisnici ga smatraju zbunjujućim ili da ne zadovoljava njihove potrebe. Kada se to dogodi, vjerojatno će se vratiti na stari način rada.
- Korisnici nisu valjano osposobljeni – U drugim slučajevima, sustav poslovne inteligencije se implementira bez odgovarajuće obuke korisnika. Ako korisnici ne razumiju alat, i kako će im pomoći, neće ga upotrebljavati.

- Korisnici ne žele promjenu – Jedan od vrlo čestih problema koji se pojavljuje gotovo u svim organizacijama. Mala grupa ljudi koji se suprotstavljaju svakoj novoj tehnologiji jer misli da stari način radi sasvim dobro.

5. Zaključak

Poslovna inteligencija i poslovna znanja od presudne su važnosti za uspješno upravljanje i ostvarenje prednosti nad konkurentima. Potrebno je imati pravodobne i kvalitetne podatke transformirane u obliku koji je pogodan za dobivanje potrebnih informacija kako bi se mogle donositi ispravne poslovne odluke a te podatke sadrže skladišta podataka.

Volumen podataka i dalje raste kako se skladišta podataka pune sve većim atomskim podacima i kako se sve učestalije ažuriraju. U početku su baze podataka bile od megabajta pa su se dalje povećavale na gigabajte, terabajte i petabajte, ali osnovni izazov skladišta podataka i poslovne inteligencije sustava ostao je izuzetno konstantan. Njihov posao je zapisivanje podataka organizacije i pravovremeno donošenje tih podataka poslovnim korisnicima.

Dimenzijsko modeliranje je široko prihvaćeno kao dominantna tehnika za prezentaciju skladišta podataka i poslovne inteligencije. Kako bi prezentacija podataka bila uspješna, stručnjaci i korisnici su uvidjeli da ona mora biti utemeljena na jednostavnosti. Jednostavnost je temeljni ključ koji omogućuje korisnicima da lako razumiju baze podataka i softver kako bi se mogli učinkovito kretati bazama podataka. Dimenzijsko modeliranje na puno načina omogućava jednostavnost.

Jedno od važnijih sredstava bilo koje organizacije je njegova informacija. Ona se gotovo uvijek koristi u dvije svrhe a to su operativno vođenje evidencija i analitičko odlučivanje. Jednostavno rečeno, operacijski sustavi su mjesto na koje se postavljaju podaci a sustavi skladišta podataka i poslovne inteligencije su mjesto gdje se podaci dobivaju.

Poslovna inteligencija temelji se na skladištenju podataka te je poslovni i korisnički pristup, a skladištenje podataka tehnološki je pristup dohvaćanja podataka, različit prikaz tih podataka kako bi se dobile potrebne informacije, omogućila analiza tih podataka i steklo poslovno znanje i poslovna inteligencija.

Mnoge uspješne tvrtke ulažu velike iznose novca u alate i tehnologije za poslovnu inteligenciju i skladištenje podataka. Vjeruju da su ažurne, točne i integrirane informacije o njihovom opskrbnom lancu, proizvodima i korisnicima ključne za njihov opstanak.

6. Literatura

1. Ćurko K. i Španić Keznan M. (2016), *Skladištenje podataka: put do znanja i poslovne inteligencije*, Sveučilište u Zagrebu, Ekonomski fakultet
2. Datawarehouse4u (2008-2009), *Star schema*, <raspoloživo na: <http://datawarehouse4u.info/Data-warehouse-schema-architecture-star-schema.html>>, [pristupljeno: 4.8.2017]
3. Houston N. (2011), *Business Intelligence 101 – A Beginner's Guide to BI Software* <raspoloživo na: <http://www.plottingsuccess.com/beginners-guide-to-bi-software-1113011>>, [pristupljeno: 17.8.2017]
4. Kimball R. i Ross M. (2013), *The Data Warehouse Toolkit, Third Edition* <raspoloživo na: <http://www.essai.rnu.tn/Ebook/Informatique/The%20Data%20Warehouse%20Toolkit,%203rd%20Edition.pdf>>, [pristupljeno: 2.8.2017]
5. Mailvaganam H. (2007), *Introduction to OLAP*, <raspoloživo na: http://www.dwreview.com/OLAP/Introduction_OLAP.html>, [pristupljeno: 17.8.2017]
6. Panian Ž. i Klepac. G (2003), *Poslovna inteligencija*, Masmedia
7. Paninan Ž. i suradnici (2007) : *Poslovna inteligencija: studije slučajeva iz hrvatske prakse*, Narodne novine
8. Ranjan, J. (2005-2009), *Business intelligence: concepts, components, techniques and benefits*, <raspoloživo na: <http://www.jatit.org/volumes/research-papers/Vol9No1/9Vol9No1.pdf>>, [pristupljeno: 18.8.2017]
9. Seiner R. (2007), *Four Ways to Build a Data Warehouse*, <raspoloživo na: <http://tdan.com/four-ways-to-build-a-data-warehouse/4770>>, [pristupljeno: 9.8.2017]
10. Stangrone J. (2015), *5 problems that create „Unintelligent“ Business Intelligence*, <raspoloživo na: <http://www.mrc-productivity.com/blog/2015/09/5-problems-that-create-unintelligent-business-intelligence/>>, [pristupljeno: 22.8.2017]
11. Tutorialspoint (2014), *DATA MINING – data pattern evaluation*, <raspoloživo na: https://www.tutorialspoint.com/data_mining/data_mining_tutorial.pdf>, [pristupljeno: 10.8.2017]

12. Vassiliadis P. i Simitsis A. (2009) *Extraction, transformation, and loading*, <raspoloživo na: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.450.9541&rep=rep1&type=pdf>>, [pristupljeno: 16.8.2017]
13. Yousuf Faizaan (2015), *ETL (Extract, Transform, and Load) Process & Concept*, <raspoloživo na: <http://blog.appliedinformaticsinc.com/etl-extract-transform-and-load-process-concept/>>, [pristupljeno: 16.8.2017]