

Vizualizacija podataka u sustavu poslovne inteligencije

Blagović, Luka

Master's thesis / Diplomski rad

2016

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

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:137:821985>

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

Download date / Datum preuzimanja: **2024-04-16**



Repository / Repozitorij:

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



Sveučilište Jurja Dobrile u Puli
Fakultet ekonomije i turizma
"Dr. Mijo Mirković"

LUKA BLAGOVIĆ

**VIZUALIZACIJA PODATAKA U SUSTAVU
POSLOVNE INTELIGENCIJE**

Diplomski rad

Pula, 2016.

Sveučilište Jurja Dobrile u Puli
Fakultet ekonomije i turizma
"Dr. Mijo Mirković"

LUKA BLAGOVIĆ

**VIZUALIZACIJA PODATAKA U SUSTAVU
POSLOVNE INTELIGENCIJE**

Diplomski rad

JMBAG: 0303005101, izvanredni student

Studijski smjer: Poslovna informatika

Predmet: Razvoj informacijskog sustava

Znanstveno područje: Društvene znanosti

Znanstvena grana: Poslovna informatika

Znanstveno polje: Ekonomija

Mentor: prof. dr. sc. Vanja Bevanda

Pula, srpanj 2016.



IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisani Luka Blagović, kandidat za magistra poslovne ekonomije ovime izjavljujem da je ovaj Diplomski 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 Diplomskog 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, 4.7.2016. godine



IZJAVA
o korištenju autorskog djela

Ja, Luka Blagović dajem odobrenje Sveučilištu Jurja Dobrile u Puli, kao nositelju prava iskorištavanja, da moj diplomski rad pod nazivom "**Vizualizacija podataka u sustavu poslovne inteligencije**" 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, 4.7.2016. godine

Potpis

Zahvale:

Prije svega, želio bih se zahvaliti svojoj mentorici prof.dr.sc. Vanji Bevanda koja je svojim uputama, savjetima, ali i strpljenjem omogućila izradu ovog Diplomskog rada.

Također, želio bih se zahvaliti tvrtki Coin d.o.o. iz Pule na korištenoj infrastrukturi potrebnoj za izradu ovoga rada, a osobito mom dragom kolegi Dejanu Gambinu bez čijeg znanja i danih smjernica, ovaj rad ne bi bio u mogućnosti izraditi na ovoj razini kvalitete.

Nikako ne želim zaboraviti, i posebno zahvaliti kolegi, prijatelju i mom prvom mentoru Aleksu Ačimoviću, na "hrpi" prenesenog znanja i beskonačnom strpljenju kod mojih nebrojenih upita, kao i cijeloj ekipi FWD grupe iz Pule na čelu sa Tonijem Milovanom na ukazanoj prilici da uđem u svijet web developmenta na visokoj razini i upijem nova znanja.

U konačnici, hvala cijeloj mojoj obitelji koja je uvijek bezrezervno bila uz mene te skupa sa mnom proživljavala sve moje uspone i padove, životne i akademske. Hvala i mojim dragim prijateljima na svojoj pruženoj podršci i silnom bodrenju tokom studiranja.

Veliko vam hvala svima!

Sadržaj

1. Uvod	1
2. Vizualizacija podataka kao potpora upravljanja u PIS-u	2
3. Definicija i pojam poslovne inteligencije	5
4. Podaci u sustavu poslovne inteligencije	6
4.1. Odnos količine i kakvoće podataka	6
4.2. Podaci kao trošak ili podaci kao nova vrijednost	8
5. Obilježja i dijelovi sustava poslovne inteligencije za potporu upravljanju	10
5.1. Faze razvoja sustava poslovne inteligencije	11
6. Skladištenje podataka	16
6.1. Koncept skladišta podataka	16
6.2. Zahtjevi na skladišta podataka	18
6.3. Arhitekture skladišta podataka	19
6.3.1. Dvoslojna arhitektura s jednim zajedničkim skladištem podataka	20
6.3.2. Dvoslojna arhitektura s većim brojem nezavisnih lokalnih spremišta podataka	21
6.3.3. Troslojna arhitektura sustava skladištenja podataka	22
6.4. Pristupi u izgradnji skladišta podataka	23
7. ETL procesi	27
7.1. Izrada osnovne sheme	30
7.1.1. Učitavanje podataka u dimenzijsku tablicu	30
7.1.2. Učitavanje podataka u tablicu činjenica	31
8. OLAP alati	32
8.1. Strukturiranje podataka	32
9. Rudarenje podataka	35
10. Portali poslovne inteligencije	37
11. Vizualizacija podataka	39
11.1. Pojam „Upravljačke ploče“	39
11.2. Rizik loših vizualizacija	41
11.3. Smjernice za dizajniranje kvalitetne upravljačke ploče	42
11.4. Kognitivne smjernice za dizajn upravljačkih ploča	44

12. Izrada sustava poslovne inteligencije za obrazovnu javnu instituciju	48
12.1. Korišteni podaci u sustavu	48
12.2. Izgradnja skladišta podataka	50
12.3. Punjenje dimenzijskih tablica	51
12.4. Punjenje tablica činjenica	54
12.5. Izrada Mondrian sheme	57
12.6. Izrada upravljačke ploče sa preuzetim podacima	59
12.7. Značajke kreiranih grafova za upravljačku ploču	62
12.8. Izrada ulazne stranice i korištenje vizualizacija	65
13. Zaključak	68
Sažetak	69
Abstract	71
Literatura	73

1. Uvod

Obzirom na današnju sve rašireniju potrebu za korištenjem sustava poslovne inteligencije, ovaj će se rad detaljnije baviti jednim njenim segmentom – vizualizacijom podataka, ali će objasniti i cjelokupan proces izrade istog. U radu su navedene teoretske pretpostavke koje predstavljaju znanja za izradu takvog sustava. Uz teoretski dio, rad sadrži i objašnjenja izvedbe malog (po broju podataka) sustava poslovne inteligencije koji je izrađen za potrebe javne visokoškolske obrazovne ustanove.

Cilj rada jest teoretski objasniti te praktično primijeniti postojeća znanja o razvoju sustava poslovne inteligencije u jednoj organizaciji, kroz primjenu pravila o izradi jednog takvog sustava. Hipoteza je da korištenje sustava poslovne inteligencije u organizaciji pospješuje učinkovitost u upravljanju organizacijom, a ovaj rad će tu hipotezu pokušati i potvrditi. Korištena su znanja i teoretske upute za izradu sustava poslovne inteligencije i to kroz – preuzimanje i skladištenje podataka, upotrebu ETL procesa, izrade kocki podataka, izrade upravljačkih ploča te na kraju i same vizualizacije.

Nakon uvoda, u drugom i trećem poglavlju objašnjeni su razlozi zašto bi uopće jedna organizacija gradila sustav poslovne inteligencije te koji su njeni benefiti. S tehničke strane, objašnjeni su temeljni pojmovi vezani za poslovnu inteligenciju te kako je ona uopće i nastala. U daljnjem dijelu rada opisana su obilježja sustava poslovne inteligencije, podaci koji se u njemu koriste te faze izrade takvog sustava. Obzirom da izrada sustava započinje preuzimanjem podataka i izgradnjom skladišta podataka, objašnjeni su i načini izgradnje skladišta, njegova obilježja, moguće arhitekture te pojam područnih skladišta podataka koja su korištena i u praktičnom dijelu te su stoga važna za ovaj rad. Uz skladištenje, u potpunosti su opisani ETL procesi, odnosno preuzimanje podataka, transformacija i punjenje skladišta podataka, te izrada OLAP kocki za sustav.

Kao dio cjeline, spomenuti su i objašnjeni portali poslovne inteligencije, te rudarenje podataka. Fokus rada je na pravila u vizualizaciji podataka, odnosno na smjernice za dizajniranje kvalitetnih i vizualno atraktivnih upravljačkih ploča.

U posljednjem dijelu rada, opisan je proces izrade sustava poslovne inteligencije, s postojećim podacima, te izrada vizualizacija kroz upravljačke ploče i navigacijsku stranicu. Sustav je nastao primjenom znanja spomenutih u ovome radu.

2. Vizualizacija podataka kao potpora upravljanju u PIS-u

U današnje vrijeme informacija je moć i jedan od ključnih resursa odlučivanja i upravljanja. Jedna od temeljnih funkcija informacijskog sustava je upravo ta da omogućuje brži i jednostavniji pristup potrebnim i relevantnim informacijama.

"Informacijski sustav je sustav tehničkih i ljudskih resursa koji osiguravaju pohranu, obradu, distribuciju i komunikaciju vezanu za informacije potrebne za dio ili cijelo poduzeće. Poseban oblik informacijskih sustava su poslovni informacijski sustavi, koji osiguravaju informacije potrebne za vođenje poduzeća."¹

"Poslovni informacijski sustavi služe za:

1. prikupljanje, obradu i spremanje podataka i informacija
2. stvaranje bogatih informacijskih baza, a u ovom slučaju radi se o povijesnim podacima pa se samim time koristi skladište podataka
3. omogućavanje brzog i jednostavnog pristupa navedenim bazama, u ovom slučaju kroz upravljačke ploče
4. pružanje informacijskih usluga svakom zainteresiranom korisniku
5. informatičku i informacijsku korisničku obradu podataka
6. poboljšavanje i razvijanje informacijske djelatnosti i informacijskog obrazovanja i kulture
7. uključivanje u informacijsko-komunikacijske mreže, osobito Internet, doprinos njihovu razvoju i maksimalno korištenje njihovih resursa."²

Kako bi stvorili uvjete za bolje upravljanje, razvijaju se i sustavi poslovne inteligencije, koji imaju za cilj što je moguće bolje zadovoljiti potrebe njihovih korisnika. Rukovoditelji koriste razne metode za mjerenje uspješnosti poslovanja. Najčešće spominjane jesu Balanced scorecard i Six sigma. Da bi nakon planiranog perioda bili u mogućnosti što brže i jednostavnije dobiti ispravne rezultate na ispravan način, podatke je potrebno strukturirati i servirati na način koji je jednostavan za korištenje, te ima tendenciju uštede vremena rukovodstva.

Razvoj ovakvih sustava započinje definiranjem korisničkih zahtjeva. Ukoliko su

¹ Margaret Rouse, *IS (Information system or information services)*", <http://whatis.techtarget.com/definition/IS-information-system-or-information-services>, 1.7.2016.

² Željko Panian i suradnici, *Poslovni informacijski sustavi*, Element, 2010., str. 22.

zahtjevi jasno i nedvosmisleno definirani, razvoj menadžerskih sustava je jednostavniji i točniji. Za ovaj je rad najvažnija funkcija upravljanja unutar navedenih sustava, obzirom da se sustavi poslovne inteligencije razvijaju upravo da bi služili kao potpora upravljanju. Metode za mjerenje uspješnosti koje koriste menadžeri zahtijevaju upravljanje temeljeno na znanju, odnosno generiranim podacima. U današnje je vrijeme takva vrsta odlučivanja postala standard, a sustavi poslovne inteligencije kao potpora sve su raširenija pojava. Iako i danas postoji rašireno mišljenje da su sustavi poslovne inteligencije nešto što koriste samo velike i bogate organizacije, kako zbog cijene izrade i održavanja, tako i zbog kompleksnosti, to je danas daleko od istine. Čak i najskuplji i najkompleksniji sustavi, ukoliko su dobro planirani i izvedeni, isplaćuju se u kratkim rokovima, obzirom na sve benefite koje donose funkciji upravljanja.

U današnje vrijeme menadžeri traže jednostavna i prilagodljiva rješenja za pristup informacijama koje trebaju i korištenje tih informacija na način koji žele. "U tom smislu menadžeri shvaćaju i poslovnu inteligenciju. Sa njihovog stajališta, gotovo je potpuno nevažno kako se do nje došlo, nije čak ni značajno kako se ona definira. Bez obzira kako je nastala i što ona zapravo jest, poslovna im je inteligencija potrebna zato da bi mogli na inteligentan način kontrolirati, planirati, usmjeravati i voditi cjelokupno poslovanje tvrtke."³ - ukratko, ona je u današnje vrijeme moćan alat za uspješno **vođenje** organizacije.

Da bi se udovoljilo takvim zahtjevima menadžmenta, suvremeni pristup razvoju rješenja namijenjenih pružanju podrške procesima upravljanja zagovara izgradnju inteligentnih aplikacija spajanjem, odnosno povezivanjem odgovarajućih komponenata i funkcija. Takav pristup, koji u središte pozornosti stavlja poslovne potrebe prema kojima se kroji inteligentno rješenje, stručnjaci danas nazivaju "inteligencijom menadžmenta". Na tragu navedenog, danas se razvijaju alati (kao što je i praktični dio ovoga rada) koji nude jednostavna i troškovno učinkovita rješenja za implementaciju inteligencije usmjerene na poslovanje organizacije, a koja omogućuju korištenje aplikacije bez prethodnih znanja te predstavljaju potporu u stvaranju modela za predstavljanje, simulaciju, upravljanje i sl. poslovanjem tvrtke na fleksibilan, jednostavan i učinkovit način.

³ Željko Panian i suradnici, *Poslovna inteligencija: Studije slučajeva iz hrvatske prakse*, Narodne novine, Zagreb, 2007., str. 33

3. Definicija i pojam poslovne inteligencije

Poslovna se inteligencija kao pojam može definirati na čitav niz načina, ovisno o literaturi koja se proučava. U nastavku je navedeno nekoliko definicija:

"Poslovna inteligencija nije ni proizvod niti sustav. To je arhitektura i kolekcija integriranih operativnih aplikacija i aplikacija za potporu upravljanju te baza podataka koje poslovnoj zajednici omogućuju lak pristup poslovnim podacima.

Poslovna inteligencija, ili inteligencija o poslovanju, posebnost je svake organizacije, baš kao što je slučaj i s politikama i poslovnim pravilima u skladu s kojima se upravlja tvrtkom i vodi poslovna praksa. Tu jedinstvenost treba istražiti kako bi se mogla ostvariti konkurentska prednost."⁴

"Poslovna je inteligencija 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 ljudima koji ih trebaju. Na taj način, svaki pojedinac dobiva informacije "skrojene" prema njegovim potrebama – financijski direktor financijske podatke, direktor prodaje podatke o prodaji, direktor marketinga marketinške podatke itd."⁵

"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."⁶

Obzirom na navedene definicije, lako bi se moglo zaključiti da poslovna inteligencija zapravo teži stvaranju, pohranjivanju i obradi što veće količine informacija o svim važnim, internim i eksternim utjecajima i aspektima djelovanja tvrtke koje se dotiče. To ipak nije slučaj. Pojam i cijela zamisao korištenja poslovne inteligencije zasniva se na nekim temeljnim zamislima:

4 Moss, Larissa T, Atre, Shaku. *Business Intelligence Roadmap*. Addison-Wesley, Boston, 2003., str. 4. i 31.

5 Murfitt, Steve. "Using Business Intelligence", <http://www.digitrends.net/scripts/> (3.4.2016.)

6 Badami, Vinod. "Payback on Business Intelligence", <http://information-management.com/> (3.4.2016.)

- intencija poslovne inteligencije nije stvaranje što većeg broja podataka, već stvaranje korisnijih, kvalitetnijih i, u konačnici, upotrebljivijih informacija pri upravljanju,
- prethodna zamisao je zapravo i bit poslovne inteligencije koja joj daje na značaju u procesima vođenja poslova, i stvaranju okoline u kojoj je jednostavnije i sigurnije donositi važne poslovne upravljačke odluke,
- poslovna inteligencija pruža korisnicima samo one setove informacije koji su im potrebni, ali zato pravovremeno i u točno preciziranom formatu,
- ukoliko se ispravno razvije i integrira, poslovna inteligencija će smanjiti izloženost korisnika (radnika) velikoj količini informacija, od kojih većina uopće ne pridonosi produktivnosti u radu već upravo suprotno, sa povećanjem broja beskorisnih informacija u datom trenutku, produktivnost u radu ima tendenciju pada.

Informacija, kvalitetna i pravovremena, je najvažnija pretpostavka preživljavanja tvrtke na turbulentnom tržištu. To vrijedi za gotovo sve industrijske grane i gospodarske djelatnosti.

Nastanak i razvoj poslovne inteligencije kakve danas znamo započet je 60-ih godina prošlog stoljeća, a preteća mu je bio DSS (engl. Decision Support System), odnosno sustav za potporu odlučivanju. Zbog neučinkovitosti tadašnjih sustava, javila se potreba za razvojem rješenja koje danas predstavlja poslovna inteligencija. Tadašnja se tehnologija bazirala na upotrebi papira i magnetne trake te je bila spora i neučinkovita. Podaci su se često gubili zbog strukturalne prirode medija za pohranu. Razvojem tehnologije, online aplikacije ubrzo su zauzele centralnu poziciju u funkcioniranju poslovnih sustava postavši dio direktne interakcije sa kupcima, a sa takvim aplikacijama, za koje ipak treba spomenuti da su imale pregršt mana, pomalo je stasao i pojam sustava poslovne inteligencije, do današnjeg stupnja u kojem predstavlja prevagu na tržištu i nepresušan izvor znanja.

4. Podaci u sustavu poslovne inteligencije

Ovisno o brojnosti poslovnih događaja i realizacijom poslovnih procesa, količina prikupljenih internih informacija u tvrtki ili organizaciji raste. To, dakako, ne mora značiti da je kvaliteta informacija proporcionalna njihovom broju, već se često događa upravo suprotno. Tvrtka tijekom određenog perioda sakupi određeni broj informacija, a što je taj broj veći, pretpostavlja se da je veći dio tih informacija nekvalitetan, upotrebom zakona velikih brojeva.

Razloga može biti čitav niz, no jedan je razlog gotovo istovjetan svim sustavima koje pohranjuju velike količine podataka – brojnost izvora.

"Brojnost izvora podataka eksponencijalno raste onda kada se tvrtka u svojem djelovanju odluči usmjeriti prema Internetu i konceptu elektroničkog poslovanja."⁷

4.1. Odnos količine i kakvoće podataka

Čak i pod pretpostavkom da tvrtka za svoje poslovanje ne koristi Internet, što je u današnje vrijeme gotovo nezamislivo, izvora ima mnogo (osobni kontakti, telefon, fax i sl.), a uz upotrebu Interneta tu se javljaju daljinski osobni kontakti (telekonferencije), elektronička pošta, instant poruke (chat), mobilna telefonija i sl. U uvjetima ovakve neizbježne nedosljednosti u prikupljanju podataka, najveće opasnosti jesu nepotpunost, dupliciranje i stvaranje zaliha podataka, osobito nepotrebnih ili nepotpunih. Svi ovi faktori utječu na povećanje rizika od nesvjesne upotrebe loših, zastarjelih ili nekvalitetnih podataka pri upravljanju. Neznanje je još uvijek najčešći razlog zašto se kakvoća prikupljenih podataka često zanemaruje, a na razvoju sustava poslovne inteligencije štedi. Menadžment je odgovoran za poduzimanje konkretnih koraka da bi se vrijednost prikupljenih podataka pretvorila u poslovnu korist i prednost.

Da bi se stvorila jasna slika o kakvoći prikupljenih podataka, potrebno je izvršiti detaljnu analizu svih njihovih aspekata. Podaci pohranjeni u raznim repozitorijima, čine srž poslovne inteligencije. U današnje vrijeme postoji velik broj softverskih alata kojima je svrha analiza kakvoće podataka, njihovog integriteta i sl. Njihovom upotrebom izdvajaju se kvalitetni od nekvalitetnih ili čak pogrešnih podataka te se pristupa izračunu troškova

⁷ Amor, D. *The E-business (R)Evolution*. Hewlett Packard Professional Books/Prentice Hall PTR, Upper Saddle River, 2000., str. 19.

čišćenja postojećih podataka (engl. Data cleansing) i povrata ulaganja (ROI) da bi se podaci doveli na odgovarajuću razinu kakvoće.

Problem kakvoće podataka uvelike je ovisan o obilježjima poslovanja tvrtke i često nije moguće mnogo učiniti po tom pitanju. Ipak, kakvoća podataka se može podijeliti u četiri osnovne kategorije: standardiziranost, podudarnost, verificiranost i proširivost.

U nastavku su navedeni primjeri i značenja kategorija u realnom poslovanju:

1. Standardiziranost

U međuljudskoj se komunikaciji koriste različite fraze za istu misao. I dok je to u vrijeme usmene predaje minoran problem, ako se uopće takvim može nazvati, kada se neki pojam unosi u npr. bazu podataka, fraza koja se koristi itekako dolazi do izražaja. Razlog tome je računalno nerazumijevanje, odnosno nepodudaranje unesenih pojmova, npr. "cijena danas" i "današnja cijena". Čitajući te fraze, one ljudskom mozgu govore istu stvar, ali računalo ne može tu razliku raspoznati jer se fraze ne podudaraju. Tako nastaju tzv. nestandardizirani podaci. Standardizacija takvih podataka je jedan od osnovnih uvjeta stvaranja kvalitetnih baza i skladišta podataka, a sam postupak standardizacije često može biti dug i mukotrpan.

2. Podudarnost

Podudarnost podataka je vrlo čest slučaj kod baza sa korisničkim podacima. Za primjer uzmimo korisnika koji se dva puta na dva različita mjesta pretplatio na određeni katalog ili uslugu. U jednoj je prijavi naveo adresu "Marka Marulića 20", a na drugoj "Marulićeva 20". U praksi to je ista osoba, no on u bazi čini dva entiteta, te će mu, ukoliko se ne vrši redovno ažuriranje baze, na adresu stizati dva istovjetna kataloga.

3. Verificiranost

Ukoliko se podaci o, primjerice, korisnicima usluga neke tvrtke, pribavljaju iz nekog razloga preko određenih službenih legalnih kanala (iz neke državne agencije, pošte, banke i sl.), uvijek postoji šansa da dostavljeni podaci iz nekog razloga nisu točni (korisnik je zaboravio prijaviti promjenu adrese ili prezimena). Stoga se, kada je god to moguće, preporuča usporedba sa nekim verificiranim izvorom takvih podataka (vlastitom bazom). Ukoliko se podaci poklapaju, sa velikom vjerojatnošću se može pretpostaviti da su podaci iz eksternih izvora točni te da korisnik ima sve ispravne podatke u bazi. Ukoliko ne dolazi do

poklapanja, postoji veća mogućnost da će određeni broj, primjerice u slučaju slanja određenih pisama, biti vraćen na pošiljateljevu adresu.

4. Proširivost

Proširivost, jednostavno rečeno, znači da se podaci mogu nadograđivati novosaznalim informacijama. To je osobito korisno u marketingu, gdje se kod klijenta, uz osobne podatke, mogu nadopunjavati i dodavati podaci o geografskoj lokaciji, potrošačkim preferencijama, interesima, kupovnoj moći itd.

4.2. Podaci kao trošak ili podaci kao nova vrijednost

Podaci koji su pohranjeni u memoriju nekog lokalnog ili udaljenog računala, te u praksi ne služe ničemu, odnosno nitko se njima aktivno ne koristi, za tvrtku predstavljaju isključivo trošak. Pod tim se podrazumijeva nabavka i plaćanje održavanja hardvera na kojem su spomenuti podaci pohranjeni. Ukoliko se podaci ne koriste, oni generiraju stalne i sve veće troškove kroz vrijeme. Razlog tome je, uz navedeni hardver, i činjenica da su u tom slučaju jedini korisnici tih podataka informatičari koji moraju o njima brinuti zajedno sa hardverom.

Kroz sustav poslovne inteligencije ti se podaci stavljaju u uporabu te se od dotadašnjih beskorisnih podataka stvara dodana vrijednost i povećava produktivnost, ukoliko se takvi oblikovani podaci isporuče pravovremeno i pravim korisnicima.

Transformacijom navedenih podataka oni dobivaju na svojoj korisnosti. Daljnji uvjet za kapitaliziranje takvih podataka je umreženost tvrtke ukoliko to već nije slučaj te dodjela pristupnih ovlasti. Primjera radi, prodajnom odjelu nisu potrebni marketinški podaci, te će se stoga svakom odjelu otvoriti oni podaci koji su važni za njihovo funkcioniranje i planiranje.

U fazi usvajanja sustava, podaci se sve više otvaraju prema cijeloj organizaciji da bi različiti odjeli mogli promatrati sve podatke iz različitih perspektiva te time lakše razumjeti funkcioniranje drugih odjela. Istovremeno, menadžment dobiva mogućnost uvida u svaki segment poslovanja tvrtke ili organizacije, često prateći rezultate i promjene na dnevnoj bazi.

Usvajanjem ovakvog sustava u svakodnevno poslovanje, to ne mora biti kraj razvoja. U tom je slučaju organizacija u situaciji da može intenzivirati svoju komunikaciju sa klijentima, dobavljačima, konkurentima, stakeholderima, te na taj način prikupljati

eksterne podatke važne za funkcioniranje organizacije. Na taj se način interni podaci dodatno obogaćuju te stvaraju jasniju sliku o poziciji tvrtke, kako u internom poslovanju, tako i odnosima prema vanjskim subjektima.

U konačnici, prikupljeni podaci u sustavu kroz određeni period mogu postati i izvor prihoda. Kroz akumuliranu podatkovnu inteligenciju, unutar podataka je pohranjeno određeno iskustvo, bio to način pristupa naplati, marketinških aktivnosti i sl. koje druga tvrtka koja se susreće sa sličnim preprekama, može vidjeti kao iznimno značajno ulaganje u svoje poslovanje.

5. Obilježja i dijelovi sustava poslovne inteligencije za potporu upravljanju

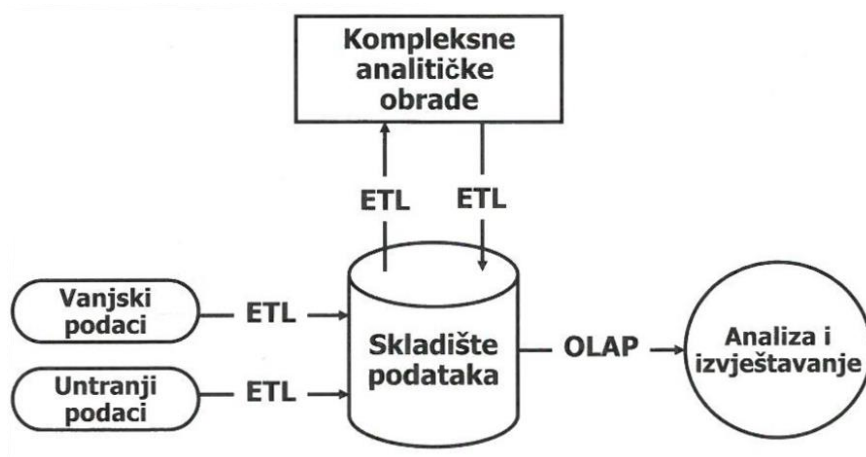
Da bi bili uspješni, sustavi poslovne inteligencije, kako je otprije navedeno, trebali bi zahvaćati unutarnje i vanjske podatke. Dok se unutarnji podaci najčešće kroz automatizirane procedure rutinski prikupljanju i pohranjuju u spremišta podataka (engl. Data Warehouse), vanjske je podatke potrebno podrobnije proučiti, u smislu filtriranja korisnih od nepotrebnih te ih je potrebno, slično kao i kod unutarnjih, kroz specijalizirane procedure, pohraniti u centralnu lokaciju, kako bi se mogli integrirati i zajednički koristiti sa internim podacima tvrtke ili organizacije.

Vanjski se podaci najčešće dijele na podatke o djelatnosti i o konkurenciji.

Najčešći vanjski podaci koje organizacije koriste jesu tzv. granski prosjeci, odnosno prosjeci određenih mjerljivih vrijednosti unutar gospodarske djelatnosti. Do takvih je informacija u pravilu prilično jednostavno doći te ih je lako implementirati u postojeće skladište podataka, ponajviše zbog sličnog ili ponekad identičnog formata podataka kao što sadrže interni. Povezivanje se često vrši raznim šifrarnicima, ukoliko se radi o velikom broju podataka koji se ne mijenjaju u svojoj strukturi kroz godine.

Do eksternih podataka o konkurenciji je znatno teže doći, jer takvi podaci su samo u iznimnim situacijama dostupni, odnosno konkurenti ih drže u tajnosti. Ovakva zbivanja su puno rjeđa kod neprofitnih organizacija, kojima često državne regulative nalažu otvaranje podataka, što je u skladu s načelima EU vezanim za transparentnost u javnom sektoru.

Kao što je spomenuto, unutarnje je podatke jednostavnije pribaviti i implementirati, te gotovo svaka organizacija koja u svom sustavu posjeduje podatke od dvije do nekoliko desetaka povijesnih godina, koristi sustav za praćenje trendova određenih stavki ili poslovanja uopće kroz određeni period. Time se može jednostavnije odrediti stupanj uspješnosti organizacije.



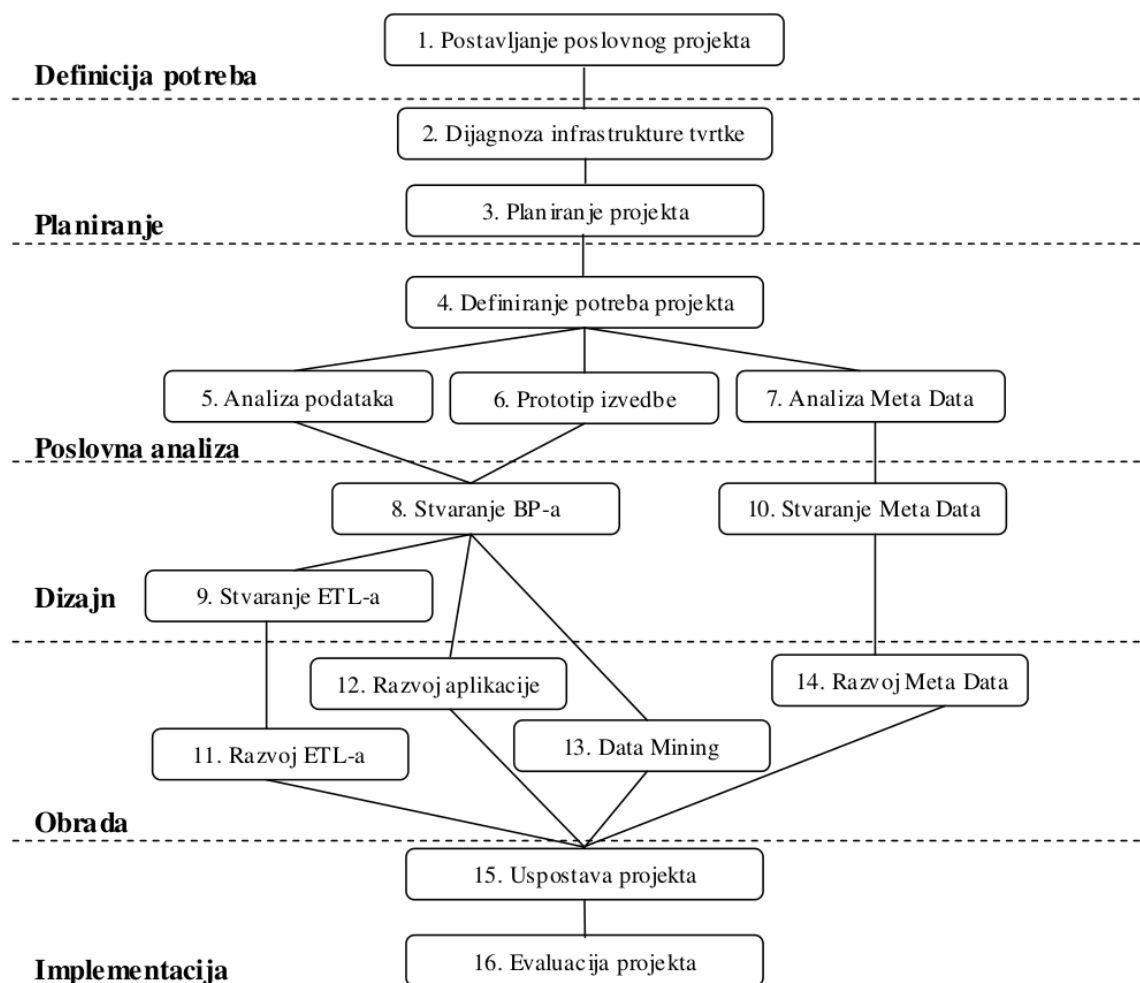
Slika 1: Shema sustava poslovne inteligencije

Izvor: Panian Ž., Klepac G., „Poslovna inteligencija“, Masmedia, Zagreb, 2003., str. 33

S tehnološkog aspekta, sustavi poslovne inteligencije sadrže nekoliko komponenti. U te komponente spadaju spremišta ili skladišta podataka, ETL (engl. Extract/Transform/Load) procedure, OLAP alati i ekspertni sustavi. Sve navedene komponente u svojoj jezgri sadrže podatke, te shodno tome izvori koje koriste moraju biti isti za svaku od navedenih komponenti. To je važno iz razloga što svaka od komponenti može biti specifična i produkt različitih proizvođača. Iz tog razloga, izvori moraju biti isti da bi u konačnici kroz te komponente mogli dobiti najbolje odgovore na pitanja pri donošenju upravljačkih odluka.

5.1. Faze razvoja sustava poslovne inteligencije

Projektiranje sustava poslovne inteligencije često je u raznim izvorima literature organizirano kroz šest razina s ukupno 16 koraka koje su zajedničke svim razvojnim projektima, te koje tumače kompletan životni ciklus aplikacije potpore odlučivanju.



Slika 2. Koraci po fazama razvoja sustava PI

Izvor: Moss L.T. Atre S.,2003, *Business Intelligence road map (na hrvatskom jeziku)*

Razine razvoja:

1. Definicija potreba: definiraju se potrebe ili nužde koje su potaknule menadžera da odobri pothvat razvoja sustava

- *Postavljanje poslovnog projekta* - predlažu se rješenja u vidu sustava poslovne inteligencije, shodno poslovnoj prilici ili eventualnom problemu. Prijedlozi se jasno definiraju, odnosno definira se njihova svrha. To znači da budući sustav može služiti rješavanju problema ili biti orijentiran ka iskorištavanju neke poslovne prilike. Uz definiranje same svrhe, u obzir se obavezno uzima i financijski aspekt razvoja, odnosno njegova isplativost.

2. Planiranje: osmišljavanje i definiranje strateških i operativnih planova te smjernica koje određuju smjer odvijanja razvoja projekta

- *Dijagnoza infrastrukture tvrtke* - detekcija stanja u organizaciji, odnosno određivanje stanja infrastrukture koja bi morala služiti kao potpora poslovnoj inteligenciji. U većini slučajeva dijelovi infrastrukture su već razvijeni i koristili su se puno prije ideje o razvoju poslovne inteligencije. Oni se smatraju osnovom, a ostale komponente infrastrukture se razvijaju skupno sa razvojem BI sustava.

Infrastrukturu dijelimo na dvije komponente:

- a) Tehničke komponente (hardver, softver, middleware, baze podataka, operativni sustavi...)
 - b) Ne-tehničke komponente (standardi meta podataka, kontrole procesa, istraživanja, modeli...)
- *Planiranje projekta* - detaljiziran proces koliko je god to moguće, iz razloga što sustav potpore odlučivanju uključuje čitav niz faktora koji utječu na konačni proizvod (BI sustav), a tu je uključen budžet, tehnologija, zaposlenici, menadžment itd.

3. Poslovna analiza: podrazumijeva analizu poslovnih potreba i mogućnosti kao i detekciju eventualnog problema koji se treba riješiti, radi što bolje detaljizacije proizvoda i lakšeg pronalaženja rješenja.

- *Definicija potreba projekta* - vrlo kompleksan korak koji zahtijeva angažiranost, kako projekatanta, tako i krajnjih korisnika. Radi se o definiranju opsega projekta, pa se podrazumijeva da će kroz proces razvoja projektni timovi imati zahtjeve za promjenama, a krajnji korisnici i stakeholderi bolje razumijevanje o potencijalu, ali i ograničenjima sustava poslovne inteligencije.
- *Analiza podataka* - jedan od najnezahvalnijim dijelova procesa razvoja. Tvrtke često kroz dugogodišnje poslovanje nakupljaju velike količine podataka koje su nepotpune ili nekonzistentne, ne računajući na njihovo eventualno iskorištavanje u budućnosti. Analiza i ispravljanje ("čišćenje") ovakvih podataka je zamoran i često dugotrajan proces, te nije rijetkost da ovaj korak zauzima najveći vremenski udio u cjelokupnom razvoju sustava.
- *Prototip* - kao i za razvoj bilo kakvih sličnih softvera, prototip je sredstvo kojeg

programeri izrađuju radi dobivanja odgovora na pitanje koje alate koristiti, kombinirajući mogućnosti koje imaju sa postavljenim dizajnom sustava. Sa druge strane, korisnicima prototip pruža informacije o mogućnostima i ograničenjima budućeg sustava u praksi te im pruža mogućnost racionalnijih postavljanja zahtjeva sukladno postojećim uvjetima.

- *Analiza Meta podataka* - obzirom da meta podaci (ili kako se još nazivaju "podaci o podacima") nastaju, kako sa tehničke strane upotrebom mnogobrojnih alata, tako i sa poslovne strane. Takva količina podataka procesuirana se određenim softverskim rješenjima, razvijenim samostalno ili kupljenim gotovim licenciranim rješenjima.

4. Dizajn: osmisлити sustav koji će moći riješiti postojeći problem ili stvoriti uvjet za korištenje poslovne prilike

- *Stvaranje baze podataka* - cilj ovakvih baza je pohrana velike količine podataka, što detaljnije i na određeni način grupiranih, ovisno što se želi koristiti u sustavu
- *Stvaranje ETL-a (engl. Extract/Transform/Load)* – odnosi se na preuzimanje podataka iz baze ili vanjskih izvora, te vršenja transformacije podataka u željeni oblik. Puno se vremena može izgubiti ukoliko se koriste izvorni podaci loše kvalitete, zbog potrebe uređivanja i pročišćavanja istih.
- *Stvaranje Meta podataka* - ovisno o odluci o kupnji softverskog rješenja ili izradi vlastitog, meta podaci definiraju način na koji dizajn zadovoljava logički meta model

5. Obrada: kroz zacrtani vremenski period razvija se sustav koji će u konačnici omogućiti i po mogućnosti osigurati povrat uložених sredstava

- *Razvoj ETL-a* - za razvoj procedura, kako jednostavnijih, tako i složenijih, za transformaciju podataka, koriste se različiti alati, ovisno kolika je čistoća podataka, što se utvrđuje pri koraku Analiza podataka, te potvrđuje i rješava u koraku Stvaranje ETL-a.
- *Razvoj aplikacije* - samom razvoju se pristupa u trenutku kada se zadovolje funkcionalni zahtjevi na sustav kroz dizajn prototipa. Samo razvijanje odnosi se na posao programiranja operativnog prototipa, ili kako se još može nazvati – demo. Ovakav prototip se priprema za buduće kompliciranije transformacije i izrade vizualnih prikaza.
- *Rudarenje podataka* - korištenjem specijaliziranih alata za rudarenje po podacima,

uz puno znanja, moguće je detektirati skrivene informacije među podacima (uzorci i sl.), na koje se rijetko obraća pozornost kod klasičnog izvještavanja u BI-u koje je okrenuto unaprijed definiranim rješenjima i projekcijama.

- *Razvoj meta podataka* - samo u slučaju da se tvrtka odluči ne za licencirano rješenje, već izgradnju vlastite pohrane meta podataka. Ovakav pristup najčešće zahtjeva zaduženje posebne grupe zaposlenika koja se bavi samo tom problematikom.

6. Implementacija: radi se o implementaciji proizvoda te praćenju njegove učinkovitosti, kao i vremenskog perioda do povrata sredstava u investiciju

- *Uspostava projekta* - po objavljivanju sustava, a nakon temeljitog testiranja, vrši se obuka korisnika za rad sa sustavom i eventualnom pohranom meta podataka.
- *Evaluacija projekta* - vrši se ocjenjivanje projekta, od vremenskih prekoračenja, prekoračenja troškova i sl. Također, preispituju se korišteni alati, procedure i postupci koji nisu bili korisni, te ih se preispituje ili čak odbacuje.

Važno je napomenuti da se nije nužno striktno pridržavati navedenih koraka, te u stvarnosti to često i nije slučaj. Neki od navedenih koraka su u međuzavisnosti (posloženi vertikalno na skici) te je njihov slijed sasvim logičan. Koraci koji se nalaze na istoj razini najčešće se mogu izvršavati istovremeno.

6. Skladištenje podataka

Poslovna se inteligencija izvodi iz podataka koje nastaju tijekom redovitog poslovanja tvrtke, a obzirom da je tih podataka mnogo i nastaju na različitim mjestima u različitim uvjetima, putem različitih kanala, javlja se potreba da se takva gomila podataka organizira na primjeren način.

Za operativno vođenje poslovanja koriste se transakcijske baze podataka. Specifičnost transakcijskih baza podataka jest u tome da im je glavna zadaća da moraju odražavati najnovije, trenutno stanje poslovnog sustava. Radi se o bazama u kojima se stare vrijednosti trenutačno zamjenjuju novima odmah po unosu. Ukratko, sadržavaju podatke o stanju poslovanja ili određenog aspekta poslovanja u datom trenutku. Za razliku od operativnog poslovanja, u sustavima poslovne inteligencije, brisanje i zamjena starih podataka novima proizvodi obrnuti efekt. Iz povijesnih podataka i praćenja vremenskog redoslijeda zbivanja, kroz poslovnu inteligenciju, zapravo se stvaraju nova znanja i prepoznatljivi uzorci koji mogu doprinijeti funkciji upravljanja. Također, sa aspekta izravnih korisnika takvog sustava, koji su različiti, cijeli je proces decentraliziran te zahtjeva drukčije oblike organiziranja podataka u informacijskim sustavima.

6.1. Koncept skladišta podataka

Prema konceptu i postavljenoj definiciji Williama H. Inmona skladište podataka je "oblik baze podataka kojem su svojstvena sljedeća četiri obilježja:

1. usmjerenost predmetima (funkcionalnim područjima)
2. sadržajna nepromjenjivost
3. integriranost
4. vezanost uz vrijeme (vremenska određenost)"⁸

1. Usmjerenost predmetima

Funkcionalna usmjerenost znači da se podaci organiziraju na način da se slažu oko predmeta, odnosno određenih funkcionalnosti, za razliku od operativnih baza podataka koje su usmjerene aplikacijama, jer su organizirane oko poslovnih aplikacija.

⁸ Inmon, William H. *Building the data warehouse*. John Wiley-QED, New York(NJ), 1992.,str. 9-10

2. Sadržajna nepromjenjivost

Kod operativnih baza podataka postoji vrlo mala vjerojatnost da korisnik sa istim upitom u određenom vremenskom periodu dobije iste rezultate, jer se u vremenu između upita sadržaj baze mijenja, odnosno ažurira kao produkt poslovnih transakcija.

Na suprotan način funkcionira skladište podataka. Kada se podaci jednom pohrane u skladište, uglavnom se ne mijenjaju. To znači da je nevažan broj i učestalost postavljanja istog upita od strane korisnika, jer će rezultat tog upita uvijek biti isti. To je smisao sadržajne nepromjenjivosti.

3. Integriranost

Obzirom da se podaci u transakcijske baze podataka često unose nekonzistentno, ovisno o aplikaciji ili samom autoru podataka, cilj integracije je zapravo unificiranje podataka u smislu postavljanja ispravnog formata ili ispravaka manjih grešaka u unosu, na izlazu iz transakcijske baze podataka i ulazu u skladište podataka.

4. Vremenska određenost

Obzirom da su povijesni podaci vrlo važni za poslovnu inteligenciju, u skladištima podataka se moraju nalaziti i povijesni podaci, za razliku od operativnih baza podataka, koje odražavaju samo trenutno stanje. Jednostavan razlog tome leži u činjenici da se predviđanje budućih događaja i procesa ne može predvidjeti bez poznavanja povijesnih podataka. Ukratko rečeno, skladišta podataka služe da bi predvidjeli budućnost, ali njihov sadržaj odražava prošlost.

Meta podaci ili "podaci o podacima" opisuju informacije ili podatke unutar skladišta, integriraju dolazeće podatke, predstavljaju alat za redefiniranje i ažuriranje određenog modela data warehouse-a. Uključuju opise elemenata podataka, kao što su opisi tipova podataka, opisi atributa, opisi domena, nazivi, veličine i dozvoljene vrijednosti elemenata podataka. Meta podaci strukturiraju informacije u skladištu u kategorije, grupe, hijerarhije i sl. Oni zapravo služe da bi pružili informacije o podacima smještenim u skladište podataka.

Imena tablica i kolona, njihovi detaljni opisi, veze sa poslovnim značenjem, nedavna punjenja, broj ulogiranih korisnika, način indeksiranja – sve su to meta podaci.

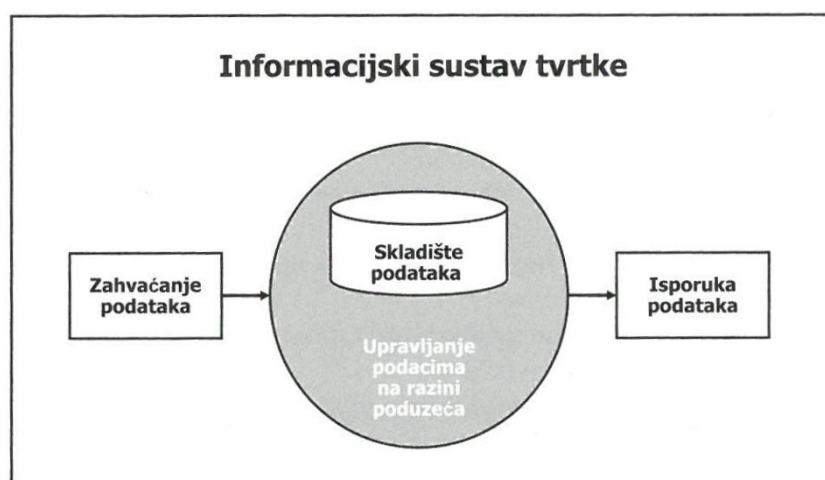
Meta podaci imaju sljedeće karakteristike:

- Predmetno su orijentirani
- Definiraju način na koji će se transformirani podaci interpretirati (npr. 5/10/07 biti će 5. listopada 2007. i sl.)
- Pružaju informacije o srodnim podacima u skladištu podataka
- Čuvaju izračunate vrijednosti i predefinirane formule kako bi se izbjegla loša interpretacija

Sa stanovišta administratora skladišta, meta podaci predstavljaju dokumentaciju o sadržaju i procesima unutar skladišta. Sa druge strane, sa strane korisnika meta podaci predstavljaju mapu za kretanje kroz informacije.

Meta podaci se smatraju "vezivnim tkivom" skladišta podataka i sustava poslovne inteligencije te sadrže sve što je bitno za cijeli proces skladištenja podataka.

6.2. Zahtjevi na skladišta podataka



Slika 3: Mjesto skladišta podataka u informacijskom sustavu tvrtke

Izvor: Panian Ž., Klepac G., „Poslovna inteligencija“, Masmedia, Zagreb, 2003., str. 84

1. U skladištu podataka mora postojati velika količina detaljnih podataka, od telefonskih poziva do bankovnih transakcija. Razlog tome je težnja tvrtke da u bilo kojem trenutku, planiranom ili neplaniranom, te podatke može smisljeno i brzo iščitati te donijeti kvalitetnu upravljačku odluku.
2. Skladište podataka je potrebno kontinuirano ažurirati "svježim" podacima. I dok je to do

prije desetak godina značilo otprilike svakih mjesec dana, u današnje vrijeme elektroničkog poslovanja, i ažuriranja na dnevnoj bazi često se smatraju nedostatnim. Stoga je potrebno, ovisno o potrebama tvrtke, podatke ažurirati što je moguće češće.

3. Skladište podataka mora biti dostupno većem broju ljudi, a ne samo menadžmentu, pojedinom odjelu ili čak pojedincu. Primjer je upravljanje odnosima sa klijentima koje se često vrši kroz više odjela ili organizacijskih jedinica, a ne samo kroz pojedini.

4. Potrebno je oblikovati skladište podataka bez pretjeranog pokušaja predviđanja scenarija korištenja, već na način da ono može poslužiti čak i nepredvidivoj svrsi, ukoliko se stvori takva situacija sa nepredviđenim okolnostima.

5. Dostupnost u svakom trenutku, obzirom da si niti jedna tvrtka ne može dopustiti da takvi podaci u potrebitom trenutku budu nedostupni jer time mogu doći u situaciju da riskiraju pružanje nedostatne ili nekvalitetne usluge klijentu, što posljedično može dovesti do gubljenja klijenata.

6. Skladišta podataka moraju biti proširiva, osobito radi aktivnosti kao što su marketing koje neprestano gomilaju nove setove podataka i zahtijevaju prilično detaljne odgovore na upite. Proširenje je važno i zbog eventualnih odluka o promjenama strategija, čime se npr. tvrtka može naglo proširiti na drugo tržište čime znatno povećava broj podataka koje svakodnevno poslovanje generira.

7. Kao posljednji, ali nikako manje važan zahtjev, odnosi se na što jače mjere zaštite integriteta podataka, a u nekim slučajevima i kompletne tajnosti. Ovaj je zahtjev važan zbog cyber kriminala koji je razvojem tehnologije u stalnom porastu te klijenti često iz straha od krađe podataka, nisu voljni ustupiti podatke u ovakva skladišta. Stoga tvrtke moraju ulagati ponekad i značajna sredstva u sigurnost skladišta podataka.

6.3. Arhitekture skladišta podataka

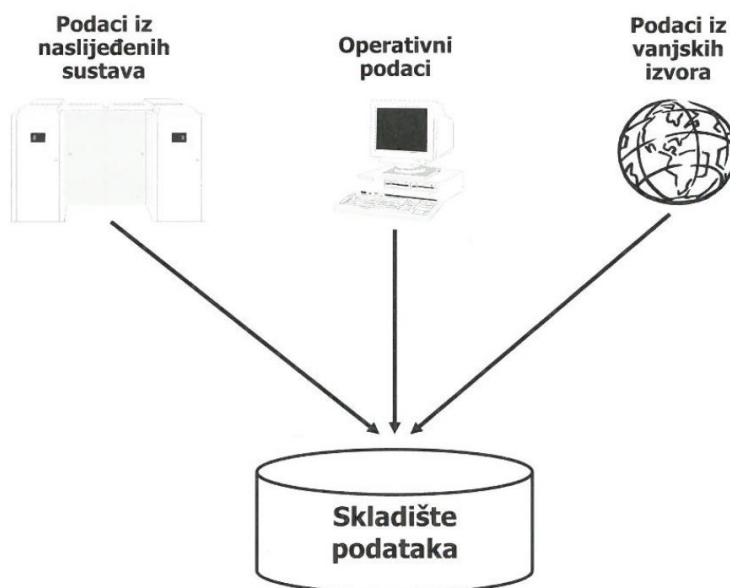
Danas su u upotrebi tri osnovne arhitekture skladišta podataka:

1. Dvoslojna arhitektura s jednim zajedničkim skladištem podataka
2. Dvoslojna arhitektura s većim brojem nezavisnih lokalnih spremišta podataka
3. Troslojna arhitektura s jednim zajedničkim skladištem podataka i većim brojem povezanih lokalnih spremišta podataka

"U praksi se mogu naći dobri primjeri primjene sva tri modela skladišta podataka. Svaki od njih ima određenih kvaliteta koje dolaze do izražaja u određenim konkretnim

poslovnim situacijama i uvjetima. Može se ustvrditi kako u posljednje vrijeme izrazito raste popularnost trećeg modela – onog zasnovanog na troslojnoj arhitekturi s jednim zajedničkim skladištem podataka (Data Warehouse) i većim brojem povezanih lokalnih spremišta podataka (Data Mart). Zato se može s priličnom vjerojatnošću pretpostaviti da je upravo taj model onaj koji će obilježiti praksu skladištenja podataka u bližoj budućnosti."⁹

6.3.1. Dvoslojna arhitektura s jednim zajedničkim skladištem podataka



Slika 4: Dvoslojna arhitektura s jednim zajedničkim skladištem podataka

Izvor: Panian Ž., Klepac G., „Poslovna inteligencija“, Masmedia, Zagreb, 2003., str. 95

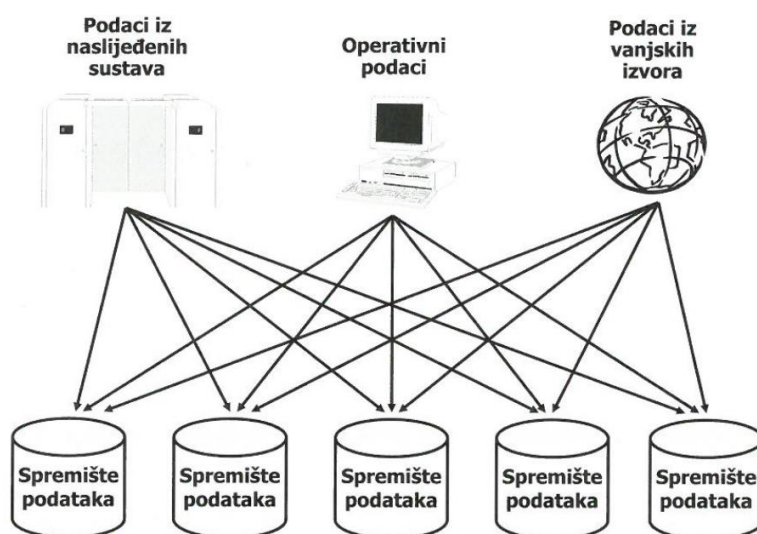
Ovaj se način smatra tradicionalnim, a funkcionira na način da centralno skladište zahvaća i preuzima podatke iz više izvora podataka:

- ranije razvijeni, a potom naslijeđeni sustavi koji održavaju baze podataka nekih ranije korištenih arhitektura
- operativni podaci iz raznih transakcijskih programa implementiranih u poslužitelja
- vanjski izvori podataka, dostupni putem Interneta i drugih komunikacijskih kanala ili mreža

⁹ Panian Ž., Klepac G., „Poslovna inteligencija“, Masmedia, Zagreb, 2003., str. 94.

U ovakvim se arhitekturama skladište velike količine različitih podataka te je stoga potrebno da sheme sadržavaju više aplikacijskih rješenja. To u praksi znači bitno kompliciranje izrade sheme, te iziskuju dosta vremena dizajnerima i programerima. Budući da je ovakva shema vrlo centralizirana i velika, tvrtke često, zbog nemogućnosti izračuna povrata ulaganja, posežu za novijim rješenjima, poput dvoslojne arhitekture s većim brojem lokalnih spremišta podataka, radi decentralizacije arhitekture.

6.3.2. Dvoslojna arhitektura s većim brojem nezavisnih lokalnih spremišta podataka



Slika 5: Dvoslojna arhitektura s većim brojem nezavisnih lokalnih spremišta podataka

Izvor: Panian Ž., Klepac G., „Poslovna inteligencija“, Masmedia, Zagreb, 2003., str. 97.

Ovakva je metoda postala iznimno popularna alternativa centraliziranoj arhitekturi. Temelji se na kombiniranju većeg broja lokalnih spremišta podataka (engl. Data Mart). Primjene ovakvog pristupa nije bez nedostataka. Ovakva lokalna spremišta se najčešće implementiraju da bi podržala određene aplikacije u određenim organizacijskim jedinicama, što dovodi do situacije u kojoj svaki sustav zahvaća specifične podatke iz transakcijskih sustava. Dugoročno, ovakva arhitektura dovodi do znatnijeg povećanja troškova održavanja.

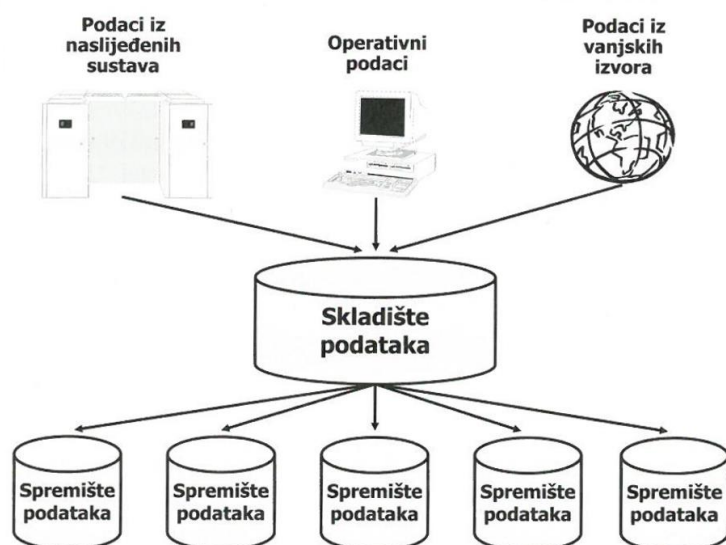
Obzirom na tako "raspršeni" sustav, javljaju se neki problemi svojstveni ovoj arhitekturi:

- "Povećano radno opterećenje
- Ograničena skalabilnost
- Otežana implementacija novih aplikacija

- Poteškoće u komunikaciji među organizacijskim jedinicama tvrtke
- Poteškoće pri održavanju konzistentnosti podataka"¹⁰

Kao i u slučaju centralizirane arhitekture, navedeni nedostaci bili su povod razvoj nove arhitekture zasnovane na tri sloja sa jednim zajedničkim skladištem i više povezanih lokalnih spremišta.

6.3.3. Troslojna arhitektura sustava skladištenja podataka



Slika 6: Troslojna arhitektura sustava skladištenja podataka

Izvor: Panian Ž., Klepac G., „Poslovna inteligencija“, Masmedia, Zagreb, 2003., str. 99.

Nakon što su detektirani svi nedostaci u prethodne dvije vrste arhitekture, kao novo rješenje nametnula se troslojna arhitektura, koja lokalna spremišta podataka povezuje u jedno zajedničko skladište podataka.

U takvo se skladište pohranjuju sve relevantne informacije iz operativnih sustava tvrtke koji se potom isporučuju svim lokalnim spremištima. Takve podatke, potrebne za svaku jedinicu zasebno, preuzimaju operativne aplikacije.

Troslojna arhitektura sadrži neke prednosti u odnosu na prethodno navedene dvije:

- Smanjeno radno opterećenje

¹⁰ Panian Ž., Klepac G., „Poslovna inteligencija“, Masmedia, Zagreb, 2003., str. 98.

- Lakše dodavanje novih aplikacija
- Povećana skalabilnost
- Povećana konzistentnost podataka
- Otklanjaju se poteškoće u komunikaciji među organizacijskim jedinicama

Najveći problem su, dakako, veći troškovi povezivanja takvog zajedničkog skladišta podataka na razini tvrtke. I taj je problem rješiv upotrebom metodologije "inkrementalne izgradnje sustava skladištenja podataka", kojom se ovaj rad neće detaljnije baviti.

6.4. Pristupi u izgradnji skladišta podataka

U procesu nastajanja ili izgradnje skladišta podataka, moguće je primijeniti dva različita pristupa: od vrha prema dolje (engl. Top-down) ili od dna prema gore (engl. Bottom-up).

1. Od vrha prema dolje

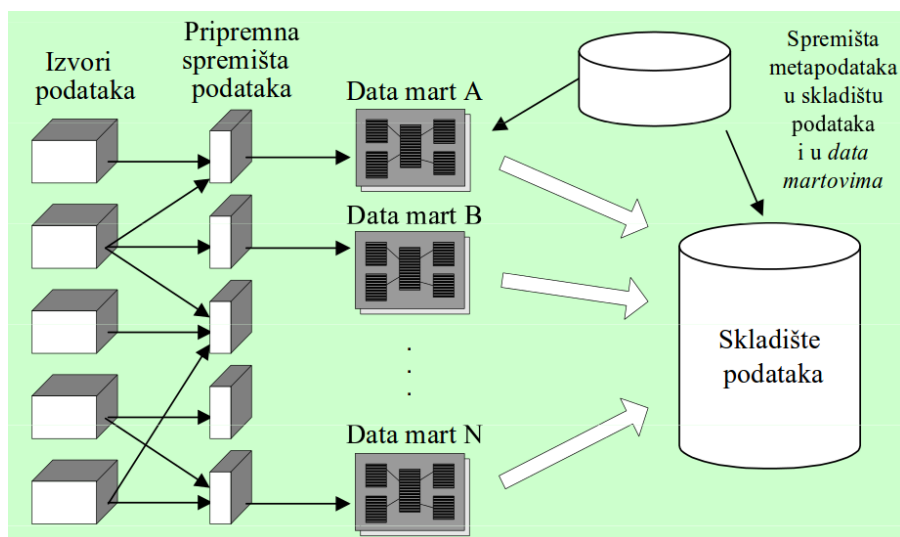
Ovaj pristup je definirao Bill Inmon i podijelio ga u dvije faze:

- Izgradnja centralnog repozitorija - Enterprise data warehouse (EDW) – odlike EDW-a jesu da su podaci normalizirani da bi se izbjegle nekontrolirane redundancije te se omogućava uspostava jedna verzija tablice činjenica (engl. Fact table). Također, podaci se bilježe na što nižoj razini agregacije čime se dobiva više detalja, ali su zato kompleksniji za rukovanje te zahtijevaju više prostora i procesorske snage.
- Izgradnja područnih repozitorija iz EDW-a – podaci se nalaze u denormaliziranoj formi u shemi zvijezde (engl. Star schema) radi lakše i brže uporabe od strane krajnjih korisnika. Podaci su sumirani sukladno zahtjevima korisnika.

Ovakav se pristup često primjenjuje kada su jasno definirani zahtjevi prema budućem skladištu podataka te kada imamo na raspolaganju potrebno vrijeme i novac za modeliranje i implementaciju.

Među prednosti top-down pristupa svakako spada činjenica da se izgrađuje centralizirani repozitorij podataka koji sadrži jednu nepromjenjivu verziju činjenica. Najveći nedostatak ovog pristupa, uz potrošnju značajnih sredstava i vremena, jest i činjenica da vrijeme izgradnje obuhvaća vrijeme izgradnje EDW-a, potom područnih repozitorija te se

tek tada može pristupiti izvješćivanju.



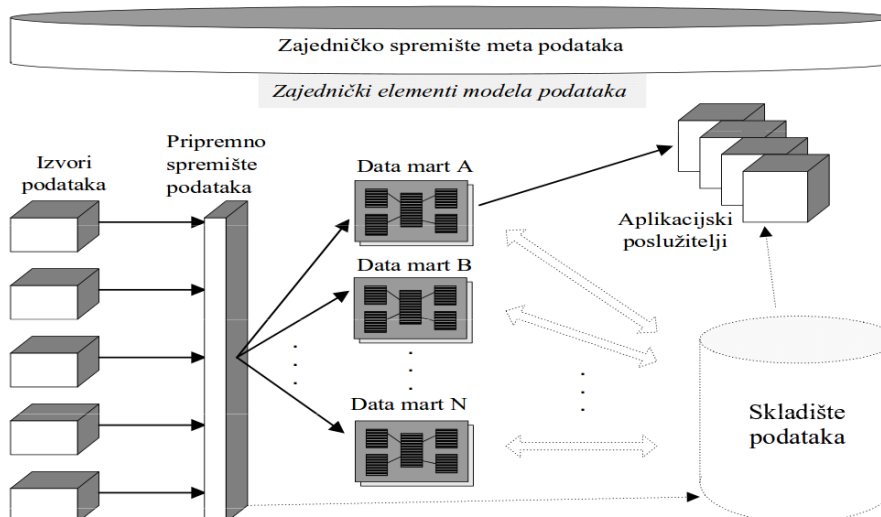
Slika 7: Shema top-down pristupa

Izvor: <https://www.fer.unizg.hr>

2. Od dna prema gore

Pristup je razvio Ralph Kimball i koristi iterativni pristup izgradnji skladišta podataka, te podrazumijeva pojedinačno definiranje područnih repozitorija u različitim vremenskim razdobljima kako i kada se pojavi potreba za takvim podacima. Data martovi se kombiniraju i integriraju radi izgradnje skladišta podataka te se kombiniraju usuglašenim dimenzijama i činjenicama. Te usuglašene dimenzije se prvotno odnose na ključeve dimenzija, nazive atributa i vrijednosti atributa.

Razlika između ovog i top-down pristupa jest da ovakav iterativni pristup podrazumijeva da nije potrebno imati definirane sve potrebe glede skladišta podataka. Bottom-up pristup se primjenjuje kada su inicijalna sredstva i vrijeme ograničeni te kada nema jasne vizije ukupnosti informacijskih zahtjeva prema skladištu podataka.



Slika 8: Shema bottom-up pristupa

Izvor: <https://www.fer.unizg.hr>

Područno skladište podataka (engl. Data mart) je skup podataka dizajniran i konstruiran radi potpore odlučivanju pri čijem se dizajniranju slijede principi dizajna skladišta podataka s tim da taj skup podataka poslužuje potrebe određene skupine korisnika.

Radi se o skladištima koja najčešće pokrivaju pojedine funkcije poslovanja, male usko specijalizirane poslovne grupe ili pak pojedine odjele unutar organizacije koje imaju zajedničke poveznice u poslovanju. U praksi, takva područna skladišta podataka čine dijelove jedinstvenog skladišta podataka, primjerice jedne tvrtke ili organizacije.

Budući da se često pogrešno poistovjećuju skladišta podataka i područna skladišta podataka, važno je naglasiti da područna skladišta podataka svoj dizajn kreću analize korisničkih potreba, za razliku od skladišta podataka koja se dizajniraju naspram već postojećih prikupljenih podataka koji se skupljaju i kojima se upravlja da bi se kasnije mogli staviti u upotrebu.

Područna skladišta podataka usmjerena su na ispunjenje trenutnih potreba korisnika. Također, neki će stručnjaci reći da je područne skladište podataka zapravo sažeta i fokusiranija verzija skladišta podataka. Ipak, dok se za skladišta podataka koriste globalna pravila dizajna i upotrebe podataka, kod područnih skladišta dizajn je često detaljiziran da bi zadovoljio specifične zahtjeve ciljane grupe korisnika.

Područna skladišta podataka su višedimenzijska, te se kroz njih omogućava

pregledavanje podataka i dobivanje informacija vrlo jednostavno, korištenjem koliko god je potrebno dimenzija, ovisno o podacima koji se obrađuju.

Za potrebe praktičnog dijela rada i izrade vizualizacija podataka dizajnirano je upravo malo, jednostavno i reprezentativno područno skladište podataka, koje će biti opisano u nastavku.

7. ETL procesi

Kao što je više puta spomenuto, organiziranje, čišćenje i transformacija podataka predstavljaju vremenski i znanjem značajan dio posla, koji korisnici ne znaju niti je potrebno da znaju kako bi uspješno koristili sustav. ETL predstavlja skupno ime za skup procesa koji za zadatak imaju ekstrahirati, preoblikovati (transformirati) i puniti (unositi) skladište podataka sa podacima preuzetim iz jednog ili više transakcijskih izvora podataka, odnosno operativnih baza.

Iako korisnici ne poznaju ETL procedure, oni svojim djelovanjem i zahtjevima itekako utječu na njihovo stvaranje. Definiranjem zahtjeva i želja korisnika u vidu informacija koje će biti važne za buduće poslovanje i generiranje izvješća poslovne inteligencije, korisnici nesvjesno stvaraju kostur ETL procedura koje će eksperti razviti i automatizirati za potrebe razvoja BI sustava.

ETL proces započinje reformatiranjem, usklađivanjem i čišćenjem izvorno ekstrahiranih podataka. Reformatiranje je neophodno obzirom na nedostatak unificiranosti podataka iz različitih izvora. Ovim se postupkom podaci dovode u univerzalni format potreban za rad, i za daljnje faze razvoja sustava. Usklađivanje podataka se vrši najčešće iz potrebe da se ukloni redundancija podataka, odnosno pojava koja predstavlja pojavljivanje istih podataka na više mjesta u sustavu. Kao posljednji korak u pripremama, pristupa se čišćenju podataka. Radi se o eliminiranju podataka koji su uneseni greškom, primjerice tokom testiranja ili razvoja prototipa, pa kasnije iz nekog razloga nisu uklonjeni.

Programski alati za ekstrakciju: na prvi dojam, najlakši način ekstrakcije podataka iz transakcijskih baza bilo bi jednostavno kopiranje kompletnog sadržaja operativnih datoteka i sadržaja transakcijskih baza, koje bi se potom predale stručnjacima koji razvijaju BI sustav. No, u velikom broju slučajeva, ova opcija nije optimalno rješenje, ili upravo suprotno, nastoji se izbjeći. Razlog tome su realne potrebe za određenim podskupovima podataka, a spomenuti ekstrakt sadrži sve podatke, koji su često vrlo velikih dimenzija i puni nepotrebnih podataka, prateći sa stajališta poslovne inteligencije.

Iako bi iz perspektive poslovne inteligencije, idealna solucija bila čišćenje, sortiranje, filtriranje i sl. direktno na izvoru (transakcijske baze), na ovaj bi se način mogao omesti rad samih baza te samim time i poslovanje putem ometanja ili sprečavanja transakcija koje se kroz bazu vrše.

Optimalna rješenja su alati za ETL procese koji imaju zadatak "izvući" podatke iz operativnih baza, na siguran način i u što kraćem roku. Iako su ovakvi procesi često neprecizni, zbog redundancije podataka ili jednostavno nedovoljne detaljiziranosti budućih zahtjeva, sa navedenim se alatima nastoji zahvatiti samo one podatke koje će kasnije vjerojatno biti potrebni u aplikacijama, pri tom pazeći na preuzimanje što čistih podataka.

Programski alati za transformaciju: postoji čitav niz problema vezanih za podatke s kojima se moguće susresti pri transformaciji. Da bi se što točnije prikazala uloge transformacije, dovoljno je spomenuti tzv. "pravilo 80/20", koje se često spominje u literaturi, a koje zapravo navodi da se 80% ukupnog vremena koje se troši na ETL procese zapravo odnosi na dio sa transformacijom, a na ekstrakciju i punjenje tek oko 20% vremena.

Neki od problema sa podacima koji se mogu javiti, a kojima se alati za transformaciju bave jesu:

- *"Nekonzistentni primarni ključevi"* – primarni ključevi korišteni u izvornim datotekama i operativnim bazama često se ne podudaraju s primarnim ključevima koje pretpostavljaju alati poslovne inteligencije
- *Nekonzistentne vrijednosti podataka* – mnoge tvrtke kopiraju velike količine podataka, te podrazumijevajući pojam kopiranja ili dupliranja, pretpostavlja se da su duplani podaci potpuno istovjetni, no to nije uvijek tako
- *Različiti formati podataka* – primjera radi, žiro-račun naveden u operativnim bazama ne udovoljava zahtjevima alata poslovne inteligencije, a to je čest slučaj kod međunarodnih plaćanja
- *Netočne vrijednosti podataka* – potrebno je definirati logiku čišćenja kojom će se ispravljati netočne vrijednosti podataka, a obzirom da pri svakom novom unosu podaci ponovno moraju proći čišćenje, ta logika mora biti detaljno definirana što iziskuje mnogo vremena
- *Sinonimi i homonimi* – ovakve se situacije događaju kada isti podaci imaju različita imena (sinonimi) ili kada su različiti podaci upisani istim imenom (homonimi)
- *Ugrađena procesna logika* – obzirom da u praksi postoje softverski sustavi koji djeluju dugi niz godina, javlja se problem da nitko više ne zna kako oni rade. Obzirom da izvorni odnosi nisu dobro dokumentirani, to neizbježno može postati

opasan izvor pogrešaka"¹¹

Programski alati za punjenje: podaci koji se podvrgavaju ETL procesima su ili podaci iz transakcijskih baza ili arhivirani povijesni podaci. U slučaju korištenja povijesnih podataka potrebno je koristiti tri vrste ETL programa:

1. Programi za inicijalno punjenje
2. Programi za punjenje povijesnih podataka
3. Programi za inkrementalno punjenje

Programi za inicijalno punjenje skladišta podataka moraju u sebi sadržavati procedure za usklađivanje podataka kako bi se smanjila mogućnost ili potpuno eliminiralo stvaranje pogrešaka.

Programi za punjenje povijesnih podataka se smatraju "nastavkom" inicijalnog punjenja sa naglaskom na činjenici da su povijesni podaci statični, a ne "živi" poput onih ekstrahiranih iz transakcijskih baza podataka.

Programi za inkrementalno punjenje se pokreće nakon dva prethodno navedena. Svrha mu je, nakon što je skladište inicijalno napunjeno svim podacima, da periodički puni skladište novim željenim i relevantnim podacima.

Punjenje skladišta podataka se često opisuje pomoću dva procesa: inicijalno i inkrementalno punjenje. U teoriji inicijalno se punjenje vrši samo jednom, te se tada u skladištu podataka nalaze podaci iz svih izvora. Sa druge strane, inkrementalno punjenje podrazumijeva smještanje izmijenjenih i novih podataka u skladište. **Inkrementalno punjenje** se smatra kompleksnijim, iako upravlja sa manjom količinom podataka. Inicijalno se punjenje u svojoj definiciji, ne vrši gotovo nikada u praksi. Previše je različitih podataka, osobito u slučajevima kada se skladište puni novih podacima u kratkim vremenskim intervalima. Tada se povećava mogućnost ulaza raznolikih podataka, koji su neuređeni. Za svako punjenje potrebno je vršiti male izmjene, stoga se smatra da je u praksi zapravo svako punjenje inkrementalno.

Uzimajući u obzir proces inkrementalnog punjenja kao nužan, odvojen proces inicijalnog punjenja zapravo postaje suvišan. Jedna od ključnih aktivnosti u procesu

11 Panian Ž., Klepac G., „Poslovna inteligencija“, Masmedia, Zagreb, 2003., str. 90.

inkrementalnog punjenja jest razvoj mehanizma koji dopušta razmatranje jedino izvornih podataka koji su izmijenjeni ili su novi, čime se značajno reducira obrada zahtjeva na skladište podataka, obzirom da tada informacije koje su nepromijenjene ne mogu biti analizirane.

7.1. Izrada osnovne sheme

Proces učitavanja osnovne sheme zahtijeva učitavanje dimenzijske tablice i tablice činjenica. Tabela činjenica sadrži strani ključ dimenzijske tablice te kao takva predstavlja zavisani entitet. Logično, to znači da se dimenzijska tablica učitava prva.

Kada postoji jedinstveni izvor podataka, obično se koristi jedna ekstrakcija ili upit i podaci se obrađuju u jednom programu za punjenje. Jedan od glavnih razloga za implementiranje odvojenih procesa punjenja za svaku tabelu je održavanje.

Obzirom da se djelokrug skladišta povećava sa vremenom, neke od tablica mogu biti referencirane sa nekoliko tablica činjenica.

7.1.1. Učitavanje podataka u dimenzijsku tablicu

Učitavanje dimenzijske tablice zahtijeva obradu ulaznih podataka prema pravilima koja se uspostavljaju za vrijeme planiranja osnovne sheme. Sam proces punjenja mora preuzeti ulazne podatke, skupiti vrijednosti dimenzija te obraditi nove zapise dimenzija.

Prirodni ključ predstavlja jedinstveni identifikator, za što se najčešće koristi ID. Također, dimenzijska tablica mora sadržavati i tzv. surogat ključ, odnosno ključ koji jedinstveno identificira red u tablici. Opis korištenja ključeva biti će prikazan u opisu izrade praktičnog dijela rada u nastavku.

Samo učitavanje se vrši u nekoliko koraka. Prvi je ekstrakcija podataka, u kojoj se podaci mogu dobiti na više različitih načina, npr. relacijski podaci se dobivaju upitima na SQL server, dok se eksterni podaci dobivaju u tabličnom formatu ili se mogu pronaći online. Nakon ekstrakcije svi se značajni atributi koji se odnose na jedan slog organiziraju u jedan red. U idućem se koraku svaki red izvornih podataka prevodi u odgovarajući atribut dimenzijske tablice. Naravno, potrebno je koristiti određena pravila, npr. polja sa višestrukim atributima se dijele, null vrijednosti se zamjenjuju sa "Nepoznato" itd. Tek su takvi podaci spremni za obradu i povezivanje sa tablicom činjenica.

7.1.2. Učitavanje podataka u tablicu činjenica

U usporedbi sa učitavanjem podataka u dimenzijsku tablicu, punjenje tablice činjenica nešto je jednostavnije. Kada postoji višestruki izvor podataka, navedeni se koraci ponavljaju za svaki od njih. Izvorni se podaci, kao što je navedeno, mogu dobiti na više načina, ovisno o sustavu i raspoloživim alatima.

Za svaku dimenziju povezanu sa tablicom činjenica, neophodno je odrediti odgovarajući surogat ključ. Za podatke koje dobivamo iz izvornog sustava, jedino je primarni ključ dostupan. Korištenjem primarnog ključa, potrebno je odrediti ispravan surogat ključ, a ovaj se korak ponavlja za svaki slog. Primarni ključ traži se u dimenzijskoj tablici. Kada je odgovarajući slog pronađen na temelju primarnog ključa, njemu se dodjeljuje odgovarajući surogat ključ kada se isti unosi u tablicu činjenica.

8. OLAP alati

OLAP (engl. On-line Analytical Processing) danas predstavlja najrašireniji način kreiranja i distribucije izvještaja na svim razinama odlučivanja. OLAP predstavlja alate za analitičku obradu, a podrazumijeva tehnologije za skupljanje, upravljanje, obradu i prezentaciju. Za ovaj je rad najvažniji dio sa prezentacijom, konkretnije vizualizacijom te će o njemu biti riječi nešto više u nastavku.

Sa tehničkog aspekta, OLAP sustavi su visoko indeksirane datoteke u kojima se podaci spremaju radi njihove kasnije analize. Ovakvi se sustavi, za razliku od transakcijskih, zasnivaju na multidimenzijskoj analizi, u čemu i leži njegova prava "moć". Podaci se mogu promatrati istovremeno kroz veći broj dimenzija, ovisno o njihovoj strukturi. Sustav predstavlja jednostavno rješenje čak i za korisnike koji nisu posebno obučeni i uvježbani za obavljanje analitičkih poslova.

Važna karakteristika OLAP-a jest da se međusobno razlikuju mnogo više od međusobne razlike relacijskih baza podataka, programskih jezika i sl. Vrlo je važno da pri vrednovanju ovih alata sudjeluju i korisnici i IT stručnjaci, u svrhu pronalaska najboljeg rješenja.

Druga iznimno pozitivna karakteristika jest brzina. Korisnici kroz upite do svog odgovora mogu doći vrlo brzo, iako i to ovisi o količini podataka koji se procesuiraju. Ipak, to je svakako puno brže rješenje od sličnih alata.

Za neke je stručnjake OLAP zapravo produženje ljudske inteligencije. Kao glavni razlog navode sposobnost većeg broja dimenzija (u većim sustavima deset i više njih), što je ljudskog umu nemoguće procesuirati. Za njih to vrši OLAP, i to prilično brzo.

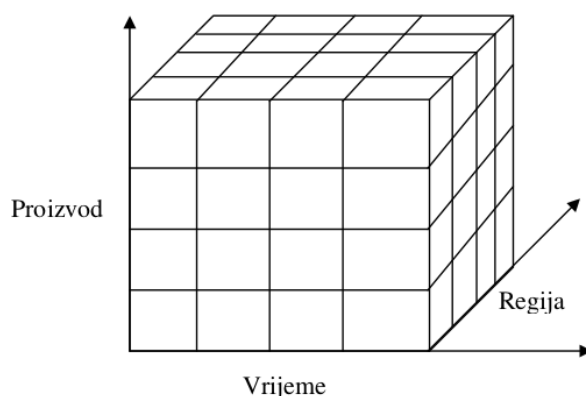
Nije na odmet napomenuti da OLAP također ima širok raspon mogućnosti, od pretraživanja do proračuna i zahtjevnijih analiza.

8.1. Strukturiranje podataka

"Podaci se strukturiraju sukladno načinu na koji ljudi prirodno provode analize određenih pojava ili procesa. Među strukturama navode se dimenzijski modeli i dimenzijske analize podataka te hijerarhijske strukture. Primjerice, skupovi podataka strukturirani po dimenzijama (proizvodima, dobavljačima, regijama), mogu se podvrći

analizi prema navedenim dimenzijama. Nužna dimenzija svake analize je vrijeme, čime dolazi do višedimenzijske matrice."¹²

Slikovno, dimenzije je najlakše prikazati kockama podataka, odnosno višedimenzijskim matricama. Unutar takvih kocki spremaju se mjere koje nas interesiraju. Osi kocke čine dimenzije po kojima želimo pratiti vrijednosti (mjere). Sam broj dimenzija nije ograničen, a u slučajevima kada se radi o više od tri dimenzije, takve kocke se nazivaju *hiperkockama*.



Slika 9: Primjer OLAP kocke sa tri dimenzije

Izvor: Panian Ž., Klepac G., „Poslovna inteligencija“, Masmedia, Zagreb, 2003., str. 235

OLAP kocka se puni vrijednostima iz relacijske baze podataka u kojoj su evidentirani željeni podaci za promatranje. Idealno rješenje je skladište podataka iz niza razloga – podaci su pripremljeni na pravi način, integrirani, pročišćeni i transformirani. Pri dimenzijskoj analizi podataka višedimenzijski skupovi mogu se analizirati horizontalno i vertikalno te se takva metoda naziva *Slice and Dice*.

Suvremeni OLAP alati daju mogućnost rotacije (engl. Pivoting) dimenzija kocki po principu "pokaži i klikni", pri čemu je os rotacije najčešće vrijeme, ali to ne mora biti nužno tako.

U postupku svrdlanja (engl. Drill down) korisniku je omogućeno da analizu započetu s vrha hijerarhije nastavi prema nižim razinama podataka, dakle od općih podataka prema detaljiziranim.

OLAP-ova glavna prednost je njegova brzina i mogućnost dobivanja ad-hoc

¹² Vitasović, V. *Sustavi poslovne inteligencije za uspješnu organizaciju 21. stoljeća*, Diplomski rad, Pula, 2009.

odgovora, umjesto klasičnih papirnatih izvješća. Željeni odgovor moguće je generirati u stvarnom vremenu, nekada i u sekundama. Ovakva vrsta analize čini temelj za izradu vizualizacija podataka i predstavlja "mozak" koji isporučuje podatke za svaki graf, tablicu ili selektor.

Danas postoji nekoliko varijanti **OLAP alata**:

- 1. Relacijski OLAP (ROLAP)** – posjeduju relacijski model baze podataka koji se može definirati na način da vrlo vjerno odražava stvarni svijet. Problem se javlja kada se želi unijeti neka specifičnost u proizvod s ciljem da se zadrži određena kvaliteta.
- 2. Multidimenzijски OLAP (MOLAP)** – skup alata koji imaju karakteristiku da koriste vlastitu bazu podataka sa n brojem dimenzija. Svi elementi tog sustava su ravnopravni, a alati su time teže prilagodljivi promjenama u veličini sustava.
- 3. Desktop OLAP (DOLAP)** – alati za stolna računala, prilagođeni uređajima skromnijih kapaciteta memorije i mogućnosti obrade.
- 4. Hibridni OLAP (HOLAP)** – hibridni proizvod pomoću kojeg je moguće provoditi multidimenzijske analize simultano iz podataka iz multidimenzionalne baze i onih iz sustava upravljanja relacijskim bazama podataka. Time se kombiniraju MOLAP i ROLAP.

9. Rudarenje podataka

Iako se metode rudarenja podataka neće koristiti u praktičnom dijelu ovoga rada, biti će navedene samo osnovne informacije i značajke termina.

"Rudarenje podataka (engl. Data Mining) je finalna manifestacija procesa skladištenja podataka. Pod pojmom rudarenja podataka podrazumijeva se proces pronalaženja skrivenih trendova, modela, zakonitosti i odnosa među podacima. Alati za rudarenje podataka pretražuju baze podataka tražeći skrivene modele i veze među podacima koje su do tada bile nepoznate.

Primjenom alata za rudarenje podataka moguće je pronaći neprepoznate matrice ponašanja, mogu se točnije i brže predvidjeti budući trendovi ponašanja, što omogućuje u poslovnom svijetu donošenje odluka temeljenih na znaju (engl. Knowledge-driven decisions)."¹³

Tehnike rudarenja podataka rezultat su dugog procesa istraživanja i razvoja statističkih algoritama na tom polju. Početak seže u vrijeme prvih skladištenja poslovnih procesa u računalu, a kontinuirano se nastavlja s unapređenjem pristupa podacima te generiranjem tehnologija koje omogućuju korisnicima navigaciju kroz podatke u realnom vremenu.

Primjena metoda rudarenja podataka odvija se prvenstveno u poslovanju poduzeća, ali i u ostalim područjima koja raspolažu velikim brojem podataka, na osnovu kojih se žele otkriti određene veze, pravilnosti i zakonitosti. Kvalitetni podaci čine preduvjet za uspješno provođenje ovih metoda, uz precizno određivanje ciljeva i problema koje je potrebno rješavati.

Kao što je spomenuto otprije, obzirom da data-mining nije primarna tema ovoga rada, postojeće metode rudarenja podataka će biti samo nabrojane, ali ne i detaljnije objašnjene:

1. Regresijska metoda
2. Klasifikacijske metode i stabla odlučivanja
3. Metode klasteriranja
4. Metode za analizu veza (asocijacija)

¹³ Vitasović, V. *Sustavi poslovne inteligencije za uspješnu organizaciju 21. stoljeća, Diplomski rad, Pula, 2009.*

"Kritični čimbenici za uspješno provođenje rudarenja podataka su pravilno integrirano skladište podataka i dobro poznavanje i razumijevanje procesa poslovanja nad kojim se želi primijeniti postupak rudarenja podataka. Primjena rudarenja podataka podrazumijeva veliko statističko predznanje, a poduzeća obično posjeduju mali broj stručnjaka koji se time bave."¹⁴

Rudarenje podataka se danas primjenjuje na mnogim poslovnim područjima. Uspješna primjena rudarenja podataka jest prilikom zadržavanja potrošača, otkrivanja prijevara s kreditnim karticama, sustava segmentacije i sl. Na temelju novih informacije dobivenih kroz rudarenje podacima, može se lakše doći do spoznaja o načinima ublažavanja navedenog problema.

¹⁴ Vitasović, V. *Sustavi poslovne inteligencije za uspješnu organizaciju 21. stoljeća, Diplomski rad, Pula, 2009.*

10. Portali poslovne inteligencije

Portali poslovne inteligencije (engl. BI portal) jesu takvi portali koji omogućavaju korisniku kreiranje upita i izradu izvješća korištenjem baza podataka na razini kompanije ili organizacije. Ovakva vrsta portala omogućava organizacijama da centraliziraju svoje izvore izvješćivanja u jedinstveni sustav koji je jednostavan za korištenje. Portale čine kombinacije različitih komponenata, od kojih svaka zadovoljava specifične potrebe korisnika, te su lako nadogradivi sa novim komponentama, ukoliko se ukaže potreba za tim. BI portali trebaju osiguravati pozadinsku sigurnost, skladište podataka, prezentaciju podataka i njihovu isporuku.

Korištenjem portala poslovne inteligencije u organizaciji, postiže se povećana produktivnost zaposlenika, maksimizirana upotreba poslovne inteligencije i visoka stopa povrata uloženog u informacijsku arhitekturu.

Da bi bio efikasan, portal poslovne inteligencije mora pružiti benefite za korisnike, kako bi maksimizirao svoju uporabu. Također, mora osigurati interaktivnu inteligenciju dobivenu iz više izvora te filtriranu na način da je fokus stavljen na interese i potrebe specifičnih korisnika. Glavna zadaća ovakvih portala jest integracija eksterne i interne inteligencije u radnu svakodnevicu zaposlenika, te bi se trebao integrirati u postojeće investicije u portalsku tehnologiju.

"**Portal** je web aplikacija koja pruža personalizaciju, mogućnost autorizacije, agregiranje sadržaja iz različitih izvora te prezentacijski sloj informacijskih sustava. Agregacija predstavlja pojam integracije sadržaja iz različitih izvora unutar web stranice."¹⁵

Funkcionalnost portala može se podijeliti u tri glavne komponente:

- 1. Kontejner portleta** (engl. Portlet container) – svaki portlet je ugrađen unutar kontejner portleta koji kontrolira životni ciklus portleta te im pruža nužne resurse i informacije o njihovoj okolini. Kontejner portleta je odgovoran za inicijalizaciju i uništenje portleta kao i za prosljeđivanje korisničkih upita i prikupljanje odgovora.
- 2. Agregator sadržaja** (engl. Content aggregator) – sakuplja sadržaj generiran iz različitih portlet aplikacija, što je ujedno i jedna od glavnih zadaća portala uopće.
- 3. Zajedničke usluge** (engl. Common services) – usluge nisu dio specifikacije portleta, ali

¹⁵ "What is portal?, <http://archive.oreilly.com>, 4.6.2016.

komercijalni portali pružaju čitav niz takvih usluga kako bi se distancirali od konkurencije. Najpoznatije takve usluge jesu:

- **Single sign-on:** omogućava jedinstven pristup svim aplikacijama odmah po autorizaciji. Nije potrebno unositi korisničke podatke za svaku aplikaciju zasebno. Ovakve se informacije štite enkripcijom te vrlo često duljinom sesije, odnosno ukoliko je korisnik neaktivan nakon n minuta ili sati, sustav ga automatski odjavljuje iz sigurnosnih razloga.
- **Personalizacija:** najčešće vrste personalizacija omogućavaju korisniku izmjenu stranice na dva načina. Prvo, omogućavaju promjene boja i vizualnog identiteta, od ikona do cijelih tema. A kao drugi način, najčešće se omogućava personalizacija prikaza ovisno o korisnikovim karakteristikama (npr. odabir da se prikazuje portlet koji sadrži informacije o Premium proizvodima samo onim korisnicima koji imaju prihode veće od iznosa n).

Portleti su web komponente koje se pokreću unutar kontejnera i imaju zadaću generiranja dinamičkog sadržaja. Portleti predstavljaju fragmente sadržaja ili komade koda koji se integriraju u portale stoga se često sami portali definiraju kao skup portleta, u kojem svaki prozor prikazuje neki portlet. Radi jednostavnijeg shvaćanja pojma, portleti često mogu prikazivati elektroničku poštu, vremensku prognozu, rasprave na forumima, novosti sa društvenih mreža i sl.

Glavne karakteristike portleta:

1. Portletima upravljaju specijalizirani kontejneri, kako je i otprije navedeno
2. Portleti generiraju dinamički sadržaj
3. Životni ciklus portleta je određen njegovim kontejnerom
4. Portleti generiraju komade markup koda, a ne cijele dokumente
5. Portleti nisu dostupni putem klasičnog URL-a, što znači da je nemoguće poslati poveznicu na portlet.

11. Vizualizacija podataka

Vizualizacija predstavlja vizualnu reprezentaciju određenog skupa podataka. Dio pojma vizualizacije koji je važan za ovaj rad jest vizualizacija podataka (engl. data visualization).

Koncept korištenja slikovnih prikaza za bolje razumijevanje podataka postoji već stoljećima, od mapa u 17. stoljeću do prve pojave tzv. "pita" grafova početkom 19. stoljeća. Nekoliko desetljeća kasnije, nastao je jedan od prvih kompleksnih statističkih grafičkih prikaza. Charles Minard vizualno je prikazao Napoleonovu invaziju na Rusiju. Karta je prikazivala veličinu Napoleonove vojske, kao i put kojim se vojska kretala. Uz to, dodao je i poznate podatke o temperaturi i vremenske ljestvice kako bi se događaj prikazao što vjernije i razumljivije.

Ipak, razvoj tehnologije je bio pravi okidač vizualizacija podataka. Računala su napokon omogućavala obrađivanje velikih količina podataka velikom brzinom. U današnje vrijeme, vizualizacija podataka postala je brzorastuća mješavina znanosti i umjetnosti koja ima moć mijenjati poslovna tržišta i odnose snaga.

Cilj vizualizacije podataka jest stvaranje vizualne reprezentacije podataka koji su jednostavno razumljivi korisniku. Vizualna analitika je također pojam koji se usko vezuje uz sustave poslovne inteligencije.

Glavne značajke ovakvih vizualizacija jesu arhitektura, koja je upravljana tokom podataka, filtriranje i preslikavanje podataka te interaktivnost, koja se odnosi na selekciju podataka i istraživanje.

Vizualizacija podataka najčešće se vrši komponentama poput grafova i tablica, koje mogu biti statične ili interaktivne. Takve se komponente, u slučajevima poslovnih situacija, na neki način "ukalupljuju" u jednu cjelinu (sučelje) pomoću kojeg korisnik upravlja i postavlja upite svojim interakcijama, a odgovore dobiva vizualnim putem, u što kraćem roku. Što je sustav bolje postavljen, vrijeme prikaza odgovora će biti kraće.

Takva sučelja, koja korisnik koristi za interakciju, nazivaju se upravljačke ploče (engl. dashboardi).

11.1. Pojam "Upravljačke ploče"

U informacijskoj tehnologiji, upravljačka ploča predstavlja korisničko sučelje, koje se na

neki način može usporediti i sa vozačkim sučeljem u automobilu. Zadaća mu je organizirati i prezentirati informacije na način da budu dostupne i što razumljivije, iako je to subjektivna komponenta. Usporedba sa automobilskim sučeljem nije toliko precizna, jer je računalno sučelje puno interaktivnije, ali ipak može dočarati samu svrhu upravljačke ploče. Upravljačka ploča kao pojam je doživio procvat razvojem Google-ovih aplikacija (Google Docs, Google Analytics, Google Drive...). Danas pojam upravljačke ploče predstavlja univerzalan izraz za neko korisničko sučelje koje posjeduje interaktivne komponente.

Veliki broj proizvoda u današnje vrijeme koji imaju za cilj integraciju iz više izvora, prikazani na unificiranom sučelju, nazivaju se upravljačkim pločama, što zbog takvog poimanja vlastitog proizvoda, što zbog marketinških prednosti.

Primjera radi, neki proizvod može zahvaćati podatke istovremeno sa lokalnog računala iz neke trenutno aktivne aplikacije te sa nekog Web mjesta i kroz upravljačku ploču prezentirati te podatke kao da su došli iz istog izvora. Upravljačke ploče mogu biti promjenjive (engl. customizable) na mnogo načina, od načina prikaza podataka, do vizualnih efekata.

Od samog pojma upravljačke ploče, odvaja se pojam BI upravljačka ploča. BI upravljačka ploča predstavlja alat za vizualizaciju podataka koji prikazuje trenutno stanje metrike i ključnih pokazatelja ili indikatora za određenu organizaciju. Ova vrsta upravljačke ploče ima zadatak integrirati brojeve, metriku i pokazatelje uspješnosti na jednom (single) ekranu. Često se prilagođavaju određenim korisnicima ili organizacijskim jedinicama da bi se moglo precizno odrediti što koja jedinica ili korisnik vide na svojoj upravljačkoj ploči. Glavne značajke BI upravljačke ploče jesu prilagodljivo sučelje, po mogućnosti opcija da korisnik sam može birati što će se prikazivati, te mogućnost da u stvarnom vremenu povlači podatke iz višestrukih izvora.

Među vodećim proizvođačima BI upravljačkih ploča su Microsoft i Oracle, no takva se sučelja mogu razviti i kroz druge poslovne aplikacije, a u praktičnom dijelu ovoga rada koristiti će se Pentaho BI open-source platforma sa svojim CDE¹⁶ alatima za izradu upravljačkih ploča.

¹⁶ CDE (engl. Community Dashboard Editor) – komponenta sustava Pentaho koja služi za razvoj upravljačkih ploča

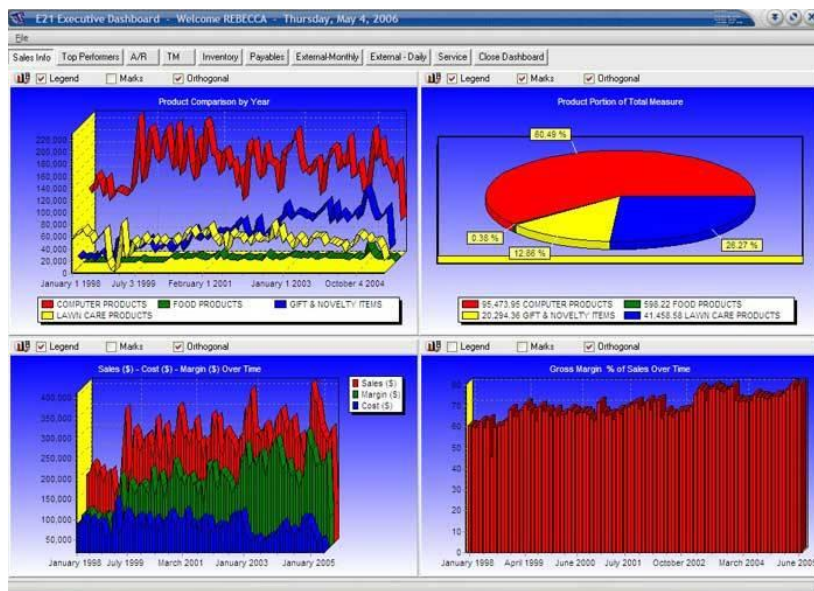
11.2. Rizik loših vizualizacija

Zanimljive su procjene koje se mogu naći na svjetskim portalima i poslovnim mrežama, a koje tvrde da 80% BI projekata nije označeno uspješnima. Iako ne postoji skupno mišljenje koji je tome razlog, mnogi se slažu da su loše vizualizacije i kompleksnost upravljačkih ploča glavni razlozi za slabije korištenje ovakvih sustava.



Slika 10: Primjer kvalitetnog rasporeda komponenti na upravljačkoj ploči

Izvor: biztechbpm.blogspot.com



Slika 11: Primjer loše dizajnirane upravljačke ploče

Izvor: www.dashboardzone.com

Na prvi dojam, BI sustavi dijele problem svih softvera u razvoju – nedostatak komunikacije između ključnih korisnika i IT-a te nedovoljno precizno određivanje specifikacija i zahtjeva. No, to nije uvijek ključni razlog, jer postoji još jedan "manje važan" i često zanemarivan – vizualizacija pohranjenih podataka. Rijetke su tvrtke koje mnogo vremena i brige vode oko vizualizacije podataka. Gledajući sa strane korisnika, savršeno izgrađeno skladište podataka, pročišćeni podaci i sl. ne znače ništa ukoliko ti podaci nisu razumljivi i ukoliko ih ni menadžment, prije svih, ne razumije. Mnoge tvrtke ne shvaćaju da se softver stvara radi korisnika i zadovoljenja njihovih potreba, a ne radi same kvalitete softvera ili njegovih značajki. Softver može imati stvari koje drugi nemaju, ali ako korisnik nema potrebu za takvim značajkama, one ne vrijede ništa u korisničkom, a u konačnici i tržišnom smislu.

Podaci se moraju prilagoditi ljudskim bićima i ljudskoj percepciji. "Vizualizacija podataka je ugrađena u ljudski DNA i preko 70% našeg tijela je u očima. Od podataka pa do izvještaja, ljudi uključeni u projekt su uglavnom jednostavno IT i to je jedan od velikih problema BI danas. Bez ispravne vizualizacije i prezentacije podataka, korisnici će zapeti na golim podacima i trošiti će sate pokušavajući ih razumjeti, ili još gore, biti će im prezentirane loše vizualizacije koje navode na krive zaključke."¹⁷

11.3. Smjernice za dizajniranje kvalitetne upravljačke ploče

"Dashboardi su često kreirani bez jasnog plana, na način da se podaci dodaju samo zato jer postoji preostali prazan prostor. Različiti ljudi u organizacijama zahtijevaju različite podatke koji će se prikazivati i ubrzo dashboard postane teško čitljiv i prepun besmislenih nepovezanih informacija. Kada se to dogodi, dashboard postaje beskoristan."¹⁸

Da bi se izbjegla ova pojava, preporuča se pridržavanje određenih pravila:

- Određivanje ciljane publike – najefikasnije upravljačke ploče namijenjene su jednoj specifičnoj grupi korisnika te prikazuju podatke specifične za pretpostavljene slučajeve korištenja. Ovaj je korak često zanemaren ili se lako previdi pa kreirane upravljačke ploče uključuju veliku mješavinu različitih podataka, od kojih su neki

17 Smolić, H. "Loše vizualizacije: zašto BI projekti propadaju, 23.2.2014., <http://www.skladistenje.com/lose-vizualizacije-zasto-bi-projekti-propadaju/> (27.4.2016.)

18 Smith, N. "Designing and building great dashboards – 6 golden rules to successful dashboard design", 21.10.2015., <https://www.geckoboard.com/blog/building-great-dashboards-6-golden-rules-to-successful-dashboard-design/#.V0ScD54vDCI> (13.5.2016.)

korisni jednoj grupaciji korisnika, a neki drugi drugoj.

- Odabir pravog tipa upravljačke ploče – tipovi upravljačkih ploča dijele se na operativne, strateške i analitičke:

1. Operativne – prikazuju podatke operativnog funkcioniranja tvrtke. Npr. ukoliko je neka organizacija svoje aktivnosti vrši putem web stranice, vrlo je važno da ta stranica bude stabilna i pouzdana, stoga će se pratiti podaci o funkcioniranju servera i njegovih performansi. Ukoliko se organizacija bavi prodajom, možda će htjeti specifičnu upravljačku ploču koja prikazuje broj narudžbi i rezervacija.

Kod operativnih upravljačkih ploča vrlo je važno da se pune podacima u realnom vremenu, ili barem što bliže navedenom.

2. Strateške – uglavnom prikazuju ključne indikatore uspješnosti (engl. Key Performance Indicators) koje menadžment prati u određenom vremenskom intervalu (tjedno, mjesečno...) . Ovaj tip upravljačkih ploča omogućuje menadžmentu kvalitetnu ukupnu sliku stanja u organizaciji sa eventualnim prilikama sa kojima se susreću.

3. Analitičke – prikazuje i operativne i strateške podatke. Česta je pojava kompliciranje funkcionalnosti kombinacijom navedenih podataka, iako to omogućava korisniku još preciznije analize. Ova "analitička komponenta" često se implementira u upravljačke ploče i kada nije potrebna, što utječe na kvalitetu i uspješnost ploča.

Jasno je da će korisnici sa različitim potrebama i u različitim trenucima trebati različite podatke. U tom je slučaju preporučljivo, ukoliko je to moguće, odvojiti cjeline u više upravljačkih ploča.

- *Logično grupiranje podataka* – dobro dizajnirana upravljačka ploča pretpostavlja da su podaci prikazani u logičkim skupinama. Ukoliko prikazujemo određene indikatore uspješnosti i primjerice količinu prodanih proizvoda, potrebno je izdvojiti to dvoje u dvije zasebne logičke grupe, kako ne bi došlo do poteškoća u iščitavanju podataka. Zanimljiva je informacija da neki dashboard "gurui" smatraju da je pozicija logotipa tvrtke na upravljačkoj ploči, a koji se najčešće nalazi u gornjem lijevom kutu, zapravo najgora. Kao razlog navode da većina zapadnih kultura čitaju od gore prema dolje i od lijevog prema desnom. Upravo bi na tom mjestu trebao biti dio ploče sa logičkom skupinom koja se smatra najrelevantnijom.
- *Pridavanje važnosti podacima prema publici* – korporativne upravljačke ploče mogu

imati čitav niz različitih korisnika kojima se podaci prezentiraju, stoga je važno voditi računa o načinu slaganja podataka. Pod time se misli na podjele po određenim skupinama, npr. cijela organizacija, odjel, zasebni zaposlenici ili dobavljači. Vrlo je važno biti precizan u određivanju ciljne publike. Dok će se u manjim tvrtkama često prikazivati svi indikatori na svim razinama, u većim organizacijama se često dizajniraju zasebne ploče za svaki odjel.

- *Prezentiranje samo važnih podataka* – upravljačke ploče su često natrpane raznim grafovima, tablicama, brojevima... Takva pretrpanost uzrokuje pomicanje fokusa na nevažne stvari. Iako postoje ploče pretrpane korisnim podacima i oni pretrpane beskorisnim podacima, razlika zapravo ne postoji. To nisu kvalitetni upravljačke ploče. Svaki tip upravljačkih ploča zahtjeva manje ili više podataka. I dok strateške ploče sa ključnim indikatorima mogu biti kvalitetne i sa 3-4 praćena podatka, operativne ploče rijetko se kada dizajniraju bez barem 20 elemenata. Nema jasnih pravila na koji način kvalitetno prikazati te podatke, već je nužno na najbolji mogući način omogućiti da prikaz bude relevantan i smislen. Nije potrebno dodavati grafove ili brojeve samo zato jer ima još slobodnog mjesta na ekranu.
- *Frekventnost osvježavanja podataka* – ispravna procjena intervala osvježavanja podataka vrlo je važna zbog postizanja optimalnih performansi upravljačke ploče. Operativni tip ploča je po ovom pitanju najzahtjevniji, jer se služi podacima u realnom vremenu, dok se strateški osvježavaju tjedno, mjesečno i sl.

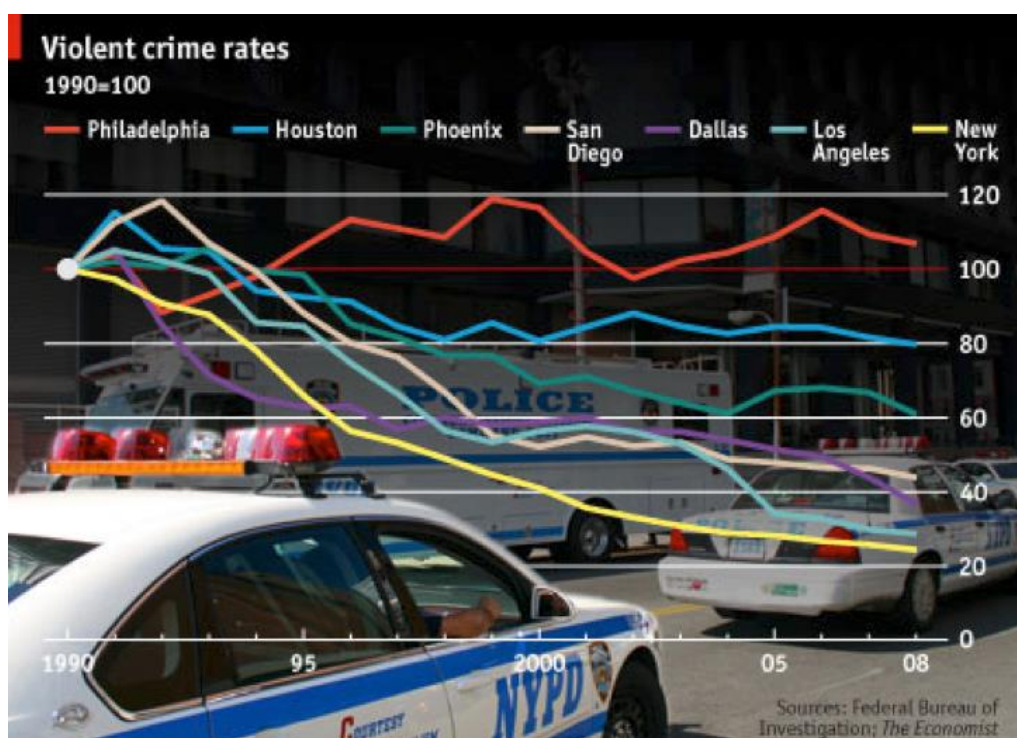
Prateći navedene smjernice, moguće je dizajnirati kvalitetnu upravljačku ploču, koji će biti vizualno upotrebljiv te će sadržavati samo relevantne podatke.

11.4. Kognitivne smjernice za dizajn upravljačkih ploča

"Efektivne upravljačke ploče smanjuju kognitivne napore korisnika, pomažu rješavanju problema i omogućuju otkrivanje važnih podataka i podatkovnih uzoraka. Budući da su nam kognitivni eksperimenti omogućili golemo znanje o mehanizmima i procesima koje čovjek koristi za percipiranje stvari i pojava, te kako ih koristi za donošenje odluka, razumno je pretpostaviti da su takve spoznaje vrlo važne i za dizajniranje

upravljačkih ploča."¹⁹ Faktori o kojima se mora voditi računa kod dizajniranja upravljačkih ploča jesu: način na koji korisnici percipiraju podatke, koliko sam prijedlog dizajna ploče zahtjeva korisničke koncentracije i napora te koje znanje korisnik mora posjedovati da bi znao efikasno iskoristiti prikazane informacije. Shodno navedenom, u većini dostupne literature spominju se neke zajedničke smjernice za dizajn svih upravljačkih ploča:

a) Postavljanje naglaska na čitljivost – da bi korisnik na ispravan način koristio ploče, on mora biti u mogućnosti iščitavati podatke brzo i jednostavno, te bi stoga sučelje moralo biti takvo da potiče vizualno procesiranje, a ne obrnuto.



Slika 12: Primjer slabo čitljive upravljačke ploče

Izvor: www.uxmag.com

Na gornjoj je slici vidljivo da pozadinska slika onemogućava jasno iščitavanje podataka bez dodatnog vizualnog napora korisnika. Da bi se ovakve poteškoće izbjegle, potrebno je koristiti vidljiviji kontrast između pozadine i elemenata upravljačke ploče, a preporuča se jednobojna pozadina, ili blagi gradijent, bez prebrzih prijelaza boja.

¹⁹ Shilpi Choudhury, *Four Cognitive Design Guidelines for Effective Information Dashboards*, <https://uxmag.com/articles/four-cognitive-design-guidelines-for-effective-information-dashboards> [18.6.2014], 1.7.2016.

b) Minimaliziranje kognitivnog opterećenja – teoretski uvid u kognitivnu arhitekturu naglašava memorijska ograničenja kod ljudi. Stručnjaci u području HCI (engl. Human-Computer interaction) slažu se da bi interaktivne sekvence korištenja alata poput upravljačkih ploča morale biti dizajnirane na način da minimiziraju punjenje kratkoročne memorije. Budući da je spoznajna memorija superiorna u odnosu na onu dugoročnu, ploča bi morala pružiti sve potrebne informacije koje korisnik treba za specifičnu analizu ispred njegovih očiju.

c) Korištenje grafičkih prikaza – ljudski je mozak stvoren za vizualnu percepciju, te se to smatra "prirodnim stanjem", neovisno radi li se o pretvaranju čistih podataka u grafove ili pak ikonice i sličice. Za čovjeka, slika zaista vrijedi tisuću riječi, i gotovo uvijek, lakše i brže ćemo percipirati slikovni prikaz nego tekstualni ili brojčani.

d) Pratiti zakone percepcije (engl. Gestalt laws) – predstavljaju niz zakona koji objašnjavaju kako čovjek percipira uzorke i na temelju njih donosi zaključke. Korištenje ovih zakona u dizajniranju vizualizacija i chartova može uvelike olakšati korisniku prepoznavanje uzoraka u podacima. Najvažniji zakoni smatraju se:

- Pragnanz zakon – radi se o temeljnom principu Gestalt psihologije koji govori o tome da ljudi nastoje organizirati svoja iskustva na način koji je kontinuiran, uređen, simetričan i jednostavan.

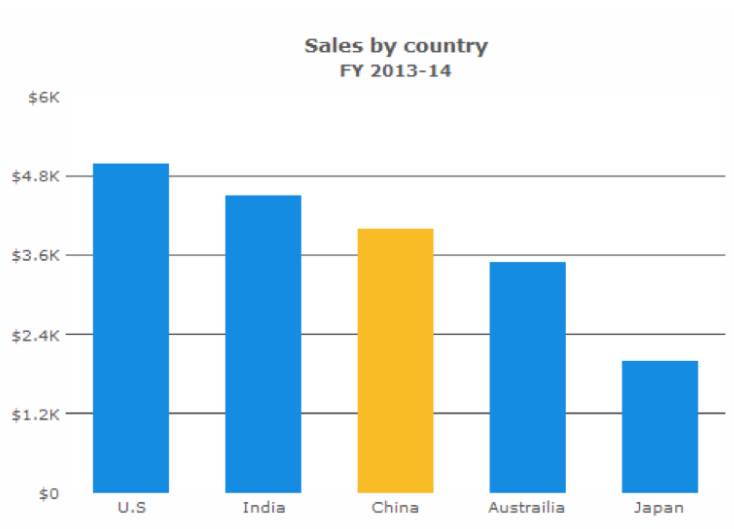


Slika 13: Primjer upotrebe Pragnanz zakona na izradi grafa

Izvor: www.uxmag.com

Na gornjoj slici korisniku je jednostavnije iščitati podatke sa desnog grafa, jer su podaci smisleno složeni, čemu ljudski mozak i teži. Ovakav se pristup preporuča ukoliko su podaci temeljeni na imenima ili kategorijama, a ne vremenskim intervalima. U praktičnom dijelu ovoga rada, glavni grafovi biti će upravo posloženi po vremenskim jedinicama (godinama), te stoga primjena Pragnanz zakona nije moguća.

- Zakon o žarišnom centru (engl. Law of focal point) – zakon govori o tome da različiti objekti unutar vizualne prezentacije odvlače ljudski fokus.



Slika 14: Primjer upotrebe Zakona o žarišnom centru

Izvor: www.uxmag.com

Za primjer, ukoliko želimo korisničku pozornost da jednom zasebnom dijelu grafa, to ćemo postići promjenom njegove boje, kao na slici iznad.

12. Izrada sustava poslovne inteligencije za obrazovnu javnu instituciju

U nastavku rada biti će opisan cjelokupan proces izrade sustava poslovne inteligencije za Fakultet ekonomije i turizma "Dr. Mijo Mirković" iz Pule.

Unutar procesa izrade, korišteni su različiti alati i metode transformacija podataka, kao i različiti programski jezici poput Javascripta, u različitim fazama izrade.

Treba naglasiti da se radi o prilično malom sustavu s malom količinom podataka, ali dovoljno kompleksnim procesom izrade kako bi se vjerno mogao replicirati proces izrade većih sustava, te u konačnici pokazati kako jedan takav sustav može doprinijeti u smislu efikasnosti upravljanja jedne javne obrazovne ustanove kao što je fakultet.

12.1. Korišteni podaci u sustavu

Podaci koji su podloga za odlučivanje i koji jesu dio sustava poslovne inteligencije jesu podaci o analizi učinkovitosti studenata na FET-u, statistici diplomiranih studenata na FET-u te trendovima studiranja u Republici Hrvatskoj.

Podaci vezani za FET, dobiveni su izravno od strane korisnika, u klasičnom excel (xls) tabličnom formatu. Podaci o trendovima u RH, također su tabličnom (xls) formatu, no oni su preuzeti sa Portala otvorenih podataka Republike Hrvatske (data.gov.hr).

Promatra li se kvaliteta dobivenih podataka, može se zaključiti da su oni prilično kvalitetni i "zahvalni" u pogledu dovođenja u ovakav sustav. Ipak, jedina zamjerka dobivenim podacima jest nestandardiziranost u određenim slogovima – primjerice "Financije, računovodstvo i revizija" i "FRR/Financijski management", koje predstavljaju isti studijski smjer, ali prateći trend kroz godine, u vizualizaciji će biti prikazani kao različiti studijski smjerovi.

Objašnjenje procesa izrade sustava biti će objašnjeno na komponenti nazvanoj "Analiza učinkovitosti" jer se radi o najkompleksnijim podacima u sustavu. U donjoj slici moguće uočiti kako su izgledali izvorni podaci u tabličnom formatu.

Podaci sadrže sljedeće informacije:

1. Akademska godina
2. Studijski smjer
3. Godina studija
4. Status studenta
5. Status upisa (upis prvi put, ponavljači ili apsolventi)
6. Studij (prediplomski ili diplomski)
7. Broj studenata

2005/06 prediplomski					
SMJER	I GODINA				
	REDOVNI		IZVANREDNI		
	I put		I put		
Ekonomija		41		24	
Financije, računovodstvo i revizija		68		47	
Marketinško upravljanje		68		41	
Management i poduzetništvo		91		50	
Turizam		100		60	
Poslovna informatika		96		45	
Σ		464		267	

2006/07 prediplomski						
SMJER	I GODINA				II GODINA	
	REDOVNI		IZVANREDNI		REDOVNI	IZVANREDNI
	I put	Ponavljači	I put	Ponavljači	I put	I put
Ekonomija	46	25	49	8	31	21
Financije, računovodstvo i revizija	72	41	42	28	51	33
Marketinško upravljanje	102	36	47	24	40	23
Management i poduzetništvo	109	46	42	21	53	28
Turizam	107	59	48	27	39	29
Poslovna informatika	75	53	53	25	35	26
Σ	511	260	281	125	249	160

2007/08 prediplomski										
SMJER	I GODINA				II GODINA				III GODINA	
	REDOVNI		IZVANREDNI		REDOVNI		IZVANREDNI		REDOVNI	IZVANREDNI
	I put	Ponavljači	I put	Ponavljači	I put	Ponavljači	I put	Ponavljači	I put	I put
Ekonomija	51	23	42	24	26	24	29	22	25	
Financije, računovodstvo i revizija	70	28	55	29	69	38	26	22	32	17
Marketinško upravljanje	57	40	35	28	53	30	28	21	27	
Management i poduzetništvo	128	47	73	28	73	37	33	22	37	9
Turizam	88	57	60	36	69	28	34	7	30	12
Poslovna informatika	73	44	43	29	45	24	23	7	31	13
Kultura i turizam	46	20	45	0	0	0	0	0	0	0
Σ	513	259	353	150	335	181	173	87	182	

Slika 15: Izvorni podaci o analizi učinkovitosti (vidljive samo prve tri promatrane godine)

Izvor: ISVU sustav

Napomena: Podaci su modificirani u cilju zaštite privatnosti korisnika

Obzirom da se radi o maloj količini podataka, radi lakšeg preuzimanja, izvršene su određene preinake u samom tabličnom formatu. U slučaju većeg broja podataka, takve bi preinake dakako vršila određena programska procedura. U ovom slučaju, to nije bilo potrebno, obzirom da se podaci nisu mijenjali, već samo prilagođavali na način da se prikaz promijenio iz "stepenastog" i hijerarhijski (vidljivo usporedno na gornjoj i donjoj slici).

2005/06	
Prediplomski	
1 godina	
Redovni	
1 put	
Ekonomija	41
Financije, računovodstvo i revizija	68
Marketinško upravljanje	68
Management i poduzetništvo	91
Turizam	100
Poslovna informatika	96
Izvanredni	
1 put	
Ekonomija	24
Financije, računovodstvo i revizija	47
Marketinško upravljanje	41
Management i poduzetništvo	50
Turizam	60
Poslovna informatika	45
2006/07	
Prediplomski	
1 godina	
Redovni	
1 put	
Ekonomija	46
Financije, računovodstvo i revizija	72
Marketinško upravljanje	102
Management i poduzetništvo	109
Turizam	107
Poslovna informatika	75
Ponavljajući	
Ekonomija	25
Financije, računovodstvo i revizija	41
Marketinško upravljanje	36
Management i poduzetništvo	46
Turizam	59

Slika 16: Ručno prilagođeni izvorni podaci za preuzimanje

Izvor: izradio autor

Napomena: Podaci su modificirani u cilju zaštite privatnosti korisnika

12.2. Izgradnja skladišta podataka

Za skladištenje podataka, korištena je PostgreSQL baza podataka, a za prikaz podataka u grafičkom sučelju korišten je PgAdmin III. Cjelokupan sustav razvijan je na Linux Ubuntu 16.06. sustavu.

Za potrebe sustava, izgrađena su tri područna skladišta podataka, po jedno za svaku komponentu, što znači da je korišten bottom-up pristup bez centralnog skladišta podataka. Korištena je dvoslojna arhitektura s većim brojem nezavisnih područnih spremišta podataka.

Kao što je spomenuto, za potrebe objašnjenja procesa izrade sustave, nadalje u radu biti će objašnjeno područno skladište za Analizu učinkovitosti te sve korake – od preuzimanja do vizualizacije.

Unutar PgAdmin-a izrađeno je šest dimenzijskih tablica, po jedna za svaku dostupnu informaciju, te tablica činjenica. Svaka od dimenzijskih tablica, sadrži ID i promatranu dimenziju, kao što je vidljivo na donjoj slici.

	id [PK] serial	smjer character varying
1	9	EKONOMIJA
2	10	FINANCIJE, RAČUNOVODSTVO I REVIZIJA
3	11	MARKETINŠKO UPRAVLJANJE
4	12	MANAGEMENT I PODUZETNIŠTVO
5	13	TURIZAM
6	14	POSLOVNA INFORMATIKA
7	15	KULTURA I TURIZAM
8	16	ZNANOST O MORU
9	17	DC UMAG
10	18	FRR/FINANCIJSKI MANAGEMENT
11	19	INFORMATIKA
*		

Slika 17: Dimenzijska tablica studijskih smjerova

Izvor: autor (PgAdmin III)

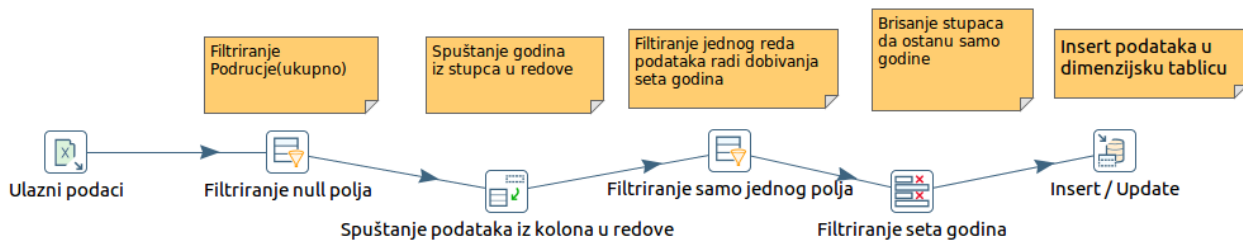
12.3. Punjenje dimenzijskih tablica

Za potrebe ovog rada, u ovom će se poglavlju objasniti jedna od procedura za punjenje dimenzijske tablice, one koje sadrže dimenziju akademske godine.

Za izradu procedura za preuzimanje i pohranjivanje podataka u bazu korišten je softver koji je dio Pentaho paketa za podatkovnu integraciju (još se naziva i Kettle), a zove se Spoon. Softver je iznimno moćan i posjeduje sve karakteristike koje jedan alat za podatkovnu integraciju mora postojati. Podržava čitav niz mogućnosti vezanih za ulazne podatke, od formata do različitih izvora, a isto vrijedi i za umetanje ili ažuriranje podataka u izlaznoj bazi, u ovom slučaju područnom skladištu podataka. Iako je baziran na Java programskom jeziku, kroz ponuđene korake (engl. Steps), za manipulaciju podacima mogu se koristiti kako npr. SQL upiti, pa sve do Javascripta. Programske procedure za preuzimanje podataka unutar Spoona nazivaju se transformacije (engl. Transformations).

Osim transformacija, moguće je kreirati i "jobove", koji su zapravo transformacije koje unutar svojih koraka pozivaju druge, ili eksterne, transformacije kreirane za posebne namjene. Za potrebe ovoga rada korištene su samo transformacije.

Na donjoj slici prikazana je cjelokupna transformacija podataka za stvaranja dimenzije za akademske godine. U nastavku će biti pojašnjen svaki korak unutar transformacije zasebno.

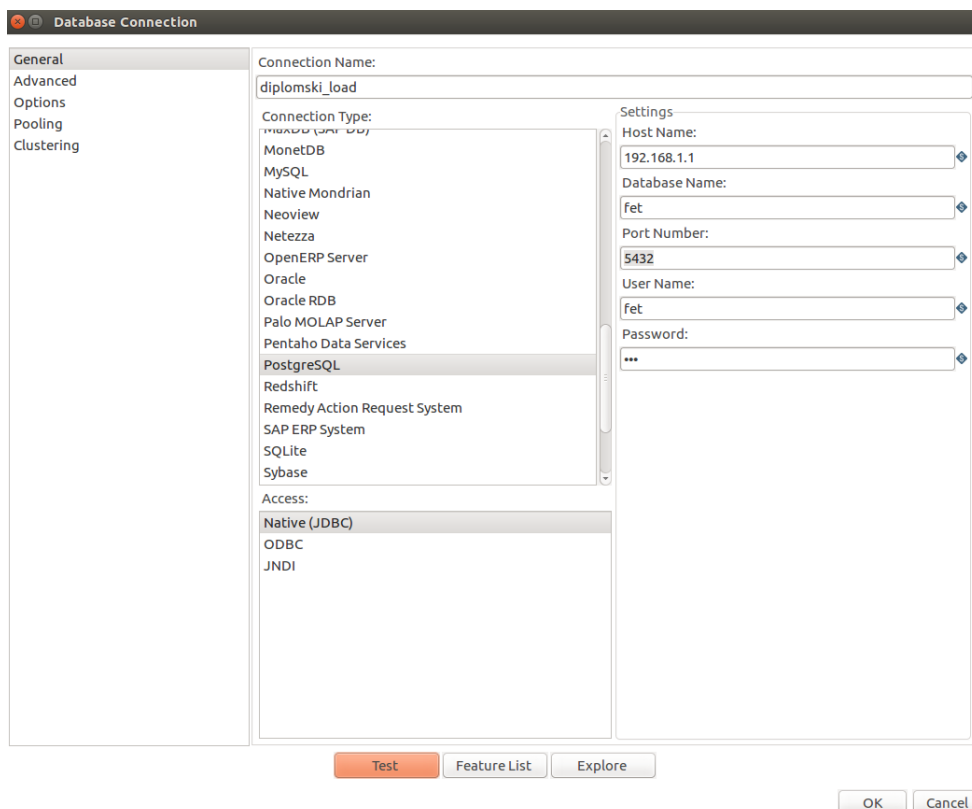


Slika 18: Primjer jednostavne transformacije u Spoon-u

Izvor: autor (Spoon)

Kao korak za umetanje svih ulaznih podataka korišten je korak Microsoft Excel input, koji omogućava umetanje klasičnih .xls ili .xlsx tablica. Osim samog umetanja tablice, korak nudi čitav niz mogućnosti manipulacije ulaznih podacima kako bi se podaci odmah na ulazu mogli što bolje pripremiti za daljnje korake. Primjerice, korak nudi mogućnosti učitavanja više tabličnih datoteka odjednom, kao i spreadsheetova. Također, moguće je birati od kojeg reda će se početi učitavati podaci, koje kolone će biti uključene i u kojem formatu (decimale, Regex i sl.). Ono što korak nudi, a može biti itekako korisno jest mogućnost automatskog uključivanja polja sa imenom datoteke(engl. filename) ili putanjom (engl. Path), a iz kojih se često mogu jednostavno nekoliko Javascript naredbi izvući bitni podaci, poput godina, grada, odjela i sl.

Kao završni korak za sve transformacije korištene u izradi sustava za ovaj rad, koristi se "Insert/Update". Radi se o koraku koji zapravo puni ili ažurira podatke u skladištu podataka. U ovom slučaju, obzirom da se radilo o prvom punjenju podataka, nije bilo potrebe za ažuriranjem. Da bi transformirane podatke mogli unijeti u skladište, potrebno je stvoriti konekciju na skladište sa ispravnim parametrima te autorizacijom.



Slika 19: Kreiranje konekcije na područno skladište podataka unutar Spoon-a
Izvor: autor (Spoon)

Po spajanju na bazu, određuju se kolone koje će se umetati u otprije kreiranu dimenzijsku tablicu u PostgreSQL klijentu. Informacije radi, tablice su kreirane jednostavnim SQL naredbama za kreiranje tablica. U slučajevima svih korištenih dimenzijskih tablica, kao primarni ključ postavljena je kolona "ID", tipa "serial", što znači da po umetanju novog reda u tablicu, pgAdmin automatski generira prvi idući slobodni "ID" broj. Nakon umetanja, rezultat izgleda ovako:

	id [PK] serial	ak_godina character varying
1	1	2015/16
2	2	2014/15
3	3	2013/14
4	4	2012/13
5	5	2011/12
6	6	2010/11
7	7	2009/10
8	8	2008/09
9	9	2007/08
10	10	2006/07
11	11	2005/06
*		

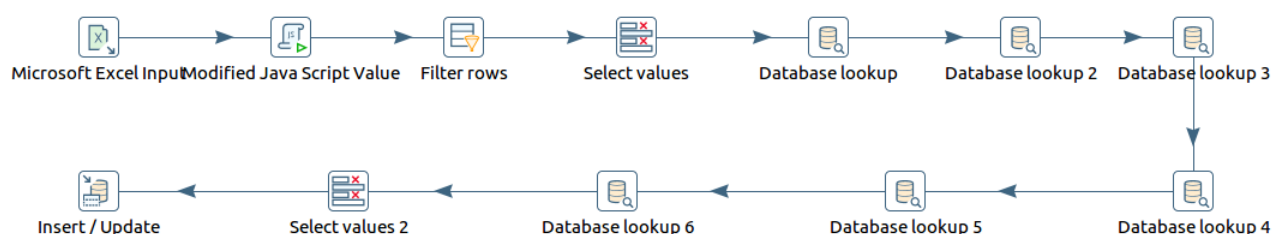
Slika 20: Izrađena dimenzijska tablica za akademske godine
Izvor: autor (pgAdmin III)

Dimenzijske tablice inače mogu sadržavati i do nekoliko tisuća podataka, no u ovom slučaju radi se samo o jedanaest redova, obzirom na broj promatranih akademskih godina, što je ipak dovoljno da se predoči postupak izrade i punjenja ovakvih tablica.

12.4. Punjenje tablica činjenica

Nakon što su se na isti način napunile sve potrebne dimenzijske tablice, pristupilo se izradi tablice činjenica za Analizu učinkovitosti studenata.

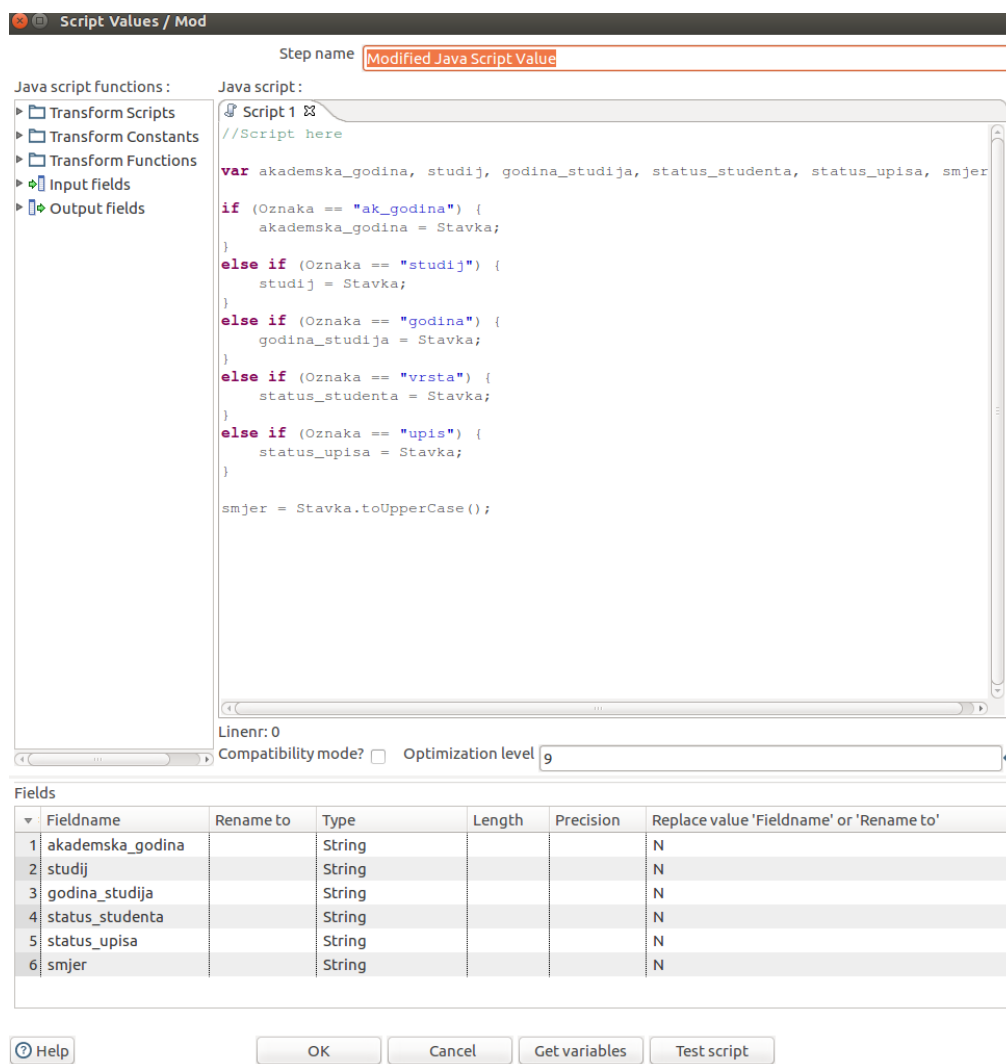
Tablica činjenica se izradila na način da se na ulazu koristila glavna excel tablica, a kroz korake su se radili tzv. pogledi u bazu (engl. Database lookup), odnosno tražila su se podudaranja sadržaja ćelija iz dimenzijskih tablica, na izlazu su se vraćali ID-evi iz dimenzijskih tablica. Ukratko, pomoću vršenja usporedbi vrijednosti glavne tablice i kolona u dimenzijskim tablicama, generirala se tablica vraćenih primarnih ključeva iz svih dimenzijskih tablica koje čine ključ za generiranje OLAP upita. U tablici činjenica dobiveni ID-evi iz dimenzijskih tablica čine skupa primarni ključ koji na upit vraćaju vrijednost kao mjeru, a u ovom slučaju postoji samo jedna mjera – broj studenata.



Slika 21: Procedura za izradu tablice činjenica za Analizu učinkovitosti studenata

Izvor: autor (Spoon)

Transformacija počinje ulaznim korakom i učitavanjem glavne tablice. Takvi se podaci šalju u drugi korak, gdje se konkretno može vidjeti uporaba "Modified Java Script Value", odnosno manipulacija podacima pomoću Javascript programskog jezika. U ovom slučaju, radi se o jednom o najvažnijih koraka u ovoj transformaciji.



Slika 22: Prikaz upotrebe Javascript koraka kod formiranja tablice činjenica

Izvor: autor (Spoon)

Obzirom da su sve stavke (pogled na sliku ulazne excel tablice) u jednoj koloni naziva "Stavka", potrebno je podatke razložiti u više kolona koje odgovaraju vrijednostima sa kojima će se uspoređivati u dimenzijskim tablicama. Kreiranjem varijabli, nastaju nove kolone sa datim imenom, a koje se koriste dalje u transformaciji.

U idućem se koraku filtriraju eventualni prazni redovi ili redovi sa null vrijednošću, te se odabiru vrijednosti (kolone) koje će biti vidljive i dostupne za vršenje pogleda u bazu. Tada na red dolaze pogledi u bazu, ključni korak u kreiranju tablice činjenica. Pogled u bazu kao korak vrši se za svaku dimenziju zasebno. U slučaju ove konkretne tablice činjenica, pogled se radio prema vrijednosti polja u koloni glavne tablice i pripadajuće

dimenzijske tablice. Pa stoga, ako se uzme za primjer pogled u dimenzijsku tablicu akademskih godina, za polje sa vrijednosti "2015/16", iščitati će se pripadajući ID (u ovom slučaju "1"). Isti je slučaj i sa bilo kojom drugom dostupnom godinom, pa će se za vrijednost "2005/06", vratiti ID vrijednosti "11".

Važno je napomenuti da su ovakvi pogledi vršeni na identičan način za sve dimenzijske tablice (studijski smjerovi, status studenta itd.). Nakon izvršenih pogleda, filtriraju se konačne kolone koje se moraju naći u tablici činjenica, a radi se o svim dimenzijskim ID-evima te pripadajuće vrijednosti kombinaciji ključeva (broj studenata).

	smjer_id [PK] integer	ak_god_id [PK] integer	tip_studija_id [PK] integer	godina_studija_id [PK] integer	status_id [PK] integer	upis_id [PK] integer	broj_studenata integer
1	9	1	1	1	1	1	28
2	9	1	1	1	1	2	6
3	9	1	1	1	2	1	16
4	9	1	1	1	2	2	2
5	9	1	1	2	1	1	22
6	9	1	1	2	1	2	9
7	9	1	1	2	2	1	10
8	9	1	1	2	2	2	4
9	9	1	1	3	1	1	18
10	9	1	1	3	1	3	10
11	9	1	1	3	2	1	3
12	9	1	1	3	2	3	28
13	9	1	2	1	1	1	3
14	9	1	2	1	1	2	0
15	9	1	2	1	2	1	4
16	9	1	2	1	2	2	0
17	9	1	2	2	1	1	2
18	9	1	2	2	1	3	5
19	9	1	2	2	2	1	3
20	9	1	2	2	2	3	3
21	9	2	1	1	1	1	33
22	9	2	1	1	1	2	2
23	9	2	1	1	2	1	16
24	9	2	1	1	2	2	4
25	9	2	1	2	1	1	21
26	9	2	1	2	1	2	10
27	9	2	1	2	2	1	9
28	9	2	1	2	2	2	2
29	9	2	1	3	1	1	7
30	9	2	1	3	1	3	11
31	9	2	1	3	2	1	2
32	9	2	1	3	2	3	37
33	9	2	2	1	1	1	3
34	9	2	2	1	1	2	0
35	9	2	2	1	2	1	4
36	9	2	2	1	2	2	1
37	9	2	2	2	1	1	4
38	9	2	2	2	1	3	9
39	9	2	2	2	2	1	2
40	9	2	2	2	2	3	2
41	9	3	1	1	1	1	31
42	9	3	1	1	1	2	7
43	9	3	1	1	2	1	9
44	9	3	1	1	2	2	1

Slika 23: Ispunjena tablica činjenica za Analizu učinkovitosti studenata

Izvor: autor (pgAdmin III)

12.5. Izrada Mondrian sheme

Mondrian shema (engl. Mondrian schema) je izrađena ručno, odnosno bez korištenja dodatnih alata. Radi se o ručnom pisanju podatkovne datoteke u izvornom (.xml) formatu prema pravilima koja koristi Mondrian OLAP engine, a koja su dostupna na službenoj Mondrian stranici (<http://community.pentaho.com/projects/mondrian/>). Mondrian je također dio Pentaho sustava, a služi za stvaranje OLAP kocki te time omogućava rukovođenje velikim količinama podataka strukturiranih u dimenzije.

```
<?xml version="1.0" encoding="UTF-8"?>
<Schema name="fet" measuresCaption="Measures" defaultRole="Authenticated">
  <!-- ===== DIMENZIJA PO SMJEROVIMA ===== -->
  <Dimension name="smjerovi" caption="SMJEROVI">
    <Hierarchy name="smjerovi" hasAll="true" allMemberName="SVI SMJEROVI" primaryKey="id" caption="SMJEROVI">
      <Table name="smjerovi"/>
      <Level name="Smjerovi" column="smjer" captionColumn="smjer" type="String" uniqueMembers="false">
      </Level>
    </Hierarchy>
  </Dimension>
  <!-- ===== DIMENZIJA PO AKADEMSKIM GODINAMA ===== -->
  <Dimension name="godine" caption="GODINE">
    <Hierarchy name="godine" hasAll="true" allMemberName="SVE GODINE" primaryKey="id" caption="GODINE">
      <Table name="godine"/>
      <Level name="Godine" column="ak_godina" captionColumn="ak_godina" type="String" uniqueMembers="false">
      </Level>
    </Hierarchy>
  </Dimension>
  <!-- ===== DIMENZIJA PO GODINI STUDIJA ===== -->
  <Dimension name="godina_studija" caption="GODINA STUDIJA">
    <Hierarchy name="godina_studija" hasAll="true" allMemberName="SVE GODINE STUDIJA" primaryKey="id" caption="GODINE STUDIJA">
      <Table name="godina_studija"/>
      <Level name="Godina_studija" column="godina_studija" captionColumn="godina_studija" type="String" uniqueMembers="false">
      </Level>
    </Hierarchy>
  </Dimension>
  <!-- ===== DIMENZIJA PO PODRUČJU I VLASNIŠTVU USTANOVE ===== -->
  <Dimension name="podrucja_po_vlasnistvu" caption="BROJ STUDENATA PO PODRUČJIMA I VLASNIŠTVU">
    <Hierarchy name="podrucja_po_vlasnistvu" hasAll="true" allMemberName="PODRUČJA PO VLASNIŠTVU" primaryKey="id" caption="PODRUČJA PO VLASNIŠTVU">
      <Table name="podrucja_po_vlasnistvu"/>
      <Level name="Područje" column="podrucje" captionColumn="podrucje" type="String" uniqueMembers="false">
      </Level>
      <Level name="Vlasništvo" column="vlasnistvo" captionColumn="vlasnistvo" type="String" uniqueMembers="false">
      </Level>
    </Hierarchy>
  </Dimension>
  <!-- ===== DIMENZIJA PO STATUSU STUDENTA ===== -->
  <Dimension name="status_studenta" caption="STATUS STUDENTA">
    <Hierarchy name="status_studenta" hasAll="true" allMemberName="UKUPNO STATUS" primaryKey="id" caption="STATUS STUDENTA">
      <Table name="status_studenta"/>
      <Level name="status_studenta" column="status_studenta" captionColumn="status" type="String" uniqueMembers="false">
      </Level>
    </Hierarchy>
  </Dimension>
  <!-- ===== DIMENZIJA PO STATUSU UPISA ===== -->
  <Dimension name="status_upisa" caption="STATUS UPISA">
    <Hierarchy name="status_upisa" hasAll="true" allMemberName="SVI STATUSI UPISA" primaryKey="id" caption="STATUSI UPISA">
      <Table name="status_upisa"/>
      <Level name="Status upisa" column="status_upisa" captionColumn="status_upisa" type="String" uniqueMembers="false">
      </Level>
    </Hierarchy>
  </Dimension>
</Schema>
```

Slika 23: Definiranje dimenzija unutar Mondrian sheme

Izvor: izradio autor

Na gornjoj su slici vidljive definirane dimenzije koje se koriste u određenoj kocki. Dimenzije, kao i kocke, mogu sadržavati čitav niz atributa, pa i specifičnih mjera i oznaka, poput kalkuliranih mjera direktno unutar kocke. U ovom slučaju nije bilo potrebe za upotrebom tih elemenata, već su se koristili osnovni potrebni atributi. U definiranju dimenzije u shemi koristili su se atributi poput "name" i "caption", koje je veoma lako zamijeniti. To je vrlo važno za napomenuti, iz jednostavnog razloga što se "name" koristi

za definiranje kocki unutar sheme, dok je "caption" zapravo naziv za dimenziju koristi u analitičkom sučelju kod slaganja upita na skladište.

Dimenzije sadrže hijerarhije kao svoje "dijelove". Hijerarhije su sastavni dio dimenzija te njihovo definiranje ovisi o planiranom sustavu, podacima, ali i preferencijama osobe koja razvija sustav. Obzirom da i same hijerarhije imaju svoje niže članove – razine (engl. Level), jasno je da često ima prostora slobodi u pisanju dimenzija.

Od važnijih korištenih atributa za hijerarhiju izdvojeni su i "hasAll" i "allMemberName". Oba atributa se zapravo odnose na zbroj vrijednosti iz tablica, s napomenom da "hasAll" određuje hoće li se zbrojeni član koristiti, a "allMemberName", kao i u slučaju "caption", predstavlja ime koje će biti prikazano u analitičkom sučelju ili vizualizaciji, ukoliko se spomenuti član koristi. Da bi bilo lakše shvatljivo, najčešće se radi o članu imena "Ukupno" ili "Total", ali to nikako nije pravilo. Kao obavezan dio hijerarhije oblatorno je navesti tablicu u skladištu iz koje se povlače potrebni podaci. Kada je definirana tablica kao izvor, unutar svake hijerarhije moguće je kreirati razine. Na gornjoj slici, u dimenziji "Po području i vlasništvu ustanove", vidljivo je da se koristila samo jedna hijerarhija sa dvije razine. Upotreba razina ovisna je o podacima, odnosno ako su podaci koncipirani na način da nije moguće dobiti razine, onda se one dakako, neće upotrebljavati. U ovom slučaju razine su nazvane "Područje" i "Vlasništvo" sa svim pripadajućim atributima razina, poput definiranja kolone tablice u kojoj se razina nalazi, njen "captionColumn" koji definira ime razine unutar analitičkog sučelja, primarnog ključa (najčešće ID) te "uniqueMembers" koji definira da li će se eventualni istovjetni retci zbrajati da bi imali samo unikatne redove. Od velike je važnosti razumjeti da su razine definirane unutar iste hijerarhije neraskidivo povezani te će se unutar "analitike" oni umetati ili skupno ili nikako. Stoga je potrebno dobro isplanirati upotrebu razina, za razliku od hijerarhija koje su mnogo fleksibilnije za strukturiranje.

```
<!-- ===== KOCKA PO USTANOVAMA I GRADOVIMA ===== -->
<Cube name="analiza_ucinkovitosti" caption="Analiza učinkovitosti">
  <Table name="fact_au"/>
  <DimensionUsage name="smjerovi" source="smjerovi" foreignKey="smjer_id"/>
  <DimensionUsage name="godine" source="godine" foreignKey="ak_god_id"/>
  <DimensionUsage name="tip_studija" source="tip_studija" foreignKey="tip_studija_id"/>
  <DimensionUsage name="godina_studija" source="godina_studija" foreignKey="godina_studija_id"/>
  <DimensionUsage name="status_studenta" source="status_studenta" foreignKey="status_id"/>
  <DimensionUsage name="status_upisa" source="status_upisa" foreignKey="upis_id"/>
  <Measure name="Broj studenata" caption="Broj studenata" column="broj_studenata" visible="true" aggregator="sum" type="String" formatString="#"/>
</Cube>
```

Slika 24: Definiranje OLAP kocke unutar Mondrian sheme

Izvor: izradio autor

Na gornjoj je slici vidljiv definiran primjer jedne OLAP kocke, konkretno za Analizu učinkovitosti. Kocka sama po sebi ne koristi ne koristi mnogo atributa, samo osnovne, iz jednostavnog razloga što se najbitniji atributi definiraju upravo u dimenzijama. U kocki se zapravo povezuju sve kreirane dimenzije i povezuju sa mjerama, odnosno vrijednostima koje posjedujemo iz glavne tablice. Kod razine kocke definira se njeno ime i "caption", a unutar nje poveznica na tablicu činjenica otprije pohranjenu u skladište. Po definiranju tablice, u kocku se uključuju sve potrebne dimenzije (također vidljivo na gornjoj slici).

Upotreba dimenzija u kocki se također označava imenima, ali atributi koji su zaista bitni u ovom slučaju jesu "source" i "foreignKey". Unutar "source" se definira ime definirane dimenzije koja će se koristiti, a unutar "foreignKey" definira se strani ključ, odnosno ključ za datu dimenziju koji se nalazi u tablici činjenica, a pomoću kojeg se tablica činjenica povezuje na primarni ključ potrebne dimenzije te tako dobiva povratnu informaciju i radi "križanja" podataka. Kao posljednji dio kocke u pravilu se navodi mjera (engl. "Measure").

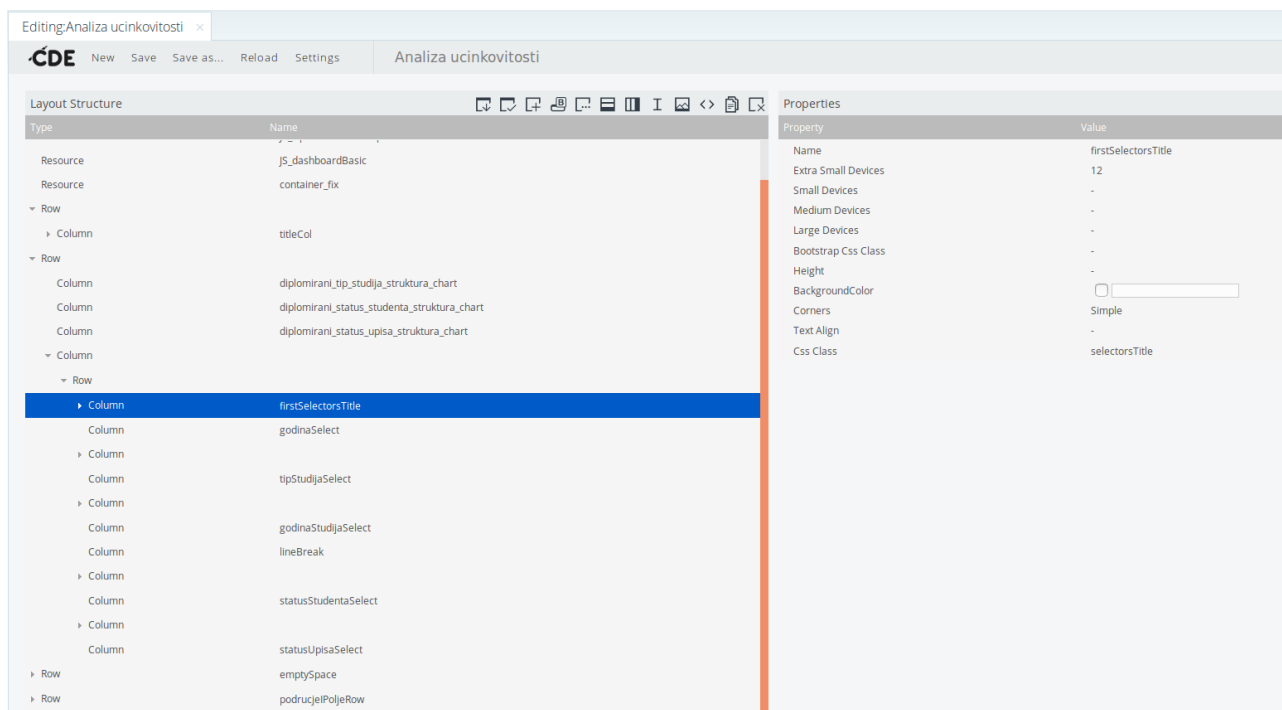
U ovom slučaju postoji samo jedna mjera koja se prati – broj studenata, a upotrebom primarnih ključeva tablice činjenica dobijamo odgovarajuće polje koje sadrži broj studenata. Mjeri se može pridružiti niz atributa, poput "name" i "caption", kolone iz koje se dobijaju podaci "column", formata, tipa podataka itd.

12.6. Izrada upravljačke ploče sa preuzetim podacima

Za izradu sučelja i upravljačke ploče korišten je Pentaho CDE alat, koji je također dio Pentaho platforme i otvorenog je koda. Kod ovoga je alata važno za napomenuti da ima vlastiti sustav autorizacije (sistem korisničkog imena i lozinke) sa kojom se prijavljuje u sustav te pristupa sučelju za izradu upravljačkih ploča.

Sučelje je podijeljeno u tri glavna dijela: Layout Panel, Components Panel i Datasources Panel.

U Layout panelu definira se struktura stranice. Obzirom da CDE alat kao zadan framework koristi Bootstrap, vrlo je jednostavno čak i laicima sastaviti osnovnu strukturu upravljačke ploče, uz mogućnost praktički potpune personalizacije upotrebom pluginova i običnom CSS ili Javascript koda, kroz interne skripte ili eksterne datoteke.



Slika 25: Struktura upravljačke ploče korištenjem Layout Panela

Izvor: autor (CDE)

Uz postavljanje "kostura" stranice, za svaki dio strukture moguće je kroz grafičko sučelje odrediti parametre poput širine, visine, pozadinske boje elementa i sl., uz mogućnost definiranja vlastitih HTML klasa i id-eva za svaki element, za kasnije eventualno personaliziranje.

U Components Panelu definiraju se sve komponente koje će se koristiti unutar upravljačke ploče – grafovi, izbornici, parametri, datumi, widgeti i sl. Popis dostupnih komponenata je zaista širok te je dostupno preko 20-ak vrsta grafova i još toliko izbornika, widgeta i sl.

Odabirom komponente, otvara se mogućnost podešavanja njenih značajki, a među najvažnijima su izvor podataka, naslov komponente, korišteni parametri iz upita, HTML objekt u koji će komponenta biti smještena, parametri koje određeni element sluša (engl. Listeners) definirane boje kod grafova, vrste legendi i još mnogo toga. Mogućnosti ima dovoljno da se svaka komponenta može podesiti točno na način na koji to razvojnici želi.

Analiza ucinkovitosti			
Type	Name	Property	Value
▼ Group	Generic	Name	tip_studija_chart
Simple parameter	tipStudijaParam	Title	KRETANJE UČINKOVITOSTI STUDENATA
Simple parameter	statusStudentaParam	Listeners	[tipStudijaParam', (...]
Simple parameter	statusUpisaParam	Parameters	[["tip_studija","tip (...]
Simple parameter	godinaStudijaParam	Datasource	diplomirani_trend_chart_query
Simple parameter	godinaParam	Height	-
▼ Group	Selects	Width	-
Select Component	tipStudijaSelector	HtmlObject	diplomirani_trend_chart
Select Component	statusStudentaSelector	clickable	True
Select Component	statusUpisaSelector	clickAction	
Select Component	godinaStudijaSelector	compatVersion	2
Select Component	godinaSelector	crosstabMode	True
▼ Group	Charts	legend	True
CCC Bar Chart	tip_studija_chart	seriesInRows	False
CCC Pie Chart	tip_studija_struktura_chart	timeSeries	False
CCC Pie Chart	status_studenta_struktura_chart	timeSeriesFormat	%Y-%m-%d
CCC Pie Chart	status_upisa_struktura_chart		

Slika 26: Definirane komponente u upravljačkoj ploči "Analiza učinkovitosti"

Izvor: autor (CDE)

Datasource Panel je zapravo temelj na kojem se bazira upravljačka ploča. Unutar ove komponente definiraju se upiti prema skladištu podataka potrebni za pokretanje svih komponenti. Za kreiranje upita može se koristiti MDX ili klasičan SQL, s tim da sa MDX načinom je moguće dobiti puno kompleksnije upite, a i mnogo je lakše implementirati parametre koji su promjenjivi (primjerice odabrane stavke u izbornicima). Za potrebe ovoga rada izrađeni su MDX upiti za komponente.

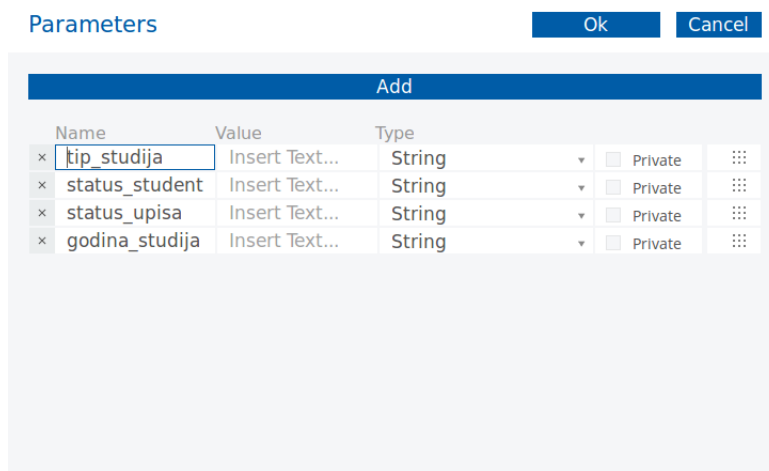
Analiza ucinkovitosti			
Datasources		Properties	
Type	Name	Property	Value
▼ Group	MDX Queries	Name	diplomirani_trend_chart_query
mdx over mondrianJndi	diplomirani_trend_chart_query	Access Level	Public
mdx over mondrianJndi	tip_studija_selector_query	Jndi	fet
mdx over mondrianJndi	status_studenta_selector_query	Mondrian schema	fet
mdx over mondrianJndi	status_upisa_selector_query	Query	SELECT NON EMPTY ([s (...]
mdx over mondrianJndi	godina_studija_selector_query	Parameters	[["tip_studija","", "(...]
mdx over mondrianJndi	godina_selector_query	Banded Mode	Compact
mdx over mondrianJndi	diplomirani_tip_studija_struktura_query	Calculated Columns	[]
mdx over mondrianJndi	diplomirani_status_studenta_struktura_query	Columns	[]
mdx over mondrianJndi	diplomirani_status_upisa_struktura_query	Output Options	[]
		Output Mode	Include
		Cache Keys	[]
		Cache Duration	3600
		Cache	True

Slika 27: Definirani upiti unutar Datasource Panela

Izvor: autor (CDE)

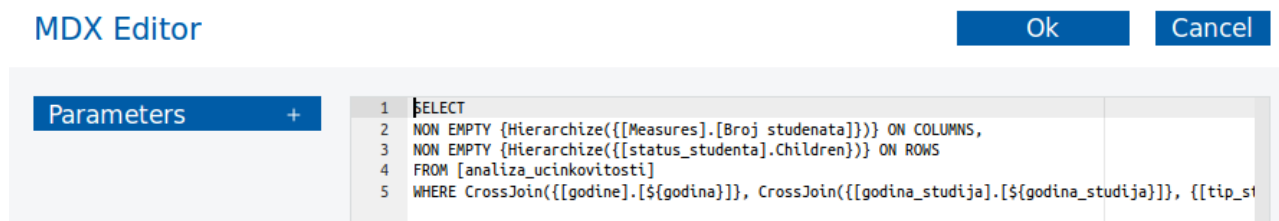
Kod upita moguće je definirati niz značajki potrebnih da bi se upit normalni izvršavao, od predviđanja parametara, odabira Mondrian sheme koja je otprije definirana i

koja je zapravo pozadina svakog napravljenog upita te u konačnici i pisanja samog upita pomoću MDX sintakse. Valja napomenuti da je MDX ima prilično kompleksnu sintaksu te je početniku potrebno određeno vrijeme da počne lakše vladati kreiranjem upita. Promjenjivi parametri (najčešće dobiveni iz izbornika), definiraju se na idući način - $\$ \{parametar_xy\}$.



Slika 28: Upotreba i definiranje parametara kod izrade upita

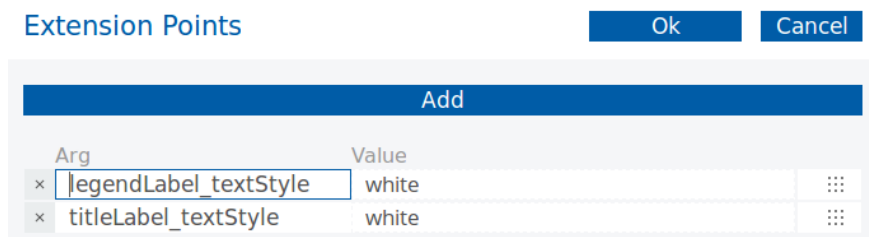
Izvor: autor (CDE)



Izvor: autor (CDE)

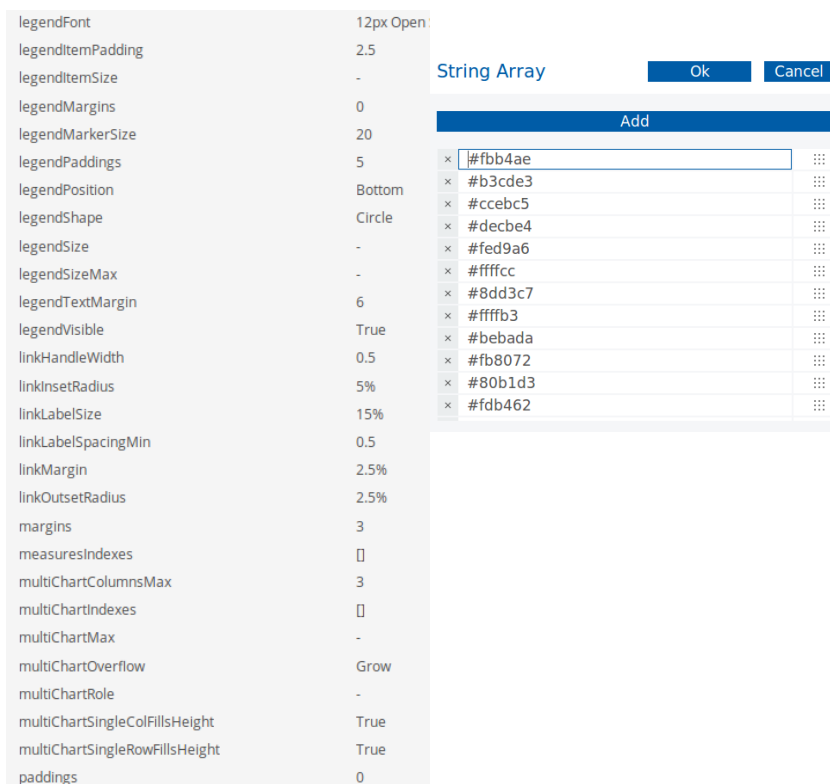
12.7. Značajke kreiranih grafova za upravljačku ploču

Za potrebe ovoga rada i demonstriranja mogućnosti CDE alata, dio značajki grafova i upravljačke ploče uopće definiran je kroz ponuđene opcije u sučelju Components Panela, a dio manualnim pisanjem koda kroz eksterne .css i .js datoteke. Unutar sučelja definirane su boje grafova, boje osi, veličina i lokacija legende, format tooltipa, font, hover opcija, interaktivnosti i pomicanje dijelova grafa prelaskom miša itd. Opcija značajki ima previše da bi se mogle prikazati vizualno. Za svaki graf, samo kroz sučelje, moguće je podešavati više od stotinu značajki.



Slika 30: Upotreba ekstenzija CDE alata

Izvor: autor (CDE)



Slika 31: Primjer postavljanja zadanih značajki i boja za grafove

Izvor: autor (CDE)

```

1 function f() {
2   var series = godinaParam;
3   var category = this.getCategoryLabel();
4   var v = this.getValue();
5   return "<div class='godina'>Promatrana akademska godina<br><b>" + series + "</b></div>" +
6         "<div class='godina'>Promatrana godina studija<br><b>" + godinaStudijaParam + "</b></div>" +
7         "<div class='godina' style='height:1px; width:95%; margin:0 auto; background:#b09a54'>" +
8         "<div class='kategorija'>Tip studija<br><b>" + category + "</b></div>" +
9         "<div class='vrijednost'>Broj studenata<br><b>" + v + "</b></div>";
10 }

```

Slika 32: Javascript funkcija za formatiranje tooltipa za grafove

Izvor: autor (CDE – Javascript editor)

Drugi dio upravljačke ploče definiran je ručnim pisanjem koda. Za upravljačke ploče u ovom radu korištena se četiri eksterna resursa, od kojih dva predstavljaju srž izgleda i ponašanja upravljačke ploče – "JS_dashboardBasic" i "CSS_dashboardGeneral".

Layout Structure	
Type	Name
Resource	CSS_dashboardGeneral
Resource	JS_OpenSansFontScript
Resource	JS_dashboardBasic
Resource	container_fix

Slika 33: Korišteni resursi za upravljačke ploče

Izradio: autor (CDE)

"CSS_dashboardGeneral" je stylesheet datoteka u kojoj se nalaze definicije elemenata koji nisu grafovi – pozadina, naslov, selektori, boje i sl.

"JS_dashboardBasic" je Javascript datoteka u kojoj su definirana ponašanja HTML elemenata poput izbornika, dodavanja potrebnih klasa na klik ili prelazak miša itd.

Radi se o dvije datoteke koje sadrže naredbe za dodatno personaliziranje upravljačke ploče koje inače ne bi bilo moguće kroz predefinirano sučelje.

```
.tipsy-inner {
  background: black;
  font-family: 'Open Sans', sans-serif;
  border: 2px solid #b09a54;

  -webkit-box-shadow: 0px 0px 4px rgba(0, 0, 0, .4);
  -msie-box-shadow: 0px 0px 4px rgba(0, 0, 0, .4);
  -moz-box-shadow: 0px 0px 4px rgba(0, 0, 0, .4);
  -o-box-shadow: 0px 0px 4px rgba(0, 0, 0, .4);
  box-shadow: 0px 0px 4px rgba(0, 0, 0, .4);

  padding: 8px 15px 9px;
  color: white;
  max-width: 480px;
}
.tipsy-s .tipsy-arrow {
  background-position: center bottom;
  bottom: 0;
  left: 50%;
  margin-left: -4px;
}
.tooltipTitle {
  color: #454545;
  display: block;
  text-align: center;
  font-size: 14px;
  margin-bottom: 5px;
  line-height: 1.5;
  padding: 5px;
  font-weight: 300;
}
```

Slika 34: Dio koda iz "CSS_dashboardGeneral" datoteke

Izvor: autor (CDE – CSS editor)

```

/***** IZRAČUN VISINE KOD CHARTOVA *****/
dashboard.calculateChartHeight = function(chartObj) {

    $('.container-fluid').css("height", "100vh");

    var containerHeight = $(".container-fluid").height();
    var upperCharts = containerHeight * 0.35;
    var lowerCharts = containerHeight * 0.45;

    var titleHeight = $(".container-fluid").height() * 0.04;
    $('#title').css({'height':titleHeight, 'padding-top':titleHeight * 0.3});

    var emptySpace = $(".container-fluid").height() * 0.02;
    $('.emptySpace').css('height', titleHeight);

    chartObj.chartDefinition.height = lowerCharts;

};

/***** IZRAČUN VISINE KOD PIE CHARTOVA *****/
dashboard.calculatePieChartHeight = function(chartObj) {

    $('.container-fluid').css("height", "100vh");

    var containerHeight = $(".container-fluid").height();
    var upperCharts = containerHeight * 0.37;
    var lowerCharts = containerHeight * 0.47;

    var titleHeight = $(".container-fluid").height() * 0.05;
    $('#title').css('height', titleHeight);

```

Slika 35: Dio koda iz "JS_dashboardBasic" datoteke

Izvor: autor (CDE – Javascript editor)

Na slikama 34. i 35. prikazani su djelići CSS i Javascript koda korištenih u izradi upravljačkih ploča, te pokazuje na koji se način oni mogu koristiti pri postizanju vizualne atraktivnosti i dinamičnosti elemenata.

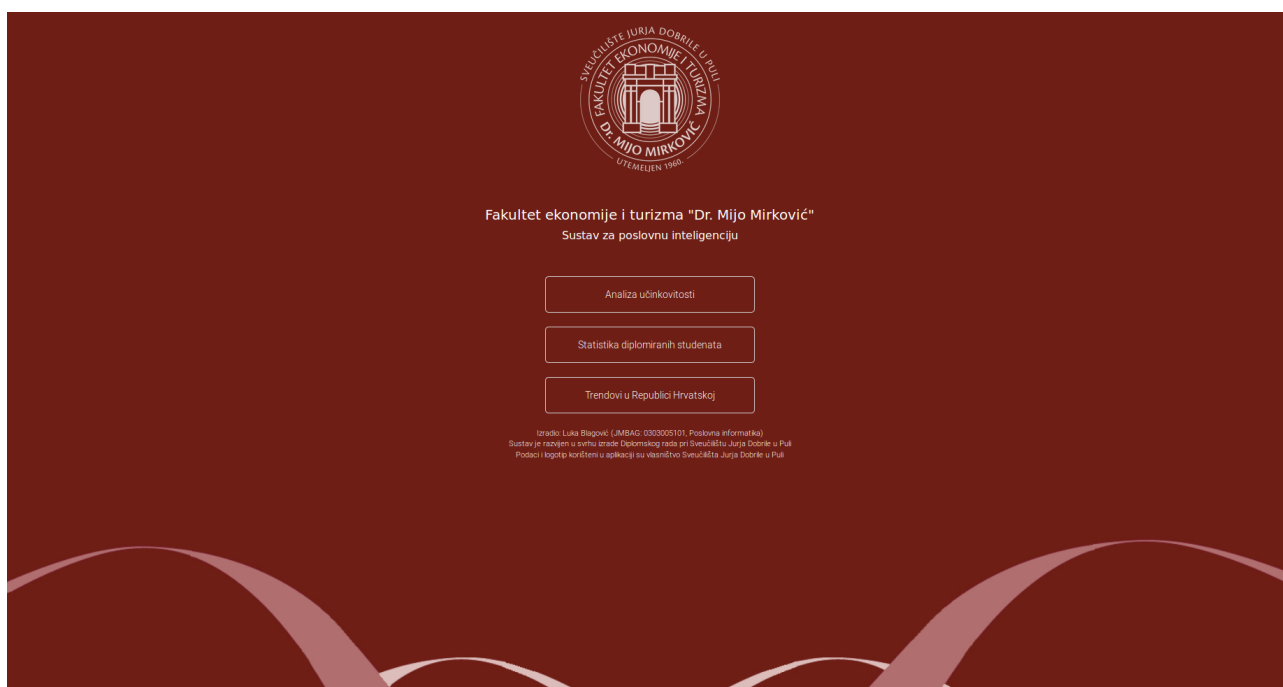
12.8. Izrada ulazne stranice i korištenje vizualizacija

Za izradu ulazne stranice sustava korištena je klasična kombinacija HTML-a za strukturu, CSS-a za izgled te Javascripta za definiranje dinamičkih elemenata. Uz to, još je korišten Bootstrap framework i jQuery library kao eksterni resursi.

Kompletna uvodna stranica je kodirana manualno u HTML editoru "Brackets", a usklađena je sa vizualnim identitetom FET-a, po zahtjevu korisnika.

U izborniku su ponuđene opcije ulaska u tri moguće upravljačke ploče:

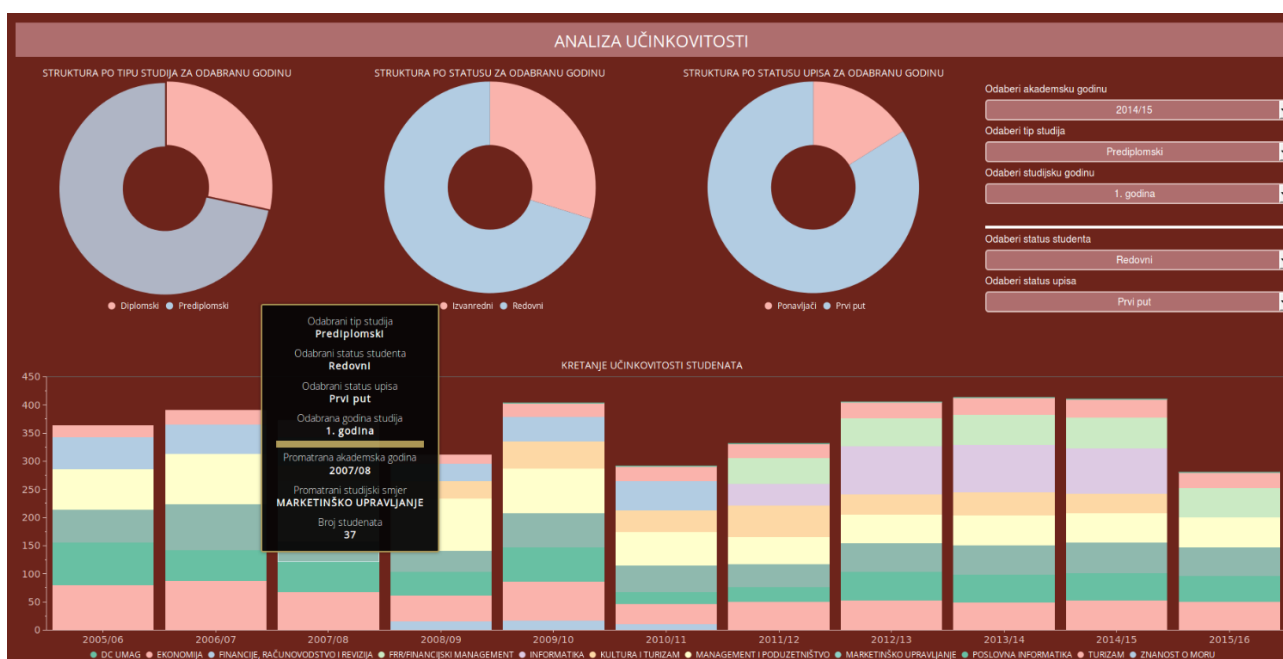
- 1) Analiza učinkovitosti
- 2) Statistika diplomiranih studenata
- 3) Trendovi u Republici Hrvatskoj



Slika 36: Ulazna stranica BI sustava

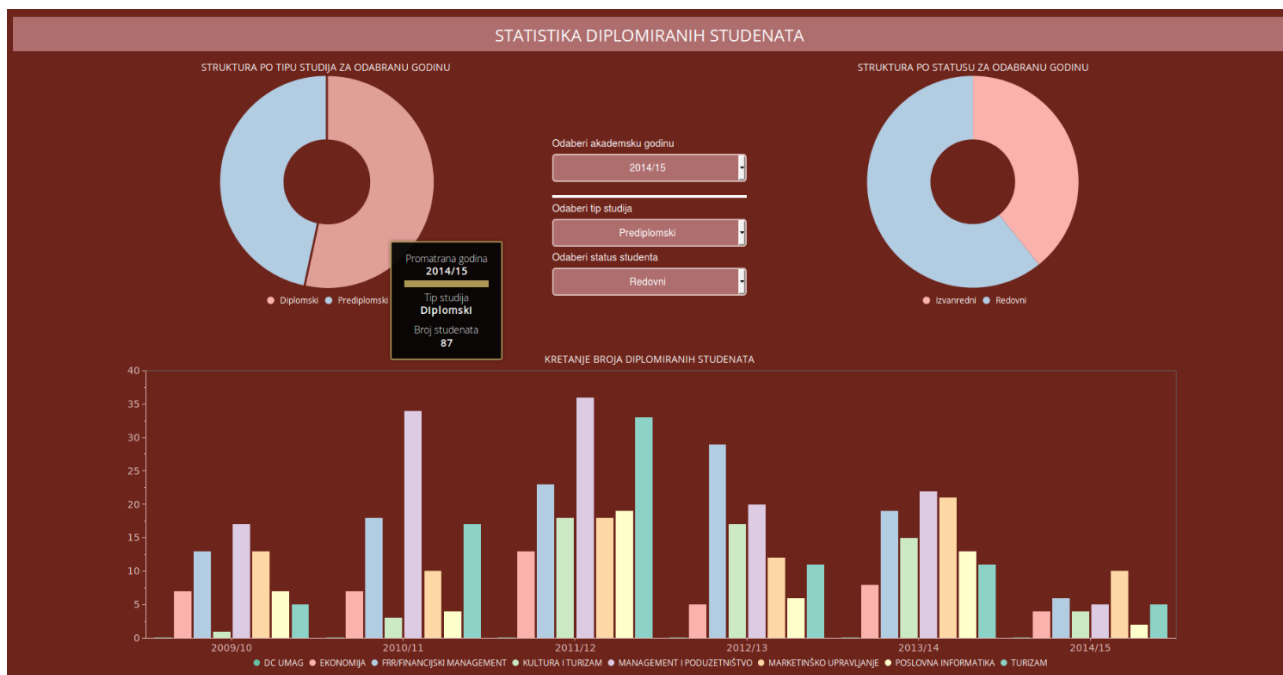
Izvor: autor

Link (coin.hr/lblagovic/fet-bi/index.html)



Slika 37: Upravljačka ploča – "Analiza učinkovitosti"

Izvor: autor



Slika 38: Upravljačka ploča – "Statistika diplomiranih studenata"

Izvor: autor



Slika 39: Upravljačka ploča – "Trendovi u Republici u Hrvatskoj"

Izvor: autor

U ovom je poglavlju objašnjena procedura izrade upravljačkih ploča s vizualizacijom podataka. Od preuzimanja, transformiranja, izrade skladišta i OLAP kocki, definiranje komponenti i personaliziranja vizualizacija, sve su to koraci izrade jednog sustava za poslovnu inteligenciju s naglaskom na vizualizaciju podataka i praćenje trendova ili strukture.

13. Zaključak

Nakon opisa postupka izrade sustava poslovne inteligencije, te dokazivanja da sustav poslovne inteligencije, i vizualizacija kao njegovih važnih dijelova, itekako utječu na učinkovitost poslovanja, jasno je vidljivo da ovako koncipirani podaci mogu stvarati prevagu na tržištu i među konkurencijom. Velike količine podataka koje se poslovanjem neminovno generiraju, kroz ovakve sustave dobivaju potpuno novi smisao. Upravo je nevjerojatna činjenica da istraživanje nepreglednih dokumenata i tablica u potrazi za određenim podatkom ili podatkovnim trendom, u stvarnosti može trajati danima, tjednima ili mjesecima, a da kroz ovakav sustav taj podatak moguće je dobiti kroz nekoliko sekundi pomoći nekoliko klikova mišem. Ako se usporede navedeni vremenski gabariti, nije potrebno posebno naglasiti koliko ovakvi sustavi donose funkciji upravljanja u organizaciji. Sustavi poslovne inteligencije više nisu pojam rezerviran za velike i bogate organizacije, već je dostupan svima koji shvaćaju da je takvo ulaganje zapravo isplativo ulaganje. Nije samo vrijeme dimenzija na koju utječe upotreba sustava, tu je i učinkovitost rukovoditelja i osoblja, korištenjem jednostavno generiranih podataka pretvorenih u smislene interaktivne grafove sa "user-friendly" sučeljima. Na taj se način izrađuju sustavi potencijalno namijenjeni bilo kome.

Samo kreiranje sustava jeste dugotrajan i ponekad mukotrpan proces za dizajnere i programere, no planiranje i priprema sustava čine ključ za uspješno korištenje sustava i zadovoljavanje korisničkih potreba na duge staze. Ispravnim postavljanjem sustava u početku, stvara se temelj za kvalitetno korištenje sustava u vidu praćenja učinkovitosti i postavljenih ciljeva. Vrlo je važno u početku znati jasno definirati što je u tom sustavu potrebno, a što nije, i na koji način mora biti koncipiran.

Sustavi poslovne inteligencije već duže vrijeme nisu nikakva nepoznanica, već široko rasprostranjen pojam, a danas vjerojatno i ne postoji veća tvrtka ili organizacija koja ne koristi neku verziju takvog sustava, u smislu vlastitog razvoja ili implementacije gotovih rješenja.

Sažetak

U ovom radu obrađena je primjena postojećih znanja o razvoju sustava poslovne inteligencije u jednoj organizaciji na praktičnom primjeru. Rad teoretski objašnjava sve važne segmente razvoja – preuzimanje i skladištenje podataka, upotrebu ETL procesa, izrade podatkovnih kocki, planiranje i izradu upravljačkih ploča, te kao poseban dio naglašava izradu vizualizacija podataka koje djeluju kao komponente upravljačkih ploča.

U današnje vrijeme informacija je moć, te predstavlja jedan od ključnih resursa odlučivanja i upravljanja. Na tragu navedenog, razvijaju se sustavi poslovne inteligencije koji imaju za cilj zadovoljenje korisničkih potreba u segmentu upravljanja organizacijom. Razvoj ovakvih sustava započinje definiranjem korisničkih zahtjeva, a što su spomenuti zahtjevi jasniji i nedvosmisleno definirani, razvoj menadžerskih sustava je jednostavniji i precizniji. Poslovna inteligencija u tom smislu predstavlja način dostavljanja pravih informacija u pravom formatu, u prave ruke, u pravom trenutku.

Podaci su ključni segment poslovne inteligencije, te je njihova kvaliteta obrnuto proporcionalna uloženom trudu i vremenu u razvoj sustava poslovne inteligencije čiji su temelj. Kvaliteta podataka se očituje kroz standardiziranost, podudarnost, verificiranost i proširivost. Podaci koji se gomilaju svakodnevnim poslovanjem, korištenjem sustava poslovne inteligencije, stvaraju dodatnu vrijednost organizaciji, umjesto dodatne troškove koji bi se generirali njihovim skladištenjem i nedostatkom korištenja.

Da bi se navedeni podaci stavili u upotrebu kroz sustav poslovne inteligencije, prvo je potrebno izgraditi skladište podataka sa potrebnim karakteristikama ovisno o potrebama korisnika. Idući korak je upotreba ETL procesa, u koje spada preuzimanje i čišćenje podataka iz transakcijskih baza podataka ili vanjskih izvora, te njihovo transformiranje na način da se mogu unijeti u skladište podataka i u konačnici koristiti u sustavu. Po završetku punjenja skladišta podataka, pristupa se izgradnji kocki podataka (OLAP), u kojima se definiraju atributi za sve tablice (dimenzijske i tablice činjenica), koje se spajaju primjenom primarnih i stranih ključeva. Što su veći sustavi sa kompleksnijim zahtjevima, izrada kocki biti će zahtjevnija. Kocke podataka čine "mozak" sustava poslovne inteligencije. Tek nakon navedenih koraka, pristupa se izradi upravljačkih ploča čiji su dio i vizualizacije podataka, sa svim kognitivnim smjernicama koje čine korisničko iskustvo upotrebe upravljačkih ploča kvalitetnim ili nekvalitetnim.

Kreiranje ovakvih sustava ponekad može biti dugotrajan proces, ali ispravno

planiranje i definiranje korisničkih zahtjeva može znatno skratiti vrijeme razvoja sustava. Sustavi poslovne inteligencije već duže vrijeme nisu nepoznanica, niti ih koriste i razvijaju samo velike i razvijene organizacije. U malim organizacijama oni su također počeli biti važan dio poslovanja i načina upravljanja, te predstavljaju alat koji može olakšati poslovanje i time potencijalno organizaciji donijeti prevagu na tržištu.

Abstract

This paper shows the application of gathered knowledge on developing business intelligence system in an organization. Paper theoretically explains all relevant development stages – downloading and storing data, usage of Extract/Transform/Load processes and, as a specific area, data visualization as a part of dashboard components. In today's era, information is power, and as such is considered one of the key elements of management. On that note, efforts are made to develop business intelligence systems as a part of organization management which are user friendly. Development starts with identifying user requests and, making them as clear and defined as possible, allowing for the development of management systems to be easy and concise. Business intelligence in that sense is a way of delivering the right information in a right format, to a specific user, in a specific time.

Data is the key element of business intelligence, and its quality is inverse to time and effort spent in developing business intelligence system, of which they're the basis. Data quality is shown through standardization, compatibility, verification and upgradeability. Data is constantly being added and expanded through everyday use of business intelligence system, adding value to the organization instead of creating extra costs of storing.

The presumption of using and upgrading data through the business intelligence system is creating a data storage based on the specific characteristics of users. Next step is the usage of EXTRACT/TRANSFORM/LOAD processes to download, extract and cleanse data from transactional data bases or external sources, and their loading to data storage and use in the business system itself. After completing data storing, ONLINE ANALYTICAL PROCESSING is performed, which defines the attributes for all database tables (dimension table and fact table), connected by primary and foreign keys. The bigger and complex the business system is, the development of OLAP cubes will be more demanding. OLAP cubes are the "brains" of the business intelligence system. After performing mentioned steps, we can start with forming dashboards, and data visualization as one of their parts, with all the cognitive guidelines that secure quality user experience.

Developing these systems often consumes a lot of time, but well planned and defined user demands can significantly shorten the process. For some time now, business intelligence systems are not reserved only for big developed organizations, but have

become an important part of decisions making process in smaller organizations. They are a tool that facilitates business planning and potentially brings leverage in the competitive business environment.

Literatura

1. Amor, D., *The E-business (R)Evolution*, Hewlett Packard Professional Books/Prentice Hall PTR, Upper Saddle River, 2001.
2. Panian Ž. i suradnici, *Poslovni informacijski sustavi*, Element Zagreb, 2010.
3. Panian Ž. i suradnici, *Poslovna inteligencija: Studije slučajeva iz hrvatske prakse*, Narodne novine, Zagreb, 2007.
4. Moss, Larissa T, Atre, Shaku, *Business Intelligence Roadmap*, Addison-Wesley, Boston, 2003.
5. Panian Ž., Klepac G., *Poslovna inteligencija*, Masmedia Zagreb, 2003.
6. Inmon, William H., *Building the data warehouse*, John Wiley-QED, New York(NJ), 1992.
7. Vitasović, V., *Sustavi poslovne inteligencije za uspješnu organizaciju 21. stoljeća*, Diplomski rad, Pula, 2009.
8. Rouse, M., "IS (information system or information services)", veljača 2015., <http://whatis.techtarget.com/definition/IS-information-system-or-information-services> , pristupljeno 1.7.2016.
9. Murfitt, S., "Using Business Intelligence", <http://www.digitrends.net/scripts/> , pristupljeno 3.4.2016.
10. Badami, V., "Payback on Business Intelligence", <http://information-management.com/> , pristupljeno 3.4.2016.
11. Smolić H., "Loše vizualizacije: zašto BI projekti propadaju?", 23.2.2014., <http://www.skladistenje.com/lose-vizualizacije-zasto-bi-projekti-propadaju/>, pristupljeno 27.4.2016.
12. Smith, N., "Designing and building great dashboards – 6 golden rules to successful dashboard design", 21.10.2015., <https://www.geckoboard.com/blog/building-great-dashboards-6-golden-rules-to-successful-dashboard-design/#.V0ScD54vDCI> , pristupljeno 13.5.2016.
13. Choudhury S., "Four Cognitive Design Guidelines for Effective Information Dashboards", 18.6.2014., <https://uxmag.com/articles/four-cognitive-design-guidelines-for-effective-information-dashboards> , 1.7.2016.