

**UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA**

DIPLOMSKO DELO

**IZGRADNJA PODATKOVNEGA SKLADIŠČA
IN KOCKE OLAP NA PRIMERU PEC**

Ljubljana, november 2009

ALJOŠA TOMIČ

IZJAVA

Študent Aljoša Tomič izjavljam, da sem avtor tega diplomskega dela, ki sem ga napisal pod mentorstvom dr. Jurija Jakliča, in da dovolim njegovo objavo na fakultetnih spletnih straneh.

V Ljubljani, dne 1.11.2009

Podpis: _____

KAZALO

UVOD.....	1
1 SISTEMI ZA PODPORO ODLOČANJU	2
1.1 PODATKOVNO SKLADIŠČE.....	2
1.2 PODROČNA PODATKOVNA SKLADIŠČA	3
1.3 TOKOVI PODATKOV.....	4
1.4 PROCES ETL.....	4
1.4.1 <i>Ekstrakcija</i>	5
1.4.2 <i>Transformacija</i>	5
1.4.3 <i>Nalaganje</i>	5
2 SPROTNA ANALITIČNA OBDELAVA PODATKOV	6
2.1 SISTEMI OLTP IN OLAP.....	7
2.2 HEMA KOCKE	9
2.2.1 <i>Zvezdna shema</i>	10
2.2.2 <i>Snežinkasta shema</i>	10
2.2.3 <i>Primerjava zvezdne in snežinkaste sheme</i>	10
2.2.4 <i>Dimenzije</i>	10
2.2.5 <i>Mere</i>	11
2.2.6 <i>Agregacije</i>	11
2.2.7 <i>Operacije znotraj sistema OLAP</i>	12
2.2.8 <i>Modeli shranjevanja podatkov</i>	13
3 IZGRADNJA PODATKOVNEGA SKLADIŠČA IN OLAP KOCKE NA PRIMERU PEC.....	15
3.1 PREDSTAVITEV PEC	15
3.2 ŠTUDENTSKE ANKETE	16
3.2.1 <i>Potek elektronskega anketiranja</i>	17
3.3 ANALIZA OBSTOJEČEGA STANJA IN ZAHTEV S STRANI PEC-A	17
3.3.1 <i>Struktura ankete</i>	18
3.3.2 <i>Uporabniki</i>	18
3.3.3 <i>Analize</i>	18
3.3.4 <i>Posebnosti podatkov in podatkovnega modela</i>	19
3.3.5 <i>Struktura anket</i>	19
3.4 ARHITEKTURA SISTEMA	20
3.5 STRUKTURA PODATKOV NA PRODUKCIJSKI BAZI	21
3.6 PODATKOVNO SKLADIŠČE PEC.....	25
3.6.1 <i>Zvezdna shema PREDMET</i>	25
3.6.2 <i>ETL proces predmeti</i>	25
3.6.3 <i>Zvezdna shema OSEBE</i>	28
3.6.4 <i>ETL proces osebe</i>	29
3.7 HEMA KOCKE OLAP	31

3.7.1 Predmeti.....	31
3.7.2 Osebe.....	32
SKLEP.....	33
LITERATURA IN VIRI	36

KAZALO SLIK

Slika 1: Tok podatkov od operativnega vira skozi proces integracije do podatkovnega skladišča in različnih analitičnih orodji.....	4
Slika 2: Proces ETL, znotraj sistema za podporo odločanju.....	6
Slika 3: Tok podatkov od produkcijskih baz, preko podatkovnega skladišča, olap kocke do končnega uporabnika.....	21
Slika 4: Podatkovni model baze na oracle strežniku.....	22
Slika 5: Podatkovni model baze na mssql strežniku.....	24
Slika 6: Zvezda shema predmeti.....	25
Slika 7: Zvezdna shema osebe.....	28
Slika 8: Podatkovni vir, mere in hierarhije kocke oseba.....	32
Slika 9: Poizvedba nad kocko osebe	33

KAZALO TABEL

Tabela 1: Podatki, ki se črpajo iz tabele anketeodgovori.....	22
Tabela 2: Podatki, ki se črpajo iz tabele oddaje	22
Tabela 3: Podatki, ki se črpajo iz tabele oddajeodgovori.....	23
Tabela 4: Podatki, ki se črpajo iz tabele oddajeodgovoriosebe.....	23
Tabela 5: Podatki, ki se črpajo iz tabele anketevprasanja	23
Tabela 6: Podatki, ki se črpajo iz tabele sodelavci.....	24
Tabela 7: Podatki, ki se črpajo iz tabele predmeti.....	24
Tabela 8: Podatki, ki se črpajo iz tabele r3p2predmet	24

UVOD

Učinkovita komunikacija je ključnega pomena pri napredku na vseh področjih človeškega razvoja. Vse od začetka razvoja informacijske družbe, pospešeno narašča količina podatkov, ki jih posamezni subjekti zbirajo. Prav tako narašča tudi potreba po analiziranju zbranih podatkov.

Podjetja in organizacije se srečujejo z ogromnimi operativnimi transakcijskimi bazami podatkov, kamor se vsakodnevno zapisuje kopica podatkov, ki so posledica sprotnega delovanja. Tako zbrani podatki so sami po sebi bistvenega pomena za tekoče delovanje podjetja oziroma organizacije, vendar so za sprejemanje pravih odločitev popolnoma neuporabni. Za sprejemanje odločitev, odločevalci potrebujejo prave podatke ob pravem času v pravi obliki (Miller & Fuld, 2001, str. 1-31).

Odgovor na zgoraj opisano težavo, analiziranja ogromne količine zbranih podatkov, nudijo informacijske rešitve v obliki sistemov za podporo odločanju. Te sistemi ponujajo enostavno, hitro in fleksibilno analizo velike količine podatkov iz različnih podatkovnih virov.

Prav zaradi aktualnosti področja sistemov za podporo odločanju sem se odločil, da v diplomski nalogi proučim možnosti pridobivanja kvalitetnih informacij iz obilice podatkov, ki sami po sebi nimajo nikakršne vrednosti.

Cilj diplomskega dela v prvi fazi, je preučiti teoretično področje sistemov za podporo odločanju in njihove posamezne sklope. V drugi fazi pa spoznanja iz teorije aplicirati na dejanski primer izgradnje sistema za podporo odločanju.

Diplomsko delo je sestavljeno iz dveh delov. V prvem delu bom proučil potek izgradnje sistema za podporo odločanja in različnih modelov, ki se jih lahko uporabi pri zbiranju, obdelavi in shranjevanju podatkov znotraj sistema. Poleg same izgradnje, bom proučil tudi procese, ki potekajo znotraj njih. Prvi del, ki je v večji meri teoretična analiza, bo služil kot pomoč in osnova pri izboru najprimernejšega modela za izgradnjo dejanskega sistema za podporo odločanja, ki bo podrobneje predstavljen v drugem delu tega diplomskega dela.

Drugi del diplomskega dela je namenjen predstavitvi izgradnje sistema za podporo odločanju na praktičnem primeru. Gre za primer izgradnje sistema za podporo odločanja, s pomočjo katerega bo lahko Pedagoški center Ekonomske fakultete analiziral podatke, pridobljene s študentskimi anketami, ki se navezujejo na delo posameznih predavateljev in predmetov na Ekonomski fakulteti. Praktični del je predstavitev izgradnje dejanskega prototipa sistema za podporo odločanja in vsebuje vse faze razvoja od začetne analize uporabniških zahtev, do končnega cilja, kocke OLAP.

1 SISTEMI ZA PODPORO ODLOČANJU

Sistem za podporo odločanju je koncept, ki pokriva množico različnih tehnologij in orodij. Rešitve s področja sistemov za podporo odločanju najpogosteje pokrivajo širok spekter procesov, programske opreme in tehnik, vse od vira podatkov pa do koristnih informacij končnemu uporabniku. Cilj sistemov za podporo odločanju je razumevanje poslovanja s pomočjo zbiranja informacij in pretvorbo teh na način, ki omogoča boljše analizo. Poleg tega morajo zagotoviti orodja, ki uporabniku omogočajo hitro pridobivanje informacij bistvenih za proces odločanja (Business Intelligence, 2009).

Celoten proces sistemov za podporo odločanju lahko razbijemo na sledeče korake (Business Intelligence, 2009):

- Odkrivanje problemov, ki jih bomo reševali s pomočjo podatkovnega skladišča in podatkov s pomočjo katerih bomo reševali probleme
- Odkrivanje virov vseh potrebnih podatkov in zajem podatkov iz njih
- Pretvorba podatkov iz različnih virov v konsolidirano in konsistentno celoto
- Shranjevanje podatkov v centralno lokacijo
- Izgradnja podatkovnega skladišča s podatki iz centralne lokacije
- Povezava analitičnih aplikacij, ki omogočajo dostop do podatkov v kocki

Med programsko opremo, ki jo uporabljamo za podporo odločanju, največkrat srečamo (Jaklič, 2002, str. 160; DataManagement, 2009):

- Orodja za izdelavo poročil in poizvedovanje, tudi orodja za sprotno analitično obdelavo podatkov (OLAP)
- Programske pakete za rudarjenje v podatkih, na primer SAS, Clementine, MS Analysis Services
- Programske pakete za delo s preglednicami, na primer Excel, OpenOffice Calc
- Programske pakete za statistično analizo, na primer SPSS
- Programske pakete za podporo vodenju projektov, na primer MS Project, Super Project
- Ekspertne sisteme, na primer XpertRule KBS

1.1 Podatkovno skladišče

Pred podrobno analizo lastnosti podatkovnega skladišča, ga je potrebno umestiti znotraj arhitekture informacijskega sistema. Podatkovno skladišče za svoj obstoj potrebuje vire podatkov, iz katerih črpa osnove podatke. Podatkovno skladišče črpa podatke iz enega ali več virov, kateri so lahko znotraj enega ali več sistemov. Prav zaradi te raznolikosti vhodnih podatkov, je potrebno pred polnjenjem podatkovnega skladišča z njimi, izvršiti proces integracije podatkov. Podatkovno skladišče s svojimi podatki omogoča uporabnikom, da z uporabo različnih analitičnih orodij najdejo odgovore na vprašanja, ki se porajajo pri poslovanju organizacij (Kimball & Ross, 2002, str. 358-362).

Podatkovno skladišče je baza podatkov zasnovana tako, da nudi podporo pri procesih odločanja. Podatki se v podatkovno skladišče iz produkcijskih podatkovnih baz običajno prenašajo na koncu dneva, tedna ali pa daljšega časovnega obdobja. Intervali so med prenosi lahko tudi krajši, odvisno od potreb posameznih organizaciji in obremenitvijo sistema, ki jo prenos povzroči. Podatkovna skladišča tako običajno vsebujejo ogromno količino zgodovinskih in statičnih podatkov. Struktura podatkovnega skladišča je zasnovana tako, da omogoča različne analize, upoštevajoč tudi kompleksne poizvedbe na večjem številu podatkov. Pomembna razlika med operativno podatkovno bazo in podatkovnim skladiščem je tudi ta, da so podatki v podatkovnem skladišču definirani za točno določen interval v času in da so stabilni. To pomeni, da se podatki v podatkovnem skladišču praviloma ne brišejo oziroma ne posodablajo (Definition of data warehouse, 2009).

1.2 Področna podatkovna skladišča

Razlika med podatkovnimi skladišči in področnimi podatkovni skladišči izhaja iz začetne analize. Pri izgradnji področnih podatkovnih skladišč je osnovno vodilo analiza uporabnikovih potreb. Za razliko od področnih podatkovnih skladišč, se izgradnja podatkovnega skladišča začne z analizo, kateri podatki so na voljo in kako jih zbrati na način, da bodo uporabni v prihodnje (Inmon, 2000).

Podatkovna skladišča zajemajo podatke celotnega poslovanja organizacije in služijo kot centralna zbirka podatkov organizacije. Področno podatkovno skladišče pa je prirejano za lažje razumevanje ožjega problemskega področja. Prav zaradi teh razlik v strukturi podatkov so podatkovna skladišča primerna za pripravo analiz na strateškem nivoju. Za pripravo analiz na nižjih nivojih oz. posameznih oddelkih pa se praviloma uporabljajo področna podatkovna skladišča. Področna podatkovna skladišča so podmnožice podatkovnih skladišč, tako je njihov edini vir podatkov prav podatkovno skladišče (Inmon, 2000).

Področna podatkovna skladišča so prilagojena za potrebe analitikov, ki najpogosteje neposredno dostopajo do podatkov v njih. Zato morajo biti podatki, za razliko od podatkovnih skladišč, v njih nekoliko prilagojeni (Jaklič, 2002, str. 20):

- Imeti morajo strukturo, ki je enostavna za razumevanje. V večini primerov to pomeni, da so podatki organizirani večdimenzionalno, kar uporabniku omogoča poljuben pogled na podatke
- Vsebovati morajo že preračunane in sumirane podatke, kar pospeši čas poizvedbe

Področna podatkovna skladišča imajo v primerjavi s podatkovnimi skladišči tudi svoje pomanjkljivosti. Bistveni sta manjša fleksibilnost, zaradi prilagojenosti točno določeni uporabi in pa relativno visoki stroški razvoja (Jaklič, 2002, str. 20).

1.3 Tokovi podatkov

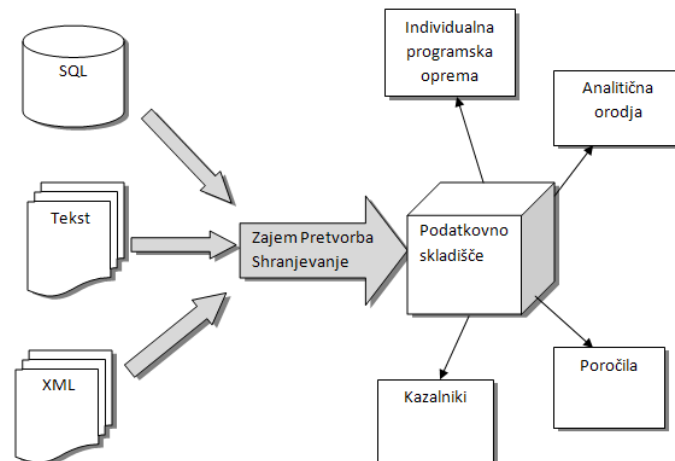
Tokovi podatkov se pojavljajo, ko se podatki prenašajo iz enega v drugi vir. Pomembno za vse prenose podatkov je, da niso sami sebi namen, ampak da vsak prenos iz enega v drugi vir pomeni tudi kvalitativne spremembe v prenesenih podatkih (Jaklič, 2002, str. 22).

V procesih integracije se podatkom s pomočjo kompleksne programske opreme dodaja vrednost tako, da se jih (Jaklič, 2002, str. 22):

- Integrira
- Prečisti
- Transformira
- Analizira, t.j ugotavlja neveljavne vrednosti
- Preverja kakovost podatkov (pravilnost in kakovost)
- Izračuna nove (predvsem sumarne) podatke
- Drugače strukturira

Na sliki 1 je prikazan tok podatkov od operativnega vira, skozi procese integracije v podatkovno skladišče, do različnih analitičnih orodij oz. aplikacij, ki služijo kot ključni podporni dejavnik pri sprejemanju odločitev tako na operativnem, taktičnem kot tudi strateškem nivoju.

Slika 1: Tok podatkov od operativnega vira skozi proces integracije do podatkovnega skladišča in različnih analitičnih orodji.



Vir: R. Kimball & M. Ross, The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 2002, str. 358-362.

1.4 Proces ETL

Ekstrakcija, transformacija in nalaganje (ang. *Extract Transform Load, ETL*) je proces znotraj podatkovnega skladišča. Proces ETL je v večini primerov precej kompleksen.

Zaradi kompleksnosti lahko prihaja do operativnih težav pri nepravilno zasnovanih sistemih ETL. Zato je pred začetkom izdelave sistema ETL potrebna podrobna analiza podatkov, ki vstopajo v proces ETL (Atre & Moss, 2003, str. 211-281).

Podatkovno skladišče se običajno polni nesinhrono iz različnih podatkovnih virov, ki služijo različnim namenom. ETL je ključni proces, ki združuje podatke iz heterogenih in asinhronih virov v homogeno okolje (Kimball & Caserta, 2004, str. 3-28).

1.4.1 Ekstrakcija

Ekstrakcija podatkov iz izvornega sistema je prvi del procesa ETL. Pri večini podatkovnih skladišč poteka ekstrakcija podatkov iz več različnih sistemov. To pomeni, da lahko vsak sistem iz katerega se vrši ekstrakcija, uporablja različne zapise oziroma oblike podatkov. Prav tako imajo lahko izvorni sistemi relacijske ali nerelacijske podatkovne baze, kot je na primer IMS. Tako mora ekstrakcija pretvoriti podatke v enotno obliko, ki je potrebna za naslednji proces, proces transformacije.

Zelo pomemben del ekstrakcije je tudi razčlenjevanje pridobljenih podatkov. Rezultat je pregled podatkov, s katerim se preveri ali se podatki skladajo s predpisano obliko. V primeru, da se ne, se podatke zavrne v celoti (Kimball & Caserta, 2004, str. 50-83).

1.4.2 Transformacija

V procesu transformacije se, s pomočjo pravil ali funkcij izbrane podatke, iz njihovega vira naloži v ciljni sistem. Kako zahteven je proces manipulacije s podatki, je odvisno predvsem od podatkovnega vira ter poslovnih in tehničnih zahtev (Kimball & Caserta, 2004, str. 113-160).

V procesu transformacije se (Kimball & Caserta, 2004, str. 113-160):

- Za nalaganje lahko izbere samo določene stolpce iz podatkovnih virov
- Prečisti in prevede podatke pridobljene iz podatkovnih virov
- Združi podatke iz različnih podatkovnih virov
- Določi vrednosti nadomestnih ključev
- Transponira izbrane vrstice ali stolpce
- Razcepi izbrane stolpce v več ločenih stolpcev
- Izvede proces vrednotenja podatkov

1.4.3 Nalaganje

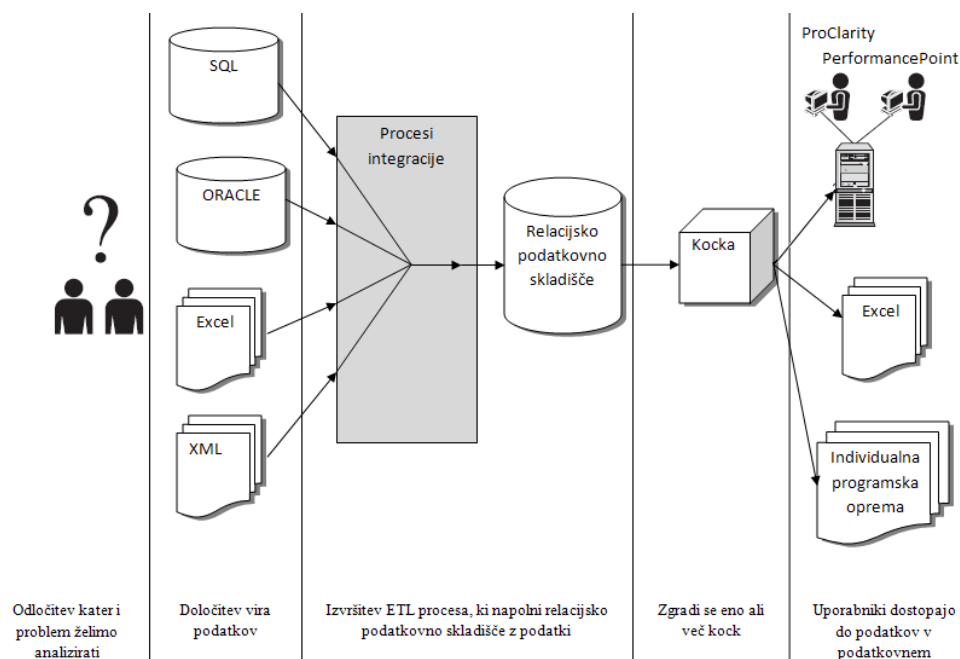
Proces nalaganja je zadnja faza v procesu ETL. V tej zadnji fazi se podatke naloži v podatkovno skladišče. Proces nalaganja podatkov v podatkovno skladišče se od

organizacije do organizacije precej razlikuje. Razlogi za to ležijo v različnih zahtevah posameznih organizacij.

Pomembno vlogo pri zagotavljanju kakovosti in zmogljivosti procesa ETL v fazi nalaganja podatkov v podatkovno skladišče, igrajo tudi shema podatkovnega skladišča in pa sprožilci, ki se sprožijo pri zapisovanju podatkov v bazo (Kimball & Ross, 2002, str. 358-362).

Na spodnji sliki je predstavljen umestitev procesa ETL, znotraj sistema za podporo odločanju.

Slika 2: Proces ETL, znotraj sistema za podporo odločanju



Vir: R. Kimball & M. Ross, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling* 2002, str. 358-362.

2 SPROTNA ANALITIČNA OBDELAVA PODATKOV

V poslovnem svetu so med vodstvenimi kadri zelo priljubljene po meri narejene preglednice, ki vsebujejo točno izbrane podatke in imajo veliko analitično moč. Problem takih preglednic je v tem, da je za njihovo pripravo potrebno imeti usposobljene ljudi in veliko časa, saj je potrebno zbrati in primerno urediti veliko količino podatkov. Vzrok je hitro in pogosto spreminjanje zahtev, kaj točno naj preglednice sploh vsebujejo. Posledica tega pa je, da preglednice pogosto vsebujejo preveč ali premalo podatkov za kvalitetno in hitro analizo. Zato se pogosto dogaja, da so take preglednice za ključnega uporabnika delno ali v celoti neuporabne, saj vsebujejo preveč, premalo ali pa sploh ne vsebujejo podatkov, ki jih uporabnik potrebuje (Pendse, 2009).

Rešitev na zgoraj opisano težavo ponujajo orodja, ki omogočajo sprotno analitično obdelavo podatkov (angl. *On Line Analytical Processing, OLAP*). OLAP pomeni programsko orodje za analiziranje zbranih podatkov organizacije. OLAP se lahko v širšem pomenu besede uporablja tudi kot sinonim za podatkovno skladiščenje in večdimenzionalno analiziranje podatkov podatkovnega skladišča (Online analytical processing, 2009).

Orodja OLAP ponujajo logično, hierarhično in strukturirano pregledovanje podatkov. OLAP omogoča vpogled v podatke po časovni dimenziji, agregatih in v nekaterih premerih, tudi v transakcijske podatke. Glavne prednosti sistema OLAP so: izredno hitra odzivnost, možnost nadgrajevanja in spreminjanja ter preprosta uporaba. Omogočeno je hitro prehajanje iz nižjih nivojev na agregatne podatke in obratno. Podprto je tudi dodatno preračunavanje in napovedovanje. Z dodatnimi analitičnimi orodji se lahko iz vsake dobljene preglednice izdela tudi rešitev v grafični obliki (Pendse, 2009).

Tehnologija OLAP omogoča ključnim uporabnikom hitro in zanesljivo analiziranje podatkov iz različnih zornih kotov. Prav tako omogoča obdelavo velike količine podatkov in pripravo analiz poslovnih procesov, ki so ključni vir strateških informacij v procesu odločanja. Zato je OLAP ključni dejavnik podpore v procesu planiranja in identifikaciji morebitnih težav v prihodnosti (Krsnik, 2001, str. 65).

Za implementacijo tehnologije OLAP najprej potrebujemo podatkovno skladišče, v katerem skladiščimo analitične podatke. Orodja OLAP nam omogočajo izgradnjo kock, v katerih organiziramo in preračunavamo podatke, na način, ki omogoča čim učinkovitejše analitične poizvedbe. Kocke tako omogočajo hitre analize velike količine podatkov brez predhodnega znanja strukturiranega povpraševalnega jezika za delo s podatkovnimi bazami (angl. *structured query language, SQL*) (Isemenger, 2001, str. 235).

Vsak dober sistem OLAP mora delovati znotraj določenih standardov. Sistem mora uporabnikom omogočati relativno hitre poizvedbe. Prav tako jim mora omogočati izdelavo čim večjega števila relevantnih analiz, ki jih uporabnik potrebuje za podporo pri odločanju. Poleg tega mora sistem zagotoviti skupno rabo podatkov znotraj skupine uporabnikov, ki te podatke potrebujejo in istočasno zagotoviti ustrezno mero varnosti, teh istih podatkov. Najosnovnejši in poleg tega tudi najpomembnejši pogoj, ki ga mora ustrezen sistem OLAP zagotoviti, je večdimenzionalni pogled na podatke. Zadnji in nič manj pomemben pogoj pa je, da omogoča dostop do informacij, pri čemer se meri, koliko informacij je sistem sposoben obdelati in ne koliko jih lahko hrani (Pendse, 2009).

2.1 Sistemi OLTP in OLAP

Glavni namen sistemov za sprotno obdelavo transakcij (angl. *Online transaction processing, OLTP*) je avtomatizacija poslovnih procesov. Transakcijski sistemi so nujni za uspešno poslovanje podjetij v modernem poslovnem okolju. Glavna značilnost

transakcijskih sistemov je operativna baza podatkov. Podatki so normalizirani, to pomeni, da so podatki v bazi zapisani tako, da se ne podvajajo, vendar so še vedno na voljo številnim uporabnikom z različnimi zahtevami. Podatki so shranjeni tako, da niso odvisni od programov, ki jih uporabljajo. Neodvisnost podatkov od programov dosežemo z dvonivojsko ali tronivojsko arhitekturo aplikacij (Grad & Jaklič, 1996, str. 1).

Transakcijski poslovni sistemi so namenjeni sprotnemu zbiranju podatkov o poslovnih entitetah. Njihov primarni cilj je zagotavljanje učinkovitega vnašanja, ažuriranja in poizvedovanja po manjšem številu vrstic istočasno. Značilnosti takega sistema je hiter dostop do točno določenega zapisa in hitro obdelovanje manjšega števila podatkov hkrati. Za doseganje potrebne hitrosti pri obdelavi transakcijskih podatkov morajo sistemi biti normalizirani in relacijski (Amo, 2000, str. 8).

Sheme relacijskih baz so v večini primerov zelo razvejanje in kompleksne, saj vsebujejo veliko število medsebojnih podatkovnih povezav. Relacijske baze sestavlja več medsebojno povezanih tabel. Vsaka izmed tabel, ki se nahaja znotraj baze, vsebuje podatke o določeni entiteti, parametrih, dokumentih itd. Relacija med tabelami je določena s ključi. Poznamo primarne in tuje ključe. Primarni ključ je polje, ki pripada dotični tabeli. Tuj ključ pa je polje, ki dejansko pripada neki drugi tabeli, vendar je zapisan tudi v dotični tabeli. Na ta način lahko povežemo dva zapisa, ki se nahajata v dveh različnih tabelah. Ključ je lahko sestavljen tudi iz več polj, takrat govorimo o sestavljenem ključu.

Zaradi beleženja vseh prometnih podatkov v relacijskih bazah, se v njih sčasoma nakopiči ogromno število zapisov. Rešitev težave velikega števila zapisov je v arhiviranju starih zapisov. Stare zapise se iz prometnih tabel prestavi v arhivske, v katere se dostopa po potrebi. Tako se v prometnih tabelah zmanjša število zapisov, kar omogoča hitrejše delovanje sistema.

Relacijske baze transakcijskih sistemov morajo biti normalizirane. Z normalizacijo podatkov dosežemo zmanjšanje števila podatkov v bazi, kar posledično pospeši izvajanje procedur znotraj sistema. Normalizacijo podatkov izvedemo z dodajanjem novih tabel, ki se jih poveže z relacijo ena proti mnogo. Na ta način se omogoči hitrejše in enostavnejše ažuriranje in poizvedovanje po podatkih. Za hitrejše poizvedovanje in filtriranje podatkov so polja v tabelah pogosto indeksirana (Berge, 2001, str. 875).

Transakcijski sistemi so namensko oblikovani za obdelavo velikega števila prometnih podatkov, ki nastajajo z vsakodnevnimi opravili. Vendar pa taki sistemi ne morejo podati najboljših odgovorov na vprašanja, ki si jih zastavlja vodstvo in poslovni analitiki. Za pridobivanje takšnih informacij se potrebuje drugačno zasnovo podatkovne baze. Struktura mora biti usmerjena na obvladovanje analitičnih in raziskovalnih poizvedb. Zahtevano strukturo se najde pri podatkovnih skladiščih, ki so prav posebej zasnovana za zajemanje in obdelovanje podatkov s pomočjo orodji OLAP.

Pri strukturi podatkovnih skladišč je najbolj pogosta zvezdna shema, ki je tudi najpreprostejša. Rešitve, ki jih ponujajo sistemi OLAP, se primarno usmerjajo na analiziranje velike količine podatkov. OLAP omogoča hitre in zahtevne poizvedbe, ki vključujejo velike količine podatkov, zaradi dveh značilnosti. Prva značilnost je, da je struktura shranjenih podatkov večdimenzionalna in ne relacijska. Druga značilnost pa je, da obstajajo predhodno izračunani agregati, ki omogočajo hiter odziv na uporabnikove zahteve (Berge, 2001, str. 859).

Glavna značilnost zvezdne sheme je ta, da v svoji sredini vsebuje tabelo dejstev in okoli nje dimenzijske tabele z nenormaliziranimi podatki. V nasprotju s transakcijskimi bazami se pri OLAP sistemih svetuje uporabo nenormaliziranih podatkov. Prav s podvajanjem podatkov se omogoči, da posamezne poizvedbe za svoje izvajanje uporabljajo manjše število tabel znotraj podatkovne baze. Na ta način se zmanjša obremenitev podatkovne baze in čas potreben, da se poizvedba izvede. Dodaten prostor, ki ga zasedajo nenormalizirani podatki, je bistveno manjši kot prostor, ki bi ga zasedle dodatne dimenzijske tabele, ki bi jih morali vpeljati v sklopu normalizacije podatkov (Berge, 2001, str. 862).

Ključna prednost sistemov OLAP je ta, da uporabniki za izvajanje poizvedb ne potrebujejo predznanja strukturiranega poizvedovalnega jezika SQL. Tako sistem lahko uporabljajo uporabniki, ki nimajo naprednih informacijskih znanj. Prav tako ključnim uporabnikom ni potrebno čakati na poročila, ki bi jih morali pripravljati specializirani oddelki. Na ta način odločevalci, ki niso veščji uporabljanja podatkovnih baz, dobijo možnost hitrega analiziranja podatkov in pridobivanje odgovorov na vprašanja, ki so ključna v procesu odločanja (Berge, 2001, str. 865).

2.2 Shema kocke

Shema kocke je podatkovni model, v katerem je prikazan skupek med seboj povezanih tabel, ki se nahajajo znotraj podatkovnega skladišča. Kocka podatke črpa iz podatkovnega skladišča. V središču sheme se nahaja tabela dejstev, ki je vir mer kocke. Tabele, ki se nahajajo okoli tabele dejstev imenujemo dimenzijske tabele. Dimenzijske tabele so vir dimenzij kocke in predstavljajo koordinate osi kocke.

Struktura kocke je definirana z merami in dimenzijami, ki izhajajo iz tabel podatkovnega vira kocke. Shema kocke je vedno sestavljena iz ene tabele dejstev in ene ali več dimenzijskih tabel. Mere kocke vedno izhajajo iz tabele dejstev, dimenzije pa vedno iz dimenzijskih tabel.

Poznamo dva različna tipa sheme. Prvi tip je zvezdna shema in drugi tip je snežinkasta shema (OLAP, 2008).

2.2.1 Zvezdna shema

Ime zvezdna shema izhaja iz načina razporeditve tabel podatkovnega modela. V sredini modela se nahaja tabela dejstev, okoli katere se nahajajo dimenzijske tabele. Tabela dejstev vsebuje podatke o merah, medtem ko dimenzijske tabele vsebujejo podatke o dimenzijah (OLAP, 2008).

2.2.2 Snežinkasta shema

Model snežinkaste sheme je dejansko izpeljan iz zvezdne sheme. Za razliko od zvezdne sheme, dimenzijske tabele v snežinkasti shemi vsebujejo normalizirane podatke. Normalizacija podatkov pomeni, da se prepreči nepotrebno podvajanje atributov znotraj dimenzijskih tabel. Normalizacijo podatkov v zvezdni shemi se doseže z dodatnimi tabelami. Z uporabo snežinkaste sheme se zmanjša število zapisov v tabelah, ter tako prihrani prostor za shranjevanje podatkov. Snežinkasta shema ima tudi svojo slabo plat. Največja težava snežinkaste sheme je kompleksnost poizvedb, kar je posledica kompleksnejšega podatkovnega modela. Kompleksnejše poizvedbe podaljšujejo čas izvedbe in zmanjšujejo učinkovitost analitičnega orodja (OLAP, 2008).

2.2.3 Primerjava zvezdne in snežinkaste sheme

Največja razlika med zvezdno in snežinkasto shemo se kaže pri določanju dimenzijskih tabel. V snežinkasti shemi se atributi, ki bi imeli v zvezdni shemi ponavljajoče se zapise, prenesejo v pomožne tabele, ki so preko ključev povezane z dimenzijskimi tabelami. Nadaljnja normalizacija podatkov v pomožni tabeli je omogočena preko dodatne pomožne tabele, ki je preko ključa povezana s pomožno tabelo (OLAP, 2008).

2.2.4 Dimenzije

Prikazovanje podatkov samo iz najvišje perspektive ni dovolj natančno in fleksibilno za specializirane analize. Analitiki in določevalci, ki želijo analizirati točno določen problem, potrebujejo podatke razvrščene po različnih kategorijah, ki se imenujejo dimenzije (Berge, 2001, str. 874).

Dimenzije so sestavni atributi kocke. Vsaka dimenzija ima v podatkovni bazi svojo dimenzijsko tabelo. Vsaka dimenzijska tabela vsebuje attribute, ki opisujejo posamezno dimenzijo kocke. Dimenzije lahko vsebujejo enega ali več nivojev. Več nivojev znotraj ene dimenzije omogoča dviganje ali spuščanje znotraj te dimenzije, to pa omogoča ogled in analizo podatkov na različnih hierarhičnih nivojih. Dimenzije kategorizirajo podatke kock na hierarhični način, ki je sestavljen iz več nivojev. Nivo je naziv za množico členov znotraj hierarhije dimenzije, ko so vsi člani množice enako oddaljeni od korena do hierarhije.

Posamezni členi so znotraj dimenzije urejeni hierarhično. To pomeni, da so nadrejeni členi vedno seštevki vseh njemu podrejenih členov. Posamezni nivoji, ki razvrščajo člene po hierarhiji so izpeljani iz stolpcev dimenzijskih tabel. Po posamezni dimenziji so podatki razvrščeni od najvišjega, oziroma najbolj splošnega pa do najnižjega, oziroma najbolj podrobnega.

Dimenzije predstavljajo posamezne vidike, ki so predmet analize. To omogoča pogled na podatke iz več različnih zornih kotov (OLAP, 2008).

Privatne dimenzije so dimenzije, ki se uporabljajo samo znotraj ene kocke. Kocka si privatne dimenzije ne deli z ostalimi kockami (OLAP, 2008).

Deljene oz. javne dimenzije so dimenzije, ki si jih več kock deli med seboj. V praksi si najpogosteje kocke med seboj delijo časovno dimenzijo (OLAP, 2008).

2.2.5 Mere

Namen podatkovne baze, ki ga uporablja sistem OLAP, je shranjevanje podatkov o različnih entitetah, katere se opazuje in spremlja. Podatki, ki se jih analizira s pomočjo združevanja in primerjave, se imenujejo mere (Berge 2001, str. 874).

Mera v kocki je množica vrednosti, ki temelji na stolpcu v tabeli dejstev kocke in je najpogosteje numerična. Vrednosti posamezne mere vsebujejo celice kocke, katere so najosnovnejši element kocke. Za določanje posamezne mere je podana funkcija agregacije, s pomočjo katere se določa, na kakšen način se vrednosti iz nižjih nivojev združujejo na višjih. OLAP kocka lahko vsebuje več mer (OLAP, 2008).

2.2.6 Agregacije

Agregacije predstavljajo vnaprej izračunane in shranjene podatke. Omogočajo boljšo odzivnost sistema in krajši čas, ki je potreben za izvedbe posameznih poizvedb. To pomeni, da so odgovori na nekatera vprašanja že poznani, še preden uporabnik določi parametre, ki jih želi analizirati. Agregacije podatkov omogočajo, da za vsako poizvedbo ni potrebna obdelava celotne tabele dejstev. Agregati so shranjeni kot koordinate specifičnih dimenzij v večdimenzionalni strukturi. Večja kot je količina podatkov shranjenih kot agregati, krajši so odzivni časi poizvedb. Vendar pa ima vnaprejšnji izračun velikega števila agregacij tudi svojo slabo stran, ker se povečajo prostorske zahteve za shranjevanje večjega števila podatkov. Kot pri vseh ostalih odločitvah je potrebno tudi tu upoštevati razmerje med uporabnostjo in stroški, kar se v našem primeru nanaša na razmerje med odstotkom izračunanih agregacij in zahtevami po prostoru, ki jih agregacije potrebujejo (Isemenger, 2001, str. 148).

2.2.7 Operacije znotraj sistema OLAP

Za delo z orodji OLAP ne potrebujemo kompleksnega strokovnega predznanja. Poznati je treba le osnovne operacije potrebne za pridobivanje informaciji, ki služijo podpori odločanju (Jaklič, 2008, str. 173; Jagarinec, 2005):

- **Vrtanje v globino (drill-down)**
O vrtanju v globino govorimo takrat, ko se po podatkih pomikamo navzdol po hierarhičnih nivojih dimenzije.
- **Zvijanje (roll-up ali drill-up)**
O zvijanju govorimo takrat, ko se po podatkih pomikamo navzgor po hierarhičnih nivojih dimenzije. Gre za agregacijo podatkov v nivoje povzetkov vzdolž ene dimenzije.
- **Rezanje (slice in dice)**
Kadar gre za podkocko z izbiro ene dimenzije, govorimo o operaciji »slice«. Pri operaciji »slice« naredimo dvodimenzionalni prerez za izbran nabor dimenzijskih vrednosti, pri katerem se omejimo le na en segment znotraj dimenzije. Kadar gre za podkocko z izbiro dveh dimenzij, govorimo o operaciji »dice«. Gre za enak proces kot pri operaciji »slice«, kjer podatke prerežemo po eni sami dimenziji z razliko v tem, da pri operaciji »dice« dodamo še eno dodatno dimenzijo.
- **Vrtenje (pivot)**
Gre za vizualno operacijo, ki omogoča alternativni vpogled v podatke. Spreminja se samo oblika poročila, podatki pa po vsebini ostajajo enaki.
- **Vrtanje čez (drill across)**
Vrtanje čez je operacija, ki se uporablja za navigacijo med izhodiščno kocko in ciljno točko z uporabo enakih koordinat. Omejitev pri operaciji vrtanje čez je ta, da morajo biti dimenzijske tabele v izhodiščni in ciljni kocki enake.
- **Vrtanje skozi (drill-through)**
Pri operaciji vrtanje skozi, uporabnik izbere posamezno celico kocke in prejme nabor izvornih podatkov, ki so bili uporabljeni v postopku izračuna vrednosti izbrane celice. Tabelarična predstavitev vključuje stolpce za vse nivoje vseh dimenzij, za vse mere. Pri tem pa upošteva samo zapise, znotraj tabele dejstev, ki so zajeti v funkciji agregacije pri preseku članov dimenzij, vezanim na dano celico.
- **Gnezdenje (nest)**
Pri operaciji gnezdenje, prikažemo podatke ene dimenzije, znotraj druge dimenzije.
- **Filtriranje**
S pomočjo filtra določimo kriterij po katerem se prikazujejo podatki, ki zadovoljujejo kriterij.
- **Zapiši nazaj (writeback)**
Operacija zapiši nazaj omogoča zapis informacije v kocko, s pomočjo jezika MDX. S pomočjo te funkcije lahko izvedemo analizo »kaj če« (ang. *what if*). Pri tej analizi, s spreminjanem vhodnih podatkov, simuliramo različno obnašanje opazovanega procesa in izhodne podatke, ki v tem procesu nastajajo.

Pri izvajanju operacij je potrebno budno spremljati izhodne podatke. Pri vsakem večjem odstopanju je potrebno preveriti ali gre za dejansko odstopanje ali pa je morda prišlo do napake v izračunih ali pa pri prenosu podatkov iz produkcijske baze v podatkovno skladišče.

2.2.8 Modeli shranjevanja podatkov

Informacije v kocki so lahko shranjene na več različnih načinov. Znane so tri prevladujoče različice shranjevanja podatkov v kockah. Izbrana različica vpliva na: odzivni čas poizvedb, hitrost osveževanja podatkov v kocki in prostorske zahteve na disku.

Procesiranje oziroma izgradnja kocke OLAP se izvaja na način, da se najprej zajamejo podatki iz podatkovnega vira, najpogosteje je to podatkovno skladišče. Nato se na podlagi definirane strukture vnaprej izračunajo agregati. Agregati so namenjeni predvsem skrajšanju odzivnega časa ob pregledovanju podatkov, saj sistem pri izračunavanju uporablja že predhodno preračunane agregate. Tako sistemu pri izvajanju poizvedb ni potrebno preračunavati vseh osnovnih podatkov v podatkovnem skladišču. Glede na način shranjevanja teh podatkov se ločijo tri prevladujoče vrste tehnologij OLAP ROLAP, MOLAP, HOLAP (Lalič, 2005, str. 51).

2.2.8.1 ROLAP

Relacijska sprotna analitična obdelava podatkov (angl. *Relational On Line Analytical Processing, ROLAP*) shranjuje oba tipa podatkov neposredno v relacijsko bazo. Vendar to še ne pomeni, da se lahko v ta namen uporabi kar relacijska baza sistema OLTP. V večini primerov je potrebno narediti novo relacijsko bazo, namenjeno za uporabo v sistemu OLAP, ki svoje podatke črpa iz relacijske baze OLTP. Ob dostopanju do podatkov se v sistemu vršijo SQL poizvedbe. Prav zaradi izvajanja poizvedb s pomočjo jezika SQL, je delovanje sistema upočasnjeno. Po drugi strani pa je v takem sistemu čas procesiranja bistveno krajši, saj procesiranje vhodnih podatkov ni potrebno. V trenutku, ko je podatek zapisan v podatkovnem skladišču, je že predstavljen v OLAP odjemalcu. Prednost modela ROLAP je tudi v količini podatkov, ki jih lahko obdela v relativno kratkem času. Tu gre predvsem za podatke z visoko kardinalnostjo (Pedersen & Jensen, 2001, str. 40-46).

2.2.8.2 MOLAP

Večdimenzionalna sprotna analitična obdelava podatkov (angl. *Multidimensional Online Analytical Processing, MOLAP*) shranjuje oba tipa podatkov v večdimenzionalnih strukturah, v relacijski podatkovni bazi. Izvorni podatki so zapisani v relacijski in večdimenzionalni obliki. Ker model MOLAP zahteva vnaprejšnje izračunavanje vseh možnih agregacij, omogoča hitrejši dostop do podatkov, oziroma hitrejši odzivni čas sistema. Model MOLAP ima še eno prednost in sicer, da ob pregledovanju kock ne potrebujemo direktne povezave z virom podatkov. To omogoča ustvarjanje lokalnih kock,

ki niso neposredno povezane na strežnik, kjer se nahaja sistem OLAP. Tudi brez povezave z osnovnim virom je omogočeno pregledovanje vseh nivojev podatkov (Pedersen & Jensen, 2001, str. 40-46).

2.2.8.3 Primerjava modelov ROLAP in MOLAP

Model ROLAP se bolje obnese na področjih, kjer obstaja ogromno število zapisov, vendar ima veliko pomanjkljivosti. Bistvena pomanjkljivost je ta, da je od modela ROLAP znatno počasnejši pri izvajanju poizvedb.

Večina podjetij, predvsem zaradi znižanja stroškov v svojih sistemih OLAP, uporablja model ROLAP. Mnogokrat uporabijo že obstoječe relacijske baze, ki pa niso optimizirane za uporabo OLAP sistema. Prav ta fleksibilnost orodja ROLAP, ki omogoča uporabo že obstoječih relacijskih baz, mu daje konkurenčno prednost pred orodji MOLAP. Po drugi strani pa zaradi fleksibilnosti trpi zmogljivost (Pedersen & Jensen, 2001, str. 40-46).

2.2.8.4 HOLAP

Zaradi nepriljubljene izbire med dodatnimi stroški, ki jih prinaša razširitev ETL procesa in počasnostjo sistema, večina sistemov OLAP uporablja hibridno sprotno analitično obdelavo podatkov (*angl. Hybrid Online Analytical Processing, HOLAP*). Model HOLAP omogoča oblikovalcem sistema izbiro, kateri podatki bodo shranjeni v sistemu ROLAP in kateri v MOLAP (Pedersen & Jensen, 2001, str. 40-46).

Model HOLAP je kombinacija modelov ROLAP in MOLAP. Ločita se dva pristopa k modelu HOLAP.

Prvi pristop je vertikalno particioniranje, pri katerem so vhodni podatki shranjeni v relacijski bazi, medtem ko so agregacije v večdimenzionalni strukturi. HOLAP upravlja podatke kot ROLAP in agregacije kot MOLAP. Pri pregledovanju agregatov pa omogoča hitro poizvedovanje in ob tem prihrani čas osveževanja kock in prostor na disku.

Drugi pristop je horizontalno particioniranje, pri katerem se nekatere prereze podatkov shrani v strukturi MOLAP, kar poveča zmogljivost poizvedb. V teh primerih gre skoraj vedno za novejšje podatke. Starejše podatke pa se shranjuje v strukturi ROLAP (Kaser & Lemire, 2006, str. 2279-2438).

3 IZGRADNJA PODATKOVNEGA SKLADIŠČA IN OLAP KOCKE NA PRIMERU PEC

3.1 Predstavitev PEC

Pedagoški center Ekonomske Fakultete v Ljubljani (v nadaljevanju PEC) skrbi za razvoj in spremljanje kakovosti pedagoškega procesa, ter za ustvarjanje in posredovanje novih znanj, na področju razvoja znanj in veščin učiteljev, za izvajanje pedagoškega dela (PEC, 2009).

Pri tem usmerjajo učitelje v izobraževanje na področju visokošolske didaktike, skrbijo za uvajanje novih metod dela in izobraževalnih tehnologij, svetujejo pri oblikovanju novih študijskih programov in pristopov, ter načrtujejo in organizirajo izobraževalne in svetovalne dejavnosti s področja kakovosti pedagoških procesov (PEC, 2009).

Posebno skrb namenjajo sodelovanju s poslovno javnostjo. Organizirajo sodelovanje predavateljev iz prakse, skrbijo za izvedbo obiskov in ekskurzij v podjetjih, spodbujajo pisanje poslovnih primerov, izvajanje projektov s študenti v podjetjih in institucijah, pripravljajo Študentsko poslovno konferenco. Skratka, razvijajo vse tiste oblike sodelovanja s slovensko in mednarodno poslovno javnostjo, ki pomembno prispevajo k povezovanju poslovne prakse s pedagoškimi procesi in s tem k njihovi večji kakovosti (PEC, 2009).

Pomemben del aktivnosti namenjajo razvoju celovitega sistema vrednotenja in spremljanja kakovosti pedagoškega procesa, ter svetovanju pedagogom za izboljšanje kakovosti pedagoškega procesa. Pri tem so poglavitne usmeritve in cilji (PEC, 2009):

- Skrb za obogatitev obstoječih sistemov vrednotenja in izboljševanja pedagoških procesov skladno s smernicami Komisije za kakovost na Univerzi v Ljubljani, ter z zahtevami EQUIS in AACSB akreditacij
- Dvig zavesti o pomenu evalvacije pedagoških procesov in uvajanju sistematičnih mehanizmov za njihovo izboljšanje na vseh programih študija na EF (študentske ankete, zunanja evalvacija kakovosti, evalvacija programov, usmerjevalne komisije, itd)
- Prenos izkušenj in uveljavljanje evropskih standardov v proces zagotavljanja odličnosti pedagoškega procesa na EF. Pripravljajo celovit elektronsko podprt proces osebnega spremljanja kakovosti in pedagoškega razvoja pedagogov, prav tako pa tudi elektronsko podprt sistem evalvacije s strani študentov

Pedagoški center Ekonomske fakultete izvaja naslednje aktivnosti (PEC, 2009):

- Priprava in izvedba študentskih anket o kakovosti pedagoškega dela, priprava letnih poročil o kakovosti dela

- Pospeševanje sodelovanja podjetij in ustanov v pedagoškem procesu, prenos poslovne prakse v pedagoški proces (z izvedbo poslovnih projektov, pisanjem in uporabo študijskih poslovnih primerov, poslovnih načrtov, z diplomskimi, magistrskimi in drugimi deli, ki obravnavajo neposredne izzive poslovne prakse), organizacijo in izvedbo Študentske poslovne konference, organizacijo ekskurzij v podjetja, naročanje in distribucija poslovnih primerov, sklepanje pogodb o znanstvenem, raziskovalnem in pedagoškem sodelovanju s podjetji in institucijami
- Pedagoško in andragoško izobraževanje učiteljev fakultete in zunanjih sodelavcev, ki sodelujejo pri izvajanju pedagoškega procesa

3.2 Študentske ankete

Temeljno poslanstvo študentskih anket je pripomoči k povečanju kakovosti pedagoškega procesa na Ekonomski fakulteti. V ta namen se razvijajo in tudi izvajajo različni mehanizmi ocenjevanja in spremljanja kakovosti pedagoškega dela (Študentske ankete, 2009).

Njihova glavna naloga je priprava in izvedba študentskih anket o kakovosti pedagoškega dela na različnih vrstah študija. Skozi ankete imajo študentje možnost podati svoje mnenje glede kakovosti, uporabnosti in zanimivosti posameznih predmetov, ter oceniti delo vseh izvajalcev pri predmetu. Študentske ocene služijo kot podlaga za spremembo zasnove in vsebine predmetov, ter razvoj pedagoških veščin posameznih izvajalcev. Študentsko mnenje je prav tako pomembno v procesu habilitacije pedagogov (Študentske ankete, 2009).

V preteklih letih so se ankete izvajale ob koncu semestra v papirni obliki. S poletnim semestrom študijskega leta 2005/2006 pa so uvedli informatizacijo študentskih anket. Tako so sedaj ankete bolj dostopne, enostavnejše za obdelavo, njihovi rezultati pa še bolj tehtni. Elektronske ankete se izvajajo na celotnem dodiplomskem študiju (redni, izredni in študij na daljavo), v začetku leta 2008 pa so e-ankete uvedene tudi na izrednem podiplomskem študiju (Študentske ankete, 2009).

Študentske ankete so z informatizacijo postale del rednih obveznosti pri posameznem predmetu, ki jih mora študent opraviti pred prijavo na izpit. Študentje naj bi to obveznost opravili najkasneje v enem mesecu po koncu predavanj pri posameznem predmetu. Študent mora anketo izpolniti (se evidentirati) preko Student Net-a, pri čemer ima pravico, da ne odgovori na nekatera ali vsa vprašanja. Zagotovljena je popolna anonimnost odgovorov, saj po elektronski oddaji študentske ankete, ta postane del skupne baze vseh oddanih anket. Na ta način se onemogoči ugotavljanje identitete posameznega anketiranca (Študentske ankete, 2009).

3.2.1 Potek elektronskega anketiranja

Študentske ankete se izvajajo v elektronski obliki, preko Študent Net-a. Študentom, ki nimajo lastnega dostopa do interneta, so na voljo prosti računalniki pred referatom (Uvedba elektronskih anket, 2009).

Ankete o predmetih, ki so jih poslušali v semestru, morajo študentje izpolniti pred prijavo na izpit. Izpolnjena anketa je pogoj za prijavo na izpit. Ne glede na to ali bodo študentje izpit opravljali takoj ali kasneje, so pozvani, da ankete izpolnijo najkasneje v enem mesecu po koncu predavanj. Izpolnjevanje ankete je za študente obvezno. Ni pa nujno, da odgovorijo na vsa vprašanja. Če iz kakršnega koli razloga na določeno vprašanje ne želijo odgovoriti, ga lahko preskočijo. Z vidika zagotavljanja čim boljše povratne informacije pedagogom, so študentje pozvani, da odgovorijo na čim več vprašanj. Pri izpolnjevanju anket je študentom zagotovljena popolna anonimnost. Evidentira se le, ali je anketa izpolnjena do konca. Rezultati ankete se shranjujejo ločeno, brez kakršnih koli podatkov, ki bi omogočali identifikacijo posameznika. Zaradi tega razloga ankete, ko jo enkrat izpolnjena, ni mogoče več spreminjati. Ankete študentje lahko začnejo izpolnjevati tudi pred koncem predavanj, dostopne pa postanejo mesec dni pred zaključkom predavanj v semestru (Uvedba elektronskih anket, 2009).

Elektronska anketa je sestavljena iz štirih delov. Prvi trije deli so vsebinsko popolnoma identični stari tiskani anketi (demografski podatki o anketirancu, vsebina in organizacija predmeta, pedagoško delo izvajalcev), četrti del pa obsega odprta vprašanja o predmetu in izvajalcih (Uvedba elektronskih anket, 2009).

V primeru, da so se študentje predavanj ali vaj udeleževali redko (manj kot 30%), jim je v okviru četrtega dela ankete postavljeno dodatno vprašanje o razlogih za neudeležbo (Uvedba elektronskih anket, 2009).

3.3 Analiza obstoječega stanja in zahtev s strani PEC-a

Obstoječi sistem, ki ga PEC uporablja za analizo študentskih anket, je s prehodom na bolonjske programe postal neprimeren. Zaradi potrebe po zagotovitvi učinkovite informacijske podpore analiziranju podatkov pridobljenih s študentskimi anketami je bilo potrebno uvesti sistem, ki bo te analize omogočal.

Za zagotavljanje informacijske podpore, ki služi analiziranju podatkov, pridobljenih s študentskimi anketami, je bil izbran sistem OLAP. OLAP je bil v prvi vrsti izbran zaradi svoje prijaznosti končnemu uporabniku, saj omogoča izvajanje kompleksnih poizvedb uporabnikom, ki nimajo naprednih informacijskih znaj. Poleg tega sistem omogoča nadgradnjo z novimi merami ali dimenzijami, v primeru, da se potreba po njih pokaže.

3.3.1 Struktura ankete

Anketa o pedagoškem delu je sestavljena iz sledečih sklopov:

- Demografski podatki (letnik, spol, smer, pričakovana ocena)
- Delo pri predmetu (vprašanja o organizaciji in delu pri predmetu)
- Delo pedagoga pri predmetu (kakšen se zdi pedagog, dosegljivost ipd)
- Opisni podatki (kaj bi pohvalili, grajali, izboljšali)

Samo izpolnjevanje ankete sledi takšni strukturi s tem, da je potrebno pred ocenjevanjem pedagogov le-te izbrati iz seznama nosilcev in sodelavcev pri predmetu.

3.3.2 Uporabniki

Analize pedagoških anket uporabljajo dve vrsti uporabnikov, pedagogi in strokovni delavci Pedagoškega centra EF.

Pedagogi imajo dostop do analiz, ki se nanašajo na predmete, kjer so nosilci ali sodelavci izvajanja, poleg tega pa imajo vodje usmeritev dostop do analiz za temeljne in izbirne smerne predmete smeri (predmeti tipa 1, 2, 3).

Sodelavci PEC imajo dostop do enakih analiz kot pedagogi, poleg tega pa potrebujejo še dodatne analize, ki služijo kot dokumentacija v habilitacijskem postopku, za delo usmerjevalnih komisij in za potrebe vodstva z namenom izboljšanja pedagoškega procesa. Poglobljene analize so namenjene sodelavcem PEC, zato mora biti njihova uporaba temu prilagojena.

3.3.3 Analize

Večina analiz obdeluje vsa vprašanja iz enega sklopa. Pri tem so demografski podatki manj pomembni. Glavni poudarek je na sklopih »Delo pri predmetu« in »Delo pedagoga«. Glavne možnosti izbora oziroma omejevanje podatkov, je študijsko leto, smer, vrsta študija in letnik, v katerem predmet nastopa. V odvisnosti od sklopa, so možnosti izbora oziroma omejevanja še tip predmeta, usmeritev, na katerih predmetih nastopa, pedagoga, naziva pedagoga, ter njihovo članstvo v katedri.

Osnovni podatek, ki ga je potrebno analizirati, je povprečna ocena ter število ocen, ki je bilo v izračunu uporabljeno. Vedno se upoštevajo samo veljavni odgovori. Pri vrtnanju v globino morajo biti dostopni še naslednji podatki: število veljavnih odgovorov, število neveljavnih odgovorov (takih, kjer je anketiranelec izbral odgovor »ne vem«, »nimam informacij« ali pa odgovora sploh ni izbral), srednjo vrednost in standardni odklon, ter najvišjo in najnižjo oceno.

Potrebno je zagotoviti tudi primerjavo izračunanih vrednosti v primerjavi s širšo skupino. Večinoma gre za primerjavo znotraj letnika ali znotraj katedre, v določenih primerih pa tudi znotraj usmeritve. V teh primerih nas vedno zanima njihov rang.

3.3.4 Posebnosti podatkov in podatkovnega modela

Večina posebnosti pri anketah o pedagoškem delu izhaja iz zgodovinskih razlogov ter zahtev po anonimnosti anket.

Zgodovinski razlogi se nanašajo na strukturo podatkov o predmetih in predmetnikih. Tu so strukture in podatki ločeni na stare, predbolonjske in nove, bolonjske predmete. Na dolgi rok se starih predmetov tako ali tako ne bo več izvajalo, zato jih pri poglobljenih analizah in gradnji podatkovnega skladišča nima smisla upoštevati.

Zaradi zahtev po anonimnosti ankete in anketirancev, v anketi ni natančnih splošnih in demografskih podatkov, kot so letnik študija, smer študija itd, ki pa so pomembni za izbor oziroma omejevanje podatkov. Anonimnost anket tako onemogoča primerjavo med letniki in smermi. Zato so ankete za redni, izredni in študij na daljavo ločene, enako velja za ankete podiplomskega študija.

3.3.5 Struktura anket

Za izdelavo podatkovnega skladišča je najprej potrebno dobro razumeti strukturo in vsebino študentskih anket. Prva delitev študentskih anket se nanaša na program, ki ga študent obiskuje.

Poznamo sledeče vrste študentskih anket:

- Študentska anketa o pedagoškem delu (3+)
- Študentska anketa o pedagoškem delu (stari programi)
- Študentska anketa o pedagoškem delu (+2)
- Študentska anketa o pedagoškem delu (podiplomski študij)
- Študentska anketa o pedagoškem delu (3+) (študij na daljavo)
- Študentska anketa o pedagoškem delu (3+) (izredni študij)

Vsaka izmed zgoraj navedenih študentskih anket o pedagoškem delu je sestavljena iz več vsebinskih sklopov.

Vsebinski sklopi ankete so:

- Podatki o študentu
- Vsebina in organizacija predmeta
- Pedagoško delo izvajalcev in izvajalk
- Mnenje študenta

- Komentar o obisku predavanj
- Izredni študij
- Študij na daljavo

Vsak sklop je sestavljen iz vprašanj. Obstaja več vrst vprašanj in sicer:

- Numerično brez osebe
- Numerično z osebo
- Tekstovno brez osebe
- Tekstovno z osebo

V nadaljnji analizi se bom omejil samo na vprašanja, ki dajejo numerične odgovore. Numerični odgovori so omejeni v razponu vrednosti od 1 pa do 5. Študent na vsako vprašanje odgovori z izborom vrednosti 1, 2, 3, 4 ali 5.

Znotraj numeričnih odgovorov obstajata dva zelo različna vsebinska sklopa. Prvi vsebinski sklop predstavljajo odgovori, ki se nanašajo na osebo in predmet. Drugi vsebinski sklop pa predstavljajo odgovori, ki se nanašajo samo na predmet. Takšna razdelitev je bistvena za nadaljnjo analizo podatkov.

3.4 Arhitektura sistema

Sistem za analizo anket je razdeljen na štiri nivoje. Na prvem nivoju se nahajata dve produkcijski bazi. Prva produkcijska baza se nahaja na strežniku MSSQL. V njej so shranjeni podatki o osebah in predmetih. Druga produkcijska baza pa se nahaja na strežniku Oracle in vsebuje podatke o anketah.

Na drugem nivoju se nahaja relacijsko podatkovno skladišče, ki temelji na sistemu za upravljanje podatkovnih baz MSSQL. Med prvim in drugim nivojem se izvaja proces ETL, ki polni relacijsko podatkovno skladišče. Proces ETL sproži uporabnik z dodeljenimi ustreznimi pravicami ali pa se proces proži samodejno, v vnaprej določenih časovnih intervalih.

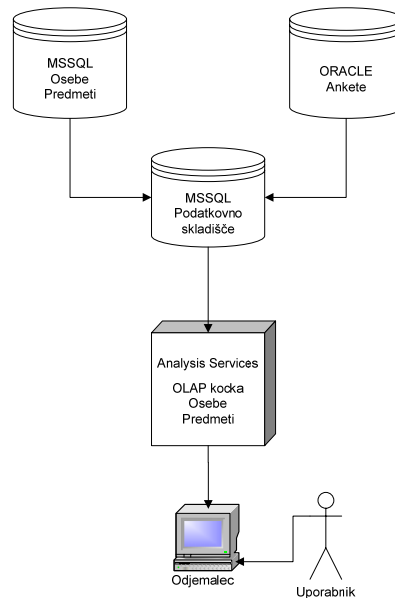
Proces vsebuje tri korake. V prvem koraku, se s pomočjo vnaprej pripravljenih stavkov SQL, iz produkcijske baze izvečejo zeleni podatki. V drugem koraku se izbrani podatki obdelajo, oziroma na podatkih se izvedejo potrebni preračuni. V tretjem koraku se obdelani podatki naložijo v relacijsko podatkovno skladišče.

Na tretjem nivoju se nahaja kocka OLAP, ki svoje osnovne podatke črpa naravnost iz relacijskega podatkovnega skladišča na drugem nivoju. Znotraj kocke OLAP se nahajajo v vnaprej definirane strukture, na podlagi katerih se v naprej izračunajo agregati. Te vnaprej izračunani agregati omogočajo krajši odzivni čas pri pregledovanju podatkov v kocki, saj pri izvajanju poizvedb ni potrebno preračunavati vseh osnovnih podatkov v podatkovnem

skladišču. Kocka OLAP je izdelana s pomočjo Microsoftovega analitičnega orodja Analysis services. S pomočjo Analysis services določimo vse strukture kocke OLAP, kot so: dimenzije, mere, funkcije agregacij, hierarhije znotraj dimenzij itd.

Na četrtem in zadnjem nivoju se nahajajo odjemalci. Preko odjemalca končni uporabniki dostopajo do želenih analitičnih podatkov znotraj kocke OLAP.

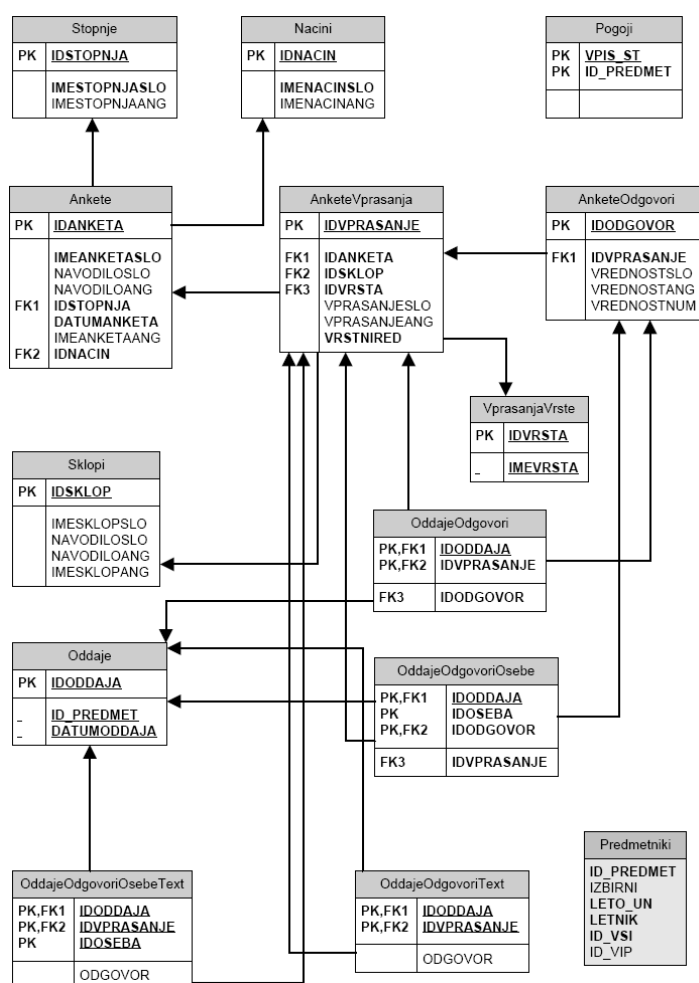
Slika 3: Tok podatkov od produkcijskih baz, preko podatkovnega skladišča, OLAP kocke do končnega uporabnika.



3.5 Struktura podatkov na produkcijski bazi

Podatki se v relacijsko podatkovno skladišče prenašajo iz dveh produkcijskih podatkovnih baz. Prva produkcijska podatkovna baza se nahaja na strežniku Oracle in vsebuje podatke o anketah. Podatkovni model na strežniku Oracle je predstavljen na Sliki 4.

Slika 4: Podatkovni model baze na Oracle strežniku.



Vir: Informatika Ekonomske fakultete, 2008.

Iz baze Oracle se črpajo podatki, ki so prikazani v tabelah 1, 2, 3, 4 in 5:

Tabela 1: Podatki, ki se črpajo iz tabele AnketeOdgovori

Polje	Opis polja
IDODGOVOR	Identifikacijska številka odgovora
VREDNOSTNUM	Numerična vrednost odgovora
IDVPRASANJE	Identifikacijska številka vprašanja

Tabela 2: Podatki, ki se črpajo iz tabele Oddaje

Polje	Opis polja
IDODDAJA	Identifikacijska številka oddaje
ID_PREDMET	Identifikacijska številka predmeta
DATUMODDAJA	Datum oddaje ankete

Tabela 3: Podatki, ki se črpajo iz tabele OddajeOdgovori

Polje	Opis polja
IDODDAJA	Identifikacijska številka oddaje
IDVPRASANJE	Identifikacijska številka vprašanja
IDODGOVOR	Identifikacijska številka odgovora

Tabela 4: Podatki, ki se črpajo iz tabele OddajeOdgovoriOsebe

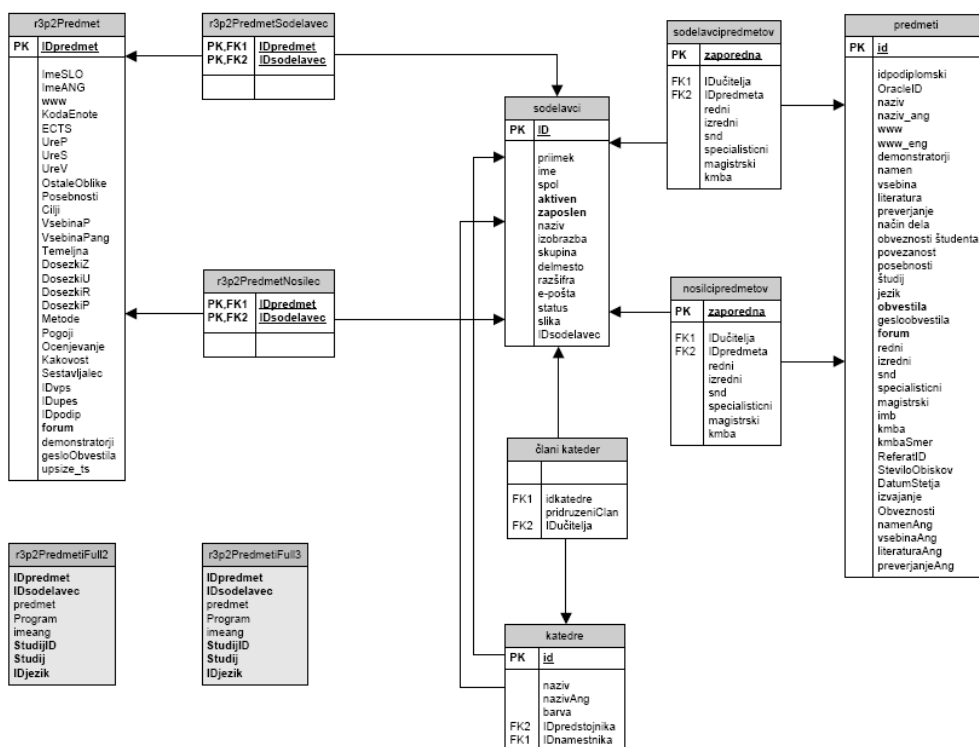
Polje	Opis polja
IDODDAJA	Identifikacijska številka oddaje
IDVPRASANJE	Identifikacijska številka vprašanja
IDODGOVOR	Identifikacijska številka odgovora
IDOSEBA	Identifikacijska številka osebe

Tabela 5: Podatki, ki se črpajo iz tabele AnketeVprasanja

Polje	Opis polja
IDVPRASANJE	Identifikacijska številka vprašanja
IDANKETA	Identifikacijska številka ankete
IDSKLOP	Identifikacijska številka sklopa ankete
IDVRSTA	Identifikacijska številka vrste ankete
VPRASANJESLO	Vprašanje v slovenskem jeziku
VPRASANJEANG	Vprašanje v angleškem jeziku
VRSTNIRED	Vrstni red vprašanj

Druga produkcijska baza pa se nahaja na strežniku MSSQL. Na tem strežniku se nahajajo podatki o osebah. Podatkovni model na strežniku MSSQL je predstavljena na Sliki 5.

Slika 5: Podatkovni model baze na MSSQL strežniku.



Vir: Informatika Ekonomske fakultete, 2008.

Iz baze MSSQL se črpajo podatki, ki so prikazani v tabelah 6, 7 in 8:

Tabela 6: Podatki, ki se črpajo iz tabele sodelavci

Polje	Opis polja
ID	Identifikacijska številka osebe
ime	Ime osebe
Priimek	Priimek osebe

Tabela 7: Podatki, ki se črpajo iz tabele predmeti

Polje	Opis polja
OracleID	Identifikacijska številka predmeta po starem programu
naziv	Naziv predmeta po starem programu

Tabela 8: Podatki, ki se črpajo iz tabele r3p2predmet

Polje	Opis polja
IDpredmet	Identifikacijska številka predmeta po novem programu (3+2)
imeslo	Naziv predmeta po novem programu (3+2)

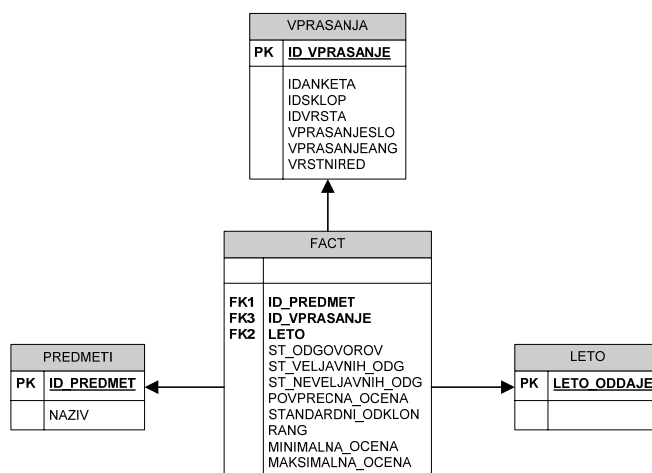
3.6 Podatkovno skladišče PEC

Relacijsko podatkovno skladišče PEC temelji na sistemu za upravljanje podatkovnih baz MSSQL. Znotraj podatkovnega skladišča se nahajajo vse tabele, ki so potrebne za izgradnjo sistema OLAP, namenjenega za analiziranje odgovorov pridobljenih s študentskimi anketami.

3.6.1 Zvezdna shema PREDMET

V središču zvezdna sheme PREDMET se nahaja tabela dejstev. S tabelo dejstev so preko ključev povezane še tri dimenzijske tabele. Dimenzijske tabele so VPRASANJA, PREDMETI in LETO. Tabela dejstev vsebuje tuje ključe vseh dimenzij in vrednostne attribute. Ključi dimenzij znotraj tabele dejstev so ID_PREDMET, ID_VPRASANJE in LETO. Vrednostni atributi znotraj tabele dejstev so ST_ODGOVOROV, ST_VELJAVNIH_ODG, ST_NEVELJAVNIH_ODG, POVPRECNA_OCENA, STANDARDNI_ODKLON, RANG, MINIMALNA_OCENA in MAKSIMALNA_OCENA.

Slika 6: Zvezda shema Predmeti.



3.6.2 ETL proces predmeti

Najprej pogledajmo proces ekstrakcije, transformacije in polnjenja za tabelo dejstev. V prvi fazi se prešteje število vseh odgovorov. To se naredi tako, da se prešteje število vseh odgovorov, ki se nanašajo na vprašanja sklopa 10 - Vsebino in organizacijo predmeta. Ker se v sklopu 10 - Vsebina in organizacija predmeta, nahajajo samo odgovori vrste 1 - Numerični in brez osebe, omejevanje po vrsti vprašanja ni potrebno. Ekstrakcijo in transformacijo se izvede s spodnjim stavkom:

```

SELECT COUNT(*) AS SteviloOdgovorov,
       case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end AS datum,
       id_predmet,
       idvprasanje
FROM up_ankete.AnketeOdgovori
INNER JOIN (up_ankete.oddaje
INNER JOIN up_ankete.oddajeodgovori
ON up_ankete.oddaje.IDoddaja = up_ankete.oddajeodgovori.IDoddaja)
ON up_ankete.AnketeOdgovori.IDodgovor = up_ankete.oddajeodgovori.IDodgovor
WHERE up_ankete.anketeodgovori.idvprasanje in (SELECT idvprasanje
FROM up_ankete.anketevprasanja
WHERE idsklop = 10 )
GROUP BY id_predmet, case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end, idvprasanje
ORDER BY id_predmet;

```

Pridobljene zapise se zapiše v stolpec ST_ODGOVOROV, v tabeli dejstev Predmeti.

V naslednjem koraku se prešteje število manjkajočih oziroma neveljavnih odgovorov. Ekstrakcijo in transformacijo se izvede s spodnjim stavkom:

```

SELECT COUNT(*) AS SteviloOdgovorov,
       case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end AS datum,
       id_predmet,
       idvprasanje
FROM up_ankete.AnketeOdgovori
INNER JOIN (up_ankete.oddaje
INNER JOIN up_ankete.oddajeodgovori
ON up_ankete.oddaje.IDoddaja = up_ankete.oddajeodgovori.IDoddaja)
ON up_ankete.AnketeOdgovori.IDodgovor = up_ankete.oddajeodgovori.IDodgovor
WHERE ((AnketeOdgovori.VrednostNUM)=0) OR ((AnketeOdgovori.VrednostNUM)=6)
AND up_ankete.anketeodgovori.idvprasanje in (SELECT idvprasanje
FROM up_ankete.anketevprasanja
WHERE idsklop = 10 )
GROUP BY id_predmet, case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end, idvprasanje
ORDER BY id_predmet;

```

Pridobljene zapise se zapiše v stolpec ST_NEVELJAVNIH_ODG, v tabeli dejstev Predmeti.

Ko je število odgovorov in število manjkajočih odgovorov znano, se lahko izračuna še število veljavnih odgovorov, ki se zapiše v stolpec ST_VELJAVNIH_ODG in kazalnike povprečne vrednosti odgovorov za predmet in vprašanje po posameznem letu, vrednost standardnega odklona za predmet in vprašanje po posameznem letu, minimalno oceno za predmet in vprašanje po posameznem letu in maksimalno oceno za predmet in vprašanje po posameznem letu.

Vseh pet kazalnikov se izračuna s pomočjo standardnih funkcij, ki jih podpira baza Oracle. Te funkcije so:

- COUNT – Funkcija prešteje število zapisov, v našem primeru gre za število odgovorov
- STDDEV – Funkcija izračuna vrednost standardnega odklona
- AVG – Funkcija izračuna povprečno vrednost
- MAX – Funkcija izračuna maksimalno vrednost

- MIN – Funkcija izračuna minimalno vrednost

Vse izračune se omeji samo na veljavne odgovore, to pomeni na takšne, ki imajo vrednosti večje 0 in manjše 6, poleg tega pa se nahajajo znotraj vprašanj sklopa 10 - Vsebinsko in organizacijo predmeta. Ker se v sklopu 10 - Vsebinsko in organizacija predmeta, nahajajo samo odgovori vrste 1 - Numerični in brez osebe, omejevanje po vrsti vprašanja ni potrebno. Ekstrakcijo in transformacijo se izvede s spodnjim stavkom:

```
SELECT distinct id_predmet,
  oddajeodgovori.idvprasanje,
    case when to_char(trunc(datumoddaja, 'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja, 'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja, 'YYYY'),'YYYY'))-1 end AS datum,
    count (ANKETEODGOVORI.vrednostnum) over (partition by id_predmet,
oddajeodgovori.idvprasanje, case when to_char(trunc(datumoddaja, 'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja, 'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja, 'YYYY'),'YYYY'))-1 end ) as steviloodgovorov,
    stddev (ANKETEODGOVORI.vrednostnum) over (partition by id_predmet,
oddajeodgovori.idvprasanje, case when to_char(trunc(datumoddaja, 'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja, 'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja, 'YYYY'),'YYYY'))-1 end ) as stddeviacija,
    avg (ANKETEODGOVORI.vrednostnum) over (partition by id_predmet,
oddajeodgovori.idvprasanje, case when to_char(trunc(datumoddaja, 'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja, 'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja, 'YYYY'),'YYYY'))-1 end ) as povprecaocena
    max (ANKETEODGOVORI.vrednostnum) over (partition by id_predmet,
oddajeodgovori.idvprasanje, case when to_char(trunc(datumoddaja, 'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja, 'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja, 'YYYY'),'YYYY'))-1 end ) as maxocena,
    min (ANKETEODGOVORI.vrednostnum) over (partition by id_predmet,
oddajeodgovori.idvprasanje, case when to_char(trunc(datumoddaja, 'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja, 'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja, 'YYYY'),'YYYY'))-1 end ) as minocena
FROM up_ankete.AnketeOdgovori
INNER JOIN (up_ankete.oddaje
INNER JOIN up_ankete.oddajeodgovori
ON up_ankete.oddaje.IDoddaja = up_ankete.oddajeodgovori.IDoddaja)
ON up_ankete.AnketeOdgovori.IDodgovor = up_ankete.oddajeodgovori.IDodgovor
WHERE up_ankete.anketeodgovori.vrednostnum > 0
AND up_ankete.anketeodgovori.vrednostnum < 6
AND up_ankete.anketeodgovori.idvprasanje in (SELECT idvprasanje
FROM up_ankete.anketevprasanja
WHERE idsklop = 10 )

ORDER BY id_predmet;
```

Pridobljene vrednosti se zapiše v stolpce STANDARDNI_ODKLON, POVPRECNA_OCENA, MAKSIMALNA_OCENA in MINIMALNA_OCENA, v tabeli dejstev Predmeti.

Zadnja vrednost, ki se jo računa, je rang. Rang se izračuna tako, da se najprej izračuna povprečno vrednost odgovorov za predmet in vprašanje po posameznih letih, po postopku, ki je že opisan zgoraj. Nato s pomočjo funkcije RANK izračuna vrednost ranga. Tudi ta izračunu je potrebno omejiti samo na veljavne odgovore, to pomeni na takšne, ki imajo vrednosti večje 0 in manjše 6, poleg tega pa se nahajajo znotraj vprašanj sklopa 10 - Vsebinsko in organizacijo predmeta. Ker se v sklopu 10 - Vsebinsko in organizacija predmeta, nahajajo samo odgovori vrste 1 - Numerični in brez osebe, omejevanje po vrsti vprašanja ni potrebno. Ekstrakcijo in transformacijo se izvede s spodnjim stavkom:

```

SELECT      id_predmet,
            povprecje,
            datum,
            idvprasanje,
            rank() OVER(ORDER BY povprecje desc) AS rank
FROM (SELECT distinct oddaje.id_predmet,
            oddajeodgovori.idvprasanje,
            case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end AS datum,
            avg (ANKETEODGOVORI.vrednostnum) over (partition by oddaje.id_predmet,
            oddajeodgovori.idvprasanje, case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end ) as povprecje
FROM up_ankete.AnketeOdgovori
INNER JOIN (up_ankete.oddaje
INNER JOIN up_ankete.oddajeodgovori
ON up_ankete.oddaje.IDoddaja = up_ankete.oddajeodgovori.IDoddaja)
ON up_ankete.AnketeOdgovori.IDodgovor = up_ankete.oddajeodgovori.IDodgovor
WHERE up_ankete.anketeodgovori.vrednostnum > 0
AND up_ankete.anketeodgovori.vrednostnum < 6
AND up_ankete.anketeodgovori.idvprasanje in (SELECT idvprasanje
FROM up_ankete.anketevprasanja
WHERE idsklop = 10 ))

Order by id_predmet;

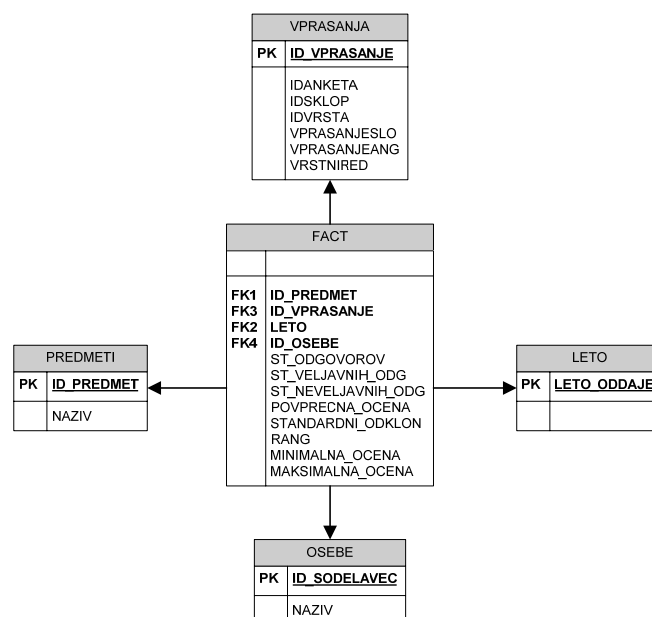
```

Pridobljene vrednosti se zapišejo v stolpec RANG, v tabeli dejstev Predmeti.

3.6.3 Zvezdna shema OSEBE

Tako kot pri zvezdni shemi PREDMET, se pri zvezdni shemi OSEBE v središču nahaja tabela dejstev, ki je preko ključev povezana z dimenzijskimi tabelami. Prve tri dimenzijske tabele VPRASANJA, PREDMETI in LETO so enake kot pri zvezdni shemi PREDMET, dodana pa je še dodatna dimenzijska tabela OSEBE. Tuji ključi dimenzij znotraj tabele dejstev so ID_PREDMET, ID_VPRASANJE, LETO in ID_OSEBE. Vrednostni atributi tabele dejstev so ST_ODGOVOROV, ST_VELJAVNIH_ODG, ST_NEVELJAVNIH_ODG, POVPRECNA_OCENA, STANDARDNI_ODKLON, RANG, MINIMALNA_OCENA in MAKSIMALNA_OCENA.

Slika 7: Zvezdna shema Osebe.



3.6.4 ETL proces osebe

Za proces ekstrakcije, transformacije in polnjenja za tabele dejstev se najprej prešteje število vseh odgovorov. To se naredi tako, da se prešteje število vseh odgovorov, ki se nanašajo na vprašanja sklopa 2 – Pedagoško delo izvajalcev in izvajalk. Poleg sklopa je potrebno vprašanja omejiti tudi po vrsti odgovorov in sicer na vrsto odgovorov 2 - Numerični z osebo. Ekstrakcijo in transformacijo se izvede s spodnjim stavkom:

```
SELECT COUNT(*) AS SteviloOdgovorov,
           id_predmet,
           idoseba,
           oddajeodgovoriosebe.idvprasanje,
           case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end AS datum
FROM up_ankete.oddaje
INNER JOIN (up_ankete.AnketeOdgovori
INNER JOIN up_ankete.OddajeOdgovoriOsebe
ON up_ankete.AnketeOdgovori.IDodgovor = up_ankete.OddajeOdgovoriOsebe.IDodgovor)
ON oddaje.IDoddaja = OddajeOdgovoriOsebe.IDoddaja
WHERE oddajeodgovoriosebe.idvprasanje in (SELECT idvprasanje
FROM up_ankete.anketevprasanja
WHERE idsklop = 2
AND IDVRSTA = 2)
GROUP BY idoseba, id_predmet, case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end, oddajeodgovoriosebe.idvprasanje
ORDER BY idoseba;
```

Pridobljene zapise se zapiše v stolpec ST_ODGOVOROV, v tabeli dejstev Osebe.

Ko je znano število vseh odgovorov, se prešteje število manjkajočih oziroma neveljavnih odgovorov. To se naredi tako, da se prešteje število odgovorov, ki imajo numerično vrednost 0 ali 6. Ekstrakcijo in transformacijo se izvede s spodnjim stavkom:

```
SELECT COUNT(*) AS SteviloOdgovorov,
           id_predmet,
           idoseba,
           oddajeodgovoriosebe.idvprasanje,
           case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end AS datum
FROM up_ankete.oddaje
INNER JOIN (up_ankete.AnketeOdgovori
INNER JOIN up_ankete.OddajeOdgovoriOsebe
ON up_ankete.AnketeOdgovori.IDodgovor = up_ankete.OddajeOdgovoriOsebe.IDodgovor)
ON up_ankete.oddaje.IDoddaja = up_ankete.OddajeOdgovoriOsebe.IDoddaja
WHERE (((AnketeOdgovori.VrednostNUM)=0) OR ((AnketeOdgovori.VrednostNUM)=6))
AND oddajeodgovoriosebe.idvprasanje in (SELECT idvprasanje
FROM up_ankete.anketevprasanja
WHERE idsklop = 2
AND IDVRSTA = 2)
GROUP BY idoseba, id_predmet, case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end, oddajeodgovoriosebe.idvprasanje
ORDER BY idoseba;
```

Pridobljene zapise se zapiše v stolpec ST_NEVELJAVNIH_ODG, v tabeli dejstev Osebe.

Po izračunu števila odgovorov in števila veljavnih odgovorov, se lahko izračuna še preostale kazalnike. Ti kazalniki so število veljavnih odgovorov, povprečna vrednost

odgovorov za osebo, predmet in vprašanje po posameznem letu, vrednost standardnega odklona za osebo, predmet in vprašanje po posameznem letu, minimalna ocena za osebo, predmet in vprašanje po posameznem letu in maksimalna ocena za osebo, predmet in vprašanje po posameznem letu.

Kazalnike se izračuna s pomočjo istih funkciji, kot pri predmetih. Izračune se omeji samo na veljavne odgovore, to pomeni na takšne, ki imajo vrednosti večjo od 0 in manjšo od 6. Poleg omejitve na veljavne odgovore se je potrebno omejiti tudi na take, ki se nahajajo znotraj vprašanj sklopa 2 - Pedagoško delo izvajalcev in izvajalk in so vrste 2 - Numerični z osebo. Ekstrakcijo in transformacijo se izvede s spodnjim stavkom:

```
SELECT distinct idoseba,
               id_predmet,
               oddajeodgovoriosebe.idvprasanje,
               case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end AS datum,
               count (ANKETEODGOVORI.vrednostnum) over (partition by idoseba, id_predmet,
oddajeodgovoriosebe.idvprasanje, case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end ) as steviloodgovorov,
               stddev (ANKETEODGOVORI.vrednostnum) over (partition by idoseba, id_predmet,
oddajeodgovoriosebe.idvprasanje, case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end ) as stddeviacija,
               avg (ANKETEODGOVORI.vrednostnum) over (partition by idoseba, id_predmet,
oddajeodgovoriosebe.idvprasanje, case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end ) as povprecaocena
               max (ANKETEODGOVORI.vrednostnum) over (partition by idoseba, id_predmet,
oddajeodgovoriosebe.idvprasanje, case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end ) as maxocena,
               min (ANKETEODGOVORI.vrednostnum) over (partition by idoseba, id_predmet,
oddajeodgovoriosebe.idvprasanje, case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end ) as minocena
FROM up_ankete.oddaje
INNER JOIN (up_ankete.AnketeOdgovori
INNER JOIN up_ankete.OddajeOdgovoriOsebe
ON up_ankete.AnketeOdgovori.IDodgovor = up_ankete.OddajeOdgovoriOsebe.IDodgovor)
ON up_ankete.oddaje.IDoddaja = up_ankete.OddajeOdgovoriOsebe.IDoddaja
WHERE anketeodgovori.vrednostnum > 0 AND anketeodgovori.vrednostnum < 6
AND oddajeodgovoriosebe.idvprasanje in (SELECT idvprasanje
FROM up_ankete.anketevprasanja
WHERE idsklop = 2
AND IDVRSTA = 2)

ORDER BY idoseba;
```

Pridobljene vrednosti se zapišejo v stolpce ST_VELJAVNIH_ODG, STANDARDNI_ODKLON, POVPRECNA_OCENA, MAKSIMALNA_OCENA in MINIMALNA_OCENA, v tabeli dejstev Osebe.

Zadnji kazalnik, ki se ga izračuna je rang. Rang se izračuna tako, da se najprej izračuna povprečno vrednost odgovorov za osebo, predmet in vprašanje po posameznem letu. Iz izračunane povprečne vrednosti odgovorov, se s pomočjo standardne Oracleove funkcije RANK, izračuna vrednost ranga. Tako kot pri prejšnjem izračunu, se je potrebno omejiti samo na veljavne odgovore, to pomeni na takšne, ki imajo vrednosti večjo od 0 in manjšo od 6. Poleg omejitve na veljavne odgovore se je potrebno omejiti tudi na take, ki se

nahajajo znotraj vprašanj sklopa 2 - Pedagoško delo izvajalcev in izvajalk in so vrste 2 - Numerični z osebo. Ekstrakcijo in transformacijo se izvede s spodnjim stavkom:

```

SELECT
            idoseba,
            id_predmet,
            povprecje,
            datum,
            idvprasanje,
            -- case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
            to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
            to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end,
            rank() OVER(ORDER BY povprecje desc) AS rank
FROM (SELECT distinct(OddajeOdgovoriOsebe.idoseba),
            oddaje.id_predmet,
            oddajeodgovoriosebe.idvprasanje,
            case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
            to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
            to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end AS datum,
            avg (ANKETEODGOVORI.vrednostnum) over (partition by idoseba, id_predmet,
            oddajeodgovoriosebe.idvprasanje, case when to_char(trunc(datumoddaja,'MM'),'MM')>'09' then
            to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY')) else
            to_number(to_char(trunc(datumoddaja,'YYYY'),'YYYY'))-1 end ) AS povprecje
            FROM up_ankete.oddaje
            INNER JOIN (up_ankete.AnketeOdgovori
            INNER JOIN up_ankete.OddajeOdgovoriOsebe
            ON up_ankete.AnketeOdgovori.IDodgovor = up_ankete.OddajeOdgovoriOsebe.IDodgovor)
            ON up_ankete.oddaje.IDoddaja = up_ankete.OddajeOdgovoriOsebe.IDoddaja
            WHERE anketeodgovori.vrednostnum > 0 AND anketeodgovori.vrednostnum < 6
            AND oddajeodgovoriosebe.idvprasanje in (SELECT idvprasanje
            FROM up_ankete.anketevprasanja
            WHERE idsklop = 2
            AND IDVRSTA = 2))
ORDER BY idoseba;

```

Pridobljene vrednosti se zapišejo v stolpec RANG, v tabeli dejstev Osebe.

3.7 Shema kocke OLAP

Kot je že opisano v predhodnih poglavjih, sem se odločili za izdelavo dveh kock OLAP. Prva kocka OLAP vsebuje podatke o predmetih in druga o osebah po predmetih.

3.7.1 Predmeti

Kocka OLAP Predmeti za svoj podatkovni vir uporablja podatkovno skladišče, bolj natančno tabele VPRASANJA, LETO, PREDMETI IN FACT (Predmeti), znotraj podatkovnega skladišča.

Mere kocke OLAP so število odgovorov, število veljavnih odgovorov, število neveljavnih odgovorov, povprečna ocena, standardni odklon, rang, minimalna ocena in maksimalna ocena. Vrednosti mer na osnovnem nivoju se črpajo naravnost iz podatkovnega skladišča, saj so bile preračunane že v toku ETL procesa. Zato nadaljnja obdelava podatkov na osnovnem nivoju ni potrebna. Potrebno je le določiti funkcije agregacij za združevanje podatkov.

Za mere število odgovorov, število veljavnih odgovorov in število neveljavnih odgovorov se za funkcijo agregacije uporabi enostavna funkcija vsote (ang. *sum*). Za ostale mere, ki

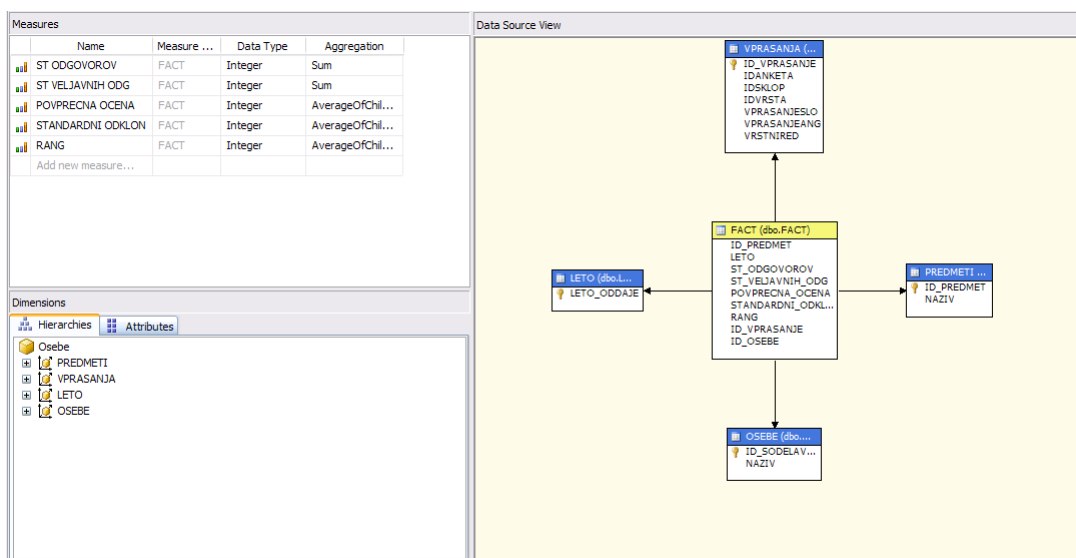
jih ne da enostavno seštevati, pa je potrebno narediti izračunane mere. To se naredi tako, da se v prvem koraku za vsako izračunano mero, najprej naredi dve meri. Prva mera za funkcijo agregacije uporablja funkcijo vsote in druga mera funkcijo štetja (ang. *count*). Izračunano mero se dobi tako, da se mero, ki uporablja funkcijo vsote deli z mero, ki uporablja funkcijo štetja. Tako se dobi izračunano mero, ki pri združevanju podatkov prikazuje povprečno vrednost združenih podatkov. Meri, ki se ju uporabi za vmesni izračun se skrije, da se ne prikazujeta v končni kocki, saj so ti podatki nerelevantni.

Dimenzije kocke OLAP določajo tabele dimenzij. V našem primeru so to tabele PREDMETI, VPRASANJA in LETO. Znotraj posamezne dimenzije so določene tudi hierarhije posamezne dimenzije. V primeru kocke Predmeti je določitev hierarhije logična samo znotraj dimenzije VPRASANJA.

3.7.2 Osebe

Kocka OLAP Osebe za svoj podatkovni vir tako kot kocka Predmeti, uporablja podatkovno skladišče. Znotraj podatkovnega skladišča uporablja tabele VPRASANJA, LETO, PREDMETI, OSEBE in FACT (Osebe). Slika 8 predstavlja podatkovni vir, mere, dimenzije in hierarhije kot so vidne v orodju MS Analysis services. V levem zgornjem kotu so vidne mere kocke, levo spodaj se nahajajo dimenzije in hierarhije. Na desni strani slike je viden podatkovni vir kocke, ki ima obliko zvezdne sheme.

Slika 8: Podatkovni vir, mere in hierarhije kocke Oseba.



Mere OLAP kocke so število odgovorov, število veljavnih odgovorov, število neveljavnih odgovorov, povprečna ocena, standardni odklon, rang, minimalna in maksimalna ocena. Vrednosti mer na osnovnem nivoju so že preračunane, zato se črpajo naravnost iz podatkovnega skladišča. Tako nadaljnja obdelava podatkov na osnovnem nivoju ni

potrebna. Določi se le funkcije agregacij, ki se uporabljajo pri združevanju podatkov. Funkcije agregaciji se določijo na enak način kot pri kocki Predmeti.

Podatki o dimenzijah OLAP kocke se črpajo iz dimenzijskih tabel PREDMETI, VPRASANJA, LETO in OSEBE. Za posamezne dimenzije je potrebno določiti tudi hierarhijo podatkov znotraj njih. V tem primeru kocke Osebe je edina logična hierarhija znotraj dimenzije VPRASANJA.

Primer uporabe kocke v orodju MS Analysis services je prikazan na sliki 9. Na levi strani slike so prikazane vse mere in dimenzije kocke. V zgornjem delu se lahko izbere dimenzije za katere se želi v kocki, po izbranih pogojih omejiti prikaz vrednosti. V spodnjem delu se nahaja tabela, v kateri se določi, katera polja bodo prikazana kot stolpci in katera kot vrstice. Za stolpce in vrstice se lahko izberejo dimenzije. V osrednjem delu tabele se določijo mere, ki so predmet analize. Primer na sliki 9 prikazuje analiziranje števila odgovorov po vprašanju, predmetu in letu. V stolpcih se nahajajo dimenzije predmeti in leto, v vrsticah pa dimenzija vprašanja. Po dimenziji vprašanj se lahko z operacijo vrtanja v globino, pomika navzdol po hierarhičnih nivojih dimenzije ali z operacijo zvijanja, navzgor po hierarhičnih nivojih dimenzije. Prikaz v tabeli se lahko omejili še tako, da se v zgornjem delu, kjer so prikazane dimenzije, z dodatnimi pogoji omejili kaj se prikazuje po posameznih dimenzijah. Na primer v dimenziji predmeti se izbere, da prikazuje samo predmet monetarna ekonomija in osnove računovodstva.

Slika 9: Poizvedba nad kocko Osebe

Dimension	Hierarchy	Operator	Filter Expression
LETO	LETO	Equal	
OSEBE	NAZIV	Equal	
PREDMETI	NAZIV	Equal	
VPRASANJA	Hierarchy	Equal	
<Select dimension>			

Drop Filter Fields Here		NAZIV ▾ LETO ▾		Monetarna ekonomija		Osnove računovodstva		Grand Total	
		2005	2006	2007	Total	2004	2005	Total	
		ST ODGOVOR	ST ODGOVOR	ST ODGOVOR	ST ODGOVOR	ST ODGOVOR	ST ODGOVOR	ST ODGOVOR	ST ODGOVOR
1 ▾	IDSI_IDVE_VPRASANJESLO								
2 ▾	Gledano v celoti, je delo izvajalca/ke kark	225	225	115	565	166	104	270	835
	Izvajalec/ka je dosegljiv/a na govornih l	22	22	9	53	3	3	56	109
	Izvajalec/ka me spodbuja k razmišljanju	547	547	437	1531	169	104	273	1804
	Izvajalec/ka podaja snov povezano in ra	225	225	115	565	173	104	277	842
	Izvajalec/ka prihaja v predavalnico dobr	547	450	437	1434	169	3	172	1606
	Izvajalec/ka zna vzbuditi zanimanje štud	225	225	115	565	167	104	271	836
	Menim, da ima izvajalec/ka korekten odn	22	22	9	53	68	3	71	124
	Menim, da ima izvajalec/ka veliko strokov	450	450	437	1337	68	3	71	1408
	Pri tem izvajalcu/ku bi želel/a poslušati še	22	22	9	53	68	3	71	124
	Total	2285	2188	1683	6156	1101	431	1532	7688
	Total	2285	2188	1683	6156	1101	431	1532	7688
	Grand Total	2285	2188	1683	6156	1101	431	1532	7688

SKLEP

V današnjem hitro spreminjajočem svetu je, za uspešno poslovanje podjetij in organizacij, pravočasno sprejemanje pravih odločitev ključnega pomena. Za hiter in učinkovit proces odločanja je pomemben hiter dostop do točnih informacij.

Podjetja dnevno obdelujejo in hranijo ogromne količine podatkov. Za hitro, enostavno in fleksibilno analizo teh podatkov, ki ne zahteva dolgotrajnih priprav podatkov ali

naprednih informacijskih znanj, potrebujemo prave informacijske rešitve oziroma orodja. Omenjenim informacijskim rešitvam pravimo sistemi za podporo odločanju.

V diplomski nalogi sem obravnaval izgradnjo sistema za podporo odločanju in različnih modelov, ki se jih lahko uporabi pri zbiranju, obdelavi in shranjevanju podatkov znotraj sistema.

V prvem, teoretičnem delu sem predstavil zakaj transakcijske ali OLTP baze niso primerne za pripravljanje analiz in zakaj so podatkovna skladišča primerna. Katera podatkovna shema je, glede na dane omejitve in želje, najbolj primerna za podatkovno skladišče. Kako s pomočjo procesa ETL podatke iz različnih zunanjih virov najprej ekstrahiramo, transformiramo in nato naložimo podatke v podatkovno skladišče. Ter kateri izmed modelov OLAP je najprimernejši za analiziranje podatkov, glede na dane omejitve in uporabniške zahteve.

V drugem, praktičnem delu je predstavljen dejanski prototip sistema za podporo odločanju, ki služi za analizo podatkov pridobljenih z anketiranjem študentov o delu predavateljev in predmetih na Ekonomski fakulteti v Ljubljani. Predstavljen je celoten proces izgradnje prototipa od uporabniških zahtev, analize obstoječega stanja, analize anket, tehničnih omejitev, analize strukture podatkov na produkcijskih transakcijskih bazah, izgradnje podatkovnega skladišča, podrobnega opisa ETL procesa, definiranja in izgradnje OLAP kocke.

Uporabniške zahteve in analiza obstoječega stanja so sestavljene iz zahtev PEC-a, o tem katere podatke želijo analizirati in omejitve, ki jih postavljajo pri uporabi podatkov. V analizi tehničnih omejitev je predstavljeno stanje dveh OLTP baz. Prva podatkovna baza se nahaja na strežniku Oracle in vsebuje podatke o anketah. Druga podatkovna baza se nahaja na strežniku MSSQL in vsebuje podatke o sodelavcih in predmetih. V analizi strukture podatkov je predstavljena struktura podatkov na obeh OLTP bazah in nabor podatkov, ki se jih potrebuje v podatkovnem skladišču. Podatkovno skladišče uporablja sistem za upravljanje podatkovnih baz MSSQL. Za podatkovno shemo podatkov v podatkovnem skladišču sem izbral zvezdno shemo. Za analiziranje podatkov se uporablja kocki OLAP, ki sta zgrajeni s pomočjo orodja MS Analysis Services.

Novi sistem za analiziranje rezultatov pridobljenih s študentskim anketami bo v pomoč strokovnim sodelavcem PEC-a, za hitro in fleksibilno analiziranje podatkov, ki bodo služili kot pomoč pri odločanju glede uspešnosti posameznih sodelavcev Ekonomske fakultete in njihovih predmetov. Prav tako bo uporaben pri pripravi raznih poročil in priporočil iz področja spremljanja kakovosti, razvoja znanj in veščin sodelavcev za izvajanje pedagoškega dela.

Poleg vseh prednosti, ki jih ponuja nov sistem za analiziranje odgovorov pridobljenih s študentskimi anketami, ima tudi veliko pomanjkljivost. Največja slabost pri predstavljenem prototipu je ta, da nima možnosti analiziranja podatkov po smereh študija

in posameznih letnikih študija. Vendar pa ta pomanjkljivost ne izvira iz tehničnih omejitev sistema, ampak iz osnovne zahteve, da ne sme obstajati nikakršna povezava med podatki o študentih in odgovori. Zaradi pomanjkanja teh kritičnih podatkov sem na žalost moral izdelati precej okrnjeno verzijo kock OLAP. Navkljub tej pomanjkljivosti mislim, da je prototip, z manjšimi spremembami, možno aplicirati na dejansko produkcijsko okolje, kjer bi ga lahko uporabljali dejanski uporabniki z namenom, za katerega je bil zasnovan – podpora odločanju.

LITERATURA IN VIRI

- 1 Amo, W, C. (2000). *Microsoft Sql Server Olap Developer's Guid*. B.k: John Wiley & Sons.
- 2 Atre, S. & Moss, L. T. (2003). *Business Intelligence Roadmap. The Complete Project Lifecycle for Decision-Support Applications*. ZDA: Addison-Wesley.
- 3 Berge, S (2001). *SQL Server and ADO programming complete*. San Francisco: Sybex.
- 4 *Business Intelligence*. Najdeno 11. oktobra 2009 na spletnem naslovu <http://bi.technologyevaluation.com/business-intelligence/>
- 5 *DataManagment*. Najdeno 19. septembra 2009 na spletnem naslovu <http://searchdatamanagement.techtarget.com/>
- 6 *Definition of data warehouse*. Najdeno 11. marca 2009 na spletnem naslovu <http://www.pcmag.com/encyclopedia>.
- 7 Grad, J & Jaklič, J.(1996). *Baze podatkov*. Ljubljana: Ekonomska fakulteta.
- 8 Informatika Ekonomske Fakultete (2006). Ljubljana: Ekonomska fakulteta.
- 9 Inmon, W. (2000). *Data Mart Does Not Equal Data Warehouse*. Najdeno 18. septembra na spletnem naslovu <http://www.information-management.com/>
- 10 Isemenger, D. (2001). *Microsoft analysis services*. B.k: Microsoft Corporation.
- 11 Jagarinec, D. (2005). *Operacije v OLAP-u*. Najdeno 24. decembra 2007 na spletnem naslovu www.mojmikro.si/articles/mi12_80-81%20podakovne.
- 12 Jaklič, J. (2002). *Upravljanje in uporaba podatkov*. Ljubljana: Ekonomska fakulteta.

- 13 Jaklič, J. (2008). *Predavanja Baze podatkov 2008/09*. Ljubljana: Ekonomska fakulteta.
- 14 Miller, J. P. & Fuld, L. (2001). *Millennium Intelligence: Understanding and Conducting Competitive Intelligence in the Digital Age*. Združene države Amerike: Information today inc.
- 15 Kaser, O. & Lemire, D. (2006). Attribute Value Reordering for Efficient Hybrid OLAP *Information Sciences*, 176 (16), 2279-2438.
- 16 Kimball, R. & Ross, M. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. B.k: Wiley.
- 17 Kimball, R. & Caserta, J. (2004). *The Data Warehouse ETL Toolkit*. Indianapolis: Wiley.
- 18 Krsnik, J. (2001). *Sprotno analitično procesiranje*. Najdeno 13. februarja 2008 na spletnem naslovu <http://www.bf.uni-lj.si/zoo/org/centre/slov/govedo/objave/olap/main.htm>.
- 19 Lalič, R. (2005). *Vpeljava sistema OLAP v podporo obveščevalnemu in analitičnemu delu policije na področju prepovedanih drog*. Ljubljana: Ekonomska fakulteta.
- 20 *OLAP*. Najdeno 16. februarja 2008 na spletnem naslovu <http://wiki.fmf.uni-lj.si/wiki/OLAP>.
- 21 *Online analytical processing*. Najdeno 1. junija 2009 na spletnem naslovu http://en.wikipedia.org/wiki/Online_analytical_processing.
- 22 *Pec*. Najdeno 1. julija 2009 na spletnem naslovu <http://www.ef.uni-lj.si/enote/pec/home.asp>.
- 23 Pedersen, T.B. & Jensen, C.S. (2001). Multidimensional database technology. *Computer Volume*, 34 (12), 40 – 46.

- 24** Pendse, N. (2009). *The OLAP Report*. Najdeno 6. maja 2009 na spletnem naslovu <http://www.olapreport.com/>.
- 25** *Študentske ankete*. Najdeno 1. julija 2009 na spletnem naslovu <http://www.ef.uni-lj.si/enote/pec/StudentskeAnkete.asp>.
- 26** *Uvedba elektronskih anket*. Najdeno 1. julija 2009 na spletnem naslovu <http://www.ef.uni-lj.si/enote/pec>

PRILOGE

Priloga 1: Seznam pogosto uporabljenih kratic

ETL	Extract Transform Load	Ekstrakcija Transformacija Nalaganje
HOLAP	Hybrid Online Analytical Processing	hibridno sprotno analitično obdelavo podatkov
MOLAP	Multidimensional Online Analytical Processing	Več dimenzionalna sprotna analitična obdelava podatkov
OLAP	On Line Analytical Processing	Sprotna analitična obdelava podatkov
OLTP	Online transaction processing	Sprotna obdelava transakcij
PEC	Pedagoški center Ekonomske Fakultete v Ljubljani	Pedagoški center Ekonomske Fakultete v Ljubljani
ROLAP	Relational On Line Analytical Processing	Relacijska sprotna analitična obdelava podatkov
SQL	Structured query language	Strukturirani povpraševalni jezik za delo s podatkovnimi bazami