

UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

DIPLOMSKO DELO

UPORABA POČASI SE SPREMINJAJOČE DIMENZIJE V
PODATKOVNIH SKLADIŠČIH

ALEŠ RAVBAR

Ljubljana, april 2003

IZJAVA

Študent _____ izjavljam, da sem avtor tega diplomskega dela, ki sem ga napisal pod mentorstvom _____ in dovolim objavo diplomskega dela na fakultetnih spletnih straneh.

V Ljubljani, dne _____ Podpis: _____

Kazalo

1. Sistemi za podporo odločanju	2
2. Podatkovna skladišča	3
2.1. Opredelitev	3
2.2. Arhitektura	5
2.3. Sestavni deli podatkovnega skladišča	7
3. Faze gradnje podatkovnega skladišča	13
4. Dimenzija v podatkovnih skladiščih	15
4.1. Počasi se spreminjajoča dimenzija	17
4.2. Hitro se spreminjajoča majhna dimenzija	19
4.3. Velika dimenzija	19
4.4. Hitro se spreminjajoča velika dimenzija	20
4.5. Degenerirana dimenzija	21
4.6. Šund dimenzija	21
4.7. Ključ	22
4.7.1. Nadomestni ključ za časovno dimenzijo	22
4.7.2. Uporaba govorečih ključev	22
4.7.3. Producerski ključ	23
5. Uporaba standardne klasifikacije v podatkovnem skladišču	23
5.1. Standardna klasifikacija dejavnosti – SKD	23
5.2. Razvojno okolje rešitve	24
5.3. Prenos podatkov v dimenzijsko tabelo	26
5.4. Izvedba dimenzijske tabele	27
5.4.1. Pretok podatkov v dimenzijsko tabelo	27
5.4.2. Podatkovno skladišče z uporabo SKD dimenzije po II. možnosti	28
5.4.3. Podatkovno skladišče z uporabo SKD dimenzije po III. možnosti	29
5.5. Primerjava izpisov po II. in III. možnosti, podatki o podjetjih se ne spreminjajo	31
5.6. Primerjava izpisov po II. in III. možnosti, podatki o podjetjih po dejavnostih se spreminjajo	34
6. Sklep	37
7. Literatura	39
8. Viri	40
9. Priloga	
10. Slovar slovenskih prevodov tujih izrazov	

Uvod

V poslovnem sistemu se v času njegovega delovanja obdelajo ogromne količine podatkov, ki se hranijo na različne načine. Danes je eden izmed možnih načinov hranjenja podatkov zbirka podatkov (Klisura, 2002, str. 4). Z njeno pomočjo zbiramo podatke preko operativne baze podatkov v podjetju, informacijskih sistemov drugod in drugih virov. V navadi je, da je taka zbirka podatkov izvedena kot relacijska baza podatkov za potrebe odločanja in kot operativna baza podatkov v podjetju, da ne moti delovanja operativne baze podatkov. Tako omogoča, da podatki preko poizvedb sestavljajo ustrezne informacije, ki omogočajo delovanje celotnega poslovnega procesa.

Za odločitve, ki so predvsem poslovnega značaja, morajo imeti informacije drugačno obliko, kot jo navadno dobimo iz operativne baze podatkov. Podatki morajo biti zato ustrezno urejeni in obdelani, kar je mogoče s pomočjo podatkovnega skladišča (angl. data warehouse).

Podatkovna skladišča vsebujejo podatke iz podatkovnih zbirk, ki so ustrezno urejeni, prečiščeni in denormalizirani (Kimball, 1996, str. 31-33), da lahko na podlagi enostavnih poizvedb omogočajo dostop do pomembnih informacij. Te informacije so potem temelj za poslovne odločitve ali analize v podjetjih, organizacijah ali ustanovah.

Statistični uradi izvajajo večinoma analize in obdelave nad podatki, ki jih hranijo različni registri. Za primerjavo pa uporabljajo podatke, ki se pridobijo na podlagi različnih raziskav, oziroma anket in popisov. Za njih je učinkovit dostop do podatkov velikega pomena, zato je možnost za uporabo podatkovnih skladišč velika.

V diplomski nalogi se bom omejil na uporabo podatkovnih skladišč v Statističnem uradu Republike Slovenije (v nadaljevanju SURS). Zaradi spremembe zakonodaje in drugih spremenljivk se v omenjenem uradu lahko zgodi, da se posamezne klasifikacije vsebinsko spreminjajo, včasih tudi zelo pogosto. S tem nastanejo težave, saj so podatkovna skladišča izdelana za daljše obdobje (Inmon, 1996, str. 17, 36) kot katerakoli operativna baza podatkov. To zahteva ustrezno rešitev v strukturi dimenzije, ki vsebuje spremenljivke in pridevke v opisni obliki in je del podatkovnega skladišča. Taka rešitev se izvede s pomočjo počasi se spreminjajoče dimenzije. Rešitev bom tudi analiziral s testnimi podatki o podjetjih, ki bodo stalni po dejavnosti in občini za vse časovne preseke, in jo primerjal z alternativno rešitvijo, ki jo navaja tudi Kimball (Kimball, 1998, str. 180). Nato bom še prikazal primer, ko pride do spremembe v podatkih o dejavnosti za posamezno podjetje in kako se to vidi na izpisih. Prednosti in slabosti bom predstavil kot rezultat analize vsake rešitve posebej.

1. Sistemi za podporo odločanju

Po definiciji B. Grošlja je odločitev (Grošelj, 1999, str. 50): »aktivnost z namenom, da se izognemo negativnim posledicam, jih zmanjšamo, ali pa izkoristimo priložnosti.«

Prvi korak v odločanju je identifikacija problema (Grošelj, 1999, str. 51). Pri tem uporabniku pomagajo (Zadnikar, 2000, str. 9) različni podatki, ki morajo biti ažurni. Nato sledi faza iskanja rešitve, kjer se definira vrsta poizvedb oziroma analiz, ki sestavljajo poročilo. Poizvedbe vsebujejo celo vrsto podatkov iz primerjalnih okolij in iz okolja, kjer izvira problem. Ta poročila se izdelajo s pomočjo sistema za podporo odločanju (angl. decision support system). Zadnja faza odločanja je izbor najboljše rešitve.

Sistemi za podporo odločanju (angl. decision support system), v nadaljevanju DSS, so (Lucas, 1994, str. 56) informacijski sistemi in spadajo v skupino direktorskih informacijskih sistemov (angl. executive information system, v nadaljevanju EIS). So informacijski sistemi, ki na podlagi enostavnih poizvedb in agregacije po podatkih omogočajo dostop do podatkov in izdelavo različnih poročil. Poročila posredujejo informacije vodstvenemu osebju oziroma končnemu uporabniku, da na podlagi subjektivne ocene izbere ustrezne poslovne odločitve, poleg tega pa omogočajo poglobljeno iskanje po podatkih. V statističnih uradih so ti uporabniki strokovnjaki - analitiki, ki so odgovorni za analize in obdelavo podatkov. Le-ti na podoben način izdelajo informacijo za javnost in jo na ustrezen način posredujejo.

Operativna baza podatkov oziroma baza podatkov na splošno, je zbirka medsebojno povezanih operativnih podatkov, ki so shranjeni v računalniškem pomnilniku, brez nepotrebnega podvajanja na način, ki omogoča njihovo uporabo različnim uporabnikom z različnimi potrebami glede uporabe. Podatki so shranjeni tako, da so neodvisni od programov, ki jih uporabljajo. Zgradimo jo na podlagi entitetno orientiranega modela (Grad, 1982, str. 1).

DSS deluje na podlagi zbirke podatkov, zato je zelo pomembno, da je podatkov dovolj. Podatki se lahko pridobijo iz različnih sistemov (Klisura, 2002, str. 5), in ravno v tem je prednost. Na primer, podatki o kmetijstvu bodo lahko izhajali iz popisa, podatki o prometu in turizmu iz programa za preglednice, podatki o prebivalstvu iz registra prebivalstva, podatki o podjetjih iz operativne baze podatkov sodnega registra ipd., izhodne podatke pa lahko vidimo sproti v enem poročilu. S pomočjo dobrega DSS približamo analitične naloge strokovnjakom, oziroma analitikom in je učinkovit, zato z njim ne izgubljajo časa (Klisura, 2002, str. 5), ko oblikujejo zahteve za informatike. S tem se lahko osredotočijo na svoje naloge, odgovornosti in na podlagi izdelanih poročil posredujejo informacije javnosti na čim bolj enostaven in razumljiv način. Poročila se lahko uporabijo v različne namene in so običajno sestavljena tako, da so najprej prikazani agregirani podatki na najvišjem nivoju. Uporabnik pa lahko izbere poglobljeno iskanje (angl. drill) po izbranem pridevku (angl. attribute) in s tem pridobi bolj podrobne informacije na enem mestu. Na določen

način lahko posebni uporabniki predložijo tudi zahteve za posebej njim prirejena poročila, ki se obnavljajo samostojno in vnaprej predvidenih terminih.

Ljudje današnjega časa potrebujejo enostaven in dinamičen sistem, ki omogoča, da si premislijo glede zahtev in jih spreminjajo, dokler jim rezultati ne ustrezajo. Ti sistemi tudi »govorijo« isti jezik (Zadnikar, 2000, str. 9), kot ga govorijo analitiki na področju demografije, regionalnih gibanj, cenovnih gibanj, turizma, gradbeništva, trgovine, kmetijstva in vinogradništva in drugih vprašanj. Celo na videz zapletene poizvedbe je lažje izpolniti, ker je mogoče do informacije priti na različne načine.

DSS slonijo na podatkovnih skladiščih, kjer so vsi podatki smiselno urejeni in dostopni in na podlagi vnaprej definiranih zahtev. Seveda podatkovno skladišče ni nujno potrebno za sistem, vendar je težko pričakovati, da bi zbiranje podatkov na drugačen način omogočilo učinkovitejšo izdelovanje poročil za poslovne odločitve (Zadnikar, 2000, str. 10).

Sam razvoj sistema za podporo odločanju je najbolj uspešen na podlagi prototipnega pristopa (Klisura, 2002, str. 11). Prototip je navadno področno podatkovno skladišče (angl. Data Mart), ki zajema določeno skupino podatkov, navadno iz enega vira in poročila. Tak prototip uporabnik testira in poda dodatne zahteve, ki izoblikujejo celovito definirano področje. Na podlagi dokončnih popravkov prototipa se izdelajo še preostali segmenti podatkovnega skladišča.

2. Podatkovna skladišča

2.1. Opredelitev

Posredno smo podatkovna skladišča že nekajkrat omenili in v grobem bi lahko rekli, da so to zbirke podatkov, kjer so vsi podatki smiselno urejeni in strukturirani, dostopni na enem mestu, da omogočajo učinkovite in ažurne poizvedbe po podatkih.

Definicij podatkovnega skladišča je več. V Leksikonu računalništva in informatike najdemo termin skladišče podatkov (Pahor, 2002, str. 523):

»1. Skladišče podatkov je v najširšem smislu oznaka za računalniški sistem, namenjen shranjevanju, pregledovanju in upravljanju velikih količin podatkov poljubne vrste. V skladiščih podatkov so ponavadi zapleteni mehanizmi za stiskanje in zapisovanje, hitro izvajanje poizvedb in uporabo zahtevnih filtrov.

2. Zbirka podatkov, pogosto na oddaljenem kraju, s prepisom najnovejših podatkov o delovanju podjetja. Načrtovalci in raziskovalci jo lahko poljubno uporabljajo, ne da bi zaradi tega upočasnili delo z glavno zbirko.«

Glavna predstavnika teorije podatkovnih skladišč sta Inmon in Kimball. Tako Inmon definira (Inmon, 1996, str. 33) podatkovna skladišča po prevodu takole (Klisura, 2002, str. 7):

»Podatkovno skladišče je vsebinsko usmerjena, integrirana, časovno odvisna statična zbirka podatkov, namenjena podpori v poslovnem odločanju.«

Vsebinska usmerjenost je v tem primeru enaka kot za poslovni sistem, na primer zavarovalnico, kjer so njena vsebinska področja delovanja razdeljena na avtomobile, nepremičnine, ljudi in njihovo zdravje (Inmon, 1996, str. 33).

Integriranost je najpomembnejša lastnost podatkovnih skladišč in se kaže že v tem, ko podatke aplikacijsko naravnane operativnega okolja prenašamo v podatkovno skladišče. Dejansko je vseeno, ali je spol v aplikaciji definiran z 0 in 1 ali z 'm' in 'ž'. Pomembno je, da se definira enotna šifra in s tem zagotovi enotnost skozi celotno podatkovno skladišče (Inmon, 1996, str. 33).

Obstojna zbirka podatkov (angl. nonvolatile), kar dejansko pomeni, da je ločena od stalnih transakcij, ki jih proizvaja poslovni sistem, kar ne moti stalnih analiz in poizvedb po podatkih (Inmon 1996, str. 35).

Časovno odvisna statična zbirka podatkov pa pomeni dejansko zbir večih časovnih stanj v eno celoto, kar omogoča analizo časovnih vrst. Časovni okvir je v podatkovnem skladišču veliko bolj dolgoročen v primerjavi z operativnimi sistemi. Tako je 5 do 10 letni okvir nekaj povsem običajnega za podatkovno skladišče (Inmon, 1996, str. 36).

Drugi teoretik Kimball (1996, str. xxiii-xxv) pa postavlja definicijo podatkovnih skladišč drugače. Po prevodu (Klisura, 2002, str. 7):

»Podatkovno skladišče je posnetek transakcijskih podatkov, strukturiranih na tak način, da je primerno za uporabo pri analizi podatkov.«

Iz navedenega je razvidno, da je že glede definicije podatkovnega skladišča v svetu veliko različnih mnenj, ki povzročajo, da je definicija izgubila svojo natančnost (Kimball, 1998, str. 5). Postopek izdelave podatkovnih skladišč je razmeroma mlad in strokovnjakov na tem področju je relativno malo. Je pa dejavnost, ki obeta, saj je prava informacija že v današnjem svetu veliko vredna in se je ne bo dalo učinkovito oblikovati drugače kot s pomočjo podatkovnih skladišč. V praksi se tudi kaže, da je zanimanje uporabnikov za podatkovna skladišča zaenkrat še zadržano, saj nimajo niti dobre predstave o tem, kaj jim podatkovna skladišča omogočajo (Zadnikar, 2000, str. 10).

V literaturi najdemo tudi termin področno podatkovno skladišče. Izraz smo že uporabili v prvem poglavju, ko smo splošno govorili o DSS. Kimball ga obravnava kot podenoto podatkovnega skladišča, ki navadno obsega rezultat dejavnosti enega poslovnega področja in sestavlja celotni poslovni proces. Podatkovno skladišče je tako unija vseh področnih podatkovnih skladišč (Kimball, 1998, str. 18). Inmon kot glavni kritik Kimballovega pristopa pa pravi, da so podatkovna skladišča kot unija področnih podatkovnih skladišč le modni trend proizvajalcev in da je nemogoče zbrati

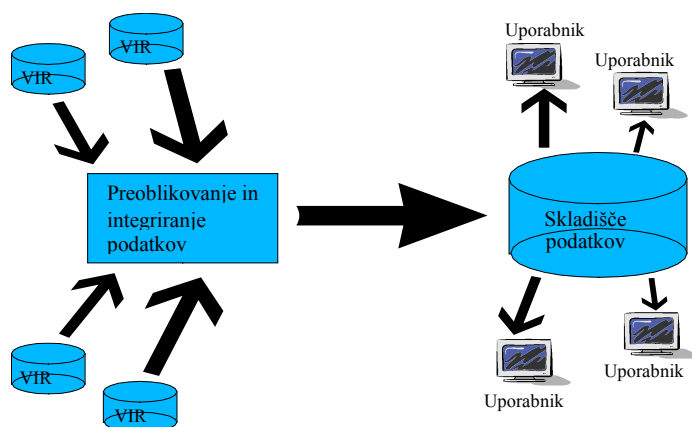
vsa področna podatkovna skladišča v eno unijo in trditi, da je to podatkovno skladišče. Poleg tega so področna podatkovna skladišča področno neodvisna in zato hitrejša. Podatkovno skladišče predstavlja le integrirano celoto področnih podatkovnih skladišč (Inmon, 1999).

Učinkovitost podatkovnih skladišč je zagotovljena tako, da je ločena od tako imenovanega transakcijskega sistema podatkov, drugače se izgubi pridobljena razumljivost in učinkovitost. Poleg tega so podatki v podatkovnih skladiščih konsistentni, kar pa za transakcijske sisteme ni vedno nujno.

2.2. Arhitektura

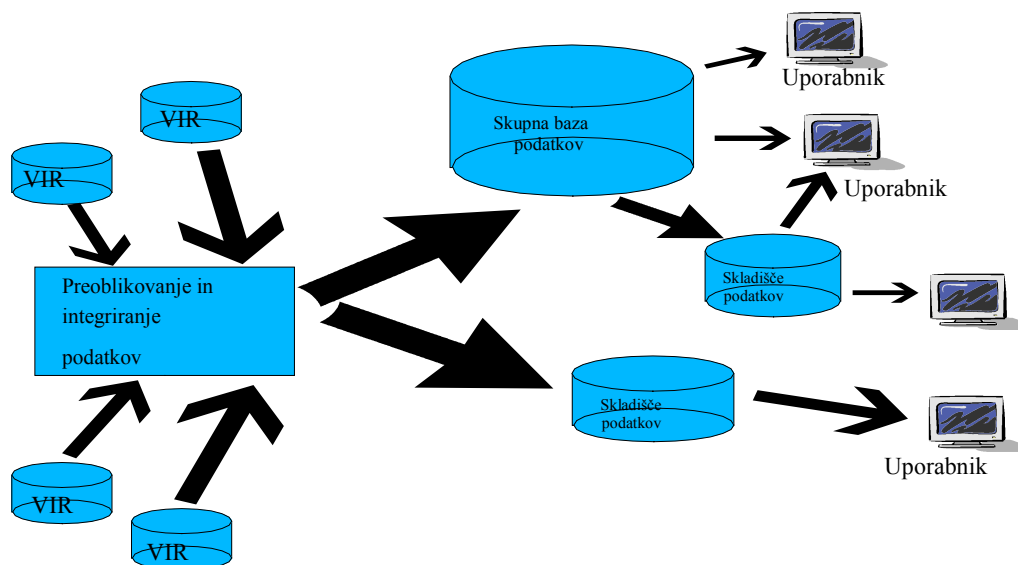
Najpomembnejša značilnost podatkovnega skladišča je integriranost, saj so podatki lahko pridobljeni na podlagi različnih virov in prečiščeni glede konsistentnosti. Poleg tega pa so podatki v skladiščih zgodovinsko odvisni. To dejansko pomeni, da imamo v podatkovnih skladiščih zajetih veliko časovnih presekov, ki se ne spreminjajo več. Tako se prenašajo v podatkovno skladišče enkrat za izbrani časovni presek in se potem samo še berejo. Na podlagi analize podatkov se odločimo za ustrezno arhitekturo podatkovnega skladišča, ki je lahko enostavna in sestavljena (Klisura, 2002, str. 10). Področna, lahko tudi manjša, podatkovna skladišča imajo značilnosti enostavne arhitekture (slika 1), podatkovna skladišča, ki so kot celota sestavljena iz več področnih podatkovnih skladišč, imajo značilnosti sestavljene arhitekture (slika 2).

Slika 1: Arhitektura enega samega skladišča podatkov



Vir: Klisura, 2002, str. 5.

Slika 2: Arhitektura več manjših podatkovnih skladišč



Vir: Zadnikar, 2000, str. 12.

Na statističnih uradih je pomen podatkovnega skladišča še večji. Za analitika je namreč pomembno, da ima na enem mestu zbrane vse potrebne podatke o izbrani temi. Ne zanima ga, kaj se dogaja s popisom prebivalstva, raziskavami na področju zaposlenih ipd. Zanimajo ga samo informacije o določeni temi. V podatkovnem skladišču so zbrani podatki iz različnih virov. Najbolj pomemben je informacijski sistem statističnega urada, kjer se hranijo podatki mnogih statističnih raziskav (Klisura, 2002, str. 12). Podatke je tako potrebno primerno obdelati, da dobimo samo tisto, kar potrebujemo v podatkovnem skladišču. Ostali viri so večinoma zunanji in jih hranijo za to usposobljeni organi in organizacije. Analitiki na podlagi navodil svetovalcev, ki dobro poznajo te podatke, izdelajo poročila.

Točnost podatkov je pomembna in ima pri gradnji podatkovnih skladišč posebno mesto. Težave s točnostjo so lahko velike ravno zaradi:

- 🌐 želje po integraciji različnih virov podatkov in
- 🌐 neskladnosti, ki je dovoljena v transakcijskih sistemih.

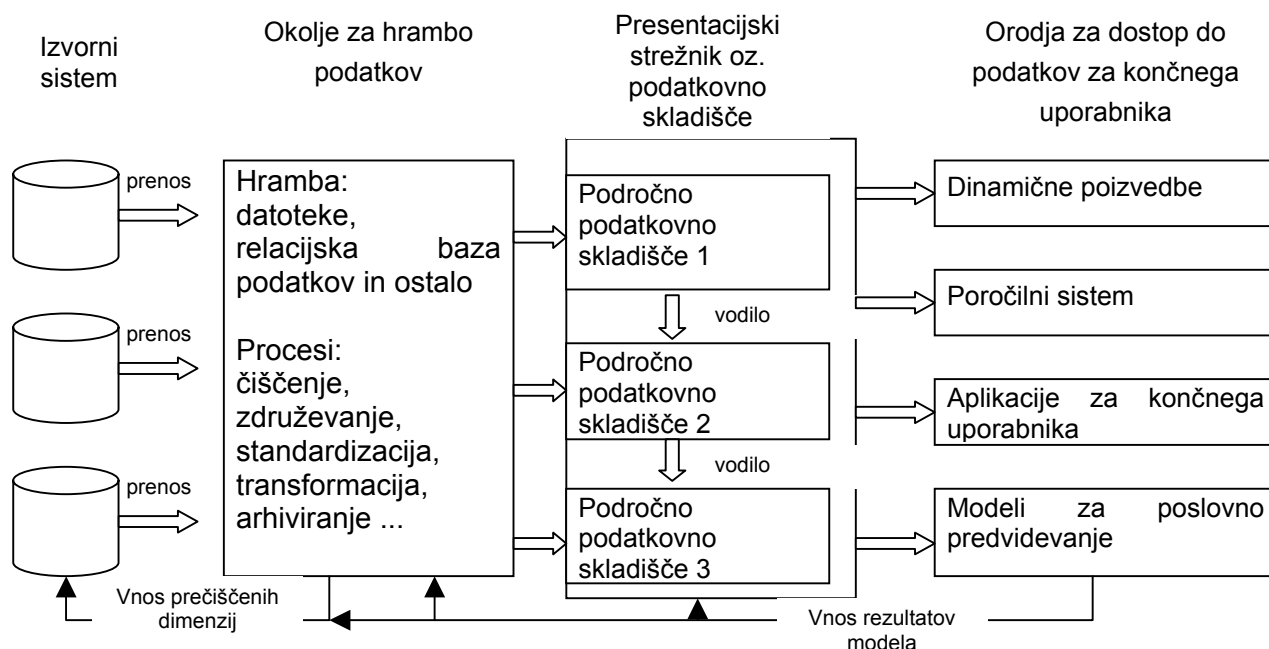
Če točnost ni zagotovljena, lahko pride do netočnih in napačno uporabljenih podatkov, kar privede do napačnih informacij. Podatke je tako treba pred vpisovanjem v skladišče očistiti, pretvoriti, združiti, delno sešteti in podobno.

Pred vnosom podatkov v podatkovno skladišče je potrebno zgraditi podatkovni model, ki ga, glede na strukturo, lahko imenujemo dimenzijski model (ang. dimension model). Dimenzijski model je odvisen od področja raziskave in strukture podatkov, ki bodo vključeni v model, in strukture izpisov, ki jih izdelajo uporabniki, da pridobijo želene informacije. Našteto lahko pridobimo na podlagi intervjuja z uporabniki (analitiki na statističnem uradu).

2.3. Sestavni deli podatkovnega skladišča


Po Kimballu (1998, str. 15) so sestavni deli oziroma osnovni elementi (angl. chess pieces) podatkovnega skladišča prikazani v sliki 3.


Slika 3: Osnovni elementi podatkovnega skladišča




Vir: Kimball, 1998, str. 15.


Posamezne dele razumemo kot:

 **Izvorni sistem** (angl. source system) – operativni sistem podatkovnih zapisov, katerih funkcija je zajeti vse transakcije v poslovnem sistemu. Glavna naloga takih sistemov je uporabnost in ažurnost. Podatki so med seboj povezani s ključi, ki zagotavljajo čim manjšo redundanco podatkov. Ti ključi so navadno produkcijski ključi ali govoreči ključi in jih obravnavamo kot atribut ali pridevek. Referenčno celovitost zagotavljamo s pomočjo tujih ključev in procesov v hrambi podatkov (razlogi bodo obravnavani kasneje). Podatki v izvornih sistemih niso nujno konsistentni, saj je pomembno, da je posamezni dogodek vsaj zapsan, čeprav ni ravno razumljiv. Pomembno je tudi, da se v takih sistemih ne izvaja poizvedb, kakršne so navadno v podatkovnih skladiščih. Poleg tega pa izvorni sistem vsebuje zelo kratkoročni časovni okvir, ki je navadno dolg od 60 do 90 dni. Izvorni sistem je navadno fizično ločen od okolja za hrambo podatkov in predstavitvenega strežnika. Podatki se tako prenesejo v hrambo podatkov v obliki datotek (angl. flat file).

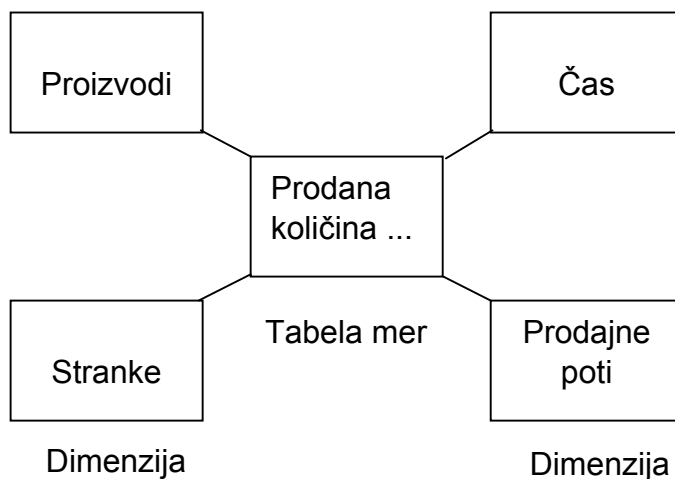
 **Okolje za hrambo podatkov** (angl. data staging area) – hrambeno okolje in sklop procesov, ki prečistijo, transformirajo, združujejo, arhivirajo, izločajo

dvojnike in omogočajo, da so podatki pripravljene za uporabo v podatkovnem skladišču. Tako je lahko okolje za hrambo podatkov vse, kar je med izvornim sistemom in predstavitvenim strežnikom. Zelo je priporočljivo, da se vse skupaj nahaja na enem mestu. Omenjeni procesi zagotavljajo referenčno celovitost (angl. referential integrity) med podatki v podatkovnem skladišču in za analizo je dovolj že izdelava večdimenzijske kocke, ki jo lahko kasneje uporabimo v orodjih za sprotno analitično obdelavo ali OLAP orodjih. V takih primerih je izdelava dimenzijskega modela odvečna. Ker pa na tej ravni ne moremo predvideti, kakšne bodo poizvedbe uporabnika, poleg tega pa je definiranje poizvedb področje predstavitvenega strežnika, je izdelava dimenzijskega modela nujna za njegovo delovanje.

 Predstavitveni strežnik (angl. presentation server) – fizični računalniški sistem, kjer je podatkovno skladišče dejansko realizirano in organizirano za direktno definiranje poizvedb in za uporabo analitikom preko raznih analitičnih orodij. Podatkovno skladišče je izvedeno v značilni zvezdasti strukturi s tabelami mer in dimenzijami na relacijski bazi podatkov, kot je lahko ORACLE, SQL SERVER in druge. Izvorni sistem, okolje za hrambo podatkov in predstavitveni strežnik so trije temeljni sestavni deli za delovanje podatkovnega skladišča.

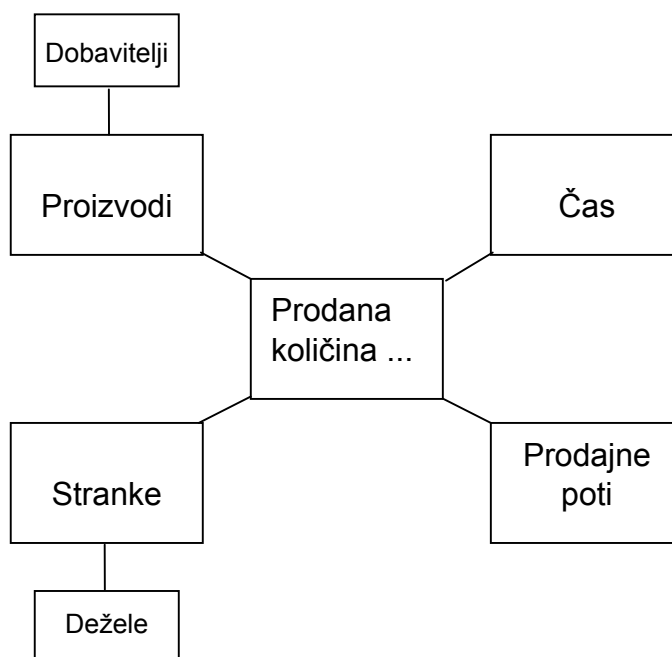
 Dimenzijski model (angl. dimensional model) – specifično modeliranje podatkov, ki je alternativa entitetno orientiranem modeliranju. Dimenzijski model vsebuje enake informacije kot entitetno orientiran model, le da so podatki v dimenzijskem modelu drugače organizirani in bolj razumljivi, kar omogoča učinkovite poizvedbe. Zanimivo, da je dimenzijsko modeliranje zapaženo pred entitetno orientiranim modeliranjem. V poznih 60-ih letih sta General Mills in Univerza Dartmouth izdelala besednjak »dejstev« in »dimenzij« (Kimball, 1998, str. 144). Kimball verjame, da je to omogočilo Nielsen Marketing Research da prenese to tehniko naprej s pomočjo preglednih podatkov o trgovini s široko potrošnjo v 70-ih in zgodnjih 80-ih. Te ideje pa prvič povzame Nielsen leta 1984. Temeljna sestavna dela dimenzijskega modela sta tabela mer ali »večključna« (angl. multipart key) tabela s številčnimi podatki, oziroma dejstvi ali merami in občutno manjše dimenzijske tabele, ki vsebujejo opisne podatke. Najbolj pogosti obliki oziroma strukturi dimenzijskega modela sta zvezda (slika 4) in snežinka (slika 5).

Slika 4: Zvezdni dimenzijski model




Vir: Oracle, 2001.

Slika 5: Dimenzijski model »snežinka«



Vir: Oracle, 2001.


 Tabela mer in dimenzijska tabela. Osnovna ideja dimenzijskega modeliranja izhaja iz tega, da je skoraj vsak tip poslovnih podatkov možno predstaviti v obliki dimenzijske kocke, kjer polja pomenijo mere, robovi kocke pa dimenzije. Seveda nismo omejeni na samo tri dimenzije, ki definirajo nam znano kocko, zato Kimball večdimenzijsko kocko imenuje tudi »hiper kocka« (angl. hypercube). Običajno


dimenzijski modeli vsebujejo 4 do 15 dimenzij, modeli z dvema ali tremi so redki, medtem ko so modeli z 20 ali več dimenzijami nepregledni. Taki modeli vsebujejo navadno tudi dimenzije, ki združujejo med seboj sorodne dimenzije v eno samo in tako naredijo model preglednejši (Kimball, 1998, str.165).

1. Tabela mer (angl. fact table) je temeljna tabela v vsakem dimenzijskem modelu. Vsebuje mere oziroma dejstva in spremenljivke, po katerih lahko te mere ocenimo. Navadno so to številčni podatki, ki vsebujejo decimalna mesta, in celoštevilčni podatki. Možno je sicer, da predstavljajo številčni podatki tudi konstanto, kot je lahko velikostni razred, vendar moramo biti o tem prepričani in navadno to definira uporabnik. Kimball (1998, str.165) definira mero kot: »Mera ali dejstvo je navadno nekaj, kar je znano vnaprej.«

Opisne spremenljivke so v tabeli mer definirane kot tuji ključ s povezavo ali relacijo na pripadajočo dimenzijo. Na tuji ključ je v tabeli mer vezan tudi indeks, ki omogoči hitrejšo poizvedbo po tabeli mer.

2. Dimenzijska tabela vsebuje spremenljivke in pridevke v opisni obliki. Pridevek Kimball opisuje kot značilnost otipljive stvari. Najbolj znani pridevki so čas, kraj, vrsta proizvoda itd., vendar niso vsi otipljivi. Dimenzijska tabela Čas ali Kraj tako vsebuje vse pridevke, ki opisujejo čas ali kraj, dimenzijska tabela Vrsta proizvoda pa vsebuje vse vrste proizvoda, ki vstopajo v poslovni proces kot polproizvod ali izstopajo iz poslovnega procesa kot končni proizvod. Dimenzijo torej sestavljajo pridevki, ki imajo medsebojno visoko korelacijo (Kimball, 1998, str.166). Dimenzija je lahko sestavljena iz ene ali več hierarhij, ki omogočajo podrobnejši ali bolj globalen vpogled. Ravno standardna klasifikacija dejavnosti, ki jo bomo opisali kasneje v 5. poglavju, je dimenzija, ki vsebuje 6 nivojev hierarhije. Poznamo tudi kombinirane ali združene dimenzije, ki združujejo dimenzije v večje skupine in tako dobi uporabnik še več uporabnih informacij. Vsaka dimenzijska tabela ima glavni ključ, ki služi za referenčno vez s tujim ključem v tabeli mer, na katero je vezana dimenzijska tabela.

 Poslovni proces ali postopek (angl. business process) – sklop večih poslovnih aktivnosti, ki dajejo uporabniku podatkovnega skladišča smiselno povezavo. Navadno je poslovni proces ali analitična tema predstavljena kot področno podatkovno skladišče.

 Področno podatkovno skladišče¹ – logična podenota podatkovnega skladišča oziroma eden izmed enakovrednih sestavov podatkovnega skladišča. Okoli tega pojma je tudi največ kritik s strani Inmona, ki zagovarja popolnoma drugačno definicijo področnega skladišča. Po njegovem je področno podatkovno skladišče povsem samostojno, unija področnih podatkovnih

¹ V strokovnih krogih je na podlagi angleškega termina slišati tudi izraz informacijska tržnica (Pahor, 2002, str. 165)

skladišč pa ne predstavlja podatkovnega skladišča, saj morajo biti le-ta primerno med seboj integrirana (Inmon, 1999). Na SURS je več oddelkov, ki pokrivajo različna področja raziskav. Ta področja so v projektu RDB v okviru Phare programa COP98 navadno predstavljena kot področna podatkovna skladišča. Le-ta dejansko kot unija predstavljajo podatkovno skladišče, imenovano SURS RDB (Pavlin, 2002). S tem je bila tudi gradnja samega podatkovnega skladišča lažja, saj so se analitiki osredotočili na definicijo samo enega poslovnega področja posebej.

- Podatkovno skladišče – Izvor podatkov za učinkovite poizvedbe. Polnimo ga iz okolja za hrambo podatkov. Upravljalec podatkovnega skladišča pa je odgovoren za oboje, podatkovno skladišče in hrambo podatkov. Na SURS na primer že prej omenjeni projekt RDB predstavlja podatkovno skladišče, ki ga sestavljajo številna med seboj enakovredna poslovna področja. To omogoča primerljivost podatkov iz različnih področij in tako obogati informacije.
- Operativni podatkovni sistem (angl. operational data store – ODS) – sistem, ki dejansko omogoči integracijo različnih virov podatkov. Je zbir procedur in funkcij, ki omogočajo, da lahko združujemo med seboj različne vire podatkov preko skupnih povezovalcev (davčna številka, matična številka organizacije, EMŠO ...). Tak sistem je lahko ločen, lahko pa je sestavni del podatkovnega skladišča.
- Sprotna analitska obdelava (angl. on line analytic processing, v nadaljevanju OLAP) – je temeljna aktivnost pri sestavljanju poizvedb in predstavitvi podatkov iz podatkovnega skladišča kot tudi poseben način izražanja poizvedb, ki se kaže v obliki OLAP »ponudb«. Ta tehnologija temelji na tako imenovanih večdimenzionalnih kockah (angl. multidimensional cube) podatkov, zato so OLAP baze podatkov znane tudi kot večdimenzionalne baze podatkov. Kimball (1998, str. 21) verjame, da OLAP področna podatkovna (angl. OLAP – style data marts) skladišča lahko povsem enakovredno sestavljajo podatkovno skladišče, če so dimenzije in mere povsem združljive s takim področnim podatkovnim skladiščem.
- Relacijski OLAP (angl. relational OLAP - ROLAP) – sklop uporabniških vmesnikov in aplikacij, ki dajejo relacijski bazi podatkov »dimenzijski pridih«. Uporablja se v primeru, ko OLAP orodje uporabimo za predstavitev na internetu. Takrat posamezne dele OLAP »ponudbe« razbijemo v dimenzijski model, uporabnik pa s pomočjo uporabniškega vmesnika in aplikacije sestavi želeno poizvedbo (Kimball, 1998, str. 407).
- Večdimenzijski OLAP (angl. multidimensional OLAP - MOLAP) – sklop uporabniških vmesnikov, aplikacij in različnih tehnologij baz podatkov, ki imajo »dimenzijski pridih«. Je nadgradnja ROLAP, izvedena tako, da so OLAP »ponudbe« že vnaprej pripravljene. Če uporabnik preko uporabniškega vmesnika sestavi tako poizvedbo, ki je že vključena v OLAP »ponudbo«,

aplikacija izdelava poročilo iz OLAP »ponudbe«, drugače aplikacija izdelava poročilo na ravni podatkov iz podatkovnega skladišča (Kimball, 1998, str. 408).

- Rešitev za uporabnika (angl. End User Application) – sklop orodij, ki proizvedejo, analizirajo in predstavljajo informacije, namenjene podpori pri poslovnem odločanju v obliki poročil. Navadno je sklop sestavljen iz orodja za dostop do podatkov za končnega uporabnika, preglednice, grafičnega vmesnika in uporabniškega vmesnika, kar poenostavi predstavitev poročila končnemu uporabniku (Kimball, 1998, str. 21).
- Orodja za dostop do podatkov za uporabnika (angl. end user data access tool) – sklop orodij, ki tečejo na strani odjemalca in omogočajo proizvodnje pod določenimi pogoji na podatkovnem skladišču, ki je na strežniku. Pogoji omogočajo izdelavo od enostavnih dinamičnih proizvodov do zapletenih izkopavanj po podatkih.
- Orodje za dinamično oziroma »ad hoc« sestavljanje proizvodov (angl. ad hoc query tool) – specifično orodje za dostop do podatkov, ki omogoča dinamično sestavljanje enostavnih proizvodov s pomočjo povezav med različnimi tabelami mer (preko skupne dimenzije, ki povezuje obe tabeli mer). Kimball trdi, da je 10% takih uporabnikov, ki bi lahko svoje proizvodnje izdelovali na tak način, ostalih 90% uporabnikov pa bi lahko izvajali dinamične proizvodnje po vnaprej pripravljenih šablonah (Kimball, 1998 str. 22).
- Aplikacije za modeliranje (angl. modeling applications) – aplikacije, ki omogočajo sestavo modelov, ki se uporabljajo pri poslovnem predvidevanju.
- Meta podatki (angl. meta data) – vse informacije o okolju podatkovnega skladišča, ki se ne uporabljajo dejansko kot podatek. Meta podatki so izjemno pomembni pri uporabi podatkovnega skladišča, v katerih so zbrani podatki iz številnih virov, uporabljajo pa jih uporabniki, ki niso seznanjeni z izvirnim pomenom vsakega podatka. V meta podatkih se skriva še podatkovni slovar, poleg tega so v njih zbrane še dodatne informacije, kot so način in časovna razporeditev osveževanja podatkov iz vira, pojasnila o tem, kako so bili podatki prečiščeni in filtrirani med prenosom v podatkovno skladišče, o tem, ali so bili združevani ali preračunani, in podobno (Pahor, 2002, str. 259). Najpogosteje so meta podatki spravljani v relacijski bazi podatkov, njihova oblika pa je odvisna od uporabljene programske opreme v podatkovnem skladišču. Pomembno je, da se v fazi izgradnje podatkovnega skladišča dokumentirajo vse oblike meta podatkov. Na podlagi dobre dokumentacije imajo taki meta podatki tudi svojo korist, ki se kaže v tem, da omogočajo s svojo vsebino olajšati uporabo podatkov tudi na področjih, ki jih ne poznamo dobro.

3. Faze gradnje podatkovnega skladišča

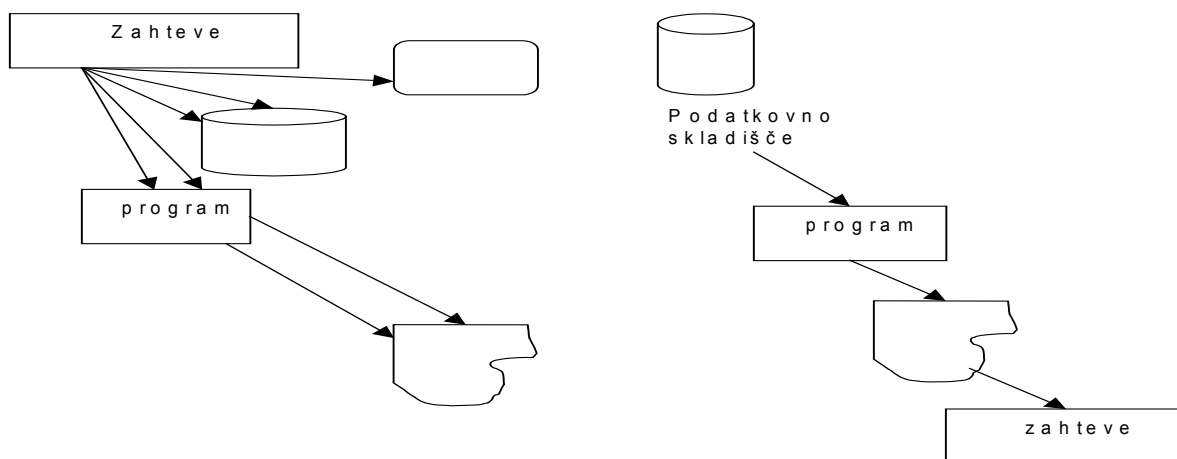
Uporabniki podatkovnega skladišča so navadno analitiki. Analitik je na prvem mestu poslovno usmerjen in šele na drugem mestu tehnično usmerjen. Poglavitna naloga analitika je definicija in odkrivanje informacij, ki so uporabljene pri poslovnem odločanju (Inmon, 1996, str. 22). Pomembno je, da se skušamo postaviti v vlogo analitika in predvideti, kako nam bo podatkovno skladišče koristilo. Analitik lahko šele na podlagi videnega odkriva možnosti, ki jih nudi podatkovno skladišče za poslovno odločanje.

Drža oziroma podpora analitika je pomembna zaradi naslednjih razlogov (Inmon, 1996, str. 22):

- je legitimna,
- je prodorna in
- ima nedoumljiv učinek na razvoj podatkovnega skladišča in sistema za uporabo podatkovnega skladišča.

Navadno se informacijski sistemi razvijajo po načinu življenjskega cikla (angl. sistem development life cycle – SDLC). Pri podatkovnih skladiščih pa je navadno obratno. V svetu analitikov so potrebe zadnja stvar za raziskovanje, s tem je tudi pristop v razvijanju podatkovnega skladišča drugačen. Inmon pravi, da ga razvijamo po načinu »obratno od življenjskega cikla« ali CLDS (obratno kot SDLC). Prvi način je posledica potreb uporabnika, medtem ko je drugi način posledica vhodnih podatkov (Inmon, 1996, str 25). Opisano je prikazano v sliki 6.

Slika 6: Razvoj sistema po »življenjskem ciklu« je za podatkovno skladišče popolnoma nasproten kot za informacijski sistem



SDLC za tradicionalne informacijske sisteme:

- zbiranje zahtev,
- analiza,
- razvoj,
- programiranje,
- testiranje,
- integracija in
- izvedba.

SDLC za podatkovno skladišče:

- razvoj podatkovnega skladišča,
- integracija podatkov,
- testiranje primernosti,
- primerjava »program proti podatkom«,
- razvoj sistema za podporo odločanju,
- analiza rezultatov in
- razumevanje zahtev

Vir: Inmon, 1996, str. 24.

Najbolj primeren postopek je tako prototipni pristop, ki omogoči uporabniku, da na podlagi grobega in enostavnega modela podatkovnega skladišča bolje definira svoje zahteve. Na podlagi dodatnih zahtev in pripomb se pilotni model prelevi v pravo podatkovno skladišče, ki na koncu zadovolji vse trenutne potrebe uporabnika. Sistem zahteva tudi vzdrževalca, ki omogoči, da je skladišče uporabno tudi v prihodnje in s tem se vanj vnesejo tudi sprotni popravki. Prototipni pristop je sestavljen iz več faz (Zadnikar, 2000, str. 16) :

1. faza: Opredelitev področja poslovanja

Najprej je treba pravilno utemeljiti potrebo po podatkovnem skladišču. Tu imajo analitik zelo pomembno vlogo, saj morajo pojasniti, kateri podatki so na voljo in kaj je možno z njimi izvesti. Narediti je potrebno analizo poslovnih zahtev, ki se sestoji iz cilja, metod za doseg cilja in preslikav teh meril v komponente, ki sestavljajo podatkovni model podjetja. Tako se definirajo poslovne zahteve, ki so mogoče. To pomeni, da izhajajo iz podatkov, ki so dejansko dosegljivi. Nikakor ne sme priti do miselnosti, da bomo zgradili podatkovno skladišče »kar tako«, potem pa bomo videli, kaj nam nudi. Podatkovno skladišče namreč ne more biti »pametnejše« od analitikov, saj je le posebna oblika organiziranosti podatkov, ki omogoča učinkovite poizvedbe po podatkih, ki so na enem mestu. Analitiku tako ne bo več potrebno mukotržno brskanje po podatkih in se bo lahko bolj ukvarjal z vsebino in analizo problema.

2. faza: Definicija projekta

Pri oblikovanju skupine za načrtovanje podatkovnega skladišča je nujno potrebno, poleg programerjev in sistemskih integratorjev, vključiti še analitike za izbrana poslovna področja in razne druge strokovnjake, kot so vzdrževalci baz podatkov, strokovnjaki za implementacijo modela strežnik/odjemalec in poznavalci sistemov za podporo odločanju. Izkušnje iz gradnje podatkovnih skladišč v podjetjih kažejo, da je postopek načrtovanja podatkovnega skladišča uspešnejši, če ga vodi analitik in ne informatik.

3. faza: Izbor strojne in programske opreme

Podatkovno skladišče zahteva celovit izbor programske in strojne opreme, saj mora biti skladišče zasnovano kot strežnik, končni uporabniki pa kot odjemalci. To se lahko zagotovi preko ustrezno izbrane infrastrukture, kot je računalniško omrežje in programska oprema za komunikacije. Programerji se morajo primerno seznaniti z uporabljeno programsko opremo in izbrati je potrebno ustrezne uporabniške vmesnike in jih razumeti, kaj nudijo.

4. faza: Implementacija

Naredi se seznam virov podatkov, ki jih je potrebno prenesti v skladišče. Pri tem je pomembno, na kakšen način se dostopa do podatkov, kolikšna je količina teh podatkov, da se na osnovi tega lahko sestavi dimenzijski model in opis meta

podatkov. Pomembno je zagotavljanje konsistentnosti izrazoslovja pri gradnji podatkovnega modela, saj je tako uporabniku omogočeno lažje prilagajanje na novo strukturo podatkov. Pomembno je tudi, da je količina podatkov taka, da zadovolji potrebe, saj prevelika količina podatkov onemogoči učinkovitost poizvedb in zmede uporabnika. Nuditi mu mora informacijo, kateri podatki so sploh dostopni. Zatem sledi fizična implementacija dimenzijskega modela in polnjenje podatkov v model.

5. faza: Delovanje

Ko je skladišče podatkov zgrajeno in napolnjeno s podatki, prevzamejo glavno vlogo uporabniki, ki preko uporabniških vmesnikov dosegaajo podatke. V tej fazi se preveri vsebinska uporabnost dimenzijskega modela in tehnične značilnosti, kot so hitrost dostopa in prepustnost računalniškega omrežja. Na podlagi dopolnitev in novih zahtev programerji dopolnijo dimenzijski model, da ugodijo zahtevam uporabnika. Ponavadi velja pravilo, da je uspešnost zaključka modela pogojena s prvim poskusom podatkovnega modela. Na podlagi prvega poskusa namreč uporabnik lahko izrazi svojo podporo in tudi izpopolni svoje zahteve. Če je prvi poskus uspešen je tudi uspeh celotnega projekta lažje dosegljiv. Če pa je prototip neuspešen, obstaja nevarnost, da bodo uporabniki obupali. Med tekočo uporabo skladišča je smiselno ugotoviti, katere poizvedbe so najbolj pogoste in jih po možnosti avtomatizirati (kot shranjene poizvedbe, lahko pa jih avtomatično polnimo v določenih časovnih intervalih). Tako lahko povečamo učinkovitost podatkovnega skladišča.

Skladišče podatkov pa ne more biti nikoli zares dokončano, ker je treba kar naprej spremljati delovanje, dopolnjevati zahteve uporabnikov, skrbeti za prostorsko razporeditev podatkov, optimizirati odzivnost in se sploh prilagajati dogodkom v zvezi s poslovanjem podjetja.

4. Dimenzija v podatkovnih skladiščih

Poleg tabele mer so dimenzije najpomembnejši gradniki podatkovnega skladišča. Vsebujejo opisne podatke, kot so vsi pridevki spremenljivk v tabelah mer. Definicijo dimenzijske tabele smo že omenili v poglavju 2.3., ko smo govorili o sestavnih delih podatkovnega skladišča, kot jih je predstavil Kimball. Najbolj pogoste dimenzije so časovna dimenzija, geografska dimenzija in dimenzija vrste proizvoda.

Ko se odločimo za gradnjo podatkovnega skladišča, najprej na podlagi podatkov določimo spremenljivke, ki jih bomo uporabili kot dimenzije. Ostale spremenljivke nam lahko služijo kot indikatorji ali kot mere v tabeli mer. Zatem s pomočjo matrične metode (Kimball, 1998, str. 266) določimo število dimenzij in število področnih podatkovnih skladišč. S pomočjo analize in izdelave matrike dobimo poimenovana vsa področna podatkovna skladišča, ki jih lahko pridobimo iz vhodnih podatkov.

Postopek poteka sledeče. Na eno stran matrike (slika 7) naneseemo spremenljivke, ki so opisne in jih imenujemo dimenzije, na drugo stran pa naneseemo spremenljivke, ki so številčne in jih imenujemo mere. Mere, ki izhajajo iz istega vira, praviloma združimo v eno področno podatkovno skladišče. Še bolje je, da področno

podatkovno skladišče predstavlja en poslovni proces, na primer »nakup«. Moč matrične metode je v tem, da lahko področna podatkovna skladišča enega vira vedno združujemo v sestavljena področna podatkovna skladišča večih virov (Kimball, 1998, str. 267).

Slika 7: Matrika področnih podatkovnih skladišč in dimenzij za nabavo

	ČAS	STRANKA	IZDELEK	ZAPOSLE NI	BLAGAJN A
Znesek prodanega blaga	X	X	X	X	X
Število strank	X				X
Število prodanih izdelkov	X	X	X	X	X
Število aktivnih blagajn	X				X

Vir: Kimball, 1998, str. 271.

Slika 7 prikazuje, da lahko iz podatkov izdelamo dve področni podatkovni skladišči. To je razvidno iz skupnih presečišč med dimenzijami in merami.

Znotraj posamezne dimenzije se lahko srečamo še z možnostjo podrobnega iskanja navznoter po posameznem pridevku. To nam omogoča hierarhija, ki vlada med posameznimi spremenljivkami znotraj dimenzije. Hierarhija je navadno definirana v orodjih za definicijo proizvodov. Proizvajalci so tako preko uporabniškega vmesnika omogočili definicijo hierarhije, s pomočjo nje pa je mogoče podrobno iskanje po različnih nivojih dimenzije. Ravno tako lahko vsak pridevek zavzame vlogo naslovnika vrstice ali stolpca v poročilih in v tem je lepota dimenzijskega modela (Kimball, 1998, str. 167).

Po Kimballu (1998, str. 182-191) podatkovna skladišča poleg običajnih dimenzij vsebujejo še počasi se spreminjajoče majhne in velike dimenzije, hitro spreminjajoče se majhne in velike dimenzije ter velike dimenzije.

Preden preidemo na dimenzije pa je najbolj pomembno, da najprej določimo obliko spremenljivke v dimenziji, ki nam bo povezovala tabelo mer in dimenzijo med seboj. Pri tem se mora oblikovalec izogibati uporabi (Kimball, 1998, str. 174):

- govorečih ključev, ki so težko uporabne kot referenčna vez na tuj ključ, saj jih je težko definirati in vzdrževati. Lahko jih uporabimo edino za povezavo na operativno bazo podatkov;
- sestavljenih ključev, ki so težko razumljivi in sestavljeni iz več besed. Le-ti se lahko uporabljajo kot naslovna vrstica, zelo težko pa kot referenčna celovitost.

Fizično je referenčna celovitost med dimenzijo in tabelo mer v dimenzijskem modelu zagotovljena s pomočjo nadomestnega ključa (angl. surrogate key), ki se ga oblikuje s pomočjo zaporedja (angl. sequence). Tehnično je zaporedje generator številčnih vrednosti, kjer določimo začetno in končno vrednost ter korak, in je vezano na dimenzijsko tabelo. To nam olajša vzdrževanje nadomestnega ključa pri vpisovanju novih vrednosti spremenljivke v dimenzijo. Vsaka vrednost nadomestnega ključa vsebuje enovit opis pridevka za določeno obdobje. Pomembno je tudi, da se pridevki, ki navadno sledijo posameznim vrednostim spremenljivke, nedvoumno razlikujejo med seboj. Vsak pridevek namreč lahko odigra vlogo naslovnika stolpca ali vrstice v poročilu.

Oblikovalci podatkovnega skladišča se morajo držati naslednjih navodil za gradnjo dimenzijske tabele (Kimball, 1998, str. 175). Pri tem morajo biti pridevki:

- izčrpn (angl. verbose);
- dobro in enovito opisani, brez manjkajočih vrednosti;
- vsebovati morajo vse vrednosti, ki so vsebovane v podatkih;
- kvalitetni (brez črkovnih napak, nemogočih, zastarelih ali »kozmetično« drugačnih vrednosti posameznega pridevka);
- indeksirani, kar pomeni ustrezno uporabo razpoložljivih indeksov;
- dostopni in
- dokumentirani (v meta podatkih, ki omogočajo obširno razlago in pojasnjujejo izvor posameznega pridevka).

4.1. Počasi se spreminjajoča dimenzija

V knjigi *The Data Warehouse Toolkit* Kimball govori o klasičnem načinu obvladovanja počasi se spreminjajoče dimenzije (Kimball, 1998, str. 180), kot je lahko geografska dimenzija ali v našem primeru dimenzija standardne klasifikacije dejavnosti (v nadaljevanju SKD). Glavna značilnost takih dimenzij je, da se šifra, ki označuje posamezni pridevek, ne spreminja, spreminja pa opis pridevka, ki se navezuje na to šifro. Problem je v izdelavi poizvedb, ki vsebujejo podatke za več časovnih presekov, s tem pa tudi več različnih verzij klasifikacije. Za uporabnika so take poizvedbe med seboj težko primerljive, kar bo razvidno tudi iz naše analize. V rešitev se nam ponujajo tri možnosti:

- I. prepis opisa pridevka z novo vrednostjo in s tem izbris stare. S tem se nam poročila avtomatsko osvežijo ob naslednji obnovi podatkov, vendar nimamo več možnosti, da dobimo vpogled v poročila, ki so bila veljavna v preteklosti;
- II. vnos novega naziva z uporabo nove vrednosti zaporedja;
- III. oblikovanje novega polja, ki bo vseboval stari opis pridevka. Pri vnosu novega opisa pridevka najprej prenesemo stari do sedaj veljavni opis pridevka v novo polje, novi opis pridevka pa vnesemo v polje za trenutno veljavni opis pridevka. S tem se vsaj deloma izognemo slabosti iz prve možnosti.

Prva možnost (I.) je uporabna v primeru, ko uporaba starega opisa pridevka nima več veljave in ga je zato potrebno brisati. Uporablja se v primeru odpravljanja napak. Druga možnost (II.) je osnovni primer tehnike sledenja sprememb opisov pridevkov v dimenziji. Priporočljivo je tudi, da oblikujemo dodatno polje za zapis datuma nastanka

zapisa in polje, kjer se vpiše končni datum veljavnosti zapisa. V primeru, da zapis trenutno velja, je to polje prazno. Referenčna vez je tako izdelana preko nadomestnega ključa v dimenzijski tabeli, ki je hkrati tudi glavni ključ, in tujega ključa v tabeli mer. Uporaba nadomestnega ključa je nujna pri predpostavki, da se produkcijski ključ ne smejo spreminjati in da ne smejo biti sestavljeni (Kimball, 1998, str. 180). Dodelitev vrednosti za tuji ključ v tabeli mer naj poteka s pomočjo poizvedbe po dimenzijski tabeli pri vpisovanju podatkov v tabelo mer. Vsekakor pa ne smemo teh polj povezovati z datumskim poljem v tabeli mer, ki opisuje časovni presek. Pri tej možnosti se srečamo s težavo, kako opisati vrednost spremenljivke, kjer primerjamo dva časovna preseka, pri čemer je prvi časovni presek vseboval drugačne opise za iste vrednosti kot drugi časovni presek. Dilema o uporabi ustreznega datuma na primeru SKD in tudi dejansko rešitev te možnosti bomo prikazali v poglavju 5.

Rešitev na način, ki ga prikazuje II. možnost, je zelo ustrezna v primeru, ko želimo prikazati določene vrednosti v skladu z njihovimi spremembami v času. Tak primer je tudi SKD, ki se je od njene ustanovitve² pa do sedaj spremenila kar nekajkrat. Te spremembe pa so obsegale tako dodajanje novih dejavnosti kot tudi spreminjanje opisa za določeno dejavnost. Tako lahko dobimo določeno skupino nadomestnih ključev, ki določajo vse dejavnosti za določeno obdobje med dvema časovnima presekom in je drugačna od skupine nadomestnih ključev, ki določajo vse dejavnosti v drugem časovnem obdobju. Poizvedba, ki vključuje dimenzijo oblikovano na tak način, izpiše več verzij nazivov za določeno dejavnost, kar dejansko tudi želimo. Če želimo izpisati samo nazive, ki so trenutno veljavni, je potrebno vnesti še ustrezen časovni presek za dimenzijo. Rešitev je zelo uporabna v velikih in sestavljenih dimenzijah, kjer je sprememb veliko. Tak primer je lahko dimenzija Stranka ali Proizvod (Kimball, 1998, str. 181), ki vsebuje veliko zapisov. Stranko kot osebo lahko vodimo od njenega nastanka, preko sprememb v njenih atributih, do njene ukinitve ali zbrisa iz registra. Vse to je zapisano v dimenzijski tabeli Stranka kot nov zapis.

Rešitev na III. način se uporablja v drugačnih situacijah kot opisana II. možnost. Navadno s to rešitvijo ne želimo natančnega sledenja spremembam v dimenziji. Lahko pride do sprememb v opisu pridevka ali na primer do sprememb v mejah geografskih enot, ki pa dejansko ne vplivajo na spremembo obravnave mere ali dejstva v tabeli mer. Tako lahko spremljamo določeno mero na »stari« ali »novi« način. Rešitev je oblikovana tako, da je ob istem nadomestnem ključu zapisan stari in novi opis pridevka. »Stari« opis enostavno prenesemo v polje kjer se hranijo »stari« opisi, »novi« opis pa pride na mesto »starega«. Tako se nam poročila, ki vsebujejo poizvedbe po teh podatkih spremenijo in osvežijo z »novimi« opisi. Če želimo poizvedbe s »starimi« opisi, jih moramo spremeniti toliko, da izpišemo polje, kjer je shranjen »stari« opis. Vendar pa ne moremo enostavno spremeniti zgodovine, kot je to možno pri II. možnosti. Pri tej možnosti se ne vpisuje novih zapisov zaradi

² Nastanek standardne klasifikacije sega v leto 1995, in sicer na podlagi uredbe v uradnem listu RS št. 34/94.

sprememb pri pridevku, niti se ne generira novih vrednosti zaporedja. Metoda je zelo uporabna v statistiki za analizo časovnih vrst.

Obstajajo tudi kombinacije možnosti. Tako je kombinacija I. in II. možnosti mogoča, ko se pojavi nekaj novih vrednosti in nekaj popravkov na nazivih za posamezno vrednost. Druga možnost pa je kombinacija možnosti II. in III. možnosti, ki se navadno izvede v želji po določeni primerjavi v času med staro in novo vrednostjo spremenljivke pri isti meri. Take kombinacije veljajo tudi za bolj kompleksne in jih v tem delu ne bomo obravnavali.

4.2. Hitro se spreminjajoča majhna dimenzija

Termin »počasi se spreminjajoča« smo uporabili za dimenzije, ki se spreminjajo občasno, kot je v našem primeru SKD. V praksi pa se je pogosto težko opredeliti, kaj je počasi in kaj je hitro. Ali lahko to vpliva na obliko dimenzije? Lahko se zgodi, da tudi pri dimenzijah, ki so definirane kot počasi se spreminjajoče, pride do hitrih sprememb. V tem primeru je uporaba II. rešitve iz prejšnjega poglavja najbolj primerna, saj moramo za učinkovito sledenje sprememb zajeti vse spremembe na pridevku (Kimball, 1998, str. 182).

Pri projektu COP 98 je take lastnosti imela geografska dimenzija, ki se na nivoju popisnega okoliša spreminja lahko tudi večkrat v tednu. Višje ko gremo po hierarhiji dimenzije, manjša je spremenljivost. Hierarhija v geografski dimenziji je sestavljena najprej iz popisnih okolišev in so geografsko najmanjši teritorij, ki obsega določeno ozemlje in v njem vsebovane hišne številke. Ti teritoriji se naprej združujejo v naselja, naselja v občine, občine v upravne enote, upravne enote v statistične regije in statistične regije v Slovenijo. V večini primerov se meje skupine nižjih enot v hierarhiji ujemajo z mejami enote na višjem nivoju (Zadnikar, 2002). Obširnost hierarhije je dejansko še bolj podrobna, vendar so v diplomskem delu zajeti samo bistveni, splošno poznani deli. Seveda je tu tudi natančnost zajemanja teh sprememb odvisna od analitičnih potreb. Na statističnem uradu se poročila izdajajo navadno mesečno ali na vsake tri mesece, kar pomeni, da je pomembno, da so zajete vsaj spremembe na vsak zadnji ali prvi dan v mesecu, odvisno od zahteve analitika. Poleg tega pa ni vedno nujno, da vhodni podatki vsebujejo podatek o prostorskem okolišu. Lahko imamo na voljo le podatke na nivoju naselja, kar zopet zmanjša podrobnost sprememb v dimenziji.

Sledenje vsem spremembam nas lahko privede tudi do odločitve, da moramo sprejeti določene kompromise, in sicer, ko trenutna struktura dimenzije ne zadosti trenutnim željam. Njena velikost je tolikšna da jo lahko imenujemo tudi »velika dimenzija« in jo opisujemo v naslednji točki.

4.3. Velika dimenzija

Prej omenjena geografska dimenzija je na nivoju popisnega okoliša nazoren primer tako velike dimenzije, še bolj nazorno pa je, če se spustimo na nivo hišne številke. Le-ta v primeru Slovenije obsega preko 500 000 podatkov (Zadnikar, 2002), ki se s sledenjem vseh sprememb lahko povečuje do preko 1 000 000 zapisov. Drugod po

svetu pa take dimenzije obsegajo milijone zapisov. Največja je dimenzija Prebivalec, ki samo v ZDA vsebuje 100 milijonov zapisov in več. Pri tem se držimo čim bolj tradicionalnih pravil za oblikovanje takih dimenzij, saj je količina podatkov zelo velika in nadzor nad njimi zahteven. Tehnologija izgradnje indeksov in strukture mora slediti:

- možnostim hitrega iskanja, posebej za spremenljivke, ki so zelo podrobne;
- možnostim učinkovitega, če ne hitrega, iskanja posameznih vrednosti preko dimenzije;
- če je le mogoče, ne vključujemo povezav med tabelo mer in veliko dimenzijo pri gradnji poizvedb;
- prav tako izvedemo novo dimenzijo iz dimenzije, kar pomeni izločiti opise pridevkov, ki se podvajajo in jih izvesti kot samostojno dimenzijo. S tem zmanjšamo velikost in podvojenost zapisov (na primer ulica v naslovu);
- ne zapisujemo dodatnih zapisov za sledenje težav pri počasi se spreminjajočih dimenzijah. To pomeni, da se število zapisov v dimenziji ne spreminja. V dimenziji Prebivalec ima vsak prebivalec zapisano določeno kombinacijo vseh pridevkov, ki ustreza trenutnemu časovnemu preseku v podatkih. Če dimenzija zajema več časovnih presekov, je število kombinacij tolikšno, da se zajame vse možne.

4.4. Hitro se spreminjajoča velika dimenzija

Najslabša možnost in tudi najzahtevnejša od vseh primerov je sledenje hitrim spremembam v spreminjajočih se velikih dimenzijah. Dober primer je dimenzija Stranka pri večji zavarovalnici, kjer vodimo razne spremenljivke posamezne stranke, ki so zelo spremenljive. V takem primeru je pametno ločiti iz dimenzije Stranka te spremenljivke (stopnja izobrazbe, velikost dohodka, zakonski stan) in jih zapisati v posebno dimenzijo, ki vsebuje te spremenljivke, imenovano na primer dimenzija Demografija. Pri tem moramo določene številčne spremenljivke spremeniti v velikostne razrede in s tem zmanjšamo razdrobljenost. Tako uspemo vsaj v grobem slediti spremembam. Zatem kreiramo vse možne kombinacije med posameznimi spremenljivkami in jim dodelimo pripadajoče nadomestne ključe. Ti ključi povezujejo stranko, ki pri vsakem novem časovnem preseku enostavno prevzame nov nadomestni ključ nove kombinacije, če se je katera od spremenljivk spremenila. Če ima vsaka naštetna spremenljivka 10 različnih vrednosti, tako dobimo nič več kot 1000 različnih kombinacij, kar ni ravno velika številka v primerjavi s 1 000 000 zapisi, če imamo samo dimenzijo strank z vsemi spremembami. Dobra lastnost takega načina je v tem, da se nam dimenzije v velikosti nič ne spreminjajo. Sprijazniti pa smo se morali, da nimamo več natančnih vrednosti posameznega pojava. Če je dimenzija Demografija le prevelika, se lahko odločimo za izdelavo več manjših dimenzij s podobno zgradbo.

Podroben vpogled v tako strukturo pa da vedeti, da v primeru, ko ni dogodkov, ki se zapisujejo v tabelo mer, tudi ni sledenja spremembam v spremenljivkah posamezne stranke v dimenzijski tabeli. V takem primeru vpišemo dogodek, ki dejansko ne vpliva na vrednost mere, in s tem zapišemo spremembo v spremenljivkah stranke v dimenzijski tabeli.




Največja prednost takega pristopa je v tem, da lahko sledimo spremembam spremenljivk stranke brez povečevanja velikosti podatkovne hrambe ali kompleksnosti podatkov, ki izhajajo iz števila časovnih presekov. Poleg tega pa je dostopni čas do podatkov ostal zadovoljiv, kljub uporabi večih dimenzij v poizvedbi (Kimball, 1998, str. 186), to je sestavljenih poizvedb. Vendar pa imamo tudi slabosti. Morali smo namreč razbiti eno dimenzijo na več manjših dimenzij, kar zmanjšuje preglednost, poleg tega pa smo morali ločiti spremenljivke (dohodek, starost) od spremenljivk, ki so bolj stalne glede sprememb (prebivališče, rojstni datum). To nam zmanjšuje učinkovitost iskanja po podatkih. Demografski podatki so tako dostopni samo v povezavi s tabelo mer, kar podaljšuje dostopni čas do podatkov zaradi sestavljene poizvedbe.

4.5. Degenerirana dimenzija

Do degeneriranih dimenzij prihajamo v primeru, ko zapisujemo individualne podatke v tabelo mer, ki se na primer nanašajo na enega prebivalca in so »skrite« v tabeli mer. Prebivalec je v tem primeru lahko predstavljen z enotno matično številko občana, v nadaljevanju EMŠO, ki sama po sebi nima kakšne razpoznavne vsebine, razen da vsebuje rojstni datum in spol. V navezavi na EMŠO pa lahko pridobimo podatke, ki jih obravnavamo kot mere (letni dohodek, stroški za izobraževanje, akontacija dohodnine, prispevki za socialno varnost). Tak pogled na podatke nam hitro da misliti, da potrebujemo dimenzijo, ki bo lahko opisala vse te mere in bo torej vsebovala pridevke letni dohodek, stroški za izobraževanje, akontacija dohodnine in prispevki za socialno varnost. Podatek kot je EMŠO, ki je zelo podoben nadomestnemu ključu, mora biti primerno ločen od mer in nadomestnih ključev v tabeli mer. EMŠO sodi kot spremenljivka takoj za tujimi ključi in pred merami v tabeli mer. Ko smo končali z dimenzioniranjem podatkov pa ugotovimo, da ni več pridevkov v tabeli mer, ki bi izhajali iz EMŠO, čeprav je na videz zelo podoben tujim ključem. V tabeli mer imamo tako samo še mere, EMŠO in tuje ključe za referenčno vez z dimenzijsko tabelo. Za EMŠO ne izdelujemo dimenzije, ker ne bi vsebovala ničesar. Takim primerom lahko pravimo »degenerirana dimenzija«.

4.6. Šund dimenzija

V operativni bazi podatkov je zelo veliko podatkov o raznih vrednostih, imenovanih statusi ali indikatorji (angl. flag). Primer so indikatorji obravnave, stanja, uparitve z drugimi viri, ki imajo praviloma dvoje do troje različnih vrednosti. Dober oblikovalec podatkovnega skladišča se mora izogibati:

-  izpuščanja takih vrednosti, ki so nespremenjene in vključene v tabelo mer,
-  oblikovanja vsake vrednosti kot spremenljivke v svojo dimenzijo in
-  izpuščanja takih podatkov iz podatkovnega skladišča.

Rešitev, ki se nam ponuja, je podobna kot v primeru pod točko 4.4. Tudi tukaj oblikujemo posebno dimenzijo spremenljivk, ki vsebuje vse možne kombinacije teh spremenljivk. Ker gre v večini primerov za pojave da/ne, je lahko število zapisov, če je takih spremenljivk 10, kvečjemu 1024. S tem lahko pridobimo dodatne možnosti za podrobne analize podatkov.

4.7. Ključ

Vse omenjene pojme smo tekom naštevanja zvrsti dimenzij, seveda, že srečali. Glavni ključ in tuji ključ sta pri gradnji relacijskega podatkovnega modela najpomembnejša gradnika. Glavni ključ zagotavlja nepodvojenost podatkov; s pomočjo tujega ključa in glavnega ključa določimo povezavo med podatki v obeh tabelah. Ravno tako pomembno vlogo imata pri gradnji podatkovnih skladišč. Tuji ključ in glavni ključ v dimenzijski tabeli določata povezavo med tabelo mer in dimenzijsko tabelo, glavni ključ pa nepodvojenost zapisov tako v tabeli mer, kot v dimenzijski tabeli.

Pri počasi se spreminjajoči dimenziji pa smo se srečali še z nadomestnim ključem. Ta se praviloma uporablja pri II. možnosti, ki je opisana v poglavju 4.1., Kimball pa celo priporoča uporabo nadomestnega ključa v vseh dimenzijah (1998, str. 191). V tem primeru je nadomestni ključ tudi glavni ključ v dimenzijski tabeli. Nepodvojenost zapisov pa dodatno zagotavlja še edinstveni ključ (angl. unique key). Lahko poenostavimo, da je nadomestni ključ nadomestilo sestavljenemu ključu, ki ga sestavljata na primer šifra pridevka in datum začetka veljavnosti šifre pridevka v dimenzijski tabeli.

Po Kimballu je nadomestni ključ v dimenziji enostavno celoštevilčno polje, ki vsebuje zaporedne vrednosti. Ključ sam za sebe ne pove ničesar in če nimamo ustrezne šifre ali opisa za pridevek, ki je vključen v dimenzijsko tabelo, o danem podatku ne vemo ničesar. V večini primerov je definicija 4 zloge dolgega³ nadomestnega ključa zadostna, da pokrijemo vse vrednosti spremenljivke (to je 2^{32}) v eni dimenzijski tabeli (tudi za naš primer geografske in SKD).

4.7.1. Nadomestni ključ za časovno dimenzijo

Po Kimballu je uporaba nadomestnih ključev v časovni dimenziji nujna. Kimball navaja uporabo datumske oblike polja kot napačno najprej zato, ker potrebujemo za zapis datuma 8 znakov, kar je enkrat več in kot drugič, je nemogoče zajeti nedefinirane vrednosti za čas, ki jih sicer lahko zapišemo kot »neznana vrednost«, »časovni presek se še ni zgodil« in podobno. Edino opravičilo za uporabo datumske oblike je izpustitev časovne dimenzije iz dimenzijskega modela in navezovanje na datum preseka v tabeli mer direktno.

4.7.2. Uporaba govorečih ključev

Uporaba pametnih ključev je v podatkovnih skladiščih zelo kritična. Glavni ključ v dimenzijski tabeli ne sme biti sestavljen iz več polj, kot je to pametni ključ oziroma sestavljeni ključ. Zato se moramo izogibati združevanju več polj v dimenzijski tabeli v en glavni ključ in namesto njih uporabiti nadomestne ključe, kot smo to navedli na začetku poglavja. Razlog je v glavnem, na podlagi do sedanjih izkušenj, slaba učinkovitost.

³ 1 bajt je sestavljen iz 8 bitov, lahko temu rečemo tudi zlog (Pahor, 2002, str. 716).

4.7.3. Producerski ključ







Uporaba producerskih ključev je podobno kot uporaba pametnih ključev v podatkovnih skladiščih neustrezna predvsem zaradi dejstva, da se prej ali slej v proizvodnji zamenja tehnologija in s tem tudi trenutna uporaba producerskega ključa, kar nam lahko podre zasnovano za izbrano dimenzijsko tabelo. Producerski ključ so navadno šifre, ki so sestavljene iz črk in števil. V določenem časovnem obdobju zagotavljajo tudi referenčno vez na pravilni opis pridevka v dimenzijski tabeli. Vendar, ko se šifra zamenja, opis pa ostane isti, ali obratno, nam oteži ustrezno shranjevanje podatkov v tabelo mer in dimenzijsko tabelo. Poleg tega uporaba šifre privede do uporabe velikega števila zlogov za tuji ključ v tabeli mer. Še večje težave se pojavijo pri počasni se spreminjajoči dimenziji, kjer je sledenje v zgodovino brez uporabe nadomestnega ključa nemogoče.

5. Uporaba standardne klasifikacije v podatkovnem skladišču

5.1. Standardna klasifikacija dejavnosti – SKD

SKD temelji na evropski klasifikaciji dejavnosti NACE Rev 1. (Hlavaty, 2002, str. 5), ki je obvezen statistični standard Evropske unije, in je neposredno povezana z mednarodno klasifikacijo dejavnosti Združenih narodov »International Standard Industrial Classification – ISIC Rev 3.«

SKD ima hierarhično razčlenbo in določa 6 ravni dejavnosti (Hlavaty, 2002, str. 6):

-  področje (označeno z eno črko),
-  podpodročje (označeno z dvema črkama),
-  oddelek (označen z dvomestno številčno oznako),
-  skupina (označena s trimestno številčno oznako),
-  razred (označen s štirimestno številčno oznako) in
-  podrazred (označen s petmestno številčno oznako).

Med številčno oznako za oddelek in skupino stoji pika. Do predzadnje ravni razreda je razčlenba povsem usklajena z NACE Rev. 1, zadnja raven – podrazred (6. nivo) – pa je državna razčlenba, ki se uporablja za razvrščanje nacionalnih posebnosti.

SKD je predpisana z uredbo vlade RS, za usklajevanje SKD s predpisi Evropske unije in za uvajanje sprememb in dopolnitev pa je zadolžen Statistični urad Republike Slovenije.

Prva uredba o uvedbi standardne klasifikacije dejavnosti je bila objavljena v Uradnem listu RS, št 34/94. Od 1.1.1997, ko je bilo končano dvoletno prehodno obdobje od sprejema uredbe, se SKD uporablja kot obvezni statistični standard.

Klasifikacija je na Statističnem uradu Republike Slovenije hranjena in vzdrževana s pomočjo strežnika za statistične klasifikacije KLASJE. Izdelan je bil v okviru programa posodobitve državne statistike (Čemažar, 2000).

5.2. Razvojno okolje rešitve

Za razvojno okolje rešitve smo izbrali ORACLE podatkovno zbirko, verzije 8.1.7. Podpira uporabo »bitmap indeksov« in particioniranje tabel, ki so za delovanje podatkovnega skladišča zelo pomembni.

Celoten projekt SURS RDB je zasnovan (Pavlin, 2002) na tehnologiji in programskih rešitvah v okolju Oracle (Oracle Data Warehouse). Ločili smo ga na dve osnovni ravni, po katerih so podatki organizirani v:

- mikro raven, ki jo predstavljajo relacijske tabele z individualnimi statističnimi podatki. Ti podatki so urejeni za uporabo standardnih klasifikacij, shranjenih in centralno vodenih na strežniku za klasifikacije KLASJE (na SURS), ter za uporabo geografske dimenzije in dimenzije SKD kot počasi se spreminjajoči dimenziji, izgrajeni znotraj podatkovnega skladišča SURS RDB in
- makro raven, kjer so podatki agregirani na raven popisnega okoliša ali na višje prostorske ravni (skupaj 117 tabel s 4,3 Gb zapisi). Do teh podatkov uporabnik dostopa preko uporabniškega vmesnika, izdelanega s pomočjo Oracle Discoverer.

Rezultati izbranih statističnih raziskovanj so razdeljeni na vsebinska področja, ki definirajo posamezna področna podatkovna skladišča. Teh področij je 10:

- prebivalstvo (selitve, rojstva, umri, sklenitve in razveze zakonskih zvez),
- plače,
- investicije v osnovna sredstva,
- izobraževanje (podatki o srednjem, dodiplomskem in podiplomskem izobraževanju),
- promet in zveze,
- gostinstvo in turizem,
- trgovina (statistika prodajaln v trgovini na drobno),
- raziskovanje in razvoj,
- gradbeništvo in
- okolje (podatki o izpustu in prečiščenju odpadnih voda ter investicijah v varstvo okolja).

Tabele mer se v podatkovnih skladiščih z vsakim vnosom novega časovnega preseka stalno povečujejo. Nemogoče je predvideti, kakšna bo njihova velikost vnaprej, zato je potrebno stalno vzdrževanje rezerviranega prostora za podatke. S pomočjo particionirane tabele je ta postopek lažji. Ideja deluje podobno, kot da bi na primer tri tabele mer za leto 2000, 2001 in 2002 združili med seboj v eno tabelo in jo razdelili na tri »dele« ali particije. Te »dele« tudi razdelimo po letih in poimenujemo ter tako definiramo particije za leto 2000, 2001 in 2002. Tabela je ena sama in nadomešča prej tri omenjene tabele. Pri vnosu novega časovnega preseka se enostavno definira nova particija s predvidenim rezerviranim prostorom za podatke in tabela je pripravljena za polnjenje podatkov.

Indeks ali kazalo (Pahor, 2002, str. 203) je dodatek tabeli ali datoteki v zbirki podatkov, ki omogoča hitrejši dostop do izbranih zapisov glede na spremenljivko ali

skupino spremenljivk, na katero se nanaša. Indeks ima lahko različne zgradbe, na primer v obliki dvojiškega B-drevesa (angl. B-tree indeks), z uporabo zgostitvene funkcije, z razvrstitvijo obrnjenih vrednosti ali z razporedom bitnih slik. V našem primeru je »bitmap indeks« kazalo z razporedom bitnih slik. Bitne slike so matrike, ki imajo na eni osi vse vrednosti spremenljivke, na drugi osi pa številko zapisa. Če nek zapis vsebuje vrednost spremenljivke, je na tistem mestu zapisana '1', drugje so '0' (Raisinghani, 2002). Temu lahko rečemo tudi sito. Dobra lastnost takih indeksov je, da se jih da enostavno sestavljati, kot bi sestavili več različnih sit, kar nam omogoča hitro vrtnanje po podatkih.

Enako razvojno okolje smo uporabljali tudi pri razvoju projekta RDB, katerega del je dimenzija SKD.

Postopek gradnje dimenzijske tabele in njeno vključitev na poljubno tabelo mer bomo prikazali na podlagi predstavitve v publikaciji OWB PROFESSIONAL COMMUNITY, Slowly Changing Dimension (Soet, 2001), ki vključuje vse postopke za izgradnjo posameznega primera počasi se spreminjajoče dimenzije, opisane v točki 4.1. Za rešitev počasi se spreminjajoče dimenzije SKD bom uporabil II. možnost počasi se spreminjajoče dimenzije in ga primerjal z rešitvijo po III. možnosti. Nato bom še prikazal, kako se kaže sprememba v podatkih, ko eno podjetje spremeni svojo dejavnost na II. in III. možnosti. S tem bom tudi analiziral uporabnost obeh primerov pri izdelavi poročil v statistiki.

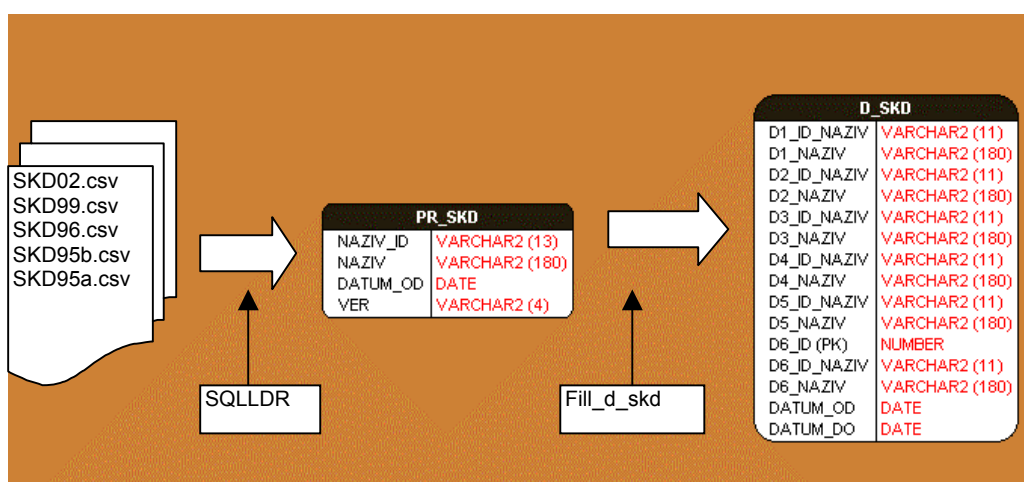
Osnovni elementi za gradnjo naše dimenzije so:

- ASCII datoteka s podatki o vsebini SKD dimenzije z določeno strukturo in obliko zapisa. V našem primeru smo imeli na voljo datoteko, ki je imela spremenljivke v zapisu ločene s podpičjem ali vejico, v primeru, če gre za opisno spremenljivko, je vrednost označena z narekovaji;
- kontrolna datoteka s podatki, ki omogočijo prenesti ASCII datoteko v ORACLE relacijsko bazo podatkov s pomočjo programa SQLLDR. Kontrolna datoteka pove SQLLDR programu, v katero tabelo na določeni shemi v ORACLE relacijski podatkovni bazi in v kakšni obliki mora prenesti podatke, ki se nahajajo v izbrani ASCII datoteki. Na drugi strani pa pove programu tudi strukturo ASCII datoteke in strukturo tabele v relacijski bazi;
- PL/SQL procedura na podlagi algoritma (slika 9) prevede podatke v tako obliko, da ustreza izbrani strukturi dimenzije;
- prehodna tabela, kjer so shranjeni podatki iz ASCII datoteke;
- končna dimenzijska tabela, ki je lahko fizično ločena od ostalih elementov tako, da se nahaja v drugi shemi relacijske baze podatkov kot sestavni del podatkovnega skladišča;
- procedura ali funkcija, ki dodeli v tabelo mer nadomestni ključ za spremenljivko glede na vnesen produkcijski ključ oziroma šifro in datum za izbran časovni presek.

5.3. Prenos podatkov v dimenzijsko tabelo

Prenos podatkov iz klasifikacije v dimenzijo bomo izvedli po II. možnosti (slika 8), omenjeni v poglavju 4.1. Pri tej odločitvi sem upošteval dejstvo, da je trenutna klasifikacija sicer že usklajena z veljavno klasifikacijo za evropsko statistiko (Hlavaty, 2002, str. 5), vendar bodo po moji oceni spremembe vse bolj pogoste, saj bo novih dejavnosti vedno več. Za primerjavo bom podal prenos podatkov po III. možnosti, ravno tako opisani v poglavju 4.1 in tako bomo dejansko lahko ugotovili prednosti in slabosti ene in druge možnosti prenosa podatkov v dimenzijsko tabelo.

Slika 8: Pretok podatkov iz ASCII datotek v dimenzijsko tabelo (II. možnost)



Vir: Relacijska podatkovna zbirka ORACLE in programsko orodje T.O.A.D.

Prenos podatkov po III. možnosti je na videz podoben, razlikuje se le v proceduri, ki prenese podatke iz prehodne tabele v dimenzijsko tabelo zaradi drugačne strukture dimenzijske tabele.

Slabosti III. možnosti in hkrati prednosti II. možnosti se kažejo v tem, da je nemogoče predvideti, koliko sprememb bo v bodoče nastalo na opisu posamezne dejavnosti. Poleg tega je težavno tudi ukinjanje dejavnosti in ponovno uveljavljanje. V tem primeru mora opis zavzeti vrednost opisa za neznano dejavnost. Za vsak tak dogodek je potrebno prenesti stari opis dejavnosti v polje za stari opis, pri čemer se moramo vprašati, ali je polje prazno ali že vsebuje kak opis od prejšnje spremembe. Če je to res, se moramo odločiti, ali ga prepíšemo in s tem zgubimo vpogled v zgodovino klasifikacije. Drugače dodamo novo polje za prepis opisa pred novim vnosom opisa dejavnosti. Pri večkratnih spremembah se nam število polj povečuje, potrebno pa je voditi dokumentacijo o obdobju, ko je bil določen opis dejavnosti v veljavi. Do spremembe lahko pride tudi v hierarhiji, ki se kaže v spremembi opisa nadrejene dejavnosti, s tem se lahko število novih polj samo še poveča, kar narekuje zahteva po ohranitvi zgodovine klasifikacije.

Prednosti III. možnosti in hkrati slabosti II. možnosti se kažejo samo v primeru pravopisnih napak, ki niso bile opažene takoj ob vnosu, v primeru spremembe dvoumnega opisa dejavnosti in »kozmetičnih« sprememb opisa dejavnosti. V takih in

podobnih primerih se starih opisov ne shranjuje v nova polja, temveč se ohrani samo do sedaj veljavni opis dejavnosti in novi opis dejavnosti. Generiranje novih zapisov s tem ni potrebno, kar ohrani velikost klasifikacije in nepotrebno novo nalaganje podatkovnega skladišča zaradi spremembe same klasifikacije.

Za nazoren prikaz prednosti in slabosti II. in III. možnosti bomo uporabili tudi analizo končnih poročil s posebej prirejenimi neresničnimi podatki o podjetjih po dejavnostih v klasifikaciji SKD.

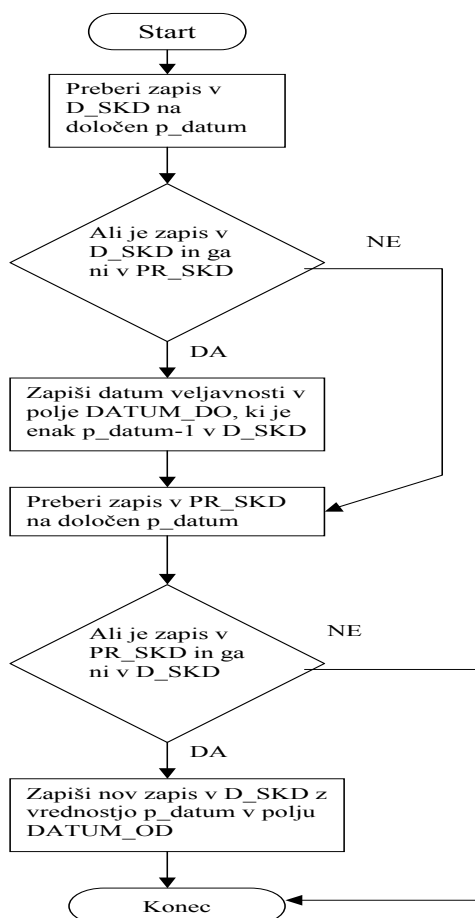
Ker je prenos podatkov izveden v ORACLE podatkovni zbirki, je pomembna uporaba za to primernih orodij. Uporabili jih bomo po točno določenem zaporedju in tako zagotovili pravilno razporeditev podatkov v dimenzijski tabeli.

5.4. Izvedba dimenzijske tabele

5.4.1. Pretok podatkov v dimenzijsko tabelo

Podatke trenutno veljavne klasifikacije prenesemo s pomočjo kontrolne datoteke in programa SQLLDR v prehodno tabelo enake strukture, kot je vhodna datoteka s podatki o klasifikaciji.

Slika 9: Algoritem zapisovanja podatkov v dimenzijo po II. možnosti.



Vir: Avtor.

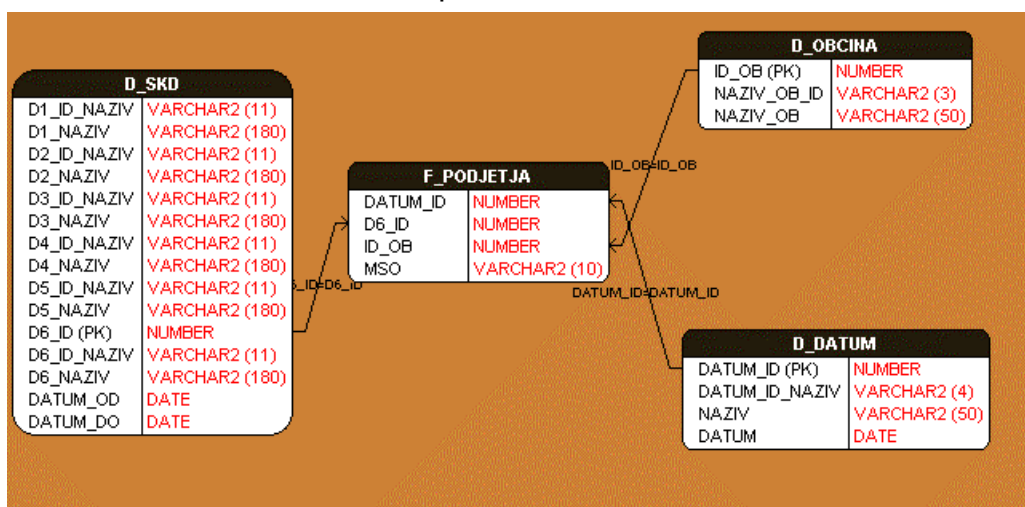
Zatem podatke s pomočjo programa, ki je izdelan po algoritmu, opisanem v nadaljevanju (slika 9), prenesemo v dimenzijsko tabelo. Najprej preverimo, ali je zapis v dimenzijski tabeli zapisan tudi v prehodni tabeli. Če to ni res, prepisemo datum konca veljavnosti zapisa s p_datum-1. Algoritemska spremenljivka p_datum predstavlja v našem primeru trenutni časovni presek, za katerega vnašamo novo verzijo klasifikacije. Zatem preberemo zapis v prehodni tabeli in preverimo, ali že obstaja zapis v dimenzijski tabeli, ki velja na p_datum. Če to drži, zapis zapišemo v dimenzijsko tabelo. Algoritem velja za samo en zapis, če je zapisov na dani časovni presek v spremenljivki p_datum več, potem se zanka ponavlja, dokler ni več zapisov na dani časovni presek. Za III. možnost je algoritem podoben, le da spremembe izvajamo na spremenljivkah za stari opis pridevka, kjer se shrani stari opis pridevka in za novi opis pridevka, kjer se prepíše stari opis pridevka z novim opisom.

5.4.2. Podatkovno skladišče z uporabo SKD dimenzije po II. možnosti

S pomočjo neresničnih in za našo analizo posebej prirejenih podatkov o številu podjetij po dejavnostih in po občinah smo definirali področno podatkovno skladišče, ki vsebuje eno tabelo mer in tri dimenzijske tabele. Slika 10 nam prikazuje strukturo področnega podatkovnega skladišča. Razvidno je, da ga sestavljajo tabela mer F_PODJETJA in dimenzijske tabele D_DATUM, D_OBCINE in D_SKD, ki ima strukturo počasi se spreminjajoče dimenzije. Struktura posameznih objektov je priložena v prilogi. V dimenzijski tabeli D_SKD se nahajajo podatki o klasifikacij SKD za časovno obdobje od 01. 01. 1995 do 31. 12. 2000.

Za polnjenje tabele F_PODJETJA smo uporabili program FILL_F_PODJETJA s funkcijo VRNI_D_SKD, ki je časovno odvisna in dodeli nadomestni ključ iz dimenzijske tabele D_SKD, dejansko veljaven za datum časovnega preseka, prebranega iz vhodnih podatkov v tabeli PR_PODJETJA. Podjetje je v tabeli merenovito predstavljeno s spremenljivko MSO, ki vsebuje matično številko organizacije.

Slika 10: Področno podatkovno skladišče, II. možnost



Vir: Programsko orodje T.O.A.D., orodje SQL Modeler.

S pomočjo orodja 'Get external data' v Excelu in pripravljenega SQL stavka prenesemo podatke v preglednico Excel in oblikujemo izpis.

SQL stavek je takle:

```
select c.NAZIV,b.NAZIV_OB,d.D6_NAZIV,count(a.MSO)
from f_podjetja a, d_obcina b, d_datum c, d_skd d
where a.ID_OB=b.ID_OB
and a.DATUM_ID=c.DATUM_ID
and a.D6_ID=d.D6_ID
group by c.naziv,b.NAZIV_OB,d.D6_NAZIV
```

Iz SQL stavka lahko razberemo, da smo za povezavo med tabelo mer in dimenzijsko tabelo izkoristili nadomestne ključe, ki so bili definirani v strukturi podatkovnega skladišča. Zaradi načina polnjenja tabele mer, kjer smo v proceduri upoštevali dejanski datum veljavnosti klasifikacije, se nam v opisih pridevkov spremenljivke SKD izpiše vedno prava dejavnost, veljavna na datum časovnega preseka.

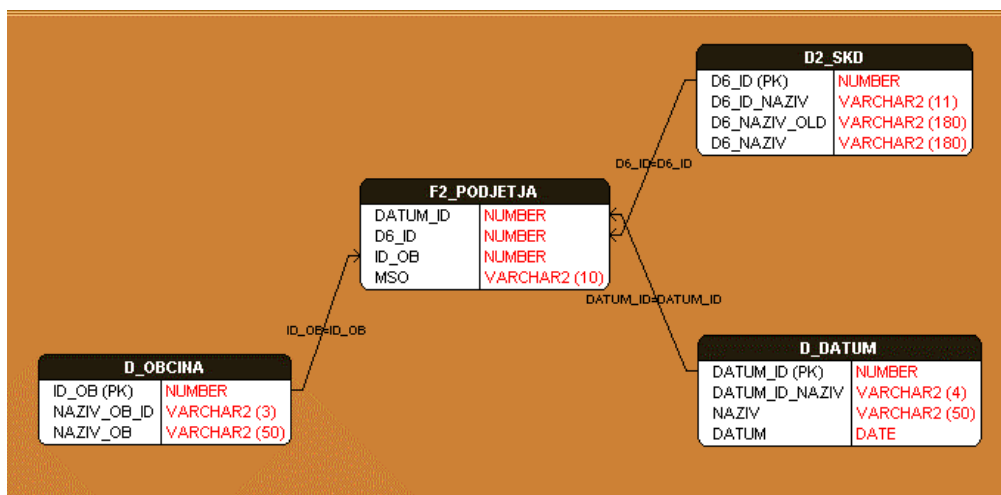
Izpise bomo primerjali kasneje in tako razložili uporabnost obeh struktur dimenzij, ki jih obravnavamo.

5.4.3. Podatkovno skladišče z uporabo SKD dimenzije po III. možnosti

Za primerjavo smo izdelali še področno podatkovno skladišče, ki bo vsebovalo enake podatke in elemente kot v prejšnjem poglavju, le da bo imela dimenzija D_SKD spremenjeno strukturo in jo bomo imenovali D2_SKD. Obravnavali bomo samo najnižji 6. nivo SKD dimenzije, s pomočjo katerega smo tudi kasneje izdelali izpise za obe možnosti. Struktura takega podatkovnega skladišča (slika 11) se v primerjavi z že opisano razlikuje v strukturi dimenzije D2_SKD, ker ne vsebuje polj za datum začetka veljavnosti in datum konca veljavnosti klasifikacije, vsebuje pa polje za stari opis dejavnosti. Tako so vpisani opisi dejavnosti iz klasifikacije SKD na dan 31. 05. 1995 v poljih za stari opis dejavnosti in opisi dejavnosti na dan 24. 08. 1997 v poljih za novi opis dejavnosti.

Polnjenje podatkov v podatkovno skladišče poteka podobno kot v prvem skladišču. S pomočjo programa z imenom FILL_F2_PODJETJA polnimo tabelo F2_PODJETJE, ki vključuje funkcijo VRNI_D2_SKD, da vpisujemo vrednosti za nadomestni ključ D6_ID.

Slika 11: Področno podatkovno skladišče, III. možnost



Vir: Programsko orodje T.O.A.D., orodje SQL Modeler.

S pomočjo orodja 'Get external data' v Excelu in pripravljene SQL stavke prenesemo podatke v preglednico Excel in oblikujemo izpis.

SQL stavek za izpis s starimi nazivi je takle:

```
select c.NAZIV,b.NAZIV_OB,d.D6_NAZIV_OLD,count(a.MSO)
from f2_podjetja a, d_obcina b, d_datum c, d2_skd d
where a.ID_OB=b.ID_OB
and a.DATUM_ID=c.DATUM_ID
and a.D6_ID=d.D6_ID
group by c.naziv,b.NAZIV_OB,d.D6_NAZIV_old
```

SQL stavek za izpis z novimi nazivi se glasi sledeče:

```
select c.NAZIV,b.NAZIV_OB,d.D6_NAZIV,count(a.MSO)
from f2_podjetja a, d_obcina b, d_datum c, d2_skd d
where a.ID_OB=b.ID_OB
and a.DATUM_ID=c.DATUM_ID
and a.D6_ID=d.D6_ID
group by c.naziv,b.NAZIV_OB,d.D6_NAZIV
```

Iz SQL stavka lahko razberemo, da smo uporabili enak način izdelave stavka, kot v prvem primeru in s tem upoštevali enako pravilo. Razlika se pokaže šele zaradi drugačne strukture dimenzije SKD, ki smo jo uporabil pri polnjenju tabele mer v proceduri. Dimenzijska tabela tako vsebuje še stolpec D6_NAZIV_OLD, kjer je shranjen opis dejavnosti, ki je veljal pred vnosom novega opisa, poleg polja D6_NAZIV. To polje pa koristno uporabimo pri izdelavi SQL stavka, in sicer odvisno, ali želimo izpis z opisi dejavnosti, ki so veljali prej in so vpisani v polju NAZIV_OLD, ali želimo izpis z opisi dejavnosti, ki veljajo sedaj in so vpisani v polju NAZIV.

Izpise bom primerjal kasneje in tako razložil uporabnost obeh struktur dimenzij, ki jih obravnavam.

5.5. Primerjava izpisov po II. in III. možnosti, podatki o podjetjih se ne spreminjajo

Izdelali smo tri izpise s pomočjo Excela in priloženih podatkov. Ti so prikazani na sliki 12.

Slika 12: Podatki o podjetjih se po dejavnosti in občini ne spreminjajo

A. Število podjetij po dejavnostih in občinah

Občina	Dejavnost podjetja, 6. nivo	Časovni presek			
		Časovni presek, dne 01. 01. 1995	Časovni presek, dne 31. 05. 1995	Časovni presek, dne 24. 08. 1997	Časovni presek, dne 31. 08. 1999
Neznana občina	Alternativne oblike zdravljenja	0	1	1	1
	Bolnišnična zdravstvena dejavnost	0	1	1	1
	Dejavnost centrov za socialno delo	0	3	3	3
	Dejavnost dobrodelnih organizacij	0	8	8	8
	Dejavnost domov za starejše	0	1	1	1
	Dejavnost invalidskih organizacij	0	3	3	3
	Dejavnost invalidskih podjetij	0	0	3	3
	Dejavnost posebnih socialnovarstvenih zavodov za odrasle	0	1	1	1
	Dejavnost socialnovarstvenih zavodov za usposabljanje otrok in mladostnikov	0	7	7	7
	Dejavnost varstveno delovnih centrov	0	1	1	1
	Druge oblike institucionalnega varstva	0	1	1	1
	Druge socialne dejavnosti	0	6	6	6
	Druge zdravstvene dejavnosti, d.n.	0	1	1	1
	Eksteritorialne organizacije in združenja	0	5	5	5
	Osnovna zdravstvena dejavnost	0	5	5	5
	Samostojne zdravstvene dejavnosti, ki jih ne opravljajo zdravniki	0	1	1	1
	Specialistična ambulantna dejavnost	0	1	1	1
	Veterinarstvo	0	5	5	5
	Zobozdravstvena dejavnost	0	1	1	1
	NEDEFINIRANO	58	11	3	8
	Nerazvrščeno	5	0	0	0
	Neznano	0	0	5	0
Neznana občina Total		63	63	63	63
Grand Total		63	63	63	63

B. Število podjetij po dejavnostih v dimenziji SKD, veljavni na dan 24. 08. 1997, in občinah

Občina	Dejavnost podjetja, 6. nivo	Časovni presek			
		Časovni presek, dne 01. 01. 1995	Časovni presek, dne 31. 05. 1995	Časovni presek, dne 24. 08. 1997	Časovni presek, dne 31. 08. 1999
Neznana občina	Alternativne oblike zdravljenja	1	1	1	1
	Bolnišnična zdravstvena dejavnost	1	1	1	1
	Dejavnost centrov za socialno delo	3	3	3	3
	Dejavnost dobrodelnih organizacij	8	8	8	8
	Dejavnost domov za starejše	1	1	1	1
	Dejavnost invalidskih organizacij	3	3	3	3
	Dejavnost invalidskih podjetij	3	3	3	3
	Dejavnost posebnih socialnovarstvenih zavodov za odrasle	1	1	1	1
	Dejavnost socialnovarstvenih zavodov za usposabljanje otrok in mladostnikov	7	7	7	7
	Dejavnost varstveno delovnih centrov	1	1	1	1
	Druge oblike institucionalnega varstva	1	1	1	1
	Druge socialne dejavnosti	6	6	6	6
	Druge zdravstvene dejavnosti, d.n.	1	1	1	1
	Eksteritorialne organizacije in združenja	5	5	5	5
	Osnovna zdravstvena dejavnost	5	5	5	5
	Samostojne zdravstvene dejavnosti, ki jih ne opravljajo zdravniki	1	1	1	1
	Specialistična ambulantna dejavnost	1	1	1	1
	Veterinarstvo	5	5	5	5
	Zobozdravstvena dejavnost	1	1	1	1
	NEDEFINIRANO	3	3	3	3
	Neznano	5	5	5	5
Neznana občina Total		63	63	63	63
Grand Total		63	63	63	63

C. Število podjetij po dejavnostih v dimenziji SKD, veljavni na dan 31. 05. 1995, in občinah

Občina	Dejavnost podjetja, 6. nivo	Časovni presek			
		Časovni presek, dne 01. 01. 1995	Časovni presek, dne 31. 05. 1995	Časovni presek, dne 24. 08. 1997	Časovni presek, dne 31. 08. 1999
Neznana občina	Alternativne oblike zdravljenja	1	1	1	1
	Bolnišnična zdravstvena dejavnost	1	1	1	1
	Dejavnost centrov za socialno delo	3	3	3	3
	Dejavnost dobrodelnih organizacij	8	8	8	8
	Dejavnost domov za starejše	1	1	1	1
	Dejavnost invalidskih organizacij	3	3	3	3
	Dejavnost posebnih socialnovarstvenih zavodov za odrasle	1	1	1	1
	Dejavnost socialnovarstvenih zavodov za usposabljanje otrok in mladostnikov	7	7	7	7
	Dejavnost varstveno delovnih centrov	1	1	1	1
	Druge oblike institucionalnega varstva	1	1	1	1
	Druge socialne dejavnosti	6	6	6	6
	Druge zdravstvene dejavnosti, d.n.	1	1	1	1
	Eksteritorialne organizacije in združenja	5	5	5	5
	Osnovna zdravstvena dejavnost	5	5	5	5
	Samostojne zdravstvene dejavnosti, ki jih ne opravljajo zdravniki	1	1	1	1
	Specialistična ambulantna dejavnost	1	1	1	1
	Veterinarstvo	5	5	5	5
	Zobozdravstvena dejavnost	1	1	1	1
	NEDEFINIRANO	11	11	11	11
Neznana občina Total		63	63	63	63
Grand Total		63	63	63	63

Vir: Avtor.

Izpisi temeljijo na neresničnih podatkih o številu podjetij po občinah in dejavnostih za obdobje od 01. 01. 1995 do 31. 08. 1999 v štirih časovnih presekih, ki so si med seboj identični. Spreminja se samo SKD dimenzija.

- A. Izpis je izdelan s pomočjo področnega podatkovnega skladišča iz poglavja 5.4.2. in prikazuje podatke o številu podjetij po dejavnostih in občinah v vseh štirih časovnih presekih. Iz tabele je razvidno, kako se vsebinsko spreminja njena informacija, kakor se je spreminjala klasifikacija SKD. Poudarek je na sivo obarvanih poljih. V tem primeru so podatki o podjetjih skozi vsa obdobja enaki, vendar pa tabela prikazuje razlike v posameznih časovnih obdobjih zaradi spreminjanja dimenzije SKD. Spremembe so razvidne pri uveljavitvi prve verzije klasifikacije, ki je veljala na dan 31. 05. 1995, vendar ne na dan 01. 01. 1995, zato so se vsi podatki o dejavnosti podjetja zvrstili kot nedefinirani. Pri uveljavitvi druge verzije klasifikacije je razviden časovni zamik, ko so nekatera podjetja imela že leta 1995 dejavnost invalidskih podjetij, čeprav je ta dejavnost postala veljavna šele po letu 1997, kar je razvidno iz podatkov za časovni presek dne 24. 08. 1997.
- B. Izpis je izdelan s pomočjo področnega podatkovnega skladišča iz poglavja 5.4.3. in prikazuje podatke o številu podjetij po dejavnostih, ki so veljale na dan 24. 08. 1997, in občinah v vseh časovnih presekih. Ker je klasifikacija dejavnosti stalna za celotno obdobje, je razvidno, da so podatki o podjetjih za vse obdobje stalni.
- C. Izpis je izdelan s pomočjo področnega podatkovnega skladišča iz poglavja 5.4.3. in prikazuje podatke o številu podjetij po dejavnostih, ki so veljale na dan 31. 05. 1995, in občinah v vseh štirih časovnih presekih. Iz tabele je

razvidno, da so podatki o podjetjih za vse obdobje enaki, vendar zaradi druge klasifikacije SKD drugačni kot v izpisu B.

Če primerjamo vse tri izpise ob predpostavki, da dobro poznamo podatke in vemo, kakšna oblika izpisa nam ustreza, lahko ugotovimo iz A. izpisa, da je dimenzija SKD po II. možnosti uporabna skozi celotno obdobje, pomembno je le stalno osveževanje klasifikacije z novo vsebino. Težava se kaže pri medsebojni primerjavi različnih časovnih presekov. Zato potrebujemo izdelavo nove tabele mer, ki bi vsebovala podatke o podjetjih po občinah in dejavnostih na točno določen časovni presek klasifikacije SKD (na primer 31. 05. 1995). Za izdelavo izpisa, ki bi vseboval klasifikacijo SKD na datum 24. 08. 1997, bi potrebovali napolniti podatke o podjetjih po občinah in dejavnostih v to novo tabelo mer za izbrani časovni presek klasifikacije SKD.

Možna je tudi primerjava podatkov med različnimi časovnimi preseki, vendar je slika zavajajoča, saj se iz podatkov razbere, da se je število podjetij po dejavnostih spreminjalo, čeprav to ni res. Boljši vpogled v podatke nam da izpis B. in C. izdelan s pomočjo III. možnosti. Če želimo izdelati izpis s klasifikacijo, ki velja za drugi časovni presek, je potrebno samo prepisati podatke o opisih dejavnosti, kot smo to opisali v poglavju 4.1. in izpis bi bil izdelan. Težava se pojavi v primeru, ko je število zapisov v stari klasifikaciji drugačno od števila zapisov v novi klasifikaciji. V tem primeru je potrebno najprej uskladiti dimenzijo z obema klasifikacijama in zatem ponovno polniti tabelo mer.

Če izpustimo predpostavko, da vemo, kakšne izpise potrebujemo, je slabost III. možnosti – nezmožnost integracije – v tem, da je težko uporabiti dimenzijsko tabelo za drugačne podatke o podjetjih brez predhodnega popravljanja dimenzije, kar pokvari našo že opravljeno raziskavo. Zato je potrebna preslikava dimenzije z ustreznimi podatki. V primeru, ko gradimo podatkovno skladišče, je seveda nemogoče predvideti, kakšne raziskave bodo želeli analitiki izvajati, zato je po izkušnjah iz projekta RDB bolje, da se posebej pripravi tabela mer za časovne vrste in izbere dimenzija po II. možnosti, saj bi se drugače dimenzije preveč množile in bi s tem otežili vzdrževanje podatkovnega skladišča. Polnjenje podatkov v, za časovne vrste prirejeno, tabelo mer pa je enako zahtevno, kot izdelati posnetek obstoječe dimenzije s klasifikacijo, ki je veljavna za drugo časovno obdobje.

Pri B. in C. izpisu se enakost v podatkih med posameznimi časovnimi preseki lepo opazi, zato je III. možnost zelo uporabna v statistiki za analizo časovnih vrst. Pomembno je samo opomniti, da klasifikacija SKD velja za izhodiščni datum. Za izdelavo izpisa na drug časovni presek klasifikacije SKD se spremeni samo poizvedba in tako izdela izpis ali pod B. točko ali pod C. točko. Iz prej razloženega problema lahko sklenemo, da je naše področno podatkovno skladišče posebej prilagojeno zahtevam po analizi časovnih vrst za te podatke o podjetjih, in sicer ob predpostavki, da poznamo obliko izpisa.

5.6. Primerjava izpisov po II. in III. možnosti, podatki o podjetjih po dejavnostih se spreminjajo

V tem primeru smo ravno tako s pomočjo Excela in priloženih podatkov izdelali tri izpise, ki so prikazani na sliki 13.

Slika 13: Podatki o podjetjih se po dejavnosti in občini spreminjajo

A. Število podjetij po dejavnostih in občinah

Število podjetij	Občina	Dejavnost podjetja, 6. nivo	Časovni presek			
			Časovni presek, dne 01. 01. 1995	Časovni presek, dne 31. 05. 1995	Časovni presek, dne 24. 08. 1997	Časovni presek, dne 31. 08. 1999
	Neznana občina	Alternativne oblike zdravljenja	0	1	1	1
		Bolnišnična zdravstvena dejavnost	0	1	1	1
		Dejavnost centrov za socialno delo	0	3	3	3
		Dejavnost dobrotelnih organizacij	0	8	8	8
		Dejavnost domov za starejše	0	1	0	0
		Dejavnost invalidskih organizacij	0	3	3	2
		Dejavnost invalidskih podjetij	0	0	3	3
		Dejavnost posebnih socialnovarstvenih zavodov za odrasle	0	1	0	0
		Dejavnost socialnovarstvenih zavodov za usposabljanje otrok in mladostnikov	0	7	6	9
		Dejavnost varstveno delovnih centrov	0	1	4	2
		Druge oblike institucionalnega varstva	0	1	1	1
		Druge socialne dejavnosti	0	6	6	6
		Druge zdravstvene dejavnosti, d.n.	0	1	1	1
		Eksteritorialne organizacije in združenja	0	5	5	5
		Osnovna zdravstvena dejavnost	0	5	1	2
		Samostojne zdravstvene dejavnosti, ki jih ne opravljajo zdravniki	0	1	1	1
		Specialistična ambulantna dejavnost	0	1	1	2
		Veterinarstvo	0	5	9	7
		Zobozdravstvena dejavnost	0	1	1	1
		NEDEFINIRANO	58	11	3	8
		Nerazvrščeno	5	0	0	0
		Neznano	0	0	5	0
	Neznana občina Total		63	63	63	63
	Grand Total		63	63	63	63

B. Število podjetij po dejavnostih v dimenziji SKD, veljavni na dan 24. 08. 1997, in občinah

Število podjetij	Občina	Dejavnost podjetja, 6. nivo	Časovni presek			
			Časovni presek, dne 01. 01. 1995	Časovni presek, dne 31. 05. 1995	Časovni presek, dne 24. 08. 1997	Časovni presek, dne 31. 08. 1999
	Neznana občina	Alternativne oblike zdravljenja	1	1	1	1
		Bolnišnična zdravstvena dejavnost	1	1	1	1
		Dejavnost centrov za socialno delo	3	3	3	3
		Dejavnost dobrotelnih organizacij	8	8	8	8
		Dejavnost domov za starejše	1	1	0	0
		Dejavnost invalidskih organizacij	3	3	3	2
		Dejavnost invalidskih podjetij	3	3	3	3
		Dejavnost posebnih socialnovarstvenih zavodov za odrasle	1	1	0	0
		Dejavnost socialnovarstvenih zavodov za usposabljanje otrok in mladostnikov	7	7	6	9
		Dejavnost varstveno delovnih centrov	1	1	4	2
		Druge oblike institucionalnega varstva	1	1	1	1
		Druge socialne dejavnosti	6	6	6	6
		Druge zdravstvene dejavnosti, d.n.	1	1	1	1
		Eksteritorialne organizacije in združenja	5	5	5	5
		Osnovna zdravstvena dejavnost	5	5	1	2
		Samostojne zdravstvene dejavnosti, ki jih ne opravljajo zdravniki	1	1	1	1
		Specialistična ambulantna dejavnost	1	1	1	2
		Veterinarstvo	5	5	9	7
		Zobozdravstvena dejavnost	1	1	1	1
		NEDEFINIRANO	3	3	3	3
		Neznano	5	5	5	5
	Neznana občina Total		63	63	63	63
	Grand Total		63	63	63	63

C. Število podjetij po dejavnostih v dimenziji SKD, veljavni na dan 31. 05. 1995, in občinah

Število podjetij	Občina	Dejavnost podjetja, 6. nivo	Časovni presek			
			Časovni presek, dne 01. 01. 1995	Časovni presek, dne 31. 05. 1995	Časovni presek, dne 24. 08. 1997	Časovni presek, dne 31. 08. 1999
	Neznana občina	Alternativne oblike zdravljenja	1	1	1	1
		Bolnišnična zdravstvena dejavnost	1	1	1	1
		Dejavnost centrov za socialno delo	3	3	3	3
		Dejavnost dobrodelnih organizacij	8	8	8	8
		Dejavnost domov za starejše	1	1	0	0
		Dejavnost invalidskih organizacij	3	3	3	2
		Dejavnost posebnih socialnovarstvenih zavodov za odrasle	1	1	0	0
		Dejavnost socialnovarstvenih zavodov za usposabljanje otrok in mladostnikov	7	7	6	9
		Dejavnost varstveno delovnih centrov	1	1	4	2
		Druge oblike institucionalnega varstva	1	1	1	1
		Druge socialne dejavnosti	6	6	6	6
		Druge zdravstvene dejavnosti, d.n.	1	1	1	1
		Eksteritorialne organizacije in združenja	5	5	5	5
		Osnovna zdravstvena dejavnost	5	5	1	2
		Samostojne zdravstvene dejavnosti, ki jih ne opravljajo zdravniki	1	1	1	1
		Specialistična ambulantna dejavnost	1	1	1	2
		Veterinarstvo	5	5	9	7
		Zobozdravstvena dejavnost	1	1	1	1
		NEDEFINIRANO	11	11	11	11
	Neznana občina Total		63	63	63	63
	Grand Total		63	63	63	63

Vir: Avtor.

Izpisi temeljijo na neresničnih podatkih o številu podjetij po občinah in dejavnostih za obdobje od 01. 01. 1995 do 31. 08. 1999. Med seboj lahko primerjamo štiri časovne preseke, ki so zajeti v področnem podatkovnem skladišču in se razlikujejo med seboj po številu podjetij po dejavnostih.

- Izpis je izdelan s pomočjo področnega podatkovnega skladišča, opisanega v poglavju 5.4.2. in prikazuje število podjetij po dejavnostih in občinah v vseh štirih časovnih presekih. Iz tabele je razvidno, kako se vsebinsko spreminja njegova informacija, kakor se je spreminjala klasifikacija SKD in tudi podatki o podjetjih. Spremembe so označene s sivo barvo.
- Izpis je izdelan s pomočjo področnega podatkovnega skladišča, opisanega v poglavju 5.4.3. in prikazuje podatke o številu podjetij po dejavnostih SKD, ki je veljala na dan 24. 08. 1997, in občini v vseh štirih časovnih presekih.
- Izpis je izdelan s pomočjo področnega podatkovnega skladišča, opisanega v poglavju 5.4.3. in prikazuje podatke o številu podjetij po dejavnostih SKD, ki je veljala na dan 31. 05. 1995, in občini v vseh štirih časovnih presekih.

Če primerjamo vse tri izpise ob predpostavki, da dobro poznamo podatke in vemo, kakšna oblika izpisa nam ustreza, lahko ugotovimo, da v tej primerjavi izpis A. dobro prikazuje spreminjanje podatkov o podjetjih in spreminjanje klasifikacije SKD skozi čas. Tako je II. možnost zelo primerna za podajanje informacije o številu podjetij po dejavnostih za posamezen časovni presek, pri čemer lahko v vsakem trenutku predstavimo podatke za katerikoli razpoložljivi časovni presek. Pri podatkih o številu podjetij po dejavnostih po posameznih časovnih presekih ugotovimo spremembe, za katere obstajata dva vzroka. Sprememba klasifikacije SKD in sprememba dejavnosti za posamezno podjetje. Težava je v tem, da teh dveh vzrokov ne moremo enostavno

ločiti. Uporabnika lahko taka slika zavaja, misleč, da je zelo veliko podjetij, ki so med enim in drugim presekom zamenjala dejavnost, kar dejansko ni res.

Če želimo izdelati analizo časovnih vrst, moramo najprej izključiti spremembe v klasifikaciji SKD. Lahko bi napolnili za časovne vrste prirejeno tabelo mer, opisano v prejšnjem poglavju, in tako izdelali izpise, kot izgledajo izpisi B. in C. , izdelani po III. možnosti. Pri izpisu B. in C. pa nam tega koraka ni potrebno izvajati, ker je struktura tako izvedena, da samo z rahlo spremembo poizvedbe izdelamo ali izpis pod točko B. ali izpis pod točko C. Tako lahko iz podatkov lepo razberemo, da se je nekaj podjetij v preseku iz dne 24. 08. 1997 registriralo pod novo dejavnostjo. Pri tem moramo biti pozorni tudi na to, da določene dejavnosti v preseku iz dne 31. 05. 1995 niso bile v veljavi, ker smo predpostavili, da je dimenzija SKD veljavna na dan 31. 05. 1995 v C. izpisu in na dan 24. 08. 1997 v B. izpisu in so zato določene dejavnosti opisane z dejavnostjo 'Neznano'. Tako smo ločili vplive sprememb klasifikacije SKD na podatke o podjetjih po dejavnostih in občinah.

Pri odločitvi za II. možnost izdelave dimenzijske tabele bomo večkrat polnili tabelo mer glede na zahteve za izdelavo izpisov. Pri odločitvi za III. možnost pa bo več dela odpadlo na polnjenje dimenzijske tabele z ustreznimi podatki in morebitne preslikave dimenzijske tabele glede na zahteve za izdelavo zapisov. Opisane prednosti in slabosti v poglavjih 5.5. in 5.6. seveda potem določijo strukturo dimenzijske in merske tabele. Če strnemo prednosti in slabosti II. in III. možnosti:

II. možnost:

- + vsebuje vse kdajkoli veljavne opise dejavnosti, ki so bili vneseni;
- + zmožnost priklučitve drugih področnih podatkovnih skladišč, s tem lažja integracija in enostavno vzdrževanje, če smo uspeli predvideti dodatno tabelo mer za analizo časovnih vrst po dejavnostih podjetij;
- stalna pomoč vzdrževalca podatkovnega skladišča, da napolni prave podatke v podatkovno skladišče za analizo časovnih vrst po dejavnostih podjetij;
- generiranje novih zapisov v dimenziji zaradi lepopisnih popravkov na opisih dejavnosti, kar povzroča povečevanje dimenzije.

III. možnost:

- + možnost analize časovnih vrst brez pomoči vzdrževalca podatkovnih skladišč;
- + vnos lepopisnih popravkov brez generiranja novih zapisov v dimenziji;
- ne omogoča vpogleda v zgodovino klasifikacije;
- ne omogoča integracije z drugimi področnimi podatkovnimi skladišči;
- povečevanje števila dimenzij;
- težje vzdrževanje, ker zahteva natančno evidenco »starih« in »novih« opisov dejavnosti SKD in pomoč pri vzdrževanju take dimenzije, ko se pojavi težava z različnim številom zapisov pri »stari« in »novi« klasifikaciji.

Osebnostno se raje nagibam k II. možnosti, ker je le-ta bolj prilagodljiva, če pride do večkratnih sprememb v klasifikaciji SKD. Podatkovno skladišče je navadno zgrajeno za daljše časovno obdobje in v tem času analitiki lahko izvajajo veliko poizvedb, ki so

si med seboj zelo različne. Zato je enostavneje, da izdelamo posebno tabelo mer za časovne vrste, dimenzija pa se v tem primeru ne spreminja. Prav tako to potrjuje dejstvo, da je bila ta rešitev uporabljena tudi v okviru projekta RDB (Zadnikar, 2002).

6. Sklep

Potreba po podatkovnih skladiščih se danes kaže večinoma v podjetjih, kjer imajo opravka z veliko količino podatkov, ki jih je potrebno stalno obdelovati. Na informacijah, ki jih s tem pridobimo, temelji poslovno odločanje, izdelava strategij in razne poslovne analize. To pa so naloge, ki jih opravlja vsako podjetje pri svojem delovanju, zato menim, da bo uporaba podatkovnih skladišč v bodoče naraščala.

V statističnih uradih pa imajo skladišča tudi pomen reorganizacije, saj se pri izdelavi mnogokrat prevetri kak postopek, ki do sedaj ni bil dobro definiran. V praksi se je namreč pokazalo, da v fazi polnjenja podatkov v podatkovno skladišče prihaja do napak, ki jih v večini analitiki niso mogli predvideti, ker niso imeli vpogleda v podatke. SURS pri nas večinoma ne vzdržuje podatkov, temveč jih naroča pri raznih organizacijah, ki so za to pooblaščen. Te organizacije pa se z analizo podatkov ne ukvarjajo. Imajo izdelane aplikacije, ki izdelajo poročila za uspešnost posameznega poslovnega procesa, medtem ko za pregled po podatkih v mesečnih, četrletnih ali letnih presekih skrbijo statistični uradi. Statistični uradi za svoje analize potrebujejo še primerjave, ki navadno temeljijo na podlagi vzorca in anket. Tako lahko približno preverijo pravilnost podatkov v organizacijah in na podlagi tega podajo razne informacije, ki so na voljo javnosti.

Med posameznimi časovnimi preseki nastajajo v podatkih spremembe. Te spremembe morajo biti zavedene v podatkovnih skladiščih na poseben način, ki omogoča razne analize. Ker se podatkovni model sestoji iz dimenzij in tabel mer, je predvsem pomembno, kako so izdelane dimenzije, ki te spremembe lahko prikažejo. Dokler se podatki spreminjajo samo na spremenljivkah, ki imajo vlogo mer, so dimenzije lahko konstantne in enostavne. Ko pa pride do sprememb na spremenljivkah, ki imajo vlogo dimenzije, se pojavi problem počasi se spreminjajoče dimenzije. V svojem diplomskem delu sem analiziral možne rešitve, ki jih predlaga Kimball v svoji literaturi o podatkovnih skladiščih. Ugotovil sem, da so odločitve glede izbire strukture težke, saj sta obe obdelani možnosti uporabni v vseh smereh z določenimi omejitvami. Te omejitve pa določajo spremembe na spremenljivkah v vlogi dimenzij. Če so spremembe množične, je ustrežnejša rešitev po II. možnosti, drugače pa je zelo uporabna rešitev po III. možnosti. Ker je gradnja podatkovnih skladišč usmerjena na dolgi rok, je odločitev o ustrezni strukturi težavna, izdelava obeh možnosti pa v veliki meri potratna.

V praksi je uporabnost podatkovnih skladišč vse večja in zato je važno, da na podlagi dobre predstavitve dimenzijskega modela prototipne rešitve, ki vključuje tudi problem počasi se spreminjajoče dimenzije, pridobiti podporo uporabnika za nadaljnje delo. Uporabnik bo tako imel motiv dobro definirati zahteve, kar nam olajša izbiro ustrezne strukture objektov v podatkovnem skladišču. SURS namreč izdeluje poročila, ki

vključujejo tako analizo časovnih vrst kot tudi analizo podatkov, na katere vplivajo vse spremenljivke, vključno z njihovimi spremembami. Vse je odvisno od oblike informacije oziroma poročila, ki ga statistični urad posreduje javnosti.

Rešitev mora biti racionalna, kar se tiče količine podatkov, saj se vsebovanost le-teh v podatkovnih skladiščih postopoma veča in morajo biti učinkovito obdelovani. S tehtanjem med uporabnostjo na eni strani in učinkovitostjo ter racionalnostjo na drugi strani se izdelovalci in uporabniki podatkovnih skladišč srečujejo in odločajo vsak dan, kar se kaže v različnih strukturah podatkovnih skladišč.

7. Literatura

1. Čemažar Sandi et al.: KLASJE – STREŽNIK ZA STATISTIČNE KLASIFIKACIJE. Ljubljana: [URL: <http://surs.gov.si/klasje/streznik.htm>], 21. 01. 2003.
2. Grad Janez: Osnove baze podatkov in njene uporabe I. del. Ljubljana: Ekonomska fakulteta, 1982. 17 str.
3. Grošelj Bojan: Informacijski sistemi za podjetnike. Ljubljana: Visoka strokovna šola za podjetništvo, 1999. 174 str.
4. Hlavaty Marija et al.: Standardna klasifikacija dejavnosti. Ljubljana: Uradni list Republike Slovenije, 2002. 294 str.
5. Inmon Bill: Data Mart Does Not Equal Data Warehouse. [URL: <http://www.dmreview.com/master.cfm?NavID=198&EdID=1675>], 1999.
6. Inmon Bill: What a Data Warehouse Is Not. [URL: <http://www.dmreview.com/master.cfm?NavID=55&EdID=3606>], 2001.
7. Inmon William H.: Building the data warehouse, 2nd edition. New York: John Wiley & Sons, 1996. 401 str.
8. Kimball Ralph: The Data Warehouse Toolkit. New York: John Wiley & Sons, 1996. 388 str.
9. Kimball Ralph: The Data Warehouse Lifecycle Toolkit. New York: John Wiley & Sons, 1998. 771 str.
10. Klisura Mirnesa: Večrazsežnostni pogled na podatke. Ljubljana: Fakulteta za matematiko in fiziko, 2002. 62 str.
11. Lane Paul: Data warehousing guide, release 1. [URL: http://otn.oracle.com/docs/products/oracle9i/doc_library/901_doc/server_901/a90237.pdf], 2001.
12. Lucas Henry C. JR.: Information systems Concepts For Management. 5th edition, San Francisco: McGraw Hill, Inc., 1994. 660 str.
13. Pahor David: Leksikon računalništva in informatike. Ljubljana: Pasadena, 2002. 786 str.
14. Pavlin Branko: Podatkovno skladišče regionalnih statističnih podatkov SURS RDB in njegova uporaba. [URL: <http://www.sigov.si/zrs/obvestil/raden02/r50.doc>], 2002.
15. Raisinghani Vijay T. in Jha Rahul: Bitmap index implemetation. [URL: <http://www.it.iitb.ernet.in/~rvijay/dbms/proj/>], 2002.
16. Soet Gerco: PC-OWB-002-Slowly Changing Dimension. Oracle, [URL: http://otn.oracle.com/sample_code/products/warehouse/files/SlowlyChangingDim_OWb.pdf], 2001.
17. Zadnikar Mojca: Primerjava orodij ORACLE DISCOVERER in ORACLE EXPRESS. Kranj: Fakulteta za organizacijske vede, 2000. 126 str.
18. Zadnikar Mojca et al.: RAZVOJ PROGRAMSKE OPREME ZA UPRAVLJANJE S PODATKI STATISTIČNEGA URADA REPUBLIKE SLOVENIJE. št. SL – 9803.02.0001.03, Ljubljana: Aster d.o.o., 2002.

8. Viri

1. Grad, Škerlj, Vitorovič: Veliki angleško – slovenski slovar. elektronska izdaja, v 1.0, 1997.
2. The Thomson Corporation and DM Review: DM Review. [URL: <http://www.dmreview.com/>], 2003.
3. SDG Computing, Inc.: The Business Intelligence And Data Warehousing Glossary. [URL: <http://www.sdgcomputing.com/glossary.htm>], 2003.
4. The Oracle8i documentation CD-ROM, 2001.

9. Priloga

1. Kontrolne SQLLDR datoteke
2. Prehodna tabela PR_SKD strukture:

Ime stolpca	Tip stolpca	Dolžina	Komentar
NAZIV_ID	VARCHAR2	13	Identifikator naziva za vrednost spremenljivke
NAZIV	VARCHAR2	180	Naziv vrednosti spremenljivke
DATUM_OD	DATE	7	Datum začetka veljavnosti klasifikacije
VER	VARCHAR2	4	Verzija klasifikacije

Vir: avtor.

3. Tabela mer F_PODJETJA

Ime stolpca	Tip stolpca	Dolžina	Komentar
DATUM_ID	NUMBER	4,0	Nadomestni ključ časovnega preseka
D6_ID	NUMBER	8,0	Nadomestni ključ dejavnosti 6.nivo
ID_OB	NUMBER	3,0	Nadomestni ključ občine sedeža podjetja
MSO	VARCHAR2	10	Matična številka podjetja

Vir: avtor.

4. Dimenzijska tabela D_SKD

Namenjena za vnos SKD po II. možnosti. To se vidi predvsem po poljih DATUM_OD in DATUM_DO, kjer vnesemo obdobje veljavnosti klasifikacije. Polja s pripono _ID se lahko uporabijo tudi za alternativno sortiranje opisov dejavnosti.

Polje D6_ID vsebuje nadomestni ključ, ki se uporablja za povezavo dimenzijske tabele in tabele mer. To polje je povezano z SKD_ID6_SEQ zaporedjem, zato moramo to zaporedje vključiti pri vpisovanju novih vrednosti v tabelo.

Ime stolpca	Tip stolpca	Dolžina	Komentar
D6_ID	NUMBER	8,0	Nadomestni ključ dejavnosti
D1_ID_NAZIV	VARCHAR2	11	Identifikator naziva vrednosti spremenljivke 1.nivo
D1_NAZIV	VARCHAR2	180	Naziv vrednosti spremenljivke 1.nivo
D2_ID_NAZIV	VARCHAR2	11	Identifikator naziva vrednosti spremenljivke 2.nivo
D2_NAZIV	VARCHAR2	180	Naziv vrednosti spremenljivke 2.nivo
D3_ID_NAZIV	VARCHAR2	11	Identifikator naziva vrednosti spremenljivke 3.nivo
D3_NAZIV	VARCHAR2	180	Naziv vrednosti spremenljivke 3.nivo
D4_ID_NAZIV	VARCHAR2	11	Identifikator naziva vrednosti spremenljivke 4.nivo
D4_NAZIV	VARCHAR2	180	Naziv vrednosti spremenljivke 4.nivo
D5_ID_NAZIV	VARCHAR2	11	Identifikator naziva vrednosti spremenljivke 5.nivo
D5_NAZIV	VARCHAR2	180	Naziv vrednosti spremenljivke 5.nivo
D6_ID_NAZIV	VARCHAR2	11	Identifikator naziva vrednosti spremenljivke 6.nivo
D6_NAZIV	VARCHAR2	180	Naziv vrednosti spremenljivke 6.nivo
DATUM_OD	DATE	7	Datum začetka veljavnosti klasifikacije
DATUM_DO	DATE	7	Datum konca veljavnosti klasifikacije

Vir: Avtor.

5. Dimenzijska tabela D_OBCINE

Ime stolpca	Tip stolpca	Dolžina	Komentar
ID_OB	NUMBER	3,0	Nadomestni ključ občine
NAZIV_OB_ID	VARCHAR2	3	Šifra naziva občine
NAZIV_OB	VARCHAR2	50	Naziv občine

Vir: Avtor.

6. Dimenzijska tabela D_DATUM

Ime stolpca	Tip stolpca	Dolžina	Komentar
DATUM_ID	NUMBER	4,0	Nadomestni ključ za časovni presek
DATUM_ID_NAZIV	VARCHAR2	4	Šifra časovnega preseka
NAZIV	VARCHAR2	50	Naziv časovnega preseka
DATUM	DATE	7	Datum časovnega preseka

Vir: Avtor.

7. Dimenzijska tabela D2_SKD

Ime stolpca	Tip stolpca	Dolžina	Komentar
D6_ID	NUMBER	8,0	Nadomestni ključ 6.nivo
D6_ID_NAZIV	VARCHAR2	11	Identifikator naziva vrednosti spremenljivke 6.nivo
D6_NAZIV_OLD	VARCHAR2	180	Naziv vrednosti spremenljivke 6.nivo, stari naziv
D6_NAZIV	VARCHAR2	180	Naziv vrednosti spremenljivke 6.nivo, novi naziv

Vir: Avtor.

Namenjena za izdelavo podatkovnega skladišča po III. možnosti. Od dimenzijske tabele D_SKD se razlikuje po dodatnem polju za vpis starih tekstov ob spremembi naziva.

8. Zaporedje SKD_ID6_SEQ:

Potrebno je za vzdrževanje vrednosti nadomestnega ključa na zadnjem nivoju dimenzijske tabele. Te vrednosti se potem dodajo v tabelo mer namesto produkcijskega ključa, ki ga srečamo v relacijski podatkovni bazi

9. Funkcija VRNI_D_SKD_VPISAN

Namenjena za primerjanje zapisov v dimenzijski tabeli D_SKD s trenutno veljavno klasifikacijo v PR_SKD. Če je dimenzijska tabela kakorkoli neskladna s trenutno veljavno klasifikacijo, funkcija vrne, da tak zapis v dimenzijski tabeli D_SKD manjka. Funkcija je primerna za vpisovanje novih zapisov v dimenzijo.




10. Funkcija VRNI_D_SKD_UKINJEN

Namenjena za primerjanje zapisov v dimenzijski tabeli z novo veljavno klasifikacijo. Če je kak zapis v dimenzijski tabeli D_SKD, ki ga ni v novi veljavni klasifikaciji v tabeli PR_SKD, se le-ta zapis označi, da ga je potrebno ukiniti. Funkcija je primerna za vpisovanje datuma v polje DATUM_DO, ki pomeni, da je določeni dejavnosti potekla veljavnost.

11. Program FILL_D_SKD

Namenjena za prenos podatkov iz prehodne tabele, kjer se nahajajo podatki o trenutno veljavni klasifikaciji, v dimenzijsko tabelo. Pri tem je en zapis sestavljen tako, da vsebuje vseh 6 nivojev klasifikacije v eni vrstici, kar omogoča lažje združevanje podatkov v skupine in obratno poglobljeno iskanje podrobnosti po klasifikaciji. Določena orodja tako strukturo učinkovito izkoriščajo, če vnaprej definiramo ustrezno hierarhijo.

Vključuje tudi uporabo funkcij VRNI_D_SKD_UKINJEN in VRNI_D_SKD_VPISAN, ki nadzorujeta:

-  zapisovanje novih zapisov, ki jih stara klasifikacija še ne vsebuje,
-  ukinjanje zapisov, ki jih nova klasifikacija ne vsebuje in
-  popravljanje zapisov, ki so se v primerjavi s staro klasifikacijo spremenili

12. Za III. možnost sem uporabil enostavno kar INSERT stavek v SQL, ki prenese vse zapise iz prehodne tabele v dimenzijsko za en časovni presek. Ob vnosu sprememb pa bi potreboval proceduro za prenos tekstov v polje NAZIV_OLD in vpis novih nazivov v polje NAZIV. V mojem primeru sem dimenzijsko tabelo napolnil s pomočjo dimenzijske tabele iz II. možnosti in naslednjega SQL stavka:


```



Insert into d2_skd
select a.d6_id,
       a.d6_id_naziv,
       nvl(b.d6_naziv, 'NEDEFINIRANO'),
       a.d6_naziv
from
(
  select d6_id, d6_id_naziv, d6_naziv
  from d_skd
  where '31.3.1996' between datum_od and nvl(datum_do,sysdate)
) a,
(
  select d6_id, d6_id_naziv, d6_naziv
  from d_skd
  where '31.3.1995' between datum_od and nvl(datum_do,sysdate)
) b
where a.d6_id_naziv=b.d6_id_naziv(+)

```

13. Funkcija VRNI_K_SKD

Namenjena je za dodelitev nadomestnega ključa za polje, ki povezuje tabelo mer z dimenzijsko tabelo. Funkcija je zelo uporabna tudi v primeru, ko želimo primerjati več časovnih presekov med seboj v eni tabeli mer. To izvedemo enostavno tako, da vnesemo v parameter ip_datum datum, ki velja kot izhodišče analize časovnih vrst in funkcija pretvori vse šifre v vseh časovnih presekih samo s klasifikacijo, veljavno na ta datum.

Vhodni parametri so:

-  Ip_skd_id, vnese identifikator naziva 6. nivoja klasifikacije dejavnosti,
-  Ip_datum, vnese datum za izračun nadomestnega ključa dejavnosti.

Funkcija vrne nadomestni ključ za klasifikacijo dejavnosti na 6. nivoju in tako omogoči, da posamezne vrednosti spremenljivke pridobijo tiste nazive, ki jim pripadajo.

14. Funkcija VRNI_K2_SKD

Namenjena je za dodelitev nadomestnega ključa za polje, ki povezuje tabelo mer z dimenzijsko tabelo. Za razliko od funkcije vrni_k_skd le-ta nima parametra Ip_datum, in tako nima časovne odvisnosti. Spremembe se tako lahko vnesejo le s prepisom tekstov v dimenzijski tabeli.

Vhodni parametri so:


-  Ip_skd_id, vnese identifikator naziva 6. nivoja klasifikacije dejavnosti,

Funkcija se razlikuje od predhodne, tako da ne vsebuje parametra za vnos časovnega preseka, ker je dimenzija D2_SKD neodvisna glede na čas.

15. Program FILL_F_PODJETJE

Namenjen za polnjenje tabele mer F_PODJETJE s podatki za izbrani časovni presek. Vključuje funkcijo VRNI_K_SKD za dodelitev nadomestnih ključev v tabeli mer.

Vhodni parametri so:

-  Ip_datum, vnese datum časovnega preseka v obliki dd.mm.lilll.

16. Program FILL_F2_PODJETJE

Namenjen za polnjenje tabele mer F2_PODJETJE s podatki za izbrani časovni presek. Vključuje funkcijo VRNI_K2_SKD za dodelitev nadomestnih ključev v tabeli mer.

Vhodni parametri so:

 Ip_datum, vnese datum časovnega preseka v obliki dd.mm.llll.

17. Del podatkov v tabeli D_DATUM

1;0001;Časovni presek, dne 01.01.2000,1.1.2000
2;0002;Časovni presek, dne 31.05.1995,31.5.1995
3;0003;Časovni presek, dne 24.08.1997,24.8.1997
4;0004;Časovni presek, dne 31.08.1999,31.8.1999
9999;9999;Nedefiniran časovni presek,
5;0005;Časovni presek, dne 01.01.1995,1.1.1995

18. Del podatkov v tabeli PR_PODJETJA

4565874554;x99.999;1.1.1995;014
4565874557;q99.000;1.1.1995;014
4565874558;q99.000;1.1.1995;014
4565874560;q99.000;1.1.1995;014
4565874564;q99.000;1.1.1995;014
4565874570;q99.000;1.1.1995;014
4565874574;x99.999;1.1.1995;014
4565874579;x99.999;1.1.1995;014
4565874583;x99.999;1.1.1995;014
4565874588;x99.999;1.1.1995;014
4565874591;n85.110;1.1.1995;018
4565874592;n85.121;1.1.1995;018
4565874594;n85.122;1.1.1995;018
4565874598;n85.130;1.1.1995;018
4565874604;n85.141;1.1.1995;018
4565874608;n85.142;1.1.1995;018
4565874613;n85.143;1.1.1995;018
4565874617;n85.200;1.1.1995;018
4565874622;n85.311;1.1.1995;065
4565874625;n85.312;1.1.1995;065
4565874626;n85.313;1.1.1995;065
4565874628;n85.314;1.1.1995;065
4565874632;n85.315;1.1.1995;065

19. Del podatkov v tabeli PR_SKD

ABBREV_TEXT; CATEGORY_CODE; DESCRIPTOR_TEXT
SKD; A; KMETIJSTVO, LOV, GOZDARSTVO
SKD; A01; KMETIJSTVO IN LOV TER Z NJIMA POVEZANE STORITVE
SKD; A01.1; Pridelovanje kmetijskih rastlin
SKD; A01.11; Pridelovanje žit in drugih poljščin
SKD; A01.110; Pridelovanje žit in drugih poljščin
SKD; A01.12; Pridelovanje vrtnin, okrasnih rastlin, semen in sadik
SKD; A01.120; Pridelovanje vrtnin, okrasnih rastlin, semen in sadik
SKD; A01.13; Vinogradništvo in sadjarstvo
SKD; A01.131; Vinogradništvo
SKD; A01.132; Sadjarstvo
SKD; A01.2; Živinoreja
SKD; A01.21; Reja govedi
SKD; A01.210; Reja govedi
SKD; A01.22; Reja ovac, koz, konj, oslov, mul in mezgov
SKD; A01.220; Reja ovac, koz, konj, oslov, mul in mezgov
SKD; A01.23; Reja prašičev
SKD; A01.230; Reja prašičev
SKD; A01.24; Reja perutnine
SKD; A01.240; Reja perutnine
SKD; A01.25; Reja drugih živali
SKD; A01.250; Reja drugih živali
SKD; A01.3; Mešano kmetijstvo
SKD; A01.30; Mešano kmetijstvo
SKD; A01.300; Mešano kmetijstvo

20. Del podatkov v tabeli PR_OBCINE

1;999;Neznana občina

10. Slovar slovenskih prevodov tujih izrazov

Tuj izraz	Slovenski prevod
ad hoc query tool	orodje za dinamično (ad hoc) sestavljanje poizvedb
attribute	pridevek
bitmap index	indeks ali kazalo z bitno sliko
b-tree index	indeks ali kazalo z dvojiškimi drevesom
business process	poslovni proces
data mart	področno podatkovno skladišče
data mining	izkopavanje podatkov
data staging area	okolje za hrambo podatkov
data warehouse	podatkovno skladišče
decision support systems	sistemi za podporo odločanju
degenerate dimension	degenerirana dimenzija
dimension model	dimenzijski model
drill	vrtanje (po podatkih)
end user application	rešitev za uporabnika
end user data access tool	orodja za dostop do podatkov za uporabnika
executive information system	direktorski informacijski sistemi
flat file	ploska datoteka
index	indeks, kazalo
junk dimension	šund dimenzija
meta data	meta podatki, ali podatki o podatkih
modelling applications	aplikacija za modeliranje
MOLAP	večdimenzijski OLAP
multidimensional cube	večdimenzijska kocka
OLAP – on line analytic processing	sprotna analitska obdelava
operational data store	operativni podatkovni sistem
presentation server	predstavitveni strežnik
rapidly changing monster dimension	hitro spreminjajoča se velika dimenzija
rapidly changing small dimension	hitro spreminjajoča se majhna dimenzija
referential integrity	referenčna vez
ROLAP	relacijski OLAP
sequence	zaporedje
slowly changing dimension - SCD	počasi se spreminjajoča dimenzija
source system	izvorni sistem
surrogate key	nadomestni ključ
variable	spremenljivka