

LATVIJAS UNIVERSITĀTE

BAKALĀURA DARBS

RĪGA 2007

LATVIJAS UNIVERSITĀTE

Ekonomikas un vadības fakultāte

Ekonomikas informātikas katedra

**VALSTS IENĒMUMU DIENESTA UN CENTRĀLĀS
STATISTIKAS PĀRVALDES INFORMĀCIJAS
APMAIŅAS SASKARNE**

**INFORMATION EXCHANGE INTERFACE OF STATE
REVENUE SERVICE AND CENTRAL STATISTICAL
BUREAU**

BAKALaura DARBS

Autors: Vadības zinību bakalaura
studiju programmas

Ekonomikas informātika

studiju virziena

4. kursa students

Renārs Supe

St. apl. Nr. (Vadz 030496)

Darba vadītājs: (M. b. a.) lektors Kārlis Praudiņš

RĪGA 2007

Anotācija

Bakalaura darbs ir veltīts tādai mūsdienās aktuālai tēmai kā informācijas apmaiņa starp dažādām valsts institūcijām un to informācijas sistēmām. Darbā ir aprakstīta un parādīta Valsts Ieņēmumu Dienesta (VID) un Centrālās Statistikas Pārvaldes (CSP) informācijas apmaiņas saskarne, kas ir bāzēta uz datu noliktavu tehnoloģijas. Ir analizētas datu noliktavas iespējas un to nepieciešamība. Pietiekami liela uzmanība veltīta datu noliktavu rīku raksturojumiem un to izmantošanas iespējām. Datu noliktavu rīku praktiskā izmantošana ir parādīta pie aktuālas VID un CSP informācijas apmaiņas saskarnes izpētes un analīzes. Darba nobeigumā ir sniegti ieteikumi esošā informācijas apmaiņas procesa pilnveidošanai.

Darbs satur 3 nodaļas, literatūras saraksts satur 10 pirmavotus un tas ir izklāstīts uz 62. lpp.

Annotation

A bachelor thesis is dedicated to information exchange between public authorities and their information systems. Work describes and demonstrates information exchange interface of State revenue service and Central statistical bureau, which is based on data warehouse technology. Work describes and analyses Data Warehouse capabilities and necessity, enough attention is devoted for Data Warehousing tools, their characteristics and utilization possibilities. The practical usages of Data Warehousing tools are represented in research and analysis process of actual system – information exchange interface of State revenue service and Central statistical bureau. Recommendations to existent information exchange process development are given at the ending.

Work has 3 chapters, literature list contains 10 sources and work is laid out on 62. pages.

SATURS

Izmantotie saīsinājumi.....	6
Ievads.....	7
1. DATU NOLIKTAVU TEHNOLOĢIJA.....	9
1.1. Datu noliktavas nepieciešamība un iespējas.....	9
1.1.1. Datu noliktavas galvenās atšķirības no parastajām informācijas sistēmām	9
1.1.2. Datu noliktavas definīcijas	12
1.1.3. Datu noliktavas tipi.....	14
1.1.4. DN galvenās funkcijas un tehnoloģiskās prasības.....	15
1.2. Datu noliktavas komponentes.....	17
1.2.1. Datubāze	17
1.2.2. Transformāciju rīki	21
1.2.3. Analīzes rīki.....	23
1.3. Datu noliktavas realizācijas metodika	26
1.3.1. Biznesa prasību noteikšana.....	27
1.3.2. Dimensionālā modelēšana	28
1.3.3. Fiziskā līmeņa projektēšana	29
1.3.4. Datu transformēšanas projektēšana un izveide.....	29
1.3.5. Lietotāju programmatūras izstrāde	32
1.3.6. Uzstādīšana un darbināšana.....	33
1.3.7. Uzturēšana un atbalsts	34
2. VISPĀRĪGS VID INFORMĀCIJAS SISTĒMU APRAKSTS.....	35
2.1. VID Informācijas sistēmas un to galvenās komponentes.....	35
2.2. VID Datu Noliktavas Sistēma	36
2.3. Datu apmaiņa ar ārējām organizācijām	38
3. VID UN CSP INFORMĀCIJAS APMAIŅAS SASKARNES RELIZĀCIJA	39
3.1. VID DNS Datu bāzes modelis.....	39
3.1.1. Izmaiņas VID DNS DB	39
3.1.2. Informācijas apmaiņas analīzes sistēmas datu bāzes modelis	40
3.2. Transformāciju procesi	42
3.3. Analīzes sistēma	45
3.3.1. Analīzes sistēmas kopējās klases un to objekti	51
3.3.2. Analīzes sistēmas individuālās klases un to objekti	52
3.3.3. Analīzes sistēmas iebūvētie filtri.....	52
3.3.4. Analīzes sistēmas konteksti	53
3.4. Pārskati	54
3.5. Pievienojumprogramma DBF failu eksportam.....	58
Secinājumi un priekšlikumi	60
Izmantotās literatūras un avotu saraksts	62

Izmantotie saīsinājumi

Saīsinājums	Skaidrojums
ASIQ	Adaptive Server IQ
BO	Business Objects - pārskatu veidošanas un analīzes programmatūra
CA	Centrālais aparāts
CDB	Centrālā datu bāze
CDN	Datu noliktavas sistēmas datu bāzes nosaukums
CMIS	Centrālā muitas informācijas sistēma
CSP	Centrālā Statistikas Pārvalde
DB	Datu bāze
DBF	Fails dBASE III formātā ar paplašinājumu „dbf”
DBVS	Datu bāzu vadības sistēma
DN	Datu Noliktava
DNS	Datu noliktavas sistēma
EK	Eiropas komisija
ES	Eiropas Savienība
ETL	Extract Transform Load - “Iegūt Transformēt Uzlādēt” transformācijas rīku darbības pamatprincips
EU-SILC	Kopienas statistika attiecībā uz ienākumiem un dzīves apstākļiem
FK	Foreign key – datu bāzes tabulas svešatslēga saitei ar citu tabulu
HTTP	<i>Hyper Text Transport Protocol</i> – Tīkla Internet standartprotokols, kas nodrošina informācijas apmaiņu globālajā tīmeklī.
HTTPS	<i>Secure HyperText Transfer Protocol</i> – Protokola HTTP paplašinājums, kas garantē drošu datu apmaiņu globālajā tīmeklī. Katra ar S-HTTP pārsūtītā datne ir vai nu šifrēta, vai ar ciparu sertifikātu.
IeIeN	Iedzīvotāju ienākuma nodoklis
IS	Informācijas sistēma
LDB	Lokālā datu bāze
LR	Latvijas Republika
NIS	Nodokļu informācijas sistēma
NM	Nodokļu maksātāji
NMR	Nodokļu maksātāju reģistrs
NN	Nenulles vērtība. Datu bāzes tabulas lauks, kurš vienmēr ir aizpildīts ar vērtību vai konstanti.
OLAP	On-Line Analytical Processing – tiešsaistes analītiskā apstrāde
OLTP	On-Line Transaction Processing – tiešsaistes transakciju apstrāde
PVN	Pievienotās vērtības nodoklis
RIN	Reģionālā iestādes nodaļa
SOAP	<i>Simple Object Access Protocol</i> – uz XML failu apmaiņu balstīts protokols, kurš nosaka kā izsaukt Web servisu un saņemt no tā rezultātu
SQL	Structured Query Language – strukturēta vaicājumvaloda
VID	Valsts ieņēmumu dienests
WSDL	<i>Web service definition language</i> - uz valodas XML bāzēts datu formāts, kas domāts tīmekļa pakalpojumu aprakstīšanai UDDI reģistros. Aprakstiem ir jābūt neatkarīgiem no konkrētiem tīklu protokoliem un datu apmaiņas formātiem

Ievads

Mūsdienās, kad strauji attīstās informāciju tehnoloģijas, dators ir kļuvis par neaizstājamu rīku ikdienas darbā. Lai uzņēmumi un valsts institūcijas varētu veiksmīgi darboties mūsdienu datorizētajā laikmetā viens no priekšnoteikumiem ir uzņēmuma informācijas sistēma, kas nodrošina uzņēmumā notiekošo biznesa vai citu procesu reģistrāciju, atskaišu un pārskatu sagatavošanu. Uzņēmuma informācijas sistēmas dati glabājas vienā vai vairākās datu bāzēs. Tā kā uzņēmuma informācijas sistēmas ir veidotas, lai atbalstītu ikdienas operāciju veikšanu (jaunu ierakstu veidošana, esošo labošana un dzēšana) tad tām pamatā tiek izmantotas OLTP datu bāzes, kuras ir tendētas uz ātru un ērtu datu izmaiņu operāciju veikšanu. Pārskati un atskaites OLTP datu bāzēs ir vairāk statistiski – to saturs ir viendabīgs, un pārskatiem ir tendence attēlot informāciju ierakstu līmenī. Sarežģītākiem pārskatiem tiek aizņemti apjomīgi serveru resursi, tie pildās samērā ilgi un citiem sistēmas lietotājiem darbības ar datiem tiek apgrūtinātas. Tāpēc lieliem uzņēmumiem, kuros nepieciešams ne vien veikt datu rediģēšanu un skatīšanu, bet arī informāciju analizēt, ir nepieciešama datu noliktava (DN) - datorizēta sistēma, kas dod iespēju biznesa lēmumu pieņēmējiem ātri piekļūt pie nepieciešamās informācijas. DN ir instruments efektīvai liela apjoma datu atlasei un analīzei, kas dod iespēju ātri un ērti sagatavot visdažādākās atskaites un pārskatus par uzņēmuma darbību. Pēdējo gadu laikā DN gūst arvien plašāku pielietojumu ne tikai biznesā, bet arī zinātne, izglītībā, valsts un pašvaldību pārvaldē un citās sfērās. Autors savā bakalaura darbā apskatīs salīdzinoši netradicionālu datu noliktavas pielietojuma veidu - datu noliktavas izmantošanu informācijas apmaiņā ar ārējiem lietotājiem.

Šobrīd Latvijā dažādām valsts institūcijām ir atsevišķas savstarpēji nesaistītas informācijas sistēmas. Lai paaugstinātu datu kvalitāti un informācijas analīzi ir svarīgi nodrošināt datu apmaiņu starp šīm informācijas sistēmām, un līdz ar to arvien biežāk ir nepieciešams risināt dažādu informācijas sistēmu sadarbības problēmas gan organizāciju iekšienē, gan informācijas apmaiņā ar citām ārējām organizācijām. Viena no iespējām risināt šādas sistēmu sadarbības problēmas ir izmantot datu noliktavu tehnoloģiju.

Starp VID un CSP pastāv starpresoru vienošanās līgums, kas nosaka, ka nepieciešamības gadījumā VID sniedz CSP tai nepieciešamo informāciju un otrādi. CSP ir atbildīgā institūcija par projekta “Kopienas statistika attiecībā uz ienākumiem un dzīves apstākļiem (saīsināti EU-SILC)” ieviešanu Latvijā. Tās ieviešanu nosaka Eiropas Parlamenta un Padomes regula (EK) Nr.1177/2003 (2003.gada 16.jūnijs) par Kopienas statistiku attiecībā uz ienākumiem un dzīves apstākļiem (EU-SILC) un to papildinošās Eiropas Komisijas 7 ieviešanas regulas. Projekts EU-SILC lielā mērā ir orientēts uz respondentu – fizisko personu

gada ienākumu datu iegūšanu to komponentu līmenī bruto un neto izteiksmē. CSP nepieciešamo informāciju vēlas saņemt DBF failu formātā.

Bakalaura darba mērķis ir VID un CSP informācijas apmaiņas saskarnes analīze, izpēte un priekšlikumu izstrāde informācijas apmaiņas procesam.

Darba izpildei ir jāveic šādi uzdevumi:

- vides jeb sākotnējā izpēte – iepazīšanās ar pašreizējo situāciju un apstrādes veidiem, pieejamās informācijas apskate un analīze u.t.t.;
- pasūtītāja prasību noteikšana – kopīgi ar pasūtītāju tiek noteikti darba mērķi un vēlamais rezultāts, kā arī iespējamie realizācijas varianti;
- konstatēto problēmu definēšana – tiek noteiktas tās problēmas, kas ierobežo risinājumu un kopīgi ar pasūtītāju tiek saskaņoti to risināšanas veidi;
- iespējamo risinājumu piedāvāšana pasūtītājam – tiek izstrādāts projektējums un tas tiek saskaņots ar pasūtītāju;
- Izstrādāt ieteikumus (rekomendācijas) informācijas apmaiņas procesa pilnveidošanai.

Bakalaura darbs sastāv no trim nodaļām:

1. nodaļā ir dots vispārīgs DN apraksts, kā arī izpētīta un novērtēta DN tehnoloģija, sīkāk apskatot konkrētas DN komponentes.
2. nodaļā ir sniegts vispārīgs VID DNS esošās vides apraksts.
3. nodaļā ir sniegts VID un CSP informācijas apmaiņas saskarnes realizācijas apraksts.

1. DATU NOLIKTAVU TEHNOLOĢIJA

1.1. Datu noliktavas nepieciešamība un iespējas

Sarunu valodā par datu noliktavu sauc jebkuru informācijas tehnoloģijās (IT) izmantotu datu arhivēšanas sistēmu. Reizēm par DN sauc datu rezerves kopijas uz cietajiem datu nesējiem (piemēram, CD-ROM diskiem vai strimeru kasetēm). Taču pat IT speciālajā literatūrā datu noliktavai nav precīzas definīcijas, kura parādītu, kas tad īsti ir DN. Tāpēc vispirms pamēģināsim noskaidrot, kas tad īsti ir DN un ar ko tā atšķiras no citām informācijas sistēmām (IS). Bez tam tiks aprakstīts, kāpēc ir nepieciešams izveidot DN un kādas tehnoloģiskās prasības tiek uzstādītas DN risinājumiem.

1.1.1. Datu noliktavas galvenās atšķirības no parastajām informācijas sistēmām

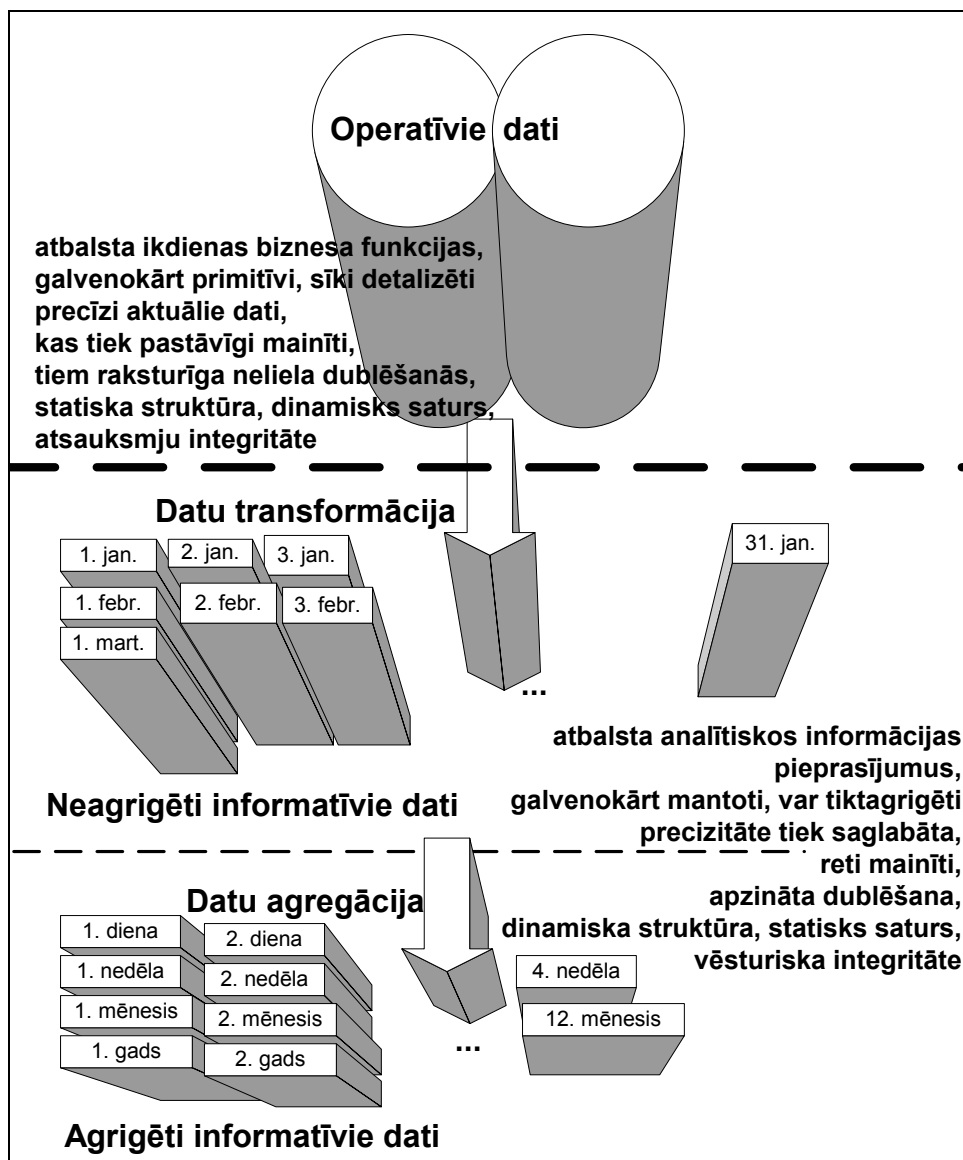
Klasiskajā IS risinājumā uzņēmumā eksistē viena liela datubāze (DB). Dati šajā datu bāzē nemitīgi tiek ievadīti. Šī DB parasti ir *On-Line Transaction Processing* (OLTP) tipa, tas nozīmē, ka tā tiek balstīta uz transakcijām, tās struktūra ir normalizēta un dati tiek klasificēti. OLTP DB ir paredzētas galvenokārt datu ievadišanas un labošanas funkcijām, nevis datu pieprasījumu apstrādes un analīzes funkcijām. Datu apjoms datu bāzē ar katru dienu tikai pieaug un ir nepieciešams regulāri palielināt servera jaudu, lai nodrošinātu pietiekami ātru datu pieprasījumu apstrādi (1, 2). Darbinieki, kuri veic datu ievadi, labošanu vai dzēšanu, veic arī datu pieprasījumus. Šie datu pieprasījumi ir viena vai vairāku ierakstu līmenī, līdz ar to šādu pieprasījumu iegūšana no DB, ne pārāk noslogo serveri. Taču ir arī cits personāls (uzņēmuma vadība, mārketinga speciālisti, analītiķi), kuram ir nepieciešami dažādi statistikas pārskati, kuri attēlo kopīgo informāciju – ļoti daudzu ierakstu līmenī, kas stipri noslogo DB serveri.

Pārnesot datus uz DN ir šādi svarīgākie ieguvumi (2, 4, 5):

- OLTP DB noslogojuma samazināšana;
- datu bāzu (avotu) datu integrēšana vienotā datu avotā (ja dati atrodas vairākās OLTP datubāzēs);
- OLTP DB “attīrīšana” no vecākiem datiem un līdz ar to - papildus slodzes noņemšana no šī DB servera;
- datu stabilitāte – datus DN datu bāzē nevar koriģēt (tikai OLAP DB gadījumā).

Parasti uzņēmuma vadībai, mārketinga speciālistiem un citam vadošajam personālam ir jāpieņem daudz svarīgus lēmumus, kuri iesaidos uzņēmuma turpmāko darbību. Šos

lēmumus viņi pieņem, balstoties uz analītiskiem datiem un uz intuīciju. Taču iegūt analītiskus datus no parastajām jeb operacionālajām informācijas sistēmām ir sarežģīti. Operacionālajās sistēmās tiek glabāti operatīvie dati. Esošos operatīvos datus ir nepieciešams pārvērst par analītiskajiem jeb informatīvajiem datiem. 1.1. att. ir parādītas atšķirības starp operatīvajiem un informatīvajiem datiem.

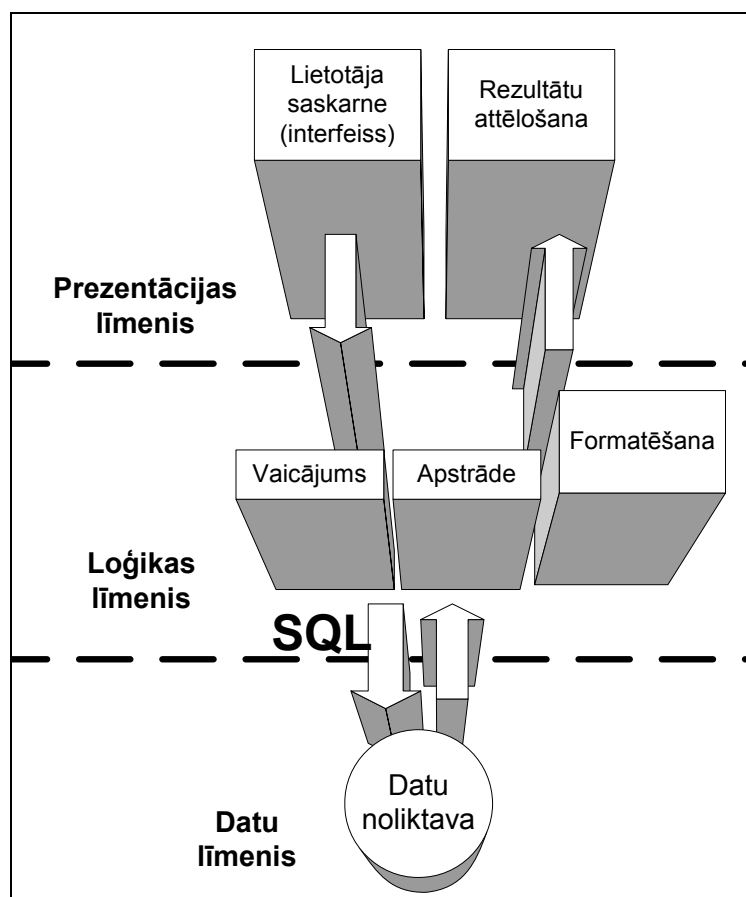


1.1. att. Operatīvo datu transformācija par informatīvajiem datiem (6, 60. lpp)

Būtiskākā DN atšķirība no parastajām IS ir informatīvo (nevis operatīvo) datu uzglabāšana atsevišķā datubāzē. Datu pārveidošanu no operatīvajiem uz informatīvajiem datiem ir iespējams veikt ar datu transformācijām. Ievietojot datus DN, datiem tiek piešķirtas analītiskas īpašības – tos ir iespējams analizēt nepieciešamajos griezumos, jo informatīvie dati ir saskaldīti līdz atsevišķu faktu līmenim. Par katru atsevišķu laika periodu (piemēram, dienu) tiek glabāta visa nepieciešamā informācija (piemēram, uzņēmuma ienākumi un izdevumi). Bieži vien šie dati tiek apkopoti pa noteiktiem laika periodiem (nedēļām, mēnešiem, gadiem),

lai atvieglotu pārskatu veidošanu par ilgākiem laika posmiem (skat. 1.1. att.). Dati var tikt sistematizēti ne tikai pa laika periodiem, bet arī pa citiem konkrētai analīzes specifikai atbilstošiem kritērijiem (ekonomiskiem, ģeogrāfiskiem, vai citiem). Šādus klasificētus analīzes kritērijus parasti sauc par dimensijām.

Otra būtiska DN atšķirība no parastajām IS ir iespēja vienkārši un ātri sagatavot jaunus, vai izmainīt esošus pārskatus dažādos griezumos, līmeņos un skatījumos. Nepieciešamo pārskatu iegūšanas process var aizņemt ievērojamu laiku, bez tam, atsevišķos gadījumos iegūstamie dati var būt ļoti slepeni un tos nedrīkst uzzināt neviens cits, kā tikai uzņēmuma vadība. Tādēļ DN ir speciāli analīzes rīki, ar kuriem darboties ir ļoti vienkārši pat bez speciālām priekšzināšanām un kuru apgūšana prasa ļoti īsu laiku. Sīkāk par analīzes rīkiem (1.2.3. nodaļā). Būtiskākā prasība DN analīzes rīkiem ir prezentācijas līmeņa nodalīšana no loģiskā līmeņa (1, 31. lpp.).



1.2. att. DN datu apstrādes un analīzes veids (1)

Kā parādīts 1.2. att., prezentācijas līmenis ietver sevī lietotāja saskarni, ar kuras palīdzību lietotājs uzdod pārskata veidošanai nepieciešamos datu atlasē nosacījumus (piemēram, pārskata perioda sākuma un beigu datumu), kā arī iegūtā pārskata rezultātu grafisko attēlojumu. Loģikas līmenī, pamatojoties uz atlasē nosacījumiem, tiek izveidots

vaicājums SQL valodā, kas tiek nodots izpildei DN datubāzē. Iegūtais rezultāts loģikas līmenī tiek apstrādāts un pēc formatēšanas atgriezts atpakaļ prezentācijas līmenī pārskata attēlošanai.

Bez divām jau minētajām DN galvenajām raksturiezīmēm, literatūras avotos (1, 6) tiek minētas vēl arī vairākas citas pazīmes, kas atšķir DN no pārējām IS:

- Tās tiek lietotas, lai vadītu un kontrolētu augsta līmeņa procesus.
- Dati ir pieejami vēsturiskā un precīzi definētā laika periodā.
- Uzsvars tiek likts uz datu apstrādes nevis datu ievades optimizāciju.
- Tās tiek lietotas vadības un speciālistu līmenī, lai pieņemtu lēmumus balstoties uz piedāvāto informāciju.

Datu analīze ar DN rīkiem sniedz šādas priekšrocības:

- analītisko jeb informatīvo datu iegūšana;
- ātras iemaņas datu pieprasījumu veidošanā;
- iespējas autonomi veikt datu analīzi, ļoti reti ir nepieciešams konsultēties pie speciālistiem.

1.1.2. Datu noliktavas definīcijas

Dažādos literatūras avotos ir sniegtas atšķirīgas DN definīcijas, tāpēc apskatīsim dažas no tām. Tanlara (1, 1. lpp.) dotā definīcija: **datu noliktava ir tehnoloģiju kombinācija, kas dod iespēju lietotājiem dinamiski radīt DB pieprasījumu, vadīt datu analīzi un noformēt rezultātus teksta vai attēla veidā.** Taču šī definīcija neietver sevī vairākus raksturojumus, kas tika minēti iepriekšējā nodaļā kā būtiskas DN iezīmes:

- esošo datu bāzu atslogošana;
- datu bāzu datu integrēšana;
- ērta piekļūšana datiem;
- analītisko, informatīvo datu iegūšana;
- datu stabilitāte.

Ne pārāk veiksmīgas ir arī citos literatūras avotos atrodamās DN definīcijas, piemēram (5): **“Datu noliktava ir no dažādām uzņēmuma sistēmām integrētu datu krātuve, kas izmantojama datu analīzei dažādos griezumos”**, vai arī koncerna “MicroLink” piedāvātā ļoti plašā definīcija: **Datu noliktava ir datubāze, kurā tiek uzkrāti resursu un finansu uzskaites dati, kas rodas veicot dokumentu apstrādi resursu un finansu uzskaites sistēmās, veicot to sagatavošanu veidā, kas ir piemērots datu tālākas apstrādes (piemēram, finansu atskaišu konsolidēšanai) un datu analīzes vajadzībām. Dati tajā tiek organizēti „informatīvo kubu” veidā, kas nozīmē, ka tie tiek sagrupēti pēc tālākās**

izmantošanas iespējām un piekārtojot tiem īpašības (dimensijas), kas ļauj tos salīdzināt, apkopot un vispārināt.

Arī lielajā terminu vārdnīcā ir atrodama LZA Terminoloģijas komisijas informācijas tehnoloģijas un telekomunikācijas apakškomisijas definīcija, ka DN ir “**dažādu uzņēmumu uzņēmējdarbības sistēmas, kurās ir savākto nozīmīgo datu centrālā glabātuve. Datu noliktava parasti tiek izveidota uzņēmuma serverī. Lai nodrošinātu datu analītisku apstrādi un saņemtu atbildes uz lietotāju vaicājumiem, dati par dažādu tiešsaistes transakciju apstrādi, kā arī dati no citiem avotiem tiek selektīvi atlasīti un sakārtoti datu noliktavas datu bāzē...**”. Arī šī definīcija ir izplūdusi un nekonkrēta.

Visi minētie trūkumi ir novērsti Inmona (6) dotajā DN definīcijā: **datu noliktava ir uz priekšmetu orientētu, integrētu, periodizētu stabilu datu kopums, kas palīdz vadības lēmumu pieņemšanas procesā.** Šī definīcija ir samērā sarežģīta un nav saprotama bez papildu komentāriem:

- uz priekšmetu orientēta - attiecas uz veidu, kādā tā ir organizēta, lai aprakstītu uzņēmējdarbības specifiku (tā saukto “biznesa loģiku”);
- integrēta – nozīmē, ka dati ir organizēti tā, lai nodrošinātu vienotu datu avotu;
- periodizēta – nozīmē, ka uzņēmējdarbības izpilde tiek mērīta ar laika mērauklu (diena, nedēļa, mēnesis) un dati tiek salīdzināti ar iepriekšējo laika periodu datiem un nepieciešamības gadījumā apkopoti pa noteiktiem laika periodiem (agrīgēti);
- stabilitāte, nozīmē to, ka dati vienreiz ievadīti DN, nevar tikt patvaļīgi mainīti un tiek atjaunoti ar stingri noteiktu regularitāti.

Tomēr, jāatzīst, ka šī ir vienīgā DN definīcija, kas precīzi atspoguļo visas būtiskākās DN iezīmes un ļauj DN atšķirt no tradicionālajām IS.

Būtiski ir arī saprast atšķirību starp DN un datuvi jeb datu tirgu (*data mart*). Saskaņā ar lielajā terminu vārdnīcā doto definīciju datuve ir datu glabātuve, kurā savākti operatīvie dati un dati, kas nepieciešami noteiktai lietotāju grupai. Šos datus var iegūt no uzņēmuma datu bāzes, datu noliktavas vai kāda cita specifiska avota. Datuves galvenais uzdevums ir nodrošināt, lai noteiktas lietotāju grupas šos datus saņemtu ērti lietojamā formā un varētu veikt ar tiem nepieciešamās darbības. Galvenā datuves atšķirība no DN ir lietotāju ierobežotās iespējas datu pieprasījuma izveidē (atšķirībā no DN, datuves lietotājam parasti tiek piedāvāts atlasīt datus pēc stingri nedefinētiem meklēšanas kritērijiem (6)). Datuvi var uzskatīt par DN attīstības formu, jeb arī par DN esošās informācijas izmantošanas veidu. Lai gan ir daudz gadījumu, kad datuves tiek izveidotas izmantojot operacionālās IS, tomēr labākais datuves izveides variants ir balstīt to uz DN datubāzi (1).

1.1.3. Datu noliktavas tipi

Nosacīti var izšķirt 3 dažādus DN tipus (1, 6.-8. lpp.):

- Finanšu DN;
- Mārketinga jeb tirdzniecības DN;
- Klientu īpašību izpētes (*behavioral*) DN.

Finanšu DN analizē biznesa darbības finansiālos rādītājus: tās satur finanšu vēstures "momentuzņēmumus", konkrēti ikdienas ieņēmumu un izdevumu datus. Finanšu dati parasti tiek atjaunoti ik mēnesi vai arī ik ceturksni vai citu pārskata periodu, kas sakrīt ar uzņēmuma finanšu kalendāru. Finanšu DN satur relatīvi mazu faktu skaitu. Bieži vien dolārs vai cita naudas vienība tiek izmantots kā ieņēmumu un izdevumu mērs. Dati glabājas noliktavā līdz ar to raksturojumiem, kuri parasti ir attiecināmi kā analīzes dimensijas. Dimensiju piemēri, kas apraksta šāda veida DN faktus, varētu būt sekojoši:

- a) konts (piemēram, ceļojumu izdevumi),
- b) struktūrvienība (piemēram, Mārketinga daļa),
- c) organizācija (piemēram, Uzņēmums),
- d) atrašanās vieta (piemēram, Valsts).

Finanšu DN lietotāju datu pieprasījumos parasti ir uzsvērts vairāklīmeņu finanšu datu salīdzinājums laika periodu griezumā. Tā kā lietotāji ir pieraduši pie tūlītējas atbildes uz veikto pieprasījumu, tad būtiska ir sistēmas ātrdarbība. Finanšu datu analīzes ir samērā vienvēidīgas, tāpēc lietotāju prasības un ar tām saistītie DN dizaina aspekti parasti ir vienkāršāki, nekā citiem DN tipiem.

Mārketinga DN tiek veidotas, lai ļautu lietotājiem novērtēt uzņēmējdarbības produktu vai pakalpojumu no daudzām atšķirīgām perspektīvām. Šim DN tipam ir raksturīgs daudz komplicētāks biznesa loģikas līmenis, jo ir nepieciešams analizēt katra darbinieka mārketinga aktivitāšu ietekmi uz ražotā vai piedāvātā produkta vai pakalpojuma noieta apjomiem. Šos datus ir jāatjauno daudz biežāk – parasti tas notiek reizi nedēļā, retāk – katras darbdienas beigās. Mārketinga DN ļauj lietotājiem analizēt datus dažādos hierarhiju līmeņos. Parasti hierarhijas tiek veidotas galvenajām dimensijām, tādām kā ģeogrāfiskā dimensija (piemēram: veikals < pilsēta(pagasts) < rajons < reģions < valsts), laika dimensija (piemēram: diena < nedēļa < mēnesis < ceturksnis < gads), vai arī produktu dimensija (piemēram: prece < veids < kategorija). Marketinga datu analīzes ir daudzveidīgas, tāpēc uzdoto datu pieprasījumu tipus ir grūti prognozēt. Daudziem rūpniecības kompānijām (piemēram, "POLAROID"), mazumtirdzniecības uzņēmumiem (piemēram, "Target Stores"), kredītiestādēm un mobilo telefonu operatoriem ir ļoti komplicētas marketinga DN, kas ir veidotas, lai atbalstītu to

aktivitātes sava tirgus sektora efektīvākā izmantošanā. Šie uzņēmumi cenšas uzlabot stratēģisko un taktisko lēmumu pieņemšanai nepieciešamās informācijas kvalitāti, lai palielinātu savu produktu (vai pakalpojumu) noietu dinamiskajā un ļoti konkurējošajā tirgū.

Klientu īpašību izpētes DN ir vērstas uz jaunu noieta tirgu apguvi un esošo klientu saglabāšanu. Šī tipa DN bieži vien satur detalizētu informāciju par ikvienu esošo vai potenciālo uzņēmuma klientu. Analītiskās DN visbiežāk ir sastopamas apdrošināšanas sabiedrībās, bankās un citās kredītiestādēs, lidostās, veselības aprūpes centros, kā arī citos uzņēmumos, kuriem ir nepieciešama iespēja pētīt un saprast savu klientu īpašības un izturēšanos, lai pēc iespējas labāk apmierinātu savu klientu vēlmes un līdz ar to nodrošinātu maksimālu piedāvātā produkta vai pakalpojuma noietu. Dati analītiskajās DN tiek atjaunoti regulāri, bet datu atjaunošanas biežums ir būtiski atkarīgs no to pielietojuma specifikas. Tā kā analītiskās DN satur detalizētu informāciju par daudzām personām, tad to apjoms parasti ir daudz lielāks, nekā finanšu un marketinga DN gadījumā. Piemēram, kompānijas "MasterCard" datu noliktavas apjoms jau 1999. gadā pārsniedza 1 terabaitu (1 000 gigabaitus) (1, 8. lpp.).

1.1.4. DN galvenās funkcijas un tehnoloģiskās prasības

Svarīgākā jebkura veida DN funkcija ir piegādāt institūciju vadītājiem un cietiem atbildīgajiem darbiniekiem analītiskus datus, kas atļauj pieņemt stratēģiskus un taktiskus lēmumus ne tikai balstoties uz personisko pieredzi un intuīciju, bet arī uz precīzu situācijas analīzi. Skatoties no tehnoloģiskā viedokļa, no DN veicamajām funkcijām var izdalīt četras būtiskākās (1, 9. lpp.):

- datu ieguve saskaņā ar sastādīto pieprasījumu;
- rezultātu statistiskā analīze;
- daudzdimensiju analīze;
- datu pieprasījumu rezultātu attēlošana.

Lai sekmīgi veiktu šīs funkcijas, DN ir jāapmierina daudzas tehnoloģiskās prasības, kuras Inmons (6, 168. lpp.) ir sistematizējis, sagrupējot pēc prioritātēm. Augstākās prioritātes grupā ir iekļautas šādas 4 prasības:

1. Spēja apstrādāt lielus datu apjomus ir pati svarīgākā tehnoloģiskā prasība, jo atšķirībā no parasto IS datubāzēm DN datu bāzē parasti tiek glabāts no terabaita (1000 gigabaitu) līdz pat petabaitam (1000000 gigabaitu) informācijas. Tik liels datu apjoms veidojās tāpēc, ka DN ir jāglabā detalizēti dati par ilgstošu laika periodu, saglabājot to precizitāti. Šī prasība parasti tiek apmierināta izmantojot elastīgu pāradresēšanu un datu indeksāciju.

2. Dažādu informācijas nesēju izmantošana nozīmē to, ka DN tehnoloģija nevar aprobežoties tikai ar tiešās piekļuves datu glabāšanas iekārtu (jeb serveru) izmantošanu. Modernajās DN tehnoloģijās tiek izmantotas gan ātrdarbīgie, bet dārgie datu nesēji (piemēram, darba stacijas operatīvā atmiņa un kešatmiņa, kas tiek izmantota galvenokārt DN prezentācijas līmenī iegūto pārskatu formatēšanai), gan lēnie, bet relatīvi lētie datu nesēji (piemēram, magnētiskā lente jeb “strimeri”, optiskie un magnetooptiskie diski, kas tiek izmantoti galvenokārt datu arhivēšanai DN datubāzes pārpildīšanās gadījumos, kā arī “BACKUP” rezerves kopiju veidošanai).
3. Iespēja datus viegli un brīvi pārraudzīt un indeksēt ir nepieciešama, lai nodrošinātu prasīto DN ātrdarbību. DN datubāzes pārraudzība un indeksēšana ir sīkāk aprakstīta 1.2.1. nodaļā.
4. Spēja saņemt un nodot datus izmantojot plašu tehnoloģiju loku ir nepieciešama gan datu ielādes procesā (pārņemot datus no operatīvajām datu bāzēm), gan pārskatu sagatavošanā.
Otrās (vidējās) prioritātes prasību grupā Inmons iekļauj vēl četras prasības:
 5. Datu izvietojuma iespēja dažādos fizikālajos līmeņos pēc izstrādātāju (programmētāju) ieskatiem.
 6. Datu paralēlās vadības iespēja.
 7. Meta datu līmeņa kontrole (meta dati ir dati par datiem, kas apraksta DN struktūru, saturu un indeksus), kas parasti tiek nodrošināta izveidojot tā sauktās “analīzes sistēmas”, kuras sīkāk tiks aprakstītas 1.2.3. nodaļā.
 8. Augsta līmeņa (“bagātas”) valodas interfeiss.
Pārējās prasības veido trešās (zemākās) prioritātes tehnoloģisko prasību grupu (6, 170.-177. lpp.):
 9. Efektīva un ātra datu ielādēšana DN, pārnesot tos no operatīvajām (izejas) datu bāzēm.
 10. Efektīva indeksu izmantošana ātras datu atlasē nodrošināšanai.
 11. Kompakts datu glabāšanas veids.
 12. Vienkāršo un salikto svešatslēgu izmantošana relāciju nodrošināšanai.
 13. Spēja efektīvi apstrādāt dažāda garuma datus (piemēram, DN datu bāzē preces vai pakalpojuma nosaukums var sastāvēt no dažiem simboliem, bet var arī veidot virkni jeb “stringu” vairāku simtu simbolu garumā).
 14. Iespēja atslēgt DB ierakstu slēgšanu jeb “loķēšanu”, kas ir nepieciešama tikai OLTP datu bāzēs labojot šo ierakstu vērtības (DN ieraksti parasti netiek laboti, bet gan dzēsti un izveidoti no jauna).

15. Ātra DN satura atjaunošana no rezerves kopijas, kas var tikt glabāta uz dažādu tipu datu nesējiem (piemēram, strimeriem, magnetooptiskajiem vai optiskajiem diskiem). Praksē ne vienmēr visas šīs prasības tiek ievērotas, jo vienmēr ir jāmeklē kompromiss starp esošo finansējumu un vēlamo rezultātu. Tāpēc, pirms uzsākt DN izveidošanu, būtu precīzi jānedefinē konkrētajai DN izvirzītās prasības, kuras parasti tiek iestrādātas DN projektējuma dokumentācijā, kas sīkāk tiks aprakstīta 1.3. nodaļā.

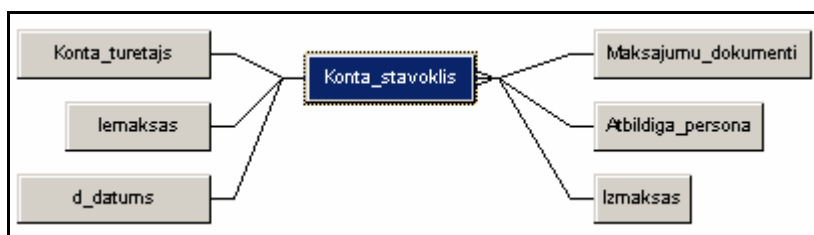
1.2. Datu noliktavas komponentes

DN tehnoloģija ietver sevī vairākas DN komponentes (jeb rīkus), kas darbu ar datu noliktavu dara ērtu un efektīvu. Ir daudz dažādi DN rīki, taču ir trīs galvenās komponentes, bez kurām DN nespēj veikt savus uzdevumus. Šīs trīs komponentes ir – datu bāze, transformācijas rīks un analīzes rīks. Katra no šīm komponentēm pilda savas noteiktas funkcijas, kuras tiks aprakstītas šajā nodaļā.

1.2.1. Datubāze

DN parasti tiek izmantota tiešsaistes analītiskā apstrādes jeb *On Line Analytical Processing* (OLAP) DB, kura būtiski atšķiras no parastām jeb OLTP datu bāzēm (darbs OLAP DB nav organizēts transakciju veidā, bet ir orientēts uz datu analīzi nolasīšanas režīmā; datu nolasīšana no OLAP DB notiek pa datu apkopojumiem, nevis pa vienam rakstam; datu labošana OLAP DB notiek dzēšot vecos un izveidojot jaunus ierakstus). Pēc autora domām OLAP datubāzes izmantošana DN nav obligāta prasība, jo atsevišķos gadījumos DN tiek izmantotas arī OLTP datubāzes. Galvenais OLAP datubāzes trūkums ir tas, ka nav iespējams nodrošināt nepārtrauktu datu aktualitāti. Tā kā operatīvie dati ik sekundi tiek papildināti un laboti, tad, pēc datu ielādes DN, tās saturs sāk ar vien vairāk atšķirties no izejas (operatīvās OLTP) datubāzes satura, un tiek sinhronizēts ar to tikai nākošās datu ielādes brīdī. OLAP datubāzes sinhronizācija ar OLTP tiešsaistes režīmā ir praktiski neiespējama. Tāpēc, gadījumos, kad DN tiek izvirzīta prasība pēc nepārtrauktas datu sinhronizācijas vai datu sinhronizācija ir jāveic biežāk, nekā reizi dienā, DN tiek izmantotas OLTP, nevis OLAP tipa DB (šādu izvēli savas DN izveide jau ir izdarījuši vairāki mobilo telefonu sakaru operatori). Tātad servera tips un DB vadības sistēma (DBVS) nav būtiskākie priekšnosacījumi DN datubāzei. Galvenās prasības ir DB tabulu izvietojums pēc zvaigznes (*Star*) vai sniegpārslas (*Snowflake*) modeļa (5) un efektīva datu indeksēšana (lai nodrošinātu pietiekami ātru datu atlasī, ir nepieciešams lai katras DN DB tabulas ikvienas kolonnas dati var tikt viennozīmīgi atlasīti tikai pēc indeksa, neanalizējot šīs kolonnas tekstuālo saturu (6)).

Zvaigznes modelis ir ļoti labs risinājums DN DB. Zvaigznes modeļa vispārējs piemērs dots 1.3. att. Tabula, kura atrodas centrā ir faktu tabula, apkārtējās ir dimensiju tabulas – pēc DN terminoloģijas.



1.3. att. Zvaigznes modeļa piemērs

Zvaigznes modelī dati ir attēloti kā fakti, ar kuriem tiek veikta analīze. Šie fakti tiek aprēķināti no operacionālo datu vērtībām, kuras tiek sasummētas pa dimensijām. Dimensija ir analīzes objekts jeb griezumus, kurā dati tiks analizēti. Zvaigznes modelim ir 2 līmeņi: faktu tabula un daudzas dimensiju tabulas. Faktu tabulām ir sekojošas pazīmes (4, 197.-198. lpp.):

- faktu tabula raksturo uzņēmējdarbības vai kāda cita biznesa procesa noteiktu rādītāju (piemēram, ieņēmumus vai izdevumus);
- fakti ir mērāmi (aprēķināmi) biznesa procesa lielumi, kuri tiek glabāti atsevišķos faktu tabulu laukos;
- faktu tabula ir savienota ar saistītajām dimensiju tabulām;
- faktu tabulām parasti nav primārās atslēgas (ja to pieļauj konkrētā DBVS), vai arī ir salikta primārā atslēga, kas sastāv no saistīto dimensiju tabulu primārajām atslēgām;
- ja tiek veikta summu agregācija, tad bieži vien faktu tabulās dati tiek dublēti (piemēram, ieņēmumi sadalījumā pa preču veidiem un ieņēmumi sadalījumā pa uzņēmuma filiālēm var tikt izvietoti divās atsevišķās faktu tabulās);
- faktu tabulas ir garas un šauras - tajās ir daudz rindu (ierakstu), bet maz lauku (kolonnu).

Dimensiju tabulām ir šādas pazīmes (4, 198.-199. lpp.):

- tās ir uzņēmējdarbības analīzes objekti, kuri attēlo dažādus griezumus, kuros var apskatīt un analizēt faktu tabulās attēlotos faktus;
- dimensiju tabulām parasti ir viena primārā atslēga;
- dimensiju tabulu datus parasti var sakārtot hierarhijās (piemēram, gads> ceturksnis> mēnesis> nedēļa> diena> stunda> minūte> sekunde), bieži vien šie dati tiek glabāti denormalizētā veidā (piemēram, pārskata perioda dimensijā var būt ieraksti, kuri atbilst dažāda garuma pārskata periodiem - mēnešiem, gadiem, ceturkšņiem);
- dimensiju tabulas ir mazas un platas, tajās ir maz rindu (ierakstu), bet daudz lauku (kolonnu);

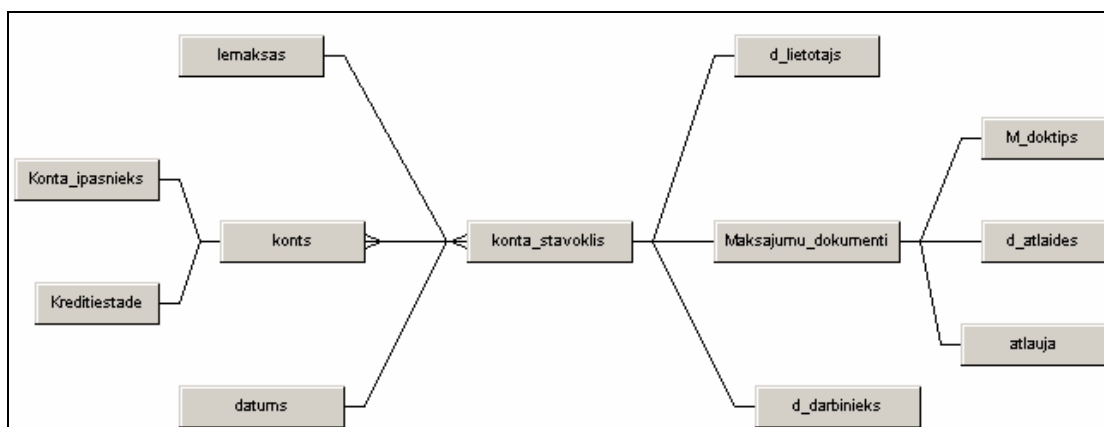
- viena no dimensijām vienmēr būs laika dimensija.

Zvaigznes modelis ir piemērots DN DB šādu iemeslu dēļ:

- tas nodrošina maksimālu elastību daudzdimensionālai datu analīzei;
- zvaigznes modeli atbalsta lielākā daļa DBVS;
- zvaigznes modeļa vienkāršība atvieglo sarežģītu datu analīžu veikšanu.

Tomēr bieži vien DN datu bāzē ir pārāk daudz dimensiju, kas izraisa problēmas datu pieprasījumu apstrādē. Lai šo problēmu risinātu var vairākas dimensijas apvienot vai arī veidot vairākas faktu tabulas pēc loģiskajām saiknēm ar dimensijām.

Otra problēma, ar ko nākas sastapties DB modeļa izveidē – dimensijā ir pārāk daudz datu un tā ir pārāk liela, kas rada problēmas ar datu pieprasījumu apstrādi. Lai risinātu šo problēmu, var dalīt dimensiju sīkākās dimensijās, vai arī veidot sniegpārslas modeli. Sniegpārslas modeļa vispārīgs piemērs dots 1.4. att.



1.4. att. Sniegpārslas modeļa piemērs

Kā redzams no attēla dimensijas var būt sīkāk sadalītas vēl citās dimensijās. Ar šo modeli tiek atrisinātas sarežģīto – pārāk lielo dimensiju problēmas.

Sniegpārslas modeļa priekšrocības salīdzinot ar zvaigznes modeli ir šādas:

- minimāla datu atkārtošanās katrā dimensijā;
- lielāka modeļa elastība – jaunas dimensijas var viegli pievienot, bez lielām faktu tabulas izmaiņām (izņemot gadījumus, kad tiek veidota sīkāka faktu detalizācija);
- noderīga sarežģītu dimensiju veidošanā.

Sniegpārslas modeļa trūkumi salīdzinot ar zvaigznes modeli ir šādi:

- ietver sevī vairāk saišu, tāpēc ir sarežģītāki datu pieprasījumi;
- iespējama slikta DB servera veiktspēja;
- sniegpārslas modeļi var nepatikt lietotājiem – lietotāju saskarne ar modeli, ja tas ir sniegpārslas, vajadzētu būt vienkāršam, lietotājam saprotamam.

Tomēr praksē bieži vien nākas saskarties ar ļoti sarežģītiem biznesa procesiem, kuriem atbilstošos datus nav iespējams izvietot ne zvaigznes, ne sniegpārslas modelim atbilstošā DB. Tāpēc bieži vien tiek veidoti sarežģītāki DB modeļi, kuri sastāv no vairākiem apvienotiem zvaigznes shēmas elementiem. Šādas DN DB tiek realizētas saliktā zvaigznes modeļa veidā. Saliktās zvaigznes modelī faktu tabula ir savienota ar saistītajām dimensiju tabulām, kuras savukārt ir savienotas ar jauno faktu tabulu.

Ļoti būtiska DN datubāzei ir pareiza un efektīva datu indeksēšana. Sīkāk apskatīsim galvenos indeksu tipus, kā piemēru izmantojot Sybase ASIQ serveri. Šā tipa serveri, tāpat kā vairums citu, izmanto bitu veida (Bit-Wise) indeksu glabāšanas tehnoloģiju. Šī indeksu glabāšanas tehnoloģija ir laba gan zemas, gan augstas kardinalitātes kolonnām. Zema kardinalitāte nozīmē, ka kolonnā ir mazāk par 1000 unikālām vērtībām, augsta – ja vairāk par 1000 unikālām vērtībām. Bit-Wise tehnoloģija nodrošina kolonnā atšķirīgo datu elementu pārveidošanu bitu veidā. Bit-Wise tehnoloģijas zemas kardinalitātes piemērs ir dots 1.5. att. Tabula kreisajā pusē rāda datus bez indeksa, tabula labajā pusē – Bit-Wise tehnoloģijas indeksu.

Krāsa	Zaļa	Zila	Sarkana
Sarkana	0	0	1
Zila	0	1	0
Zaļa	1	0	0

1.5. att. *Bit-Wise* indekss zemas kardinalitātes datiem

Kā redzams 1.5. att. Sybase ASIQ zemas datu kardinalitātes gadījumā no katras atšķirīgās kolonnas vērtības veido atsevišķu kolonnu un šajā kolonnā glabā informāciju bitu veidā – “1” šajā ierakstā glabājas pieprasītā krāsa, “0” – neglabājas.

Bit-Wise tehnoloģijas augstas kardinalitātes piemērs dots 1.6. att.

Indeksējami dati	Dati binārajā sistēmā				ASIQ indekss			
6	0	1	1	0	0	1	1	0
9	1	0	0	1	1	0	0	1
5	0	1	0	1	0	1	0	1
11	1	0	1	1	1	0	1	1
0	0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	0	1
7	0	1	1	1	0	1	1	1

1.6. att. *Bit-Wise* indekss augstas kardinalitātes datiem

1.6. att. dotajā augstas kardinalitātes gadījumā skaitļi tiek pārveidoti binārajā formātā. Sybase ASIQ katru bitu saglabā atsevišķā kolonnā un kad notiek datu atlase, tad pie datu salīdzināšanas pieprasīto vērtību pārveido binārajā formātā un līdz ar to ir zināms, kurā

kolonnā kādai bita vērtībai jāatrodas un tādēļ datu atlase notiek ļoti ātri. Tādā pašā veidā var tikt pārveidotas arī tekstuālas kolonnu vērtības. Izmantojot Bit-Wise tehnoloģiju dati gan zemas, gan augstas kardinalitātes gadījumā tiek glabāti vertikāli.

1.2.2. Transformāciju rīki

Izveidojot DN, datus ir nepieciešams pārnest no operacionālās (parasti OLTP) uz DN (parasti - OLAP) datu bāzi. Datu pārnesšanas procesā dati ir jāpārveido jeb jāpārtransformē. Transformācijas procesos ir jāņem tikai nepieciešamie dati, tiem ir jāveic pārbaudes un nepieciešamības gadījumos dati ir jākorrigē. Datu koriģēšana, galvenokārt, ir saistīta ar to, ka parastajās datu bāzēs pilnīga datu ievadīšana netiek prasīta un daudzi lauki, kurus būtu jāizmanto datu analīzē paliek neaizpildīti. Tā ir pati vienkāršākā datu transformācijas darbība – tukšumu aizpildīšana. Parasti tukšumi tiek aizpildīti ar konkrētām vērtībām, taču atsevišķos gadījumos, ir nepieciešams pēc citu lauku satura no kāda cita datu avota veikt datu piemeklēšanu un koriģēšanu.

DN struktūra būtiski atšķiras no parastās DB struktūras, tādēļ datus bieži ir nepieciešams arī grupēt pēc noteiktām pazīmēm. Tā kā dati DN tiek glabāti denormalizētā veidā, bet parastajās datu bāzēs normalizētā veidā, tad transformācijas procesos tiek veiktas denormalizācijas operācijas.

Datus DN var iegūt no vairākiem resursiem – vairākām datu bāzēm, kā arī no failiem. Taču dati dažādos resursos parasti atšķiras, bet tos ir nepieciešams apvienot vienā kopīgā bāzē. Transformācijas rīkiem jāveic dažādo datu pārveidošana uz vienotu modeli, šādas darbības parasti tiek veiktas, salīdzinot dažādu resursu datus un piemeklējot nepieciešamās vērtības pēc noteiktiem algoritmiem.

Datu transformēšanu var veikt divos veidos – var izmantot datu bāzu procedūras un var izmantot speciāli šim nolūkam paredzētus rīkus. Veikt datu transformēšanu, izmantojot datu bāzu procedūras ir sarežģīti un piedevām tiek noslogotas DB, kurās tiek veiktas šīs procedūras. Datu transformēšanas rīki nenoslogo pašas DB un datu pārveidošana notiek uz šo rīku serveriem. Tādēļ datu pārnesšanā un transformēšanā tiek izmantoti speciāli rīki, kuri tiek saukti par transformāciju rīkiem un angļiski šie rīki tiek saukti par “Extract Transform Load” (ETL) jeb “Iegūt Transformēt Uzlādēt”. Viens no ETL rīkiem ir Ascential DataStage.

DataStage tieši tika izstrādāts ar nolūku atvieglot DN aizpildes procesu. DataStage ir integrēts rīks, kas atbalsta datu iegūšanu, attīrīšanu, atšifrēšanu, transformāciju, integrāciju, agregāciju un uzlādēšanu (7). Lai gan DataStage sākumā tika izstrādāts tieši DN aizpildes procedūru veikšanai, šis rīks ir attīstījies un var pilnībā nodrošināt jebkuras datu migrācijas vai atjaunošanas procedūras. Šis transformācijas rīks ir ļoti attīstīts un ir spējīgs veikt ļoti

sarežģītus transformācijas procesus, nodrošinot un izmantojot vairāku procesoru priekšrocības.

DataStage ir šādas funkcijas, kuras palīdz un nodrošina efektīvu DN aizpildi:

- grafiski transformāciju veidošanas rīki, ar kuriem ir ērti veidot transformācijas shēmas, pēc kurām ir saprotams, kādas darbības veic transformācija;
- iegūst datus no jebkura daudzuma datu bāzēm un arī no jebkura veida datu bāzēm;
- elastīgs DB izmaiņām - jebkurā transformācijas būvēšanas brīdī var mainīt tabulu definīcijas;
- veic datu grupēšanu - pēc datu pieprasījuma ir iespējams veikt pieprasījuma pārveidošanu;
- transformē datus - DataStage satur daudz iebūvētu funkciju, ar kurām var veikt nepieciešamo datu pārveidošanu un ir iespējams veidot pašdefinētas funkcijas;
- veic datu ielādi.

DataStage ir izstrādāts kā klienta - servera aplikācija. Vairāki cilvēki vienlaicīgi var veidot transformāciju procesus. Uz klienta datora atrodas šādas DataStage komponentes:

- Projektētājs (*Designer*) - izmanto transformāciju vidi, ar kuru tiek veidoti transformāciju procesi;
- Vadītājs (*Director*) - vide, kurā tiek nodrošināta transformāciju procesu pārbaude, iepilānošana, izpilde un kontrolēšana;
- Pārvaldnieks (*Manager*) - vide, kura dod iespēju apskatīt un mainīt DataStage iekšējā repozitorija saturu;
- Administrators (*Administrator*) - vide, kurā tiek nodrošināta administratora darbību izpilde – jaunu lietotāju pievienošana, jaunu projektu izveidošana un pārvietošana.

Uz servera atrodas šādas komponentes:

- Repozitorijs – centrālā glabātuve (*Repository*), kurā metadatu veidā tiek uzglabāta visa nepieciešamā informācija, lai varētu aizpildīt datu noliktavu;
- Serveris (*Server*), kas izpilda transformāciju procesus;
- Instalācijas komplekts (*Package Installer*) ir vide, ar kuru ir iespējams pievienot – pieinstalēt, DataStage transformāciju procesus.

Detalizēts datu transformēšanas procesa projektēšanas un realizācijas apraksts ir dots darba apakšnodaļā 1.3.4.

1.2.3. Analīzes rīki

DN analīzes rīki lietotājam ir nepieciešami, lai datus pārvērstu informācijā, informāciju zināšanās un zināšanas plānos, kas novedīs pie peļņu nesošas, veiksmīgas uzņēmējdarbības (2, 6. lpp.). DN rīki nodrošina 4 veidu analīzi (1, 95.-97. lpp.):

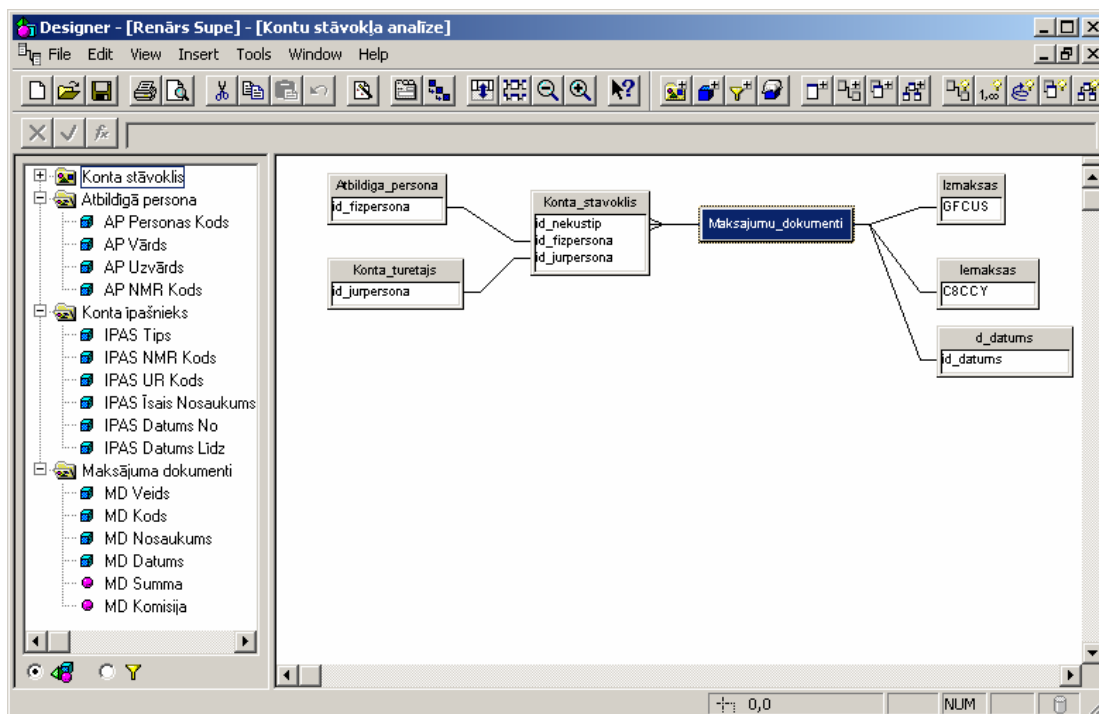
- vaicājumu apstrāde un pārskatu veidošana;
- daudzdimensionālā analīze;
- statistiskā analīze;
- datizrace - datu bāzes datu analīze, kas paredzēta apslēptu sakarību meklēšanai kādā datu grupā.

No datiem, kas atrodas DN datu bāzē, ir nepieciešams veidot pārskatus, veikt datu analīzi, datus meklēt sakarības, sakritības, lai to veiktu tiek izmantoti analīzes rīki. Analīzes rīku uzdevums ir pārveidot datus lietotājiem saprotamos terminos, lai lietotājs pats varētu veikt sev nepieciešamo datu analīzi un veidot pārskatus. Šim nolūkam tiek veidotas analīzes sistēmas, kurās DB dati tiek sadalīti loģiskās daļās un kurās atrodas lietotājam nepieciešamā informācija, lai būtu iespējams veidot pārskatus. Analīzes sistēmas ir nepieciešamas, lai metadatu veidā aprakstītu analizējamās sistēmas biznesa loģiku, kā arī tāpēc, ka DN dati var būt ļoti plaši un visus datus saliekot kopā var rasties problēmas atrast nepieciešamo informāciju. Pie pārskatu veidošanas ir iespējams izmantot vairākas analīzes sistēmas un veidot pārskatus, kas satur dažāda rakstura datus, piemēram, apvienot muitas datus ar nodokļu datiem. Viens no šādiem analīzes rīkiem ir BusinessObjects.

BusinessObjects ir integrēta vaicājumu, pārskatu veidošanas un analīzes programmatūra, kas ļauj piekļūt koplietošanas DB datiem tieši no darba stacijas, attēlot un analizēt šo informāciju BusinessObjects dokumentā (8). BO padara vienkāršāku piekļuvi datiem, jo lietotāji, sastādot pieprasījumus, strādā ar sev ierastiem jēdzieniem un elementiem tā vietā, lai veidotu vaicājumus datubāzei SQL valodā. Lai analizētu informāciju ar BO, nav nepieciešams pārzināt DB tabulu un kolonnu struktūru. Datu piekļūšanai ar BO informāciju var parādīt pārskatos vienkāršas tabulas veidā, vai kā saliktus dinamiskus dokumentus ar atlases spējīgām diagrammām. Izveidotos pārskatus ir iespējams saglabāt personīgai lietošanai, nosūtīt tos citiem lietotājiem, vai publicēt tos koplietošanas glabātuvē iespējami plašai informācijas aprītei.

Analīzes sistēmas parasti veido DN izstrādātāji, izmantojot BO projektētāju (*Designer*), vai kādu citu analogisku rīku. Izstrādātāji veido analīzes sistēmas lietotājiem pieejamas tā, ka tie var piekļūt DB interesējošajiem datiem no lietotāju saskarnes. Datubāzes lauki (tabulu kolonnas) analīzes sistēmā tiek attēloti kā objekti, no kuriem DN lietotāji var

veidot nepieciešamos pārskatus. Pēc autora domām, izstrādātāju galvenais uzdevums veidojot analīzes sistēmu ir izveidot objektiem loģiskus un viegli saprotamus nosaukumus un aprakstus (komentārus), kā arī pareizi sagrupēt tos pa piederības klasēm. BO analīzes sistēmas piemērs ir parādīts 1.7. att.



1.7. att. BusinessObjects analīzes sistēmas piemērs

Lietotāji, izmantojot analīzes sistēmas, veido BO vaicājumu, lietojot redaktoru, ko sauc par vaicājumu paneli (*Query Panel*). Vaicājumus veido, izvēloties objektus un filtrus no analīzes sistēmas. Objekti un filtri ir elementi, kas atbilst datu kopai no relāciju DB, un ir nosaukti terminos, kas attiecas uz biznesa situāciju. Katram objektam ir pievienots apraksts (komentārs), kurš paskaidro objekta būtību un izmantošanas veidu. Izpildot vaicājumu, BO savienojas ar datu bāzi un iegūst datus atlasītajiem objektiem un filtrē tos pēc izvēlētajiem filtriem.

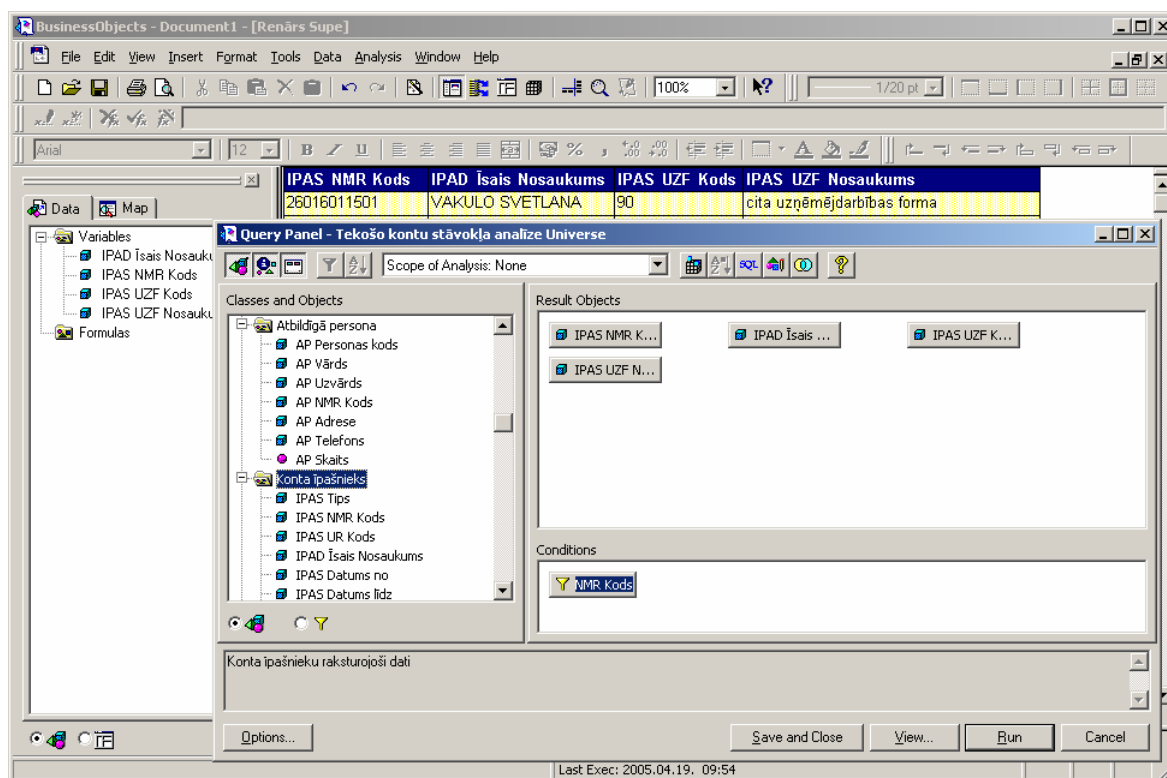
Pastāv trīs veidu objekti – dimensiju, detaļu un mēru objekti. Katrs objekta veids kalpo atšķirīgam mērķim:

- dimensiju objekti saņem datus, kas nodrošinās pārskata analīzes pamatu – tie satur simboliskā tipa datus (klientu nosaukumus, vietu nosaukumus, u.c.) vai datumus (gadus, ceturkšņus, u.c.);
- detaļu objekti parasti ir apvienoti vienā dimensiju objektā, tā nodrošinot papildus informāciju;
- mēru objekti satur aprēķinu, kuri tiek veikti ar DB datiem, rezultātus.

Lai atlasāmajai informācijai uzliktu ierobežojumus ir nepieciešami filtri. Visi filtri kalpo vienam mērķim – atlasīt nepieciešamo informāciju, nevis visu informāciju. Filtri iedalās:

- iebūvētajos – iebūvētos filtrus izveido izstrādātājs. Iebūvētie filtri iedalās vēl sīkāk:
 - a) parametrizētajos, filtri, kuriem var definēt atlasē parametra vērtību, startējot vaicājumu tiek izdota uzvedne, kurā jānedefinē filtra, piemēram, „Gads” parametra vērtība;
 - b) neparametrizētajos, filtri, kuriem nevar definēt atlasē parametra vērtību, startējot vaicājumu netiek izdota uzvedne, bet uzreiz tiek atlasīta konkrētā informācija. Parasti jau šo filtra nosaukums, piemēram, Fiktīvi NM, vai skaidrojums (Tiks atlasīti tikai fiktīvi NM) pasaka, kāda informācija tiks atlasīta.
- paša lietotāja definētajos, kurus lietotājs izveido no analīzes sistēmā esošajiem objektiem uzstādot tiem papildus nosacījumus, piemēram, Gads Equalto 2006.

1.8. att. parādīts pārskata veidošanas piemērs, izmantojot BO vaicājumu paneli.



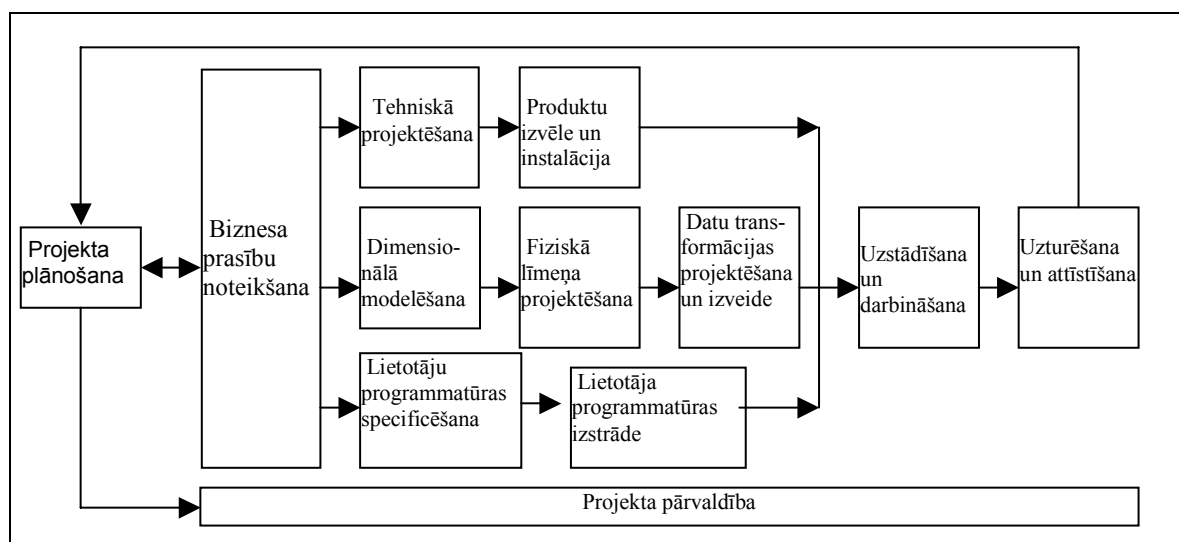
1.8. att. BusinessObjects pārskata veidošanas piemērs

Pārskatā kolonnas tiek sākotnēji izkārtotas atbilstoši vaicājuma paneļa sadaļā “*Result Objects*” ienesto objektu secībai (1.8. att. labajā pusē). Šos objektus lietotājam jāizvēlas no vaicājuma paneļa sadaļā “*Classes and Objects*” piedāvātajiem objektiem (1.8. att. kreisajā pusē). Nepieciešamības gadījumā jāizveido datu atlasē nosacījumi vaicājuma paneļa sadaļā “*Conditions*” (1.8. att. labajā apakšējā stūrī).

Ja vajadzīgie dati ir atlasīti, tos var parādīt dažādos veidos, piemēram, vienkāršā tabulā, šķērstabulā vai grafikā. Var veidot pārskatus, kas satur lielu daudzumu datu. Pārskatus var ērti izveidot un noformēt, lai parādītu interesējošo informāciju. Apskatei uz ekrāna vai izdrukāšanai, dokumentiem var pievienot attēlus, objektus un formātus. Pārskatu analīze ļauj risināt biznesa uzdevumus izmantojot datu vilkšanu un nomešanu (*drag-and-drop*), ieslēgt pārskatā aprēķinus, vai pārskatā veikt atlasītu detalizētākas informācijas iegūšanai (8).

1.3. Datu noliktavas realizācijas metodika

Šajā nodaļā tiks aprakstīta DN tehnoloģijas praktiskās realizācijas metodika. Literatūras analīze liecina, ka veiksmīgākā pieeja DN sistēmu izstrādes procesam ir pasaules praksē atzītā formālā izstrādes metodika *Business Dimensional Lifecycle* (9). Šīs metodes pamatideja ir sistēmas veidošanu aplūkot kā divpadsmit savstarpēji saistītu uzdevumu jeb etapu kopumu.

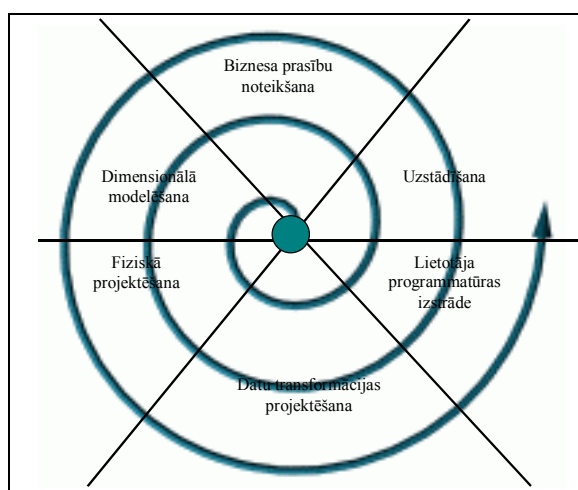


1.9. att. Datu noliktavu izstrādes metodika

Etapi jeb uzdevumi tiek variēti, atkarībā no konkrētā DN projekta. Pēc autora domām, no literatūrā (9) minētajiem etapiem būtiskākie ir septiņi pamatetapi:

- 1) biznesa prasību noteikšana;
- 2) dimensionālā modelēšana;
- 3) fiziskā līmeņa projektēšana;
- 4) datu transformācijas projektēšana un izveide;
- 5) lietotāju programmatūras izstrāde;
- 6) uzstādīšana un darbināšana;
- 7) uzturēšana un attīstīšana (atbalsts).

DN projekti vairāk kā citi programmatūras izstrādes projekti ir raksturīgi ar savu iteratīvo būtību, tāpēc vispārīgā gadījumā DN izstrāde ir jāuztver kā savstarpēji saistītu iterāciju kopums. Kopumā ir pilnīgi normāla situācija, ka ieguvumu biznesa procesa uzlabošanā īsti novērtēt un formulēt precīzākas prasības lietotāji spēj tikai līdz ar pirmās versijas darbināšanu, tādēļ jo svarīgi ir pēc iespējas ātrāk nodrošināt rezultātu atgriezenisko saiti ar lietotājiem. Vadošo izstrādātāju pieredze liecina, ka DN izmantošanas gaitā jaunas prasības rodas arī tad, ja sākotnējā prasību analīze veikta rūpīgi un kvalitatīvi. Šīs vajadzības tiek apkopotas, analizētas un tiek izstrādāta nākošā sistēmas versija, paralēli, protams, nodrošinot lietotāju darbu. 1.10. att. parādītas iterācijas DN sistēmas izstrādē.



1.10. att. Iterācijas DN sistēmas izstrādē

DN iterācija jeb cikls ir realizēts tad, kad ir nodrošināta un paveikta datu iegūšana no dažādām sākotnējām sistēmām, veikta datu integrācija, attīrīšana un transformācija un kad visas lietotāju vajadzības ir apmierinātas.

1.3.1. Biznesa prasību noteikšana

Praksē parasti ir sastopamas divas DN izstrādes shēmas:

- 1) DN veido uzņēmumā strādājošie IT speciālisti;
- 2) DN izveidi uzņēmums uztic profesionāļiem, kuri tiek pieaicināti “no malas”.

Lielos uzņēmumos parasti tiek algoti savi IT speciālisti, kuriem tad arī uztic DN izstrādi un ieviešanu. Nepieciešamības gadījumā vispirms vienu vai vairākus no šiem speciālistiem nosūta apmācībai vai pieredzes apmaiņai uz kādu no IT kompānijām, kurām ir liela pieredze darbā ar DN. Ja uzņēmumā nav pietiekoši augstas klases IT speciālistu, kuri varētu patstāvīgi izveidot DN, tad tiek pieaicināti speciālisti no vadošajām IT kompānijām. Šajā gadījumā parasti tiek noslēgts līgums par DN izveidošanu, kura būtiska sastāvdaļa ir biznesa prasību noteikšanas protokols, kā arī cita DN projekta dokumentācija.

Uzņēmuma biznesa prasības nosaka projekta virzienu, kā arī skar un ietekmē visas projekta sastāvdaļas, tāpēc ir svarīgi izprast organizācijā notiekošos biznesa procesus, un šie biznesa procesi, savukārt, kalpos par pamatu DN projektēšanai.

Biznesa analīzes galvenie mērķi ir:

- 1) noteikt atbalstāmo biznesa procesu kopu;
- 2) ieskicēt tālāko izstrādi prototipējamo nodevumu terminos;
- 3) noteikt prototipēšanas aktivitātēs iesaistāmo lietotāju klases;
- 4) piešķirt prioritātes izstrādājamās sistēmas prasībām;
- 5) precīzāk novērtēt projekta riskus (pirmā novērtēšana tiek veikta projekta uzsākšanas procesa laikā);
- 6) sagatavot stingru pamatu tālāk tehnisko risinājumu izstrādei;
- 7) noteikt nefunkcionālo prasību kopu, tai skaitā vienoties par uzturēšanas prasībām.

Iegūtās informācijas rezultātā DN speciālisti izstrādā prasību specifikāciju, kurā ietilpst:

- organizācijas raksturojums (mērķi, uzdevumi, struktūra, esošā situācija u.t.t.);
- projekta vispārīgs apraksts;
- biznesa prasības un biznesa loģikas modelis;
- datu plūsmu modelis;
- dimensiju modelis;
- projekta mērķi un uzdevumi;
- gatavo pārskatu formu un datu atlasēs nosacījumu apraksts;
- lietotāju grupas un pieejas tiesības.

Parasti biznesa prasību noteikšanas etapa dokumentācija tiek iesniegta DN izstrādes pasūtītājam apstiprināšanai, kā arī tiek noteiktas un saskaņotas veicamo uzdevumu prioritātes.

1.3.2. Dimensionālā modelēšana

Dimensionālā modelēšana ir loģiskas projektēšanas metode, kuru pielietojot, izveido dimensiju modeli - datu sakārtojumu tādā shēmā, kura nodrošina pieprasījumu ātru izpildi un ir orientēta uz nestandarta analītisku pieprasījumu veikšanu. Dimensiju modelis parasti sastāv no vienas faktu tabulas un vairākām dimensiju tabulām.

Dimensiju modeļa konstruēšana jeb dimensionālā modelēšana sastāv no četriem etapiem:

- DN matricas konstruēšanas;
- faktu tabulas shēmas konstruēšanas;

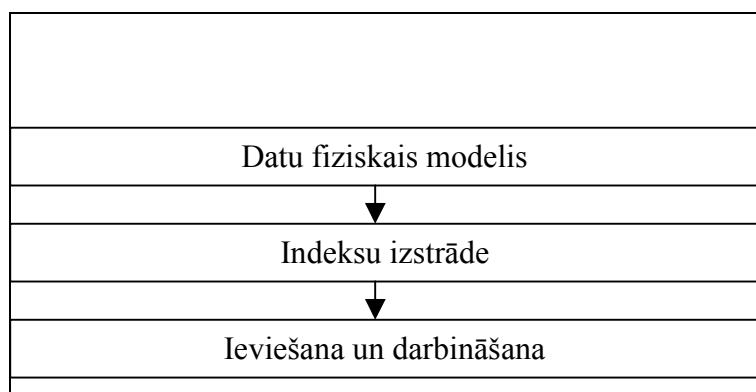
- faktu tabulas detalizētas izstrādes;
- dimensiju tabulu detalizētas izstrādes.

Dimensionālo modelēšanu noslēdz faktu tabulas un dimensiju tabulu fiziskās struktūras formālas definēšanas etaps, kurā definē arī formālos tabulu aizpildīšanas noteikumus.

Parasti šī etapa rezultātā tiek izstrādāts **dimensiju modelis**, kurš tiek iesniegts DN izstrādes pasūtītājam apstiprināšanai kā prasību specifikācijas dokumenta sastāvdaļa.

1.3.3. Fiziskā līmeņa projektēšana

Vispārīgais algoritms fiziskā līmeņa projektēšanai DN projektiem parādīts 1.11. att.:



1.11. att. Fiziskā līmeņa projektēšanas vispārīgā shēma

Reālā projektā šī shēma pēc nepieciešamības var tikt papildināta ar citiem etapiem, piemēram, agregāciju tabulu sastādīšanu, informācijas glabāšanas fizisko īpatnību izstrādi. Parasti IT speciālisti izmanto pašu izstrādātus un DN projektos praksē pārbaudītus pamatstandartus, ko izmanto fiziskā līmeņa projektēšanas laikā: datu bāzu objektu nosaukumu standartu un fizisko objektu nosaukumu un glabāšanas standartu. Šos standartus pēc nepieciešamības pielāgo konkrētā projekta vajadzībām.

Fiziskā līmeņa projektēšanu noslēdzošais etaps ir implementācija. Šajā etapā tiek fiziski izveidotas DB, tabulas un indeksi, veikta tehniskā nodrošinājuma konfigurācija un testēšana.

Šī etapa rezultātā tiks izstrādāts **datu fiziskā līmeņa modelis**, kurš tiek iesniegts DN izstrādes pasūtītājam apstiprināšanai kā prasību specifikācijas dokumenta sastāvdaļa.

1.3.4. Datu transformēšanas projektēšana un izveide

Datu transformēšana ir DN sistēmas pamats, viens no sarežģītākajiem un darbietilpīgākajiem etapiem. Transformācijas procesu nosacīti var iedalīt vairākos etapos:

- vispārīga transformācijas plāna izveide;

- dimensiju transformēšana un testēšana;
- faktu tabulas transformēšana un testēšana;
- pilna automātiska transformāciju cikla izstrāde un testēšana.

Transformācijas etaps sākas ar vispārīga plāna izveidi, kurā apraksta transformācijas procesa pamatetus:

- datu izgūšanu - datu avotus transformācijām;
- vispārīgos transformācijas procesus un algoritmus;
- datu ielādi - DB struktūru un aizpildīšanas noteikumus.

Piedāvājamo izstrādes rīku un platformu kontekstā izstrādes transformēšanas metodika ar *Ascential DataStage* būs šāda:

- datu avotu un mērķu meta datu imports vai definēšana;
- pārveidotāju un rutīnu definēšana;
- transformāciju izstrāde dimensijām un faktu tabulām:
 - 1) avotu moduļu definēšana;
 - 2) vairāku dimensiju mērķa moduļa definēšana;
 - 3) piemeklēšanu, agregāciju un sakārtojumu definēšana;
 - 4) avotu transformāciju uz mērķa shēmu (*mapping*) izveide.
- transformāciju izpildes secības un ielādes plāna definēšana un izpildes plānošana.

Datu avotu un mērķu meta datu imports vai definēšana

Ar *DataStage Manager* programmatūras palīdzību meta dati tiek importēti *DataStage* repozitorijā no avotu sistēmām. Ir iespējams importēt meta datus no šādiem avotiem:

- dažādu platformu datu bāzēm;
- XML tabulām;
- plakaniem failiem;
- *COBOL*, *Assembler*, u.c failiem.

Avotu meta datus ir iespējams definēt arī manuāli.

Pārveidotāju un rutīnu definēšana

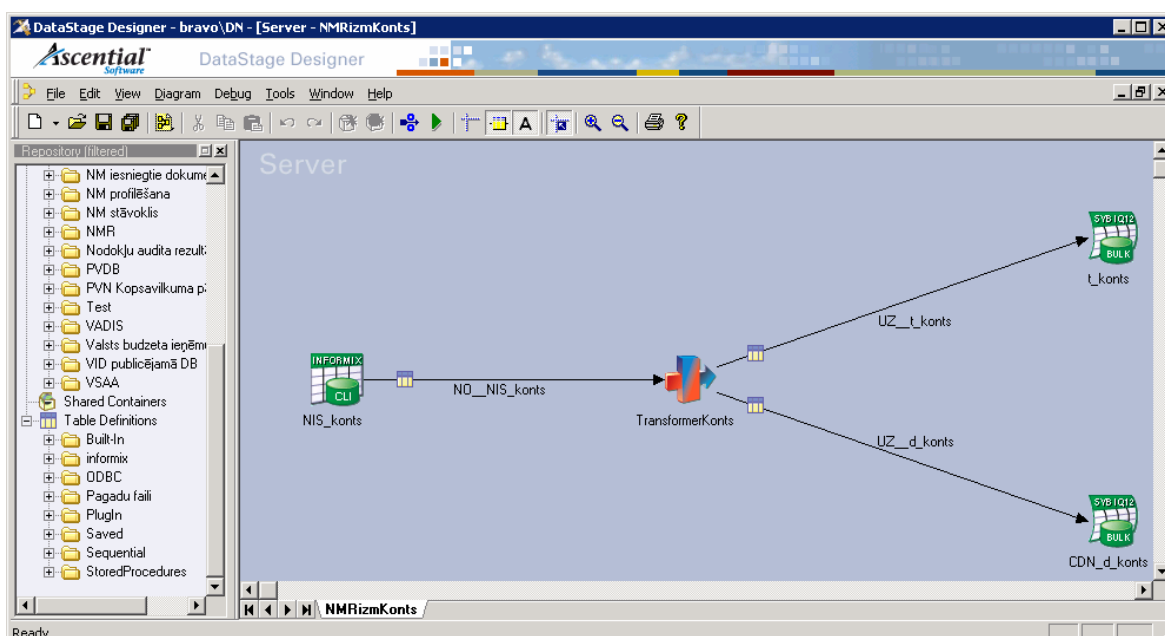
DataStage ietver plašu datu pārveidotāju un rutīnu klāstu datu transformēšanai, konvertēšanai un attīrīšanai. Nestandarta funkciju veikšanai jādefinē pārveidotāji, kas balstīti uz *DataStage* iekšējām funkcijām. Sarežģītu funkciju veikšanai, kur nepieciešama papildus programmēšana, jādefinē rutīnas.

Transformāciju izstrāde dimensijām un faktu tabulām

Dimensiju un faktu tabulām jāizstrādā atsevišķi transformācijas uzdevumi. Tā kā faktu tabulās ir lieli datu apjomi, tad katrai faktu tabulai jāizstrādā atsevišķs transformācijas uzdevums. Atsevišķs transformācijas uzdevums jāizstrādā arī faktu tabulas specifiskām dimensijām un kopējām dimensijām (piemēram, datuma un laika dimensijas). Transformācijas uzdevuma izpildes rezultāts ir sagatavoti *SybaseIQ* datu ielādes faili.

Avotu moduļu definēšana

Avotu meta dati tiek metadatu veidā glabāti *DataStage* repozitorijā un ir ērti pieejami no transformāciju izstrādes rīka *DataStage Designer*. Veidojot transformāciju, izstrādātājs izvēlas datu avotam atbilstošu ietilpinājumu un ielādē vajadzīgos meta datus (sk. 1.12. att.).



1.12. att. *DataStage Designer* datu avota definēšana

Datu avotam atbilstošajā ietilpinājumā arī jānorāda parametri, kas nepieciešami savienojuma izveidošanai ar datubāzi vai failu.

Vairāku dimensiju mērķa moduļa definēšana

Līdzīgi kā avotiem, mērķa moduļu definēšana notiek ar lietotāja grafiskās saskarnes palīdzību, ievietojot transformācijā mērķa moduļu ietilpinājumus, nokonfigurējot to parametrus un ielādējot meta datus no *DataStage* repozitorija.

Avotu transformāciju uz mērķa shēmu izveide

Jaunu transformāciju, kura apraksta veicamo operāciju virkni, veido, tās shēmu uzzīmējot ar *DataStage Designer* grafiskā redaktora palīdzību, norādot datu plūsmu no

avotiem uz mērķiem un iekļaujot tajā pārveidotājus un rutīnas. Pārveidotāji un rutīnas definē procesu, kādā veidā ieejas rindu kopums tiek apstrādāts, lai iegūtu noteikto izejas rindu kopumu.

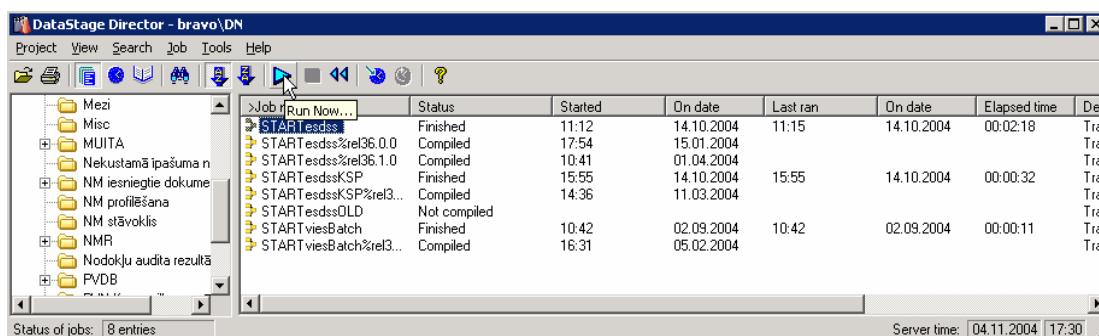
Transformācijas kompilēšanas brīdī tiek veikta automātiska verifikācija un izdoti attiecīgi paziņojumi par kļūdām.

Kad transformācijas uzdevums ir nokompilēts, jāveic tā validācija. Validācijas procesa laikā tiek izveidoti savienojumi ar datu avotiem un mērķiem, pārbaudīta meta datu atbilstība datubāzu struktūrām un failu eksistence.

Transformāciju izpildes secības un ielādes plāna definēšana un plānošana

Transformāciju izpildes secību nosaka izveidojot pakešuzdevumu. Pakešuzdevumu var izveidot ar *DataStage Designer* grafiskā redaktora palīdzību vai veicot programmēšanu. Pakešuzdevumā tiek norādīta transformāciju uzdevumu izpildes un datu ielādes secība, kā arī veikta izpildes statusu apstrāde.

Pakešuzdevumu vai transformācijas uzdevumu darbināšana un izpildes plānošana tiek veikta izmantojot *DataStage Director* (sk. 1.13. att.).



The screenshot shows the DataStage Director interface with a table of job execution details. The table has columns for Job name, Status, Started, On date, Last ran, On date, Elapsed time, and Description. The status of jobs is shown as 'Finished', 'Compiled', 'Not compiled', or 'Trar'.

Job	Run Now...	Status	Started	On date	Last ran	On date	Elapsed time	Des
START Leads		Finished	11:12	14.10.2004	11:15	14.10.2004	00:02:18	Trar
START Tesds%rel36.0.0		Compiled	17:54	15.01.2004				Trar
START Tesds%rel36.1.0		Compiled	10:41	01.04.2004				Trar
START TesdsKSP		Finished	15:55	14.10.2004	15:55	14.10.2004	00:00:32	Trar
START TesdsKSP%rel3...		Compiled	14:36	11.03.2004				Trar
START TesdsOLD		Not compiled						Trar
START viesBatch		Finished	10:42	02.09.2004	10:42	02.09.2004	00:00:11	Trar
START viesBatch%rel3...		Compiled	16:31	05.02.2004				Trar

1.13. att. Transformācijas uzdevumu darbināšana un izpildes plānošana

1.3.5. Lietotāju programmatūras izstrāde

Lietotāju programmatūras izstrāde sastāv no DN tradicionālajiem etapiem:

- analīzes sistēmu izstrādes;
- pārskatu izstrādes;
- testēšanas;
- dokumentēšanas.

Programmatūras izstrāde sākas ar analīzes sistēmu izstrādi - tiek definēta dimensiju hierarhija, izdalīti aprēķināmi un neaprēķināmi resursi utt. Nākamais etaps pēc metadatu izstrādes ir gatavo pārskatu izstrāde. Izstrādes rezultāti tiek testēti. Katrs DN izstrādes etaps tiek dokumentēts. Pēdējais posms starp DN sistēmas izstrādes pabeigšanu un nodošanu

darbināšanai ir lietotāja rokasgrāmatu izstrāde, kurās tiek aprakstīts darbs ar konkrēto *OLAP* rīku, analīzes sistēmas un gatavie pārskati.

Parasti šī etapa dokumentācija sastāv no testēšanas dokumentācijas un lietotāja dokumentācijas (rokasgrāmatām), kuras ietver:

- BO programmatūras aprakstu lietotājam nepieciešamā līmenī, ieskaitot pieslēgšanos, rekomendācijas, kā veikt analīzi, kā izmantot gatavos pārskatus, kā veidot un noformēt savus pārskatus u.c.;
- darbam nepieciešamu scenāriju aprakstus (piemēram, kā pieslēgties sistēmai, kā atvērt pārskatu);
- analīzes sistēmu aprakstus ar papildinformāciju par konkrētu objektu izmantošanu analīzei;
- informācijas iegūšanas piemērus, kuri balstīti uz reālām datu analīzes situācijām,
- pārskatu aprakstus.

Klasiskajā (operacionālajā) IS risinājumā parasti tiek paralēli uzturētas divas vides, viena no kurām tiek izmantota programmatūras testēšanai, bet otra – ekspluatācijai (tas ir, darbināšanai). Testa vidē programmētāji testē jaunās un labo esošās programmas, bet produkcijas (ekspluatācijas) vidē notiek reālie datu apstrādes procesi. Kā parāda DN izmantošanas pieredze, tad DN gadījumā testēšanas vide tiek izmantota ļoti reti, jo (6):

- 1) DN datubāzes apjoms ir tik liels, ka ir sarežģīti un dārgi uzturēt pat vienu (produkcijas) vidi, kur nu vēl divas (to var atļauties tikai pietiekama finansējuma apstākļos un tad, ja tas patiešām ir vajadzīgs);
- 2) DN izstrādes ciklisko iterāciju princips paredz iespēju nākamās iterācijas gaitā labot pirmajā iterācijā pieļautās kļūdas.

1.3.6. Uzstādīšana un darbināšana

DN uzstādīšana ir salikts, komplicēts process, kura laikā jāveic šādu etapu koordinācija un izpilde:

- sistēmas komponentu uzstādīšana un testēšana;
- programmatūras instalēšana un konfigurēšana darbam lietotāju datoros;
- lietotāju apmācība.

Pirms *OLAP* rīku uzstādīšanas lietotāju datoros, tiek pārbaudītas rīku darba spējas reālos apstākļos un veikti šādi pasākumi:

- lietotāja datora minimālās konfigurācijas noteikšana;
- instalācijas kontrole no vairākiem, dažādi konfigurētiem datoriem;

- programmatūras savietojamības pārbaude ar citām standarta programmām.

Sistēmas reālā izmantošana sākas pēc tam, kad ir izpildītas visas ieviešanas procesa nodošanas un pieņemšanas aktivitātes, ko nosaka *Nodošanas -pieņemšanas procedūra* un citi ar to saistītie dokumenti. Parasti sistēma ir nodota ekspluatācijā tad, kad ir nodotas visas komponentes, sekmīgi noritējusi akcepttestēšana, izveidots lietotāju apmācības grafiks, veikta apmācība un – vēlams - organizēta prezentācija (kopēja apskate) ar visu potenciālo sistēmas lietotāju piedalīšanos.

1.3.7. Uzturēšana un atbalsts

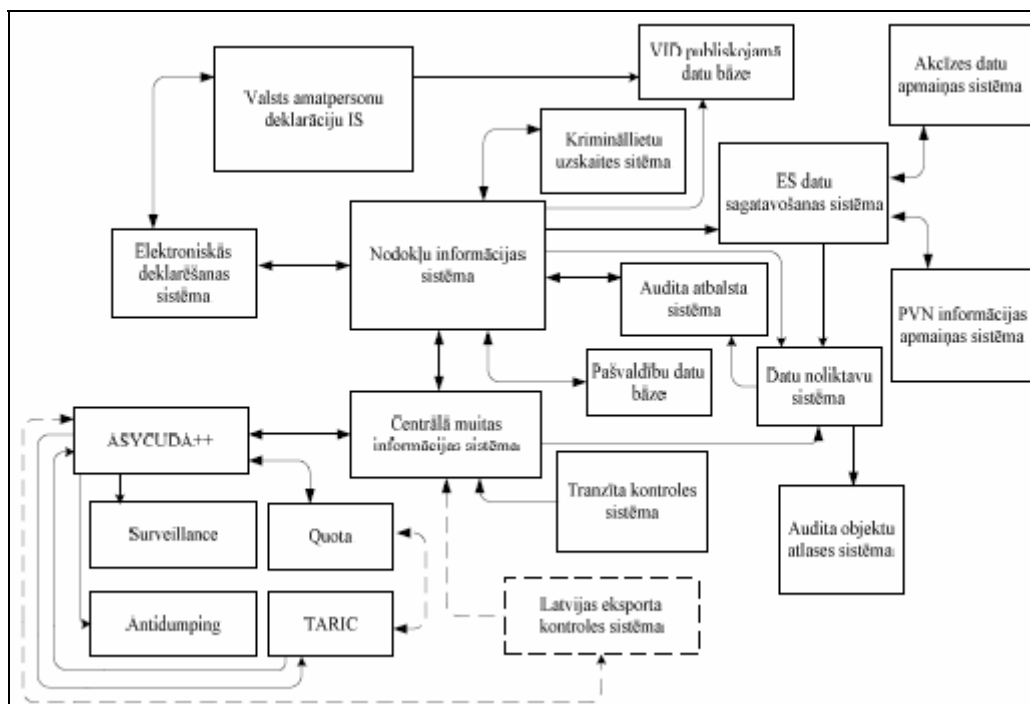
Sistēmas uzlabojuma vai kādus citus izmaiņu pieprasījumus uzturēšanas laikā parasti uzrauga un realizē atbilstoši konfigurācijas pārvaldības un izmaiņu vadības procesa prasībām līdzīgi tam, kā tas tiek veikts izstrādes laikā (tipiski regulējošie dokumenti ir *Konfigurācijas pārvaldības procedūra*, *Konfigurācijas vienumu identificēšanas vadlīnijas*, *Izmaiņu vadības procedūra* un *Ziņojumu reģistra uzturēšanas noteikumi*).

DN uzturēšana un atbalsts pēc autora domām ir pats būtiskākais DN realizācijas etaps, jo no tā efektivitātes lielā mērā ir atkarīgs, vai DN lietotāji būs apmierināti ar sistēmas darbību. Atšķirībā no klasiskās (operacionālās) IS lietotājiem, DN lietotāji pēc savas būtības ir pētnieki, un viņiem ir liela tieksme eksperimentēt. Bieži vien DN lietotāju pieeju darbam ar datiem var raksturot ar šādu frāzi: “Parādiēt man to, ko es domāju, ka gribu, un tikai tad es pateikšu, ko es patiešām gribu”. Tas nozīmē, ka tikai pēc DN pirmā etapa realizācijas, lietotāji, iepazīnušies ar reālajām DN iespējām, var korekti noformulēt savas prasības nākamā DN etapa realizācijai. Tāpēc DN uzturēšanas gaitā ir ļoti būtiski apkopot lietotāju kritiku un priekšlikumus, lai pareizi ielānotu nākamo DN realizācijas etapu (tā saukto iterāciju). Īpaši nozīmīgs šis etaps ir mārketinga un klientu īpašību izpētes DN gadījumā, jo šā tipa DN gadījumā to izstrādes sākumposmā ir grūti prognozēt ne tikai iespējamo datu pieprasījumu tipus, bet arī DB apjoma pieauguma tempus (1, 6.-8. lpp.).

2. VISPĀRĪGS VID INFORMĀCIJAS SISTĒMU APRAKSTS

2.1. VID Informācijas sistēmas un to galvenās komponentes

2.1. att. ir parādītas VID informācijas sistēmas un to komponentes.



2.1. att. VID informācijas sistēmas un to galvenās komponentes

Kā redzams 2.1. att. VID DNS nonāk dati no šādām VID IS:

- Nodokļu informācijas sistēmas (NIS);
- Centrālās muitas informācijas sistēmas (CMIS);
- ES datu sagatavošanas sistēmas.

ES datu sagatavošanas sistēma saņem datus no NIS un veic datu apmaiņu ar PVN informācijas apmaiņas sistēmu un Akcīzes datu apmaiņas sistēmu. Tāpat datu apmaiņa notiek arī starp NIS un CMIS.

Informācija, kas ir nepieciešama CSP tiek ievadīta NIS, no kuras datu bāzēm dati tālāk tiek integrēti VID DNS DB.

Dati no VID DNS nonāk Audita objektu atlases sistēmā un Audita atbalsta sistēmā, šīs sistēmas sniedz informāciju par VID IS lietotāju veiktajām darbībām.

2.2. VID Datu Noliktavas Sistēma

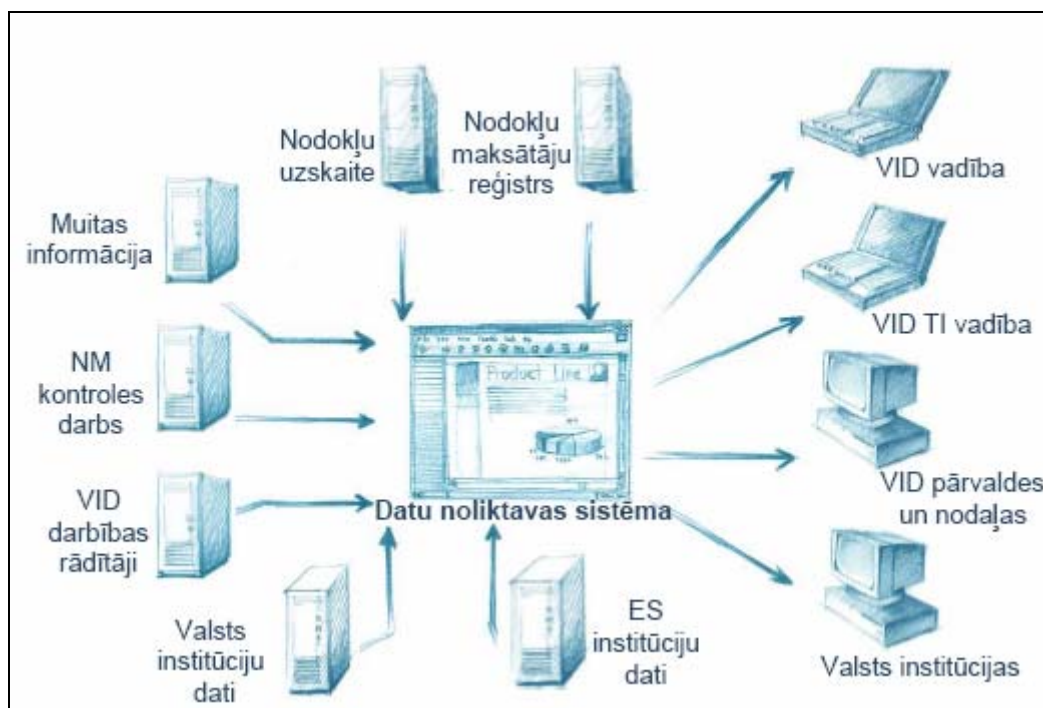
Sakarā ar datu apjoma pieaugumu lielajās VID IS un nepieciešamību redzēt vairāku sistēmu datus kopīgā skatā, 1999. gadā tika uzsākts VID Datu Noliktavas Projekts VID IS projekta ietvaros, šī projekta mērķis bija :

- Apkopot informāciju no dažādām VID datu bāzēm;
- Nodrošināt šīs informācijas analīzi gan ar gataviem pārskatiem, gan lietotāju veidotiem nestandarta pārskatiem;
- Sniegt atbalstu dažādu risku analīzes un kontroles uzdevumu veikšanā;
- Informācijas apmaiņa ar citām valsts un ES institūcijām.

2000. gadā tika uzsākta pirmo datuvju darbināšana:

- Valsts budžeta ieņēmumu informācijas bloks;
- Muitas informācijas bloks.

2.2. att. ir parādīts VID DN sistēmas pielietojums.



2.2. att. VID DN sistēmas pielietojums

VID DNS ir loģiski sadalīta informācijas blokos. Vairāku informācijas bloku realizācija notiek neatkarīgi un paralēli. Informācijas bloka realizācija satur šādus etapus:

- Prasību definīciju;
- Prototipa izstrādi;
- Prototipa darbināšanu un tās rezultātu apkopošanu;
- Pilnveidotas versijas izstrādi;
- Dokumentēšanu un lietotāju apmācību;

- Rūpnieciskās darbināšanas uzsākšanu;
- Informācijas bloka uzturēšanu.

VID DNS tiek izmantota šāda lietotāju programmatūra:

- BusinessObjects;
- BusinessQuery;
- WebIntelligence.

Pārskatu aktualizācijas programmatūra:

- BroadcastAgent

Transformācijas programmatūra (ETL procesi):

- IBM WebSphere DataStage;
- Microsoft Message Queuing.

Datu bāzes vadības sistēma:

- Sybase IQ12;
- Informix Dynamic Server.

Sistēma ir centralizēta ar kopīgu datu bāzi un repozitoriju. Uz darba stacijām ir instalēta Zero Administration BO versija. Transformāciju procesu serveris nodrošina nepieciešamās informācijas ielādi no avotu sistēmām. Lietotāju piekļuve DN sistēmai ir iespējama arī no Web pārlūkprogrammas, Centrālā aparāta, reģionālo iestāžu un TI lietotāji redz informāciju tikai par ‘saviem’ nodokļu maksātājiem.

VID DN sistēmas apjomu. Informācijas kopējais apjoms 190 Gb (Sybase IQ to uzglabā saspiestā veidā):

- Dimensiju un faktu tabulu skaits lielāks par 400;
- Ierakstu skaits apjomīgākajās tabulās līdz 270 milj.

Transformāciju procesi:

- 65 dažādi transformāciju procesi;
- Dažādos procesos iesaistītas 50 atšķirīgas datu bāzes.

Datuvju raksturojums:

- Ir izstrādātas 35 dažādas analīzes sistēmas (universe);
- Repozitorijā ir publicēti 900 koplietošanas pārskati, no kuriem lielākā daļa ir lietotāju – nodokļu inspektoru veidoti;
- Eksistē neskaitāms daudzums lietotāju veidotu pārskatu modifikāciju.

Lietotāju aktivitāte:

- Darba laikā vienlaicīgi tiek atjaunoti līdz 200 pārskatiem;
- Vismaz vienu pārskatu mēnesī sagatavo vairāk nekā 700 lietotāji;

- Mēneša laikā sagatavoto pārskatu skaits sasniedz 65 000.

Plašais vienā datu bāzē esošās informācijas apjoms, tās integrētība un lietotāju prasme ar šo sistēmu apieties rada arvien jaunus DNS pielietojumus:

- Risku analīze;
- Informācijas apmaiņa ar Valsts un ES institūcijām;
- Sinhronā vai asinhronā datu apmaiņa ar citām (t.sk. citu institūciju) IS, izmantojot Web servisu saskarnes.

2.3. Datu apmaiņa ar ārējām organizācijām

Šobrīd Latvijā dažādām valsts institūcijām ir atsevišķas savstarpēji nesaistītas informācijas sistēmas. Lai paaugstinātu datu kvalitāti ir svarīgi nodrošināt datu apmaiņu starp šīm informācijas sistēmām, un līdz ar to arvien biežāk ir nepieciešams risināt dažādu informācijas sistēmu sadarbības problēmas gan organizāciju (VID) iekšienē, gan informācijas apmaiņā ar citām ārējām organizācijām. Patlaban dažas valsts institūcijas piemēram Valsts Ieņēmumu Dienests (VID) jau saņem datus no LR Uzņēmumu reģistra, Iedzīvotāju reģistra un Valsts Zemes dienesta, savukārt dati no VID informācijas sistēmām tiek nodoti CSP, pašvaldību informācijas sistēmām, kā arī citām valsts institūcijām. Datu apmaiņai var izmantot dažādas tehnoloģijas gan datu noliktavas, gan Web servisu tehnoloģiju.

3. VID UN CSP INFORMĀCIJAS APMAIŅAS SASKARNES RELIZĀCIJA

VID un CSP informācijas apmaiņas saskarne ir bāzēta uz datu noliktavu tehnoloģijas, izmantojot šādus datu noliktavu rīkus:

- OLAP datu bāze – Sybase IQ 12.5.;
- transformācijas rīks – Ascential DataStage 7.0.1;
- analīzes rīks – BusinessObjects 5.1.8;

Informācijas apmaiņas saskarnes realizācijai ir izstrādāti:

- informācijas apmaiņas analīzes sistēmas (datuves) datu bāzes modelis;
- transformācijas procesi, kuri veic CSP nepieciešamās informācijas tabulu aizpildi;
- analīzes sistēma (datuve), kura satur CSP nepieciešamos datus;
- standarta pārskati,
- *Micrsoft VisualBasic* programma, kas pārveido ar analīzes rīku BO veidotos standarta pārskatus CSP nepieciešamajā failu formātā (DBF).

3.1. VID DNS Datu bāzes modelis

Autors iepazīsies ar CSP nepieciešamās informācijas pieprasījumu, veica izpēti VID DNS DB un secināja, ka ir nepieciešams veikt izmaiņas esošo tabulu struktūrā, papildinot tās ar jauniem laukiem.

3.1.1. Izmaiņas VID DNS DB

Esošajās VID DNS DB tabulās, kurās tika veiktas izmaiņas tiks aprakstīti tikai jaunizveidotie lauki.

Izmaiņas VID DNS DB tabulā f_iingd2_saturs

VID DNS DB f_iingd2_saturs tabulā glabājas dati par fizisko personu (rezidentu) ārvalstīs gūtajiem ienākumiem. Brīdī, kad no CSP tika saņemts informācijas pieprasījums, tabulā f_iingd2_saturs nebija CSP nepieciešamo datu par fizisko personu ārvalstīs gūto ienākumu veidiem un ienākumu izmaksas datumu, tādēļ VID DNS DB tabulā f_iingd2_saturs tika izveidoti divi jauni lauki (3.1. tabula).

Izmaiņas VID DNS DB tabulā f_iingd2_saturs

Lauka fiziskais nosaukums	Lauka formāts	Lauka apraksts
id_ienakveids	Integer, FK NN	Atsauce uz ienākumu veidu dimensiju d_ienak_veids.
datums_izm	Date, NN	Ienākumu izmaksas datums

3.1.2. Informācijas apmaiņas analīzes sistēmas datu bāzes modelis

Analīzes sistēmu datu bāžu modeļi tiek veidoti *Business Objects* izstrādātāja līmenī *Designer*. Autors izstrādātāja līmenī ir pilnveidojis informācijas apmaiņas analīzes sistēmas datu bāzes modeļi, kas sastāv no analīzei nepieciešamajām datu bāzes tabulām un saitēm starp tām. Analīzes sistēmas datu bāzes modeļa īpatnība ir tā, ka ne vienmēr ir nepieciešams realizēt visas tabulu savstarpējās saites, kādas tām ir VID DNS DB, ir jāatstāj tikai tās saites, kuras būs nepieciešamas, tāpat var norādīt saites, kādas VID DNS DB nemaz nav.

Informācijas apmaiņas analīzes sistēma ir veidota uz VID DNS DB tabulu pamata pēc saliktās zvaigznes modeļa, līdz šim tajā tika izmantota 31 VID DNS DB tabula (skat. 3.2. tabula), bet tā kā CSP bija nepieciešama informācija par IeIeN gada deklarāciju D veidlapu un pielikumu saturu, informācijas apmaiņas analīzes sistēma datu bāzes modelis tika papildināts ar 8 VID DNS DB tabulām (skat. 3.3. tabula). Katra no šīm tabulām satur informāciju par kādu no IeIeN gada deklarācijas veidlapām vai pielikumiem.

Informācijas apmaiņas analīzes sistēmā izmantotās VID DNS DB tabulas

Tabulas fiziskais nosaukums	Apraksts
d_nodokl_maks	Nodokļu maksātāju dimensija
d_vid_ti	VID teritoriālo iestāžu dimensija
d_parsk_per	Laika perioda dimensija
d_datums	Datumu dimensija. Atsauce uz laika perioda dimensiju
d_ienak_veids	Ienākuma veida klasifikators
d_dokuments	Dokumenta dimensija
d_dok_rinda	Dokumenta rindu dimensija. Dokumenta rinda ietver unikālu dokumenta rindas identifikatoru, kā arī informāciju par dokumentu rindiņu
ds_pils_pag	Teritoriālā dimensija
d_adrese	Teritoriālā adrešu dimensija
d_telefons	Telefonu dimensija
d_likvid_iesm	Likvidācijas iemeslu dimensija
ds_nace	Saimnieciskās darbības veidu klasifikators

Tabulas fiziskais nosaukums	Apraksts
d_patents	Patentu un reģistrācijas apliecību dimensija
f_darbveids	Fakti par saimnieciskās darbības veidiem
f_ddz_saturs	Informācija par Darba Devēju Ziņojumu (VSA) saturu
f_ddz_dokuments	Vispārīga informāciju par Darba Devēja Ziņojumu (VSAA) un Pašnodarbināto ziņojumu rindām.
f_an_saturs	Akcīzes nodokļa deklarāciju saturs
d_an_dokuments	Akcīzes Nodokļa deklarācijas dokumenta dimensija.
f_ugp_saturs	UGP saturs ietver informāciju, kas ir ekonomisks rādītājs par nodokļu maksātāju noteiktā periodā.
f_pvn_saturs	Faktu tabula Deklarācijas saturs ietver informāciju, kas ir ekonomisks rādītājs par nodokļu maksātāju noteiktā periodā.
d_samstat	Dimensiju tabula - sociālās apdrošināšanas maksājumu statusi
f_nin_summa	Faktu tabula, kura glabā summas par nekustamā īpašuma nodokļa maksājumiem.
f_maksasjumi	Dimensiju tabula - maksājumu veidu dati.
d_mv_dat	Maksājumu veidu dimensija.
f_zdn_saturs	Satur informāciju par darba ņēmēju statusiem, pie kāda darba devēja darba ņēmējs ir strādājis no kura līdz kuram datumam
d_fiz_persona	Fizisko personu dimensija
d_jur_persona	Juridisko personu dimensija
ds_uznemd_forma	Uzņēmējdarbības formu dimensija
d_doktips	Dimensijas tabula Dokumentu tips ietver informāciju par dokumentu tipiem
f_nin_summa	Faktu tabula, kura glabā summas par nekustamā īpašuma nodokļa maksājumiem.
f_maksajumi	Faktu tabula - maksājumi.

3.3. tabula

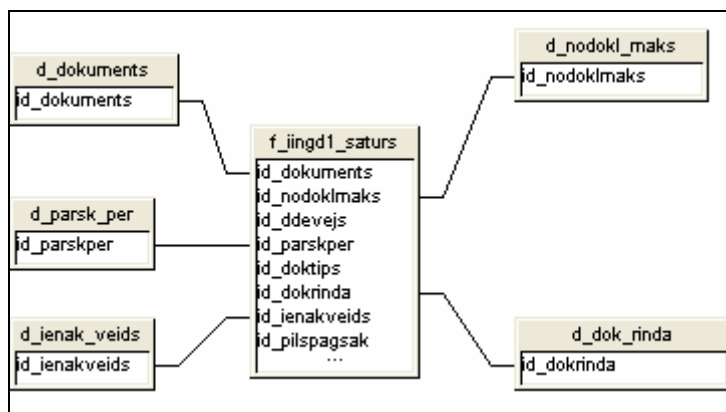
Informācijas apmaiņas analīzes sistēmai pievienotās VID DNS DB tabulas

Tabulas fiziskais nosaukums	Apraksts
f_iingd_saturs	Informācija par Iedzīvotāju ienākuma nodokļa gada deklarāciju (D veidlapa)
f_iingd1_saturs	Informācija par Iedzīvotāju ienākuma nodokļa gada deklarāciju (D1 - taksācijas gadā Latvijas Republikā gūtie ienākumi)
f_iingd2_saturs	Informācija par Iedzīvotāju ienākuma nodokļa gada deklarāciju (D2 - fizisko personu (rezidentu) ārvalstīs gūtie ienākumi)
f_iingd3_saturs	Informācija par Iedzīvotāju ienākuma nodokļa gada deklarāciju (D3 - ienākumi no saimnieciskās darbības)
f_iingd4_saturs	Informācija par Iedzīvotāju ienākuma nodokļa gada deklarāciju (D4 - attaisnotie izdevumi par izglītību un ārstnieciskajiem pakalpojumiem)
f_iinp12_saturs	Informācija par dokumentiem Paziņojums par algas nodokli (1.pielikums) un Darba devēja (ienākuma izmaksātāja) paziņojums par algas nodokli (2.pielikums)
f_iinp3_saturs	Informāciju par dokumentu Pārskats par nerezidenta gūtajiem ienākumiem un samaksāto iedzīvotāju ienākuma nodokli Latvijas Republikā (3.pielikums).

Tabulas fiziskais nosaukums	Apraksts
f_iinp45_saturs	Informācija par dokumentiem Paziņojums par personai izmaksātajām ar iedzīvotāju ienākuma nodokli apliekamajām summām, kas nav saistītas ar darba attiecībām, neieturot izmaksas vietā nodokli (4.pielikums) un Paziņojums par personai izmaksātajām ar iedzīvotāju ienākuma nodokli neapliekamajām summām.(5.pielikums).

Visas analīzes sistēmas datu bāzes modelim pievienotās faktu tabulas, kurās glabājas dati par IeIeN gada deklarāciju D veidlapu un pielikumu saturu, satur atsauces uz nodokļu maksātāju, laika perioda, dokumentu, dokumentu rindiņu un ienākumu veidu dimensijām. Faktu tabulās atrodas arī skaitliskā vērtība - Iedzīvotāju ienākuma nodokļa rindiņu summa, ko uzrāda VID.

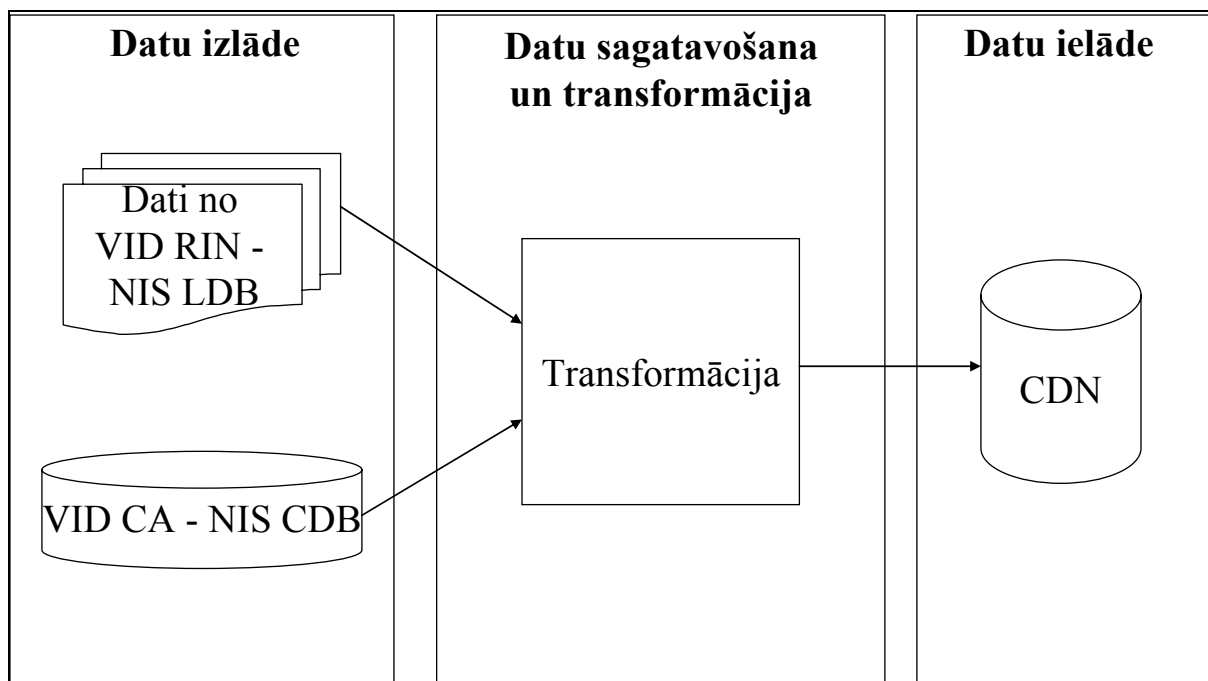
3.1. att. ir parādīta faktu tabulas f_iingd1_saturs, kuras satur datus par IeIeN gada deklarācijas D1 veidlapas saturu piesaiste pie esošajām dimensiju tabulām.



3.1. att. Faktu tabulas f_iingd1_saturs piesaiste pie dimensiju tabulām

3.2. Transformāciju procesi

Jaunizveidotos tabulas f_iingd2_saturs laukus ir nepieciešams aizpildīt. Aizpilde tiek veikta ar transformāciju procesiem. Jaunizveidoto lauku aizpildei tiks izmantots jau esošs transformāciju process, veicot tajā nepieciešamos labojumus. Transformācijas procesa vispārīga shēma ir redzama (3.2. att.).



3.2. att. Transformācijas procesa vispārīga shēma

3.2. att. ir redzams, ka VID DNS DB CDN nonāk dati gan no VID reģionālo iestāžu lokālajām NIS datu bāzēm, kā arī no VID Centrālā aparāta NIS Centrālās datu bāzes, pirms datu ielādes VID DNS DB tie tiek transformēti un sagatavoti ielādei.

Datu transformēšana sastāv no:

- 1) datu izlādes no datu avota;
- 2) datu sagatavošanas un transformācijas;
- 3) datu ielādes VID DNS datu bāzē.

Datu izlāde no datu avota jāveic šādi:

- 1) no NIS LDB jāizlādē IeIeN gada deklarāciju dati;
- 2) no NIS LDB jāizlādē papildus informācija, kas nepieciešama, lai noteiktu to VID RIN, no kuras jāņem nodokļu maksātāja dokumenti;
- 3) klasifikatoru dati no NIS CDB jālasa tiešsaistes režīmā datu sagatavošanas un transformācijas laikā, iepriekš neveicot nekādu datu izlādi;
- 4) datu izlāde no NIS LDB jāveic vienreiz nedēļā izmaiņu režīmā tajā nedēļas dienā, kas saskaņota ar Galveno nodokļu pārvaldi, taču reizi pusgadā vai arī pēc Galvenās nodokļu pārvaldes pieprasījuma ir jāveic pilna apjoma datu izlāde no NIS LDB.

Datu sagatavošana un transformācija tiek veikta šādi:

- 1) ar DNS pielietotās tehnoloģijas palīdzību tiek veikta datu sagatavošana, kas ietver izlādēto datu saarhivēšanu (ar ZIP), pārsūtīšanu un atarhivēšanu, lai tos varētu lasīt *DataStage* transformāciju procesi;
- 2) ar transformāciju rīku *DataStage* informāciju pārveido datu bāzes CDN tabulu formātos;
- 3) transformācija jāizpilda izmaiņu režīmā tajā nedēļas dienā, kas saskaņota ar Galveno nodokļu pārvaldi vai arī pilna apjoma režīmā reizi pusgadā vai pēc Galvenās nodokļu pārvaldes pieprasījuma.

Datu transformāciju procesu rezultātā ir jāiegūst faili ar paplašinājumu **.dat* un atbilstoši SQL skripti to ielādei VID DNS DB. Datu ielāde DB jāveic ar DNS jau pielietotas tehnoloģijas palīdzību, kas nodrošina automātisku datu ielādes procesa palaišanu pēc datu transformācijas beigām.

Lai būtiski samazinātu no NIS LDB pārsūtāmo IeIeN gada deklarāciju datu apjomu un transformāciju bloka izpildes laiku, IeIeN gada deklarāciju datu pārņemšana tiek veikta izmaiņu režīmā. Izmaiņu režīmā jāpārņem tie IeIeN gada deklarāciju dokumenti, kas ievadīti no jauna vai izmainīti kopš iepriekšējās datu pārņemšanas (transformēšanas) reizes.

Lai novērstu kļūdas, kuras var uzkrāties, ilgstoši darbinot transformāciju procesus izmaiņu režīmā, ir jāparedz arī pilna apjoma datu pārņemšanas režīms. Transformāciju procesiem automātiski jāpārslēdzas uz šo režīmu, ja no NIS LDB tiek saņemti pilna apjoma datu izlādes faili. Par pilna apjoma datu izlādes failiem transformāciju procesiem ir jāuzskata tie faili, kuros kā pēdējās izlādes datums norādīts 01.01.1990, tas nozīmē, ka tiek veikta pilna datu izlāde tiem IeIeN gada deklarāciju dokumentiem, kuru taksācijas periods atbilst laika periodam sākot no 1990. gada 1.janvāra.

Datu avota apraksts

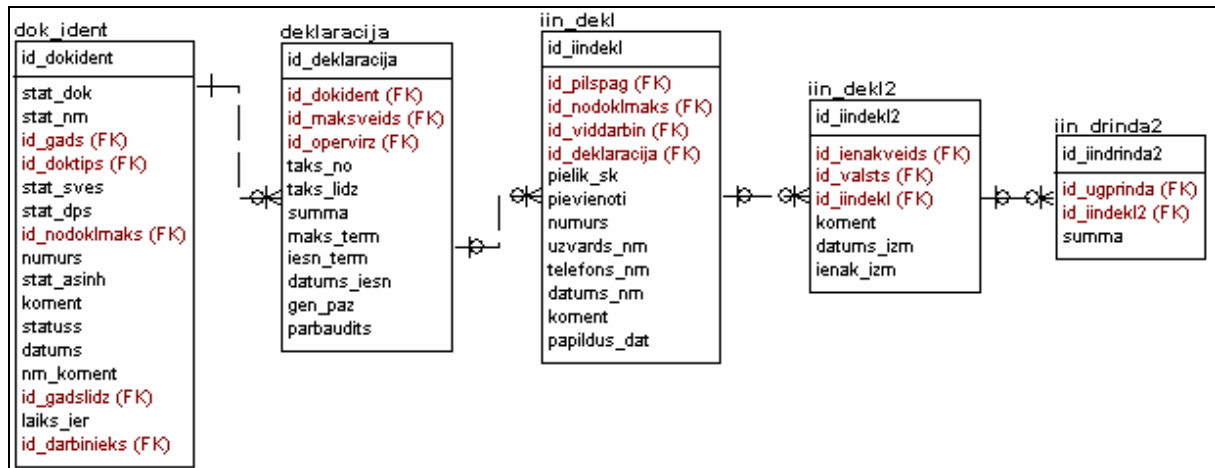
Datu avots ir NIS datu bāzes tabulas. To apraksts sniegts (3.4. tabula) un saites starp šīm tabulām attēlotas (3.3. att.).

3.4. tabula

Datu avota (NIS) DB tabulu apraksts

Tabulas fiziskais nosaukums	Apraksts
iin_dekl2	Tabula satur datus par IeIeN gada deklarācijas D2 veidlapas saturu.
iin_dekl	Tabula satur informāciju par IeIeN gada ienākumu deklarāciju.

Tabulas fiziskais nosaukums	Apraksts
iin_drinda2	Pakārtota tabulai iin_dekl2. Tabulā iin_dekl2 iekļautās informācijas izvērsums.
deklaracija	Tabula satur datus par NM deklarācijām.
dok_ident	Tabula satur dokumentu identificējošos datus



3.3. att. Saites starp datu avota tabulām

Tabulas f_iingd2_saturs aizpildīšana

Lai aizpildītu jaunus VID DNS DB tabulas f_iingd2_saturs laukus ir nepieciešams papildināt:

- 1) izlādes skripta 'nm_dok_IINGD' failu 'iingd2.dat' ar jaunām kolonnām iin_dekl2.datums_izm un iin_dekl2.id_ienakveids;
- 2) transformācijas procesu *NMfIINGD2saturs* (NIS DB tabulas *iin_dekl2* lauks 'datums_izm' jāpārnes uz DNS DB tabulas *f_iingd2_saturs* lauku 'datums_izm', bet NIS DB tabulas *iin_dekl2* lauks 'id_ienakveids' uz DNS DB tabulas *f_iingd2_saturs* lauku 'id_ienakveids').

3.3. Analīzes sistēma

Autors ir pilnveidojis analīzes sistēmu, kura nodrošina DN ievietoto CSP nepieciešamo datu atspoguļošanu. Autors ir papildinājis analīzes sistēmu ar informāciju par IeIeN gada deklarāciju D veidlapu un pielikumu saturu. Jaunizveidoto klašu, apakšklašu un objektu apraksts (3.5. tabula). Tabulas kolonnu nozīme:

- Klase – klases vai apakšklases nosaukums. Apakšklases nosaukums aiz starpsvītras /;
- Objekts – objekta nosaukums;
- Objekta skaidrojums – papildus informācija par objektu, jo objekta nosaukums nedrīkst saturēt vairāk par 35 simboliem;

- Objekta tips – objekta veids. DIM – Dimensijas objekts; MĒR – Mēra objekts (kvantitatīvs lielums).

3.5. tabula

Jauno klašu, apakšklašu un objektu apraksts

Klase	Objekts	Objekta skaidrojums	Objekta tips
Gada ienākumu deklarācija	<i>Iedzīvotāju ienākuma nodokļa gada deklarācija</i>		
Gada ienākumu deklarācija/D Saturs	<i>IeIeN gada deklarācijas D veidlapas saturs (taksācijas gadā Latvijas Republikā gūtie ienākumi)</i>		
	[NMRKOD] NM NMR Kods	Nodokļu maksātāja kods	DIM
	[DN_ATK] Dzīves vietas ATK kods	Ienākuma saņēmēja dzīvesvietas teritoriālais kods (sešzīmju)	DIM
	[IENAPL] Apliekamie ienākumi	Apliekamie ienākumi kopā	MĒR
	[IENNPL] Neapliekamie ienākumi	Neapliekamie ienākumi	MĒR
	[PRCVSA] Precizētās apr. VSAOIE	Precizētās aprēķinātās valsts sociālās apdrošināšanas obligātās iemaksas (palielinājums (+) vai samazinājums (-)), Ls	MĒR
	[IEMPPF] Iemaksas pensiju fondos	Privātajos pensiju fondos izdarītās iemaksas un apdrošināšanas prēmiju maksājumi atbilstoši dzīvības apdrošināšanas līgumam (ar līdzekļu uzkrāšanu)	MĒR
	[IETNOD] Avansā ieturētais nodoklis	Avansā samaksātais (ieturētais) nodoklis	MĒR
	[PRPIEM] Piemaksa	Pārrēķina rezultāts - piemaksa	MĒR
[PRPARM] Pārmaksa	Pārrēķina rezultāts - pārmaksa	MĒR	
Gada ienākumu deklarācija /D1 Saturs	<i>IeIeN gada deklarācijas D1 veidlapas saturs (taksācijas gadā Latvijas Republikā gūtie ienākumi)</i>		
	[NMRKOD] NM NMR Kods	Nodokļu maksātāja kods	DIM
	[IENKOD] Ienākuma veida kods	Ienākumu gūšanas veida kods	DIM
	[IENBRT] Bruto ieņēmumi	Bruto ieņēmumi	MĒR
	[IENNPL] Neapliekamie ienākumi	Neapliekamie ienākumi	MĒR
	[IZDATT] Attaisnotie izdevumi	Attaisnotie izdevumi	MĒR
	[DVSAOI] Darba ņēmēja VSAO iemaksas	Darba ņēmēja valsts sociālās apdrošināšanas obligātās iemaksas	MĒR
	[IEMPPF] Iemaksas pensiju fondos	Iemaksas privātajos pensiju fondos un dzīvības apdrošināšanas prēmijas	MĒR

Klase	Objekts	Objekta skaidrojums	Objekta tips
	[IZDAUT] Autoru izdevumi	Autoru izdevumi	MĒR
	[IZDEVU] Izdevumi saistīti ar ien.	Izdevumi, kas saistīti ar ienākumu gūšanu	MĒR
	[IENAPL] Apliedzamie ienākumi	Apliedzamie ienākumi, neatskaitot darba devēja veiktās iemaksas	MĒR
	[IETNOD] Avansā ieturētais nodoklis	Avansā samaksātais (ieturētais) nodoklis	MĒR
Gada ienākumu deklarācija /D2 Saturs	<i>IeIeN gada deklarācijas D2 veidlapas saturs (fizisko personu (rezidentu) ārvalstīs gūtie ienākumi)</i>		
	[NMRKOD] NM NMR Kods	Nodokļu maksātāja kods	DIM
	[IENKOD] Ienākuma veida kods	Ienākuma veida kods	DIM
	[IENDAT] Ienākumu saņemšanas datums	Ienākumu saņemšanas datums	DIM
	[IENSUM] Ienākumu summa latos	Summa latos	MĒR
	[SUMNOD] Nodokļa summa latos	Ārvalstīs samaksātā nodokļa summa latos	MĒR
	[NODLIK] Nodokļa likme	Nodokļa likme no ārvalstīs gūtajiem ienākumiem pēc Latvijā noteiktās likmes	MĒR
	[NODSUM] Nod. summa pēc LV likmes	Nodokļa summa latos pēc Latvijā noteiktās likmes	MĒR
Gada ienākumu deklarācija /D3 Saturs	<i>IeIeN gada deklarācijas D3 veidlapas saturs, ienākumi no saimnieciskās darbības</i>		
	[DNPERK] NMFP Reģistrācijas kods	Nodokļu maksātāja (fiziskas personas) reģistrācijas kods	DIM
	[IENLSR] Ieņēmumi no l/s ražošanas	Ieņēmumi no lauksaimnieciskās ražošanas	MĒR
	[IZDSLRL] Lauk. ražošanas izdevumi	Izdevumi, kas saistīti ar lauksaimniecisko ražošanu	MĒR
	[I_ZLSR] Ien. vai zaud. no l/s raž.	Ienākumi vai zaudējumi no lauksaimnieciskās ražošanas	MĒR
	[IENNPL] Neapl. ien. no l/s raž.	Neappliedzamie ienākumi no lauksaimnieciskās ražošanas 3000 latu apmērā gadā	MĒR
	[IENAPL] Apl. ien. no l/s raž.	Apliedzamie ienākumi no lauksaimnieciskās ražošanas	MĒR
	[I_GZLR] Iepr. g. zaud. no l/s raž.	Iepriekšējo gadu zaudējumi no lauksaimnieciskās ražošanas, kurus sedz ar taksācijas gada appliedzamajiem ienākumiem no lauksaimnieciskās ražošanas	MĒR
	[APILSR] Apl. ien. iepr. g. zaud.	Apliedzamie ienākumi no lauksaimnieciskās ražošanas, ņemot vērā iepriekšējo gadu zaudējumus no lauksaimnieciskās ražošanas	MĒR

Klase	Objekts	Objekta skaidrojums	Objekta tips
	[NEAPLI] Neapl. ienākumi	Neapliekamie ienākumi (subsīdijas, kas izmaksātas kā valsts atbalsts lauksaimniecībai vai Eiropas Savienības atbalsts lauksaimniecībai un lauku attīstībai)	MĒR
	[IENSDV] Citi ieņēmumi	Ieņēmumi no citiem saimnieciskās darbības veidiem	MĒR
	[IZSDV] Citi izdevumi	Izdevumi, kas saistīti ar citiem saimnieciskās darbības veidiem	MĒR
	[I_ZCSD] Citi ien. vai zaud.	Ienākumi vai zaudējumi no citiem saimnieciskās darbības veidiem	MĒR
	[APICSD] Apl. ien. no citiem SDV	Apliekamie ienākumi no citiem saimnieciskās darbības veidiem	MĒR
	[IGZCSD] Iepr. g. zaud. no cit. SDV	Iepriekšējo gadu zaudējumi no citiem saimnieciskās darbības veidiem, kurus sedz ar taksācijas gada apliekamajiem ienākumiem no citiem saimnieciskās darbības veidiem	MĒR
	[AP_CSD] Apl. ien. no cit. SDV zaud	Apliekamie ienākumi no citiem saimnieciskās darbības veidiem, ņemot vērā iepriekšējo gadu zaudējumus no citiem saimnieciskās darbības veidiem	MĒR
	[VSAOIM] VSAO Iemaksas	Par pašu nodokļa maksātāju aprēķinātās un veiktās Valsts sociālās apdrošināšanas obligātās iemaksas (VSAOI) (VID dati)	MĒR
	[APLISD] Apl. ien. no SDV (kopā)	Apliekamie ienākumi no saimnieciskās darbības	MĒR
	[IETNOD] Avansā ieturētais nodoklis	Avansā samaksātais nodoklis vai patentmaksa	MĒR
Gada ienākumu deklarācija /D4 Saturs	<i>IeIeN gada deklarācijas D4 veidlapas saturs (attaisnotie izdevumi par izglītību un ārstnieciskajiem pakalpojumiem)</i>		
	[NMRKOD] NM NMR Kods	Nodokļu maksātāja (iesniedzēja) kods	DIM
	[ATIZIZ] Att. izd. par izglītību	Attaisnotie izdevumi, uz kuriem attiecas Ministru kabineta noteiktā norma par izglītību (2.aile)	MĒR
	[ATIZAP] Att. izd. par ārst. pak.	Attaisnotie izdevumi, uz kuriem attiecas Ministru kabineta noteiktā norma par ārstnieciskajiem pakalpojumiem (3.aile)	MĒR
	[ATTIZD] Att. izd. izgl.,ārst. pak.	Kopā (taksācijas gadā 2. + 3.aile; pirmstaksācijas gadā 7.ailes attiecīgā rinda pirmstaksācijas gada veidlapā D4) (4. aile)	MĒR
	[IZDAPA] Izd. par ārst. pak.	Izdevumi par ārstnieciskajiem pakalpojumiem, kurus atbilstoši Ministru kabineta noteikumiem ir tiesības ietvert attaisnotajos izdevumos pilnā apmērā	MĒR
	[IZDIAP] Att. izd. ārst.,izgl.izd.	Taksācijas vai pirmstaksācijas gadā attaisnotajos izdevumos ietvertie	MĒR

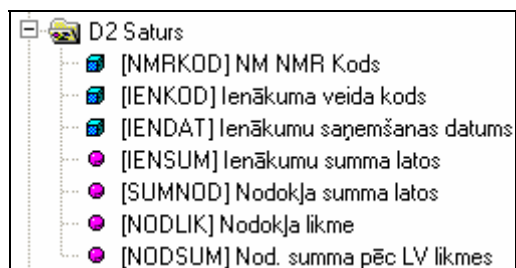
Klase	Objekts	Objekta skaidrojums	Objekta tips
		izdevumi par izglītību un ārstnieciskajiem pakalpojumiem	
	[ATIZNP] Uz nāk. taks. g. att. izd.	Uz nākamajiem taksācijas gadiem attiecināmie attaisnotie izdevumi	MĒR
Paziņojumi par algas nodokli (P1_2)	<i>Informācija par paziņojumiem par algas nodokli (IeIeN 1. - 2. pielikumi)</i>		
	[DNPERK] NMFP Personas kods	Darba ņēmēja personas kods	DIM
	[DN_ATK] Dzīves vietas ATK kods	Ienākuma saņēmēja dzīvesvietas teritoriālais kods (sešzīmju)	DIM
	[IENKOD] Ienākumu veida kods	Ienākuma veida kods	DIM
	[IEN_NO] Ienākuma gūšanas "no"	Ienākuma gūšanas datums "no"	DIM
	[IENLDZ] Ienākuma gūšanas "līdz"	Ienākuma gūšanas datums "līdz"	DIM
	[IENBRT] Bruto ieņēmumi	Bruto ieņēmumi	DIM
	[IENNPL] Neapliekamie ienākumi	Neapliekamie ienākumi	MĒR
	[NPLMIN] Neapliekamais minimums	Neapliekamais minimums	MĒR
	[ATVAPG] Atvgl. par apgādājamiem	Atvieglējumi par apgādājamiem	MĒR
	[ATVINV] Atvgl. par invaliditāti	Papildus atvieglējumi par invaliditāti	MĒR
	[ATVPOL] Atvgl. pol. repr. personām	Papildus atvieglējumi politiski represētai personai	MĒR
	[ATVNPK] Atvgl. nac. p. kust. dalīb	Papildus atvieglējumi nacionālās pretošanās kustības dalībniekam	MĒR
	[IZDVSA] VSA iemaksas	Attaisnotie izdevumi, Valsts sociālās apdrošināšanas iemaksa	MĒR
	[IZDPPF] Pensiju fondu iemaksas	Attaisnotie izdevumi, iemaksas privātajos pensiju fondos	MĒR
	[IZDATT] Attaisnotie izd., izdevumi	Attaisnotie izdevumi	MĒR
	[IENAPN] Ienākums no kura apr. nod.	Ienākumi, no kuriem aprēķināts nodoklis	MĒR
	[IETNOD] Ieturētais nodoklis	Ieturētais nodoklis	MĒR
	[PRSNOD] Budžetā pārskaitītais nod.	Likumā noteiktajā termiņā budžetā pārskaitītais nodoklis	MĒR
	Nerezidenta gūtie ienākumi (P3)	<i>Informācija par nerezidenta gūtajiem ienākumiem un samaksāto iedzīvotāju ienākuma nodokli Latvijas Republikā</i>	
[NMRKOD] NM NMR Kods		Ienākumu saņēmēja, nerezidenta nodokļu maksātāja personas (NMR) kods	DIM
[IENKOD] Ienākumu veida kods		Ienākuma veida kods	DIM

Klase	Objekts	Objekta skaidrojums	Objekta tips
	[NODLIK] Nodokļa likme	Nodokļa likme	DIM
	[IZMDAT] Izmaksas datums	Ienākumu izmaksas datums	DIM
	[IENSUM] Ienākuma summa	Ienākuma summa	MĒR
	[SAMNOD] Samaksātais nodoklis	Samaksātais nodoklis	MĒR
Izmaksas personām (P4_5)	<i>Informācija par paziņojumiem par personai izmaksātajām ar IeIeN apliekamajām summām un ar IeIeN neapliekamajām summām</i>		
	[NMRKOD] NM NMR Kods	Ienākumu saņēmēja personas (NMR) kods	DIM
	[IENKOD] Ienākumu veida kods	Ienākuma veida kods	DIM
	[IEN_NO] Ienākuma gūšanas "no"	Ienākuma gūšanas datums "no"	DIM
	[IENLDZ] Ienākuma gūšanas "līdz"	Ienākuma gūšanas datums "līdz"	DIM
	[IIZDAT] Ienākuma izmaksas datums	Ienākuma izmaksas datums	DIM
	[PAZ4_5] Pazīme par iekļaušanu GD	Pazīme, par iekļaušanu IeIeN gada deklarācijā	DIM
	[IENSAN] Saņemtais ienākums	Saņemtais ienākums (Ls)	MĒR

Lietotājam redzamā analīzes sistēmas daļa sastāv no objektiem un filtriem, kur katra objekta un filtra nosaukums ir pieskaņots vispārīgiem terminiem. Objektus un filtrus ir nepieciešams sadalīt loģiskās direktorijās, lai lietotājs, paskatoties uz direktorija nosaukumu uzreiz varētu nojaust, kādi objekti un filtri atradīsies zem tā. Analīzes sistēmās šie direktoriji tiek saukti par klasēm.

Katra klase vai apakšklase ir paredzēta kāda konkrēta DBF eksporta faila sagatavošanai. Katra no šīm klasēm vai apakšklasēm satur dimensijas, kas atbilst DBF eksporta faila kolonām. Objektu nosaukumi šajās klasēs sākas ar saīsinājumu, kas sastāv no sešiem simboliem un ir iekļauts kvadrātikavās ,piemēram, [IENAPL] Apliekamie ienākumi. Figūriekavās iekļautais saīsinājums ir DBF eksporta faila kolonas nosaukums.

BO veidojot klases, objektus un filtrus, dod iespēju kopēt tos no citām analīzes sistēmām, kā arī ir iespēja kopēt klases, objektus un filtrus pašas analīzes sistēmas ietvaros. Tas ir ļoti ērti un ietaupa daudz laika. Interesanti ir arī tas, ka BO definējot objektu tam var piešķirt formātu, kas atšķiras no objekta datu avota lauka formāta datu bāzē.



3.4. att. Analīzes sistēmas apakšklase „D2 Saturs”

3.4. att. ir redzami analīzes sistēmas apakšklases „D2 Saturs” objekti. Dimensiju objekta nosaukumam priekšā ir kubs, bet mēra objekta nosaukumam priekšā ir lode. Kā redzams, dimensiju objekti ir izvietoti klases augšpusē, bet mēru objekti tikai pēc dimensiju objektiem apakšā.

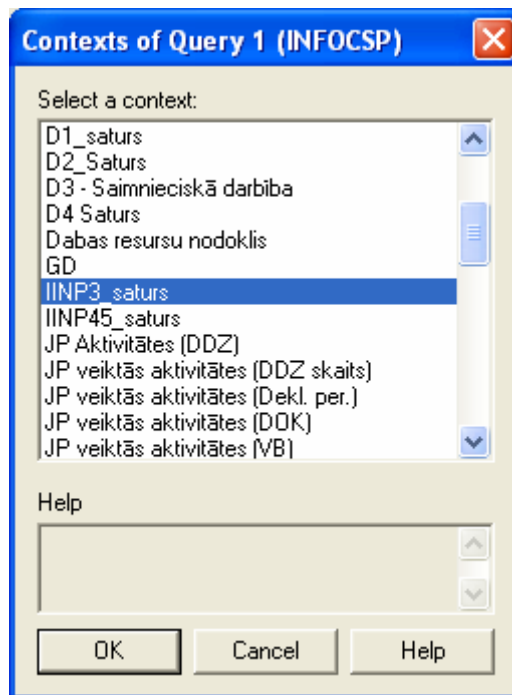
3.3.1. Analīzes sistēmas kopējās klases un to objekti

Analīzes sistēmas kopējās klases ir tādas klases, kuru objekti ir savietojami ar visu pārējo klašu objektiem.

Analīzes sistēmā “Informācija CSP” ir tādas dimensijas un tādi mēri, kas visiem objektiem ir kopēji un atrodas klasēs “Nodokļu maksātājs” un „Taksācijas periods”. Kas nozīmē to, ka lietotājs veidojot vaicājumu BO vaicājumu panelī (*Query Panel*), var izvēlēties jebkuru objektu no klasēm „Nodokļu maksātājs” un „Taksācijas periods”, pievienot klāt nepieciešamos objektus no citām klasēm vai apakšklasēm, un tiks atgriezts rezultāts.

Laika dimensijas klase („Taksācijas periods”) ir viena no klasēm, kas ir visās analīzes sistēmās, jo parasti datu analīzē tiek izmantots konkrēts laika mirklis.

Veidojot pārskatu, nav ieteicams izmantot tikai kopējo klašu objektus bez vismaz viena kādas individuālās klases objekta. Piemēram, klašu “Nodokļu maksātājs” un “Taksācijas periods” objekti ir iekļauti visos kontekstos, bet tiem nav savu atsevišķu kontekstu, tāpēc, izveidojot vaicājumu, kurā ietilpst tikai šo abu klašu objekti (piemēram, klases “Nodokļu maksātājs” objekts “[NMRKOD] NM NMR Kods” un klases “Taksācijas periods” objekts “Gads”) un nospiežot [Run], atvērsies dialoga logs “Contexts of Query...” (3.5. att.), kurā jāizvēlas, kādā kontekstā taksācijas gads ir jāsaista ar NM NMR Kodiem.




3.5. att. Konteksta izvēles logs

3.3.2. Analīzes sistēmas individuālās klases un to objekti

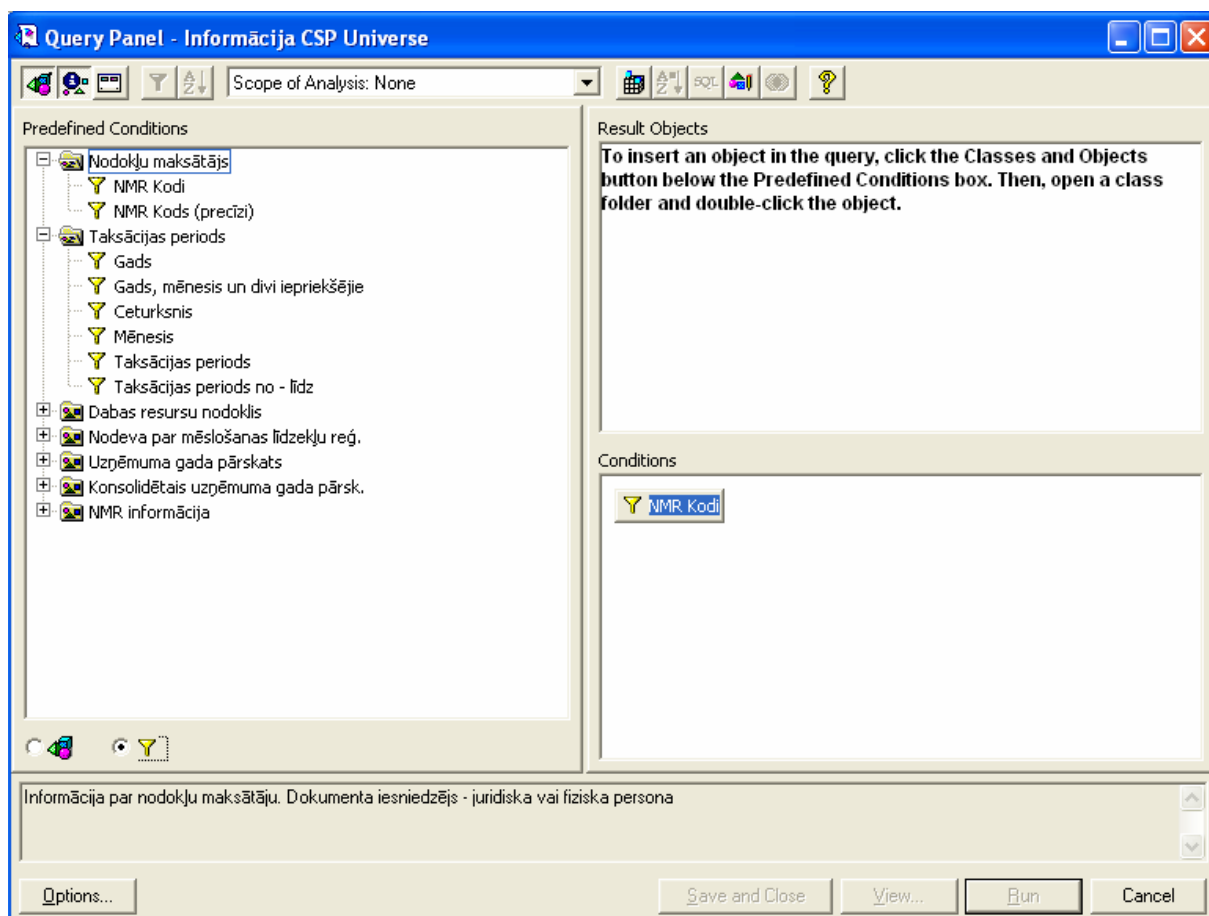
Analīzes sistēmas individuālās klases ir tādas klases, kuru objekti ir savietojami tikai ar savas klases vai ar kopējo klašu objektiem. Individuālo klašu objekti parasti tiek veidoti uz faktu tabulu bāzes, jo katrā faktu tabulā atrodas unikāli mēri, skaitliskas vērtības, pēc kurām tiek veikta analīze. Autora jaunizveidotās klases un apakšklases ir individuālās, jo šo klašu (apakšklašu) objektus var lietot kopā tikai ar savas klases (apakšklases) objektiem vai arī ar kopējo klašu „Nodokļu maksātājs” un „Taksācijas periods” klašu objektiem.

Veidojot datu vaicājumus šīs analīzes sistēmas ietvaros, jāievēro, ka apskatot objektus no dažādām individuālajām klasēm, rezultāti tiks attēloti nevis vienā, bet vairākās neatkarīgās tabulās (tabulu skaits parasti atbilst izmantoto kontekstu skaitam), jo katrs no individuālo klašu dimensiju un mēru objektiem strādā savā kontekstā, tāpēc vienā vaicājumā atlasot datus no vairākām individuālajām klasēm, būtu vēlams pievienot kādu kopējās klases „Nodokļu maksātājs” vai „Taksācijas periods” objektu, kas individuālo klašu datus saturiski sasaistītu un pārskatā ievietotu vienā tabulā.

3.3.3. Analīzes sistēmas iebūvētie filtri

Lai atlasāmajai informācijai uzliktu ierobežojumu ir nepieciešami filtri. Filtri ir nepieciešami arī tāpēc, lai atlasītu konkrētu informāciju nevis visu informāciju. Lai varētu atlasīt datus, analīzes sistēmā “Informācija CSP” ir iebūvēti gatavi filtri, kas attēloti ar ikonu . Lietotāji

var definēt paši savus filtrus, izmantojot analīzes sistēmā esošos objektus un BO Query Panel (3.6. att.) pārvietojot tos uz Conditions sadaļu.



3.6. att. Analīzes sistēmas iebūvētie filtri

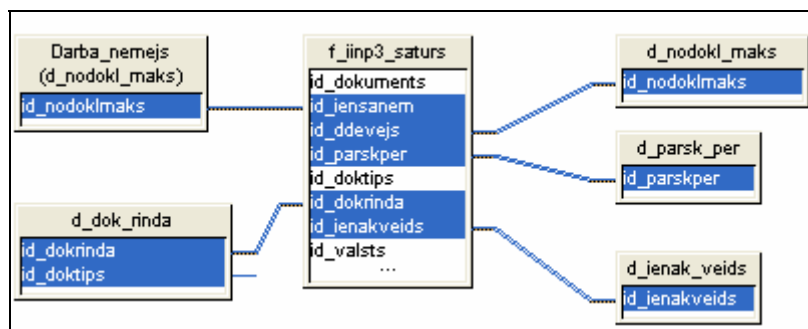
CSP nododamās informācijas pārskatos kā datu atlasē filtrus tiek izmantoti Gads: un Datums:.

3.3.4. Analīzes sistēmas konteksti

BO terminoloģijā ar vārdu “konteksti” saprot praktiski to pašu, ko literārajā valodā ar vārdu “konteksts”. Piemēram, lai noskaidrotu NM nodokļos samaksāto summu mēnesī, konteksta izvēle ir ļoti būtiska, jo iegūtie dati dažādu nodokļu kontekstos ir principiāli atšķirīgi.

Konteksti parāda tabulu savietojamību, kā arī novērš ciklus, kas var veidoties starp tabulu savstarpējām saitēm. Ar katra konteksta objektiem tiek veidots atsevišķs datu pieprasījums un pārskatā dati tiek attēloti atsevišķi.

Ievietojot analīzes sistēmā informāciju par IeIen Gada deklarāciju D veidlapu un pielikumu saturiem informācija tika izvietota 3 individuālās klasēs un 5 individuālās apakšklasēs, kas nozīmē to, ka tika izveidoti 8 jauni konteksti. Piemēram, IeIeN 3.pielikuma konteksts analīzes sistēmas datu bāzes modelī izskatās šādi (satur šādas saites starp tabulām) (skat. 3.7. att.)



3.7. att. Konteksts „IeIeNP3 saturs”

IeIeN 3.pielikums satur datus par ārvalstu pilsoņu (nerezidentu) gūtajiem ienākumiem un samaksāto IeIeN Latvijas Republikā. Tā kā IeIeN 3.pielikumā ir dati, ko uzrāda gan darba ņēmējs (nerezidents), gan darba devējs, tad tabulai f_iinp3_saturs eksistē 2 atsauces uz nodokļu maksātāju dimensiju, kurā glabājas dati gan par darba ņēmējiem, gan par darba devējiem. Gadījumos, kad viena un tā pati dimensija ir jāizmanto vairākas reizes, tiek veidoti tabulu pseidonīmi (alias) - viena tabula analīzes sistēmas datu bāzes modelī parādās vairākas reizes, bet ar dažādiem nosaukumiem. Šinī gadījumā tā būs tabula d_nodokl_maks, kurai tiks izveidots pseidonīms Darba_nemejs, lai varētu apskatīt informāciju par darba ņēmējiem, bet darba devēju raksturojoša informācija nāks no tabulas d_nodokl_maks. Pseidonīmi ir nepieciešami tāpēc, ka eksistē vairākas atsauces uz vienu dimensiju un ir nepieciešams izveidot objektus (klases), kas saturēs informāciju no katras atsauces atsevišķi. Analīzes sistēmā datu bāzes modelī ir iespējams izveidot saites ne tikai starp tabulām, bet arī vienas tabulas, vai viena tabulas lauka ietvaros. Piemēram, IeIeN 3.pielikumā ir nepieciešama informācija par nerezidenta gūto ienākumu summu un samaksāto nodokli LR, šo objektu definīcijā tiek izmantoti divu tabulu lauki, skaitliskā vērtība jeb rindiņas summa nāk no tabulas f_iinp3_saturs, bet IeIeN 3.pielikuma rindas numurs nāk no tabulas d_dok_rinda. Tā kā tabulā d_dok_rinda glabājas daudzu dokumentu (UGP, IeIeN gada deklarāciju D veidlapu saturi) rindu numuri, tad tabulas d_dok_rinda laukam (svešatslēgai) id_doktips ir jānodefinē saite, ka id_doktips = 990200087, jo šī vērtība raksturo tieši IeIeN 3.pielikumu. Un ieliekot šo saiti IeIeNP3 satura kontekstā korekti tiks atlasītas nerezidenta gūtā ienākuma summa, nodokļa likme un samaksātais nodoklis LR. Konteksts darbojas nodokļu maksātāju, laika, ienākumu veidu, dokumenta un tā rindu dimensiju griezumā.

3.4. Pārskati

Autors ir izveidojis lietotāju pieprasītos standarta pārskatus, kas ir nepieciešami, lai nodrošinātu CSP nepieciešamo datu atlasīšanu. Pārskatu veidošana ar BO ir diezgan ērta un ātri apgūstama. Pārskati tika veidoti izmantojot analīzes sistēmu „Informācija CSP”, taču pārskatu veidošanai nav obligāti jāizmanto izveidotās analīzes sistēmas, var izmantot arī citus datu

avotus. Nodefinējot datu avotu, var pieslēgties tam un veikt datu pieprasījumus un no tiem veidot pārskatus. Šajā gadījumā tiks apskatīta tikai iespēja veidot pārskatus, izmantojot analīzes sistēmu.

Lai izveidotu kādu no CSP nepieciešamajām DBF failiem, vispirms jāizveido pārskats ar BO, piemēram, BO pārskats INFOCSP_IINP3_ienakveidi.rep, paredzēts divu DBF failu IINP3_yyyymmdd.dbf un ienak_veidi_yyyy.dbf, sagatavošanai.

Autors ar analīzes rīku BO ir izveidojis 9 jaunus pārskatus, kas paredzēti 10 DBF failu izveidošanai.

Autora izveidoto pārskatu apraksts organizēts tabulas (skat. 3.6. tabula) veidā. Tabulas kolonnu nozīme ir šāda:

- „Pārskata nosaukums” - gatavā pārskata nosaukums;
- „Pārskata datnes nosaukums” - datnes nosaukums, ar kādu pārskats saglabāts;
- „Datu piegādātāja nosaukums” – pārskata datu piegādātāju nosaukumi;
- „DBF faila nosaukums” - no pārskata datu piegādātājiem veidojamo DBF failu nosaukumi. DBF faila nosaukums veidojas no pārskata datu piegādātāja nosaukuma un pārskata atjaunošanas parametriem, kur yyyy- gads, mm – mēnesis, dd – diena.

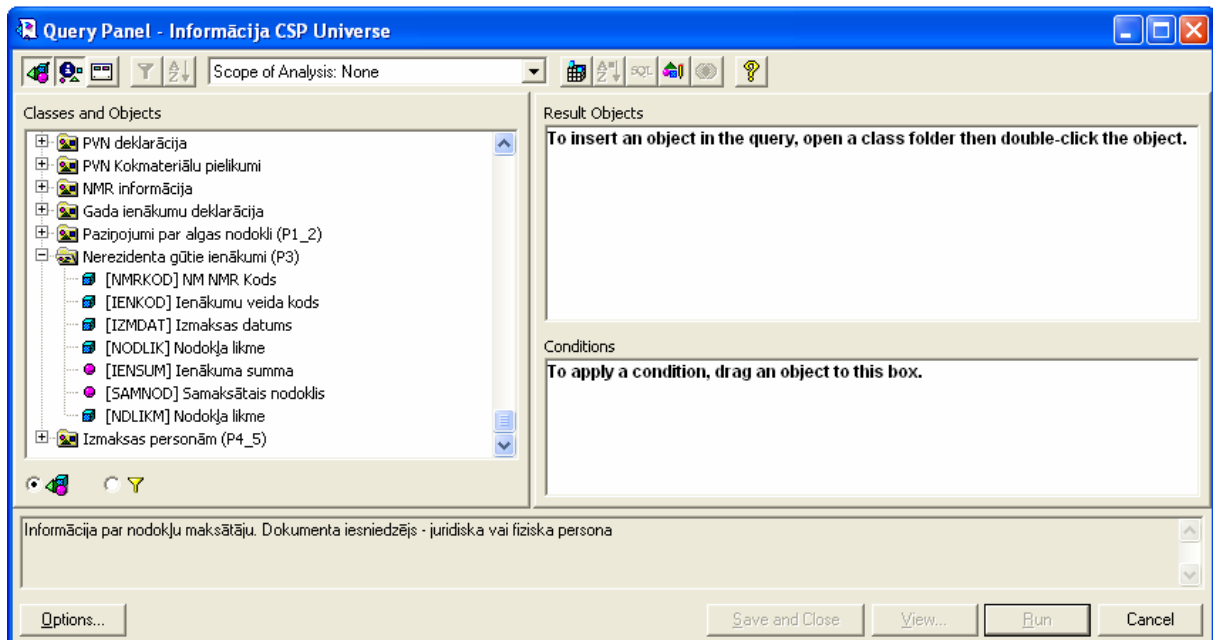
3.6. tabula

Jauno pārskatu apraksts



Pārskata nosaukums	Pārskata datnes nosaukums	Datu piegādātāju nosaukumi	DBF failu nosaukumi
Gada ienākumu deklarācija – D veidlapa	INFOCSP_D.rep	D	D_yyyy.dbf
Taksācijas gadā Latvijas Republikā gūtie ienākumi - D1 veidlapa	INFOCSP_D1_ienakveidi.rep	D1	D1_yyyy.dbf
		IENAK_VEIDI	ienak_veidi.dbf
Fizisko personu (rezidentu) ārvalstīs gūtie ienākumi – D2 veidlapa	INFOCSP_D2_ienakveidi.rep	D2	D2_yyyy.dbf
		IENAK_VEIDI	ienak_veidi.dbf
Ienākumi no saimnieciskās darbības - D3 veidlapa	INFOCSP_D_D3.rep	D_D3	D_D3_yyyy.dbf
Attaisnotie izdevumi par izglītību un ārstnieciskajiem pakalpojumiem - D4 veidlapa (kopā)	INFOCSP_D4.rep	D4	D4_yyyy.dbf
Darba devēja (ienākumu maksātāja) paziņojums par algas nodokli	INFOCSP_algas_paz_ienakveidi.rep	ALGAS_PAZ	Algas_paz_yyyymmdd.dbf

Pārskata nosaukums	Pārskata datnes nosaukums	Datu piegādātāju nosaukumi	DBF failu nosaukumi
		IENAK_VEIDI	ienak_veidi.dbf
Nerezidenta gūtie ienākumi un samaksātais IeIeN Latvijas Republikā	INFOCSP_Nerez_IIN_ienakveidi.rep	NEREZ_IIN	Nerez_IIN_YYYY.dbf
		IENAK_VEIDI	ienak_veidi.dbf
Paziņojums par personai izmaksātajām ar IeIeN apliekamajām summām	INFOCSP_IINP4_ienakveidi.rep	IINP4	IINP4_YYYY.dbf
		IENAK_VEIDI	ienak_veidi.dbf
Paziņojums par personai izmaksātajām ar IeIeN neapliekamajām summām	INFOCSP_IINP5_ienakveidi.rep	IINP5	IINP5_YYYY.dbf
		IENAK_VEIDI	ienak_veidi.dbf

Pārskati tika veidoti izmantojot BO vaicājumu redaktoru Query Panel (3.8. att.)



3.8. att. Vaicājumu redaktora logs

Vaicājumu redaktors sastāv no trim daļām: kreisajā daļā atrodas visi analīzes sistēmas izmantojamie objekti un filtri, labajā augšējā daļā tiek ievietoti pieprasījumam nepieciešamie objekti, labajā apakšējā daļā tiek ievietoti vai veidoti filtri. Starp objektiem  un iebūvētajiem filtriem  var pārslēgties, izmantojot kreisajā apakšējā loga pusē esošās divas radio pogas.

Analīzes sistēmas „Informācija CSP” objekti datu piegādātājos nodrošina kolonnu veidošanu DBF failos. Katra objekta nosaukumā kvadrātiekavās ietverts atbilstošā DBF faila kolonnas nosaukums. DBF faila nosaukumu, atbilstoši CSP prasībām, veido datu piegādātāja nosaukumi un pārskata datu atlasē parametri. Tāpēc pirms DBF failu veidošanas pārskatu dati jāatjauno ar atbilstošiem parametriem. Piemēram, ja gribam apskatīt informāciju par

fizisko personu (rezidentu) ārvalstīs gūtajiem ienākumiem 2006. taksācijas gadā mums jāatjauno BO pārskats INFOCSP_D2_ienakveidi.rep, definējot atlasas parametru Gads:2006, tad DBF faila nosaukums būs D2_2006.dbf. Katrs pārskats sastāv no viena vai vairākiem datu piegādātājiem un katra pārskata katrs datu piegādātājs nodrošina datus viena DBF faila izveidošanai. Piemēram, BO pārskats INFOCSP_D2_ienakveidi.rep, kas satur informāciju par IeIeN gada deklarācijas D2 veidlapas saturu (fizisko personu (rezidentu) ārvalstīs gūtie ienākumi norādītajā taksācijas gadā). Pārskats sastāv no 2 datu piegādātājiem D2 un IENAK_VEIDI. Šim pārskatam ir divi datu piegādātāji, jo no pārskata ir paredzēts izveidot divus DBF failus, bet datu piegādātāji vēl ir nepieciešami tāpēc, lai pārskatos ievietotu informāciju no citas analīzes sistēmas vai cita datu avota, vai ja ir nepieciešams iegūt datus no šīs pašas analīzes sistēmas citā griezumā, kas ir līdzīgi kā šai gadījumā, kad ir nepieciešams iegūt informāciju gan par fizisko personu (rezidentu) ārvalstīs gūtajiem ienākumiem, gan klasificētu informāciju par ienākumu veidiem ienākumu dimensijas griezumā. Datu piegādātājs D2 satur informāciju par gūto ienākumu veidiem un to skaitliskajām vērtībām nodokļu maksātāju griezumā, bet datu piegādātājs IENAK_VEIDI satur klasificētu informāciju par ienākumu veidiem. Pārskatā esošo informāciju var attēlot tabulas (3.9. att.), grafika vai šķērstabulas veidā.

LR VID datu noliktavas sistēma
Ienākumu veidu klasifikators

Ienākuma veida kods	Ienākuma veida nosaukums
01	Darba alga
02	Prēmija
03	Vienreizēja atlīdzība
04	Sistemātiska atlīdzība
05	Ienākumi no individuālā darba
06	Ienākumi no individuālā uzņēmuma līguma
07	Ienākumi no līgumsabiedrības
08	Ien. no nekust. īp. izm. (izn., izīr.)
09	Ien. no kustamā īpaš. izm. (izn., izīr.)
10	Pensijas neatkarīgi no to izmaksas avota
11	Citi ienākumi
12	Ienākumi no intelektuālā īpašuma
13	Slimības pabalsti ("B" daļa)
14	Savas lauks. produkcijas pārdošana
15	Koksnes pārdošana
16	Metālūžņu nodošana
17	Pakalpojumu sniegšana
18	Preču pārdošana
19	Mantojuma rezultātā gūtais ienākums
20	No fiziskās pers. saņemtais dāvinājums
21	No juridiskās pers. saņemtais dāvinājums
22	Ien. no sava īp. pārd. izg. nek. īp.
23	Procenti
24	Dividendes
25	Ienākumi no privātajiem pensiju fondiem
26	Ienākumi no nekustamā īpašuma atsavin.

3.9. att. Pārskats „Ienākumu veidu klasifikators”

3.5. Pievienojumprogramma DBF failu eksportam

Pievienojumprogramma ir nepieciešama, lai no pārskatiem (.rep failiem), kas veidoti ar analīzes rīku BO, varētu noformēt failus CSP prasītajā DBF failu formātā. Eksporta funkcionalitātes nodrošināšanai izveidota BO pievienojumprogramma BCA_SERV_CSP.

Prasības pievienojumprogrammai:


- 1) jāveic pārskata datu piegādātāju (*Data provider*) saglabāšana DBF formāta failos;
- 2) ja pārskatā ir vairāki datu piegādātāji, tad tie jā saglabā atsevišķos failos;
- 3) DBF faila nosaukums jāveido par pamatu, izmantojot datu piegādātāja nosaukumu;
- 4) DBF faila nosaukumiem jāpievieno pārskata parametri (taksācijas gads, mēnesis, utt.)
- 5) eksportam jāstrādā divos režīmos:
 - a) iteratīvais režīms (eksportu manuāli veic BO lietotājs);
 - b) plānošanas režīms (eksports tiek ieplānots noteiktā laikā, izmantojot *Broadcast agent*);
- 6) lietotājam jābūt iespējai mainīt direktoriju, kurā tiek saglabāti eksportētie DBF faili;
- 7) pievienojumprogrammas uzstādīšana jāveic no BO repozitorija.

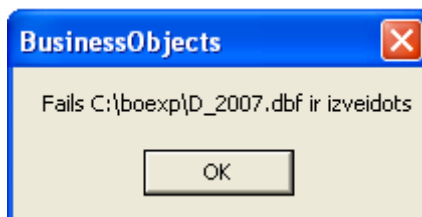
Pievienojumprogrammā ir iekļautas trīs makrokomandas un rīkjoslā. Pievienojumprogrammas makrokomandas veic šādas darbības:

- 1) Pārskata datu eksportu plānošanas režīmā (Broadcast Agent ieplānotajā laikā veiks pārskata atjaunināšanu un datu eksportu DBF failos);
- 2) Pārskata datu eksportu iteratīvajā režīmā (lietotājs atver iepriekš sagatavotus pārskatus, tos manuāli atjaunina, uzstādot datu atlasē parametrus, novērtē iegūto rezultātu un veic datu eksportu DBF failos);
- 3) DBF failu direktorijas definēšanu.

Datu piegādātāji (Data providers) tiek eksportēti atsevišķos DBF failos, kuru nosaukumiem tiek piešķirts datu piegādātāja nosaukums. Piemēram, pārskats „INFOCSP_D.rep” satur datu piegādātāju ar nosaukumu „D” un, veicot datu eksportu, dati tiks saglabāti failā, kura nosaukums sāksies ar „D”. Faila nosaukumam vēl tiks pievienoti pārskata parametri un faila paplašinājums.

Pievienojumprogramma satur informāciju par visiem pārskatu datu piegādātājiem un datu atlasēs parametriem, kuri jāpievieno failu nosaukumiem. Datu piegādātāji, par kuriem pievienojumprogrammai nav informācijas, netiks eksportēti.

Lietotāju ērtībai tika izveidota BO rīkjosla ar nosaukumu „CSP”, kurā ir podziņa , datu eksporta komandas izsaukšanai, kuru nospiežot automātiski tiks sākta datu piegādātāju eksportēšana DBF failos. Pēc datu piegādātāja eksporta pabeigšanas tiek parādīts paziņojums par DBF faila izveidošanu (skat. 3.10. att.)



3.10. att. Paziņojums par DBF faila izveidošanu

Pievienojumprogrammas uzstādīšanu veic, importējot to no BO repozitorija, kur glabājas aktuālā pievienojumprogrammas versija.

Secinājumi un priekšlikumi

Autors ir izpētījis un izanalizējis esošo VID un CSP informācijas apmaiņas saskarni un secinājis, ka informācijas apmaiņas saskarne ir bāzēta uz datu noliktavu tehnoloģijas. Visa CSP nepieciešamā informācija tiek ievadīta VID Nodokļu informācijas sistēmā, no kuras datu bāzēm dati tālāk tiek integrēti VID DNS DB.

Galvenais priekšnosacījums VID un CSP informācijas apmaiņas saskarnei ir funkcionējošu VID IS eksistence. Dati (atomāri, par katru individuālu faktu) no šādu sistēmu datu bāzēm tiek transformēti un ielādēti datu noliktavā.

Balstoties uz pieaugošo CSP nepieciešamās informācijas apjomu ir jāseko līdzi tam, vai visa CSP nepieciešamā informācija ir atrodamā VID DNS, ja nav, tad ir jānoskaidro kādas izmaiņas VID DNS ir nepieciešams veikt, lai nodrošinātu CSP ar tai nepieciešamo informāciju.

Autors piedāvā veikt uzlabojumus informācijas apmaiņas procesā, izmantojot Web servisu tehnoloģiju. Web servisu tehnoloģija ļauj komunicēt sistēmām savām starpā (10). Tā balstās uz SOAP protokolu. IT terminu vārdnīcā Web serviss ir definēts kā pakalpojums, kas tiek sniegts tīklā. Tas ir vienots veids, kā tīkla lietojumprogrammām, arī tādām, kas strādā dažādās platformās, rakstītas dažādās valodās u.t.t., savstarpēji sazināties. Ziņu pārraidei tiek izmantoti starptīkla protokoli. Web servisa klientam ir svarīgi tikai saņemt konkrētu funkcionālu resursu, savukārt Web servisa sniedzējs var jebkurā brīdī brīvi izvēlēties, kādā valodā un platformā izstrādāt un darbināt nepieciešamos algoritmus. Web servisu konkrēti īsteno aģents, kura saskarne ir aprakstīta WSDL formātā, tādējādi to bez cilvēka līdzdalības spēj interpretēt lietotāja programmatūra, kura ar šo aģentu sazinās pēc SOAP protokola.

Pārfrāzējot, Web servisi ļauj ērtā un universālā veidā apmainīties ar informāciju starp dažādām sistēmām, kuras savienotas ar Internetu vai lokālo tīklu. Informācijas apmaiņa notiek ar failu sūtījumiem, kuru struktūra ir zināma un aprakstīta Web servisa WSDL failā. Šis fails ir publiski pieejams. WSDL failā norāda kādas metodes ir Web servisam, kādi tām ir parametri, kādi ir parametru tipi, kādus ziņu pārraides protokolus atbalsta Web serviss utt. SOAP ir XML bāzēts protokols Web servisu komunicēšanai.

Web servisu tehnoloģiju, SOAP, WSDL šobrīd atbalsta gandrīz visas modernas operētājsistēmas, platformas un izstrādes rīki.

Var minēt šādas Web servisu priekšrocības:

- Uz pakalpojumiem balstīta arhitektūra;
- Vienkāršība;

- Universāls risinājums sadarbībai starp sistēmām;
- Neatkarība no platformas, operētājsistēmas un programmēšanas valodas;
- Apstrāde var būt gan sinhrona, gan asinhrona;
- Datu pārraidei visbiežāk tiek izmantots HTTP/HTTPS protokols;
- SOAP un WSDL ir pasaulē pieņemti standarti, nevis vienas komercfirmas īpašums.

Galvenais ieguvums no Web servisu tehnoloģijas izmantošanas VID un CSP informācijas apmaiņas procesā būs šī informācijas apmaiņas procesa automatizācija, noņemot noslogojumu no VID un CSP darbinieku resursiem.

Patlaban VID darbinieki līdztekus savu tiešo darba pienākumu veikšanai ar noteiktu regularitāti (reizi mēnesī) sagatavo CSP nepieciešamos pārskatus, kas ir darbietilpīgs process, jo sevī ietver:

- Pārskatu atjaunošanu ar datiem;
- Pārskatu pārveidošanu DBF. failu formātā;
- DBF. failu publicēšanu uz servera, kuram ir piekļuve CSP darbiniekiem.

Līdz ar to VID darbinieki tiek atrauti no savu tiešo darba pienākumu veikšanas. Ja tiktu izmantota Web servisu tehnoloģija, tad VID darbiniekus šajā informācijas apmaiņas procesā vairs nevajadzētu iesaistīt. Nepieciešamos pārskatus sagatavotu izstrādātājs, kurš tos publicētu VID Web servisā.

CSP galvenais ieguvums no Web servisu tehnoloģijas izmantošanas šajā informācijas apmaiņas procesā būtu iespēja sagatavot nepieciešamos pārskatus sev vēlamā laikā. Līdz šim CSP bija atkarīgi no VID darbiniekiem, kuri pārskatus sagatavoja reizi mēnesī, taču tagad CSP darbinieki pārskatus varēs sagatavot daudz biežāk (reizi nedēļā vai pat reizi dienā).

Šo iemeslu dēļ VID un CSP informācijas apmaiņas procesu ir ieteicams realizēt ar Web servisu tehnoloģiju.

Izmantotās literatūras un avotu saraksts

1. **Tanler R.** *The Intranet Data Warehouse: Tools and Techniques for Building an Intranet-Enabled Data Warehouse.* New York: John Wiley & Sons, 1997. 338 p.
2. **Loshin D.** *Business Intelligence: the savvy's manager's guide, getting onboard with emerging IT.* Amsterdam: Morgan Kaufmann Publishers, 2003. 270 p.
3. **Kimball R., Ross M.** *The Data Warehouse Toolkit: – The Complete Guide to Dimensional Modeling.* New York: John Wiley & Sons, 2002. 436 p.
4. **Moss L. T., Atre S.** *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications.* New York: Addison-Wesley Pub. Co, 2003. 576 p.
5. **Ponniah P.** *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals.* New York: Wiley-Interscience, 2001. 528 p.
6. **Inmon W.H.** *Building the Data Warehouse. – 3rd edition.* New York: John Wiley & Sons, 2002. 432 p.
7. *Ascential DataStage Documentation.* Paris: Ascential Software, 2002.
8. *BusinessObjects User's Guide.* Paris: BusinessObjects, 2003. 660 p.
9. **Kimball R. et al.** *The Data Warehouse Lifecycle Toolkit – Expert Methods for Designing, Developing and Deploying Data Warehouses.* New York: John Wiley & Sons, 1998. 771 p.
10. <http://www.service-architecture.com/>