Rīgas Tehniska universitāte

Datorzinātnes un Informācijas tehnoloģijas fakultāte

Otrais praktiskais darbs mācību priekšmetā

„Datu bāzu vadības sistēmas”

Datu vitrīnas veidošana, izmantojot OLAP datu kubu

Izstrādāja: Sergejs Terentjevs

2. kurss, 9.grupa

Apl.nr. 061RDB140

Pārbaudīja: lekt. V. Vinogradova

2006./2007. māc. g.

# ANOTĀCIJA

Praktiskajā darba tika izveidota datu vitrīna, izmantojot OLAP datu kubu. Izmantojot par datu avotu izstrādātu OLAP datu kubu, tika izstrādāti vairāki pārskati, kuri ļauj lietotājam veikt detalizētas informācijas analīzi un pieņemt lēmumus. Darba atskaitē tika iztirzāta problēma, aprakstīti būtiskākie izstrādes soļi un atrasts problēmas risinājums.

Praktiskais darbs tika izstrādāts MS Access 2007 un MS Excel 2003 uz *Pentium III*  tipā procesorā *Windows XP Proffesional* vidē.

Laboratorijas darba pārskata apjoms: 22 lappuses izklāsta tekstā, tajā ietilpst 7 nodaļas, literatūras saraksts un 27 attēli.

# SATURS

[ANOTĀCIJA 2](#_Toc198702687)

[SATURS 3](#_Toc198702688)

[UZDEVUMA NOSTĀDNE 4](#_Toc198702689)

[IZMANTOTAS TABULAS 5](#_Toc198702690)

[OLAP KUBA VEIDOŠANA 9](#_Toc198702691)

[PIVOTTABLE PĀRSKATU VEIDOŠANA 14](#_Toc198702692)

[SECINĀJUMI 21](#_Toc198702693)

[IZMANTOTA LITERATŪRA 22](#_Toc198702694)

# UZDEVUMA NOSTĀDNE

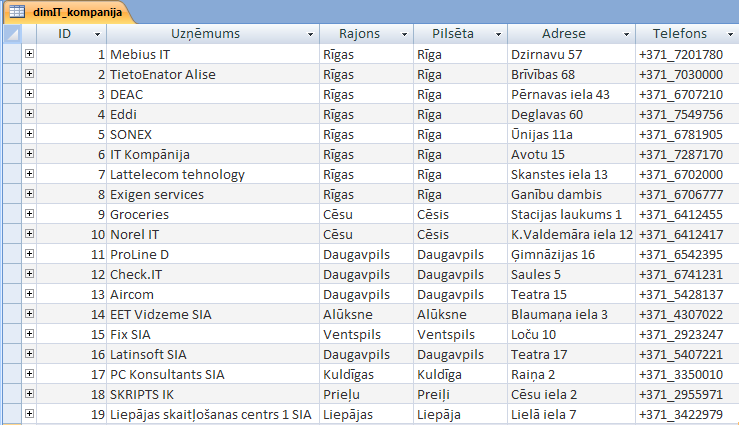
Izveidot datu vitrīnu, izmantojot OLAP datu kubu un izmantojot „zvaigznes” vai „sniegpārslas” struktūru (dimensiju skaitam jābūt vismaz četrām, faktu skaitām jābūt vismaz trīs). Pēc struktūras izveidošanas ievadīt datus un atrisināt kādu problēmu ar kuba pārveidojumiem (vismaz trīs piemēri).

Aprakstā iekļaut darba izstrādes būtiskos soļus, izvirzītu problēmu un rezultātu.

# IZMANTOTAS TABULAS

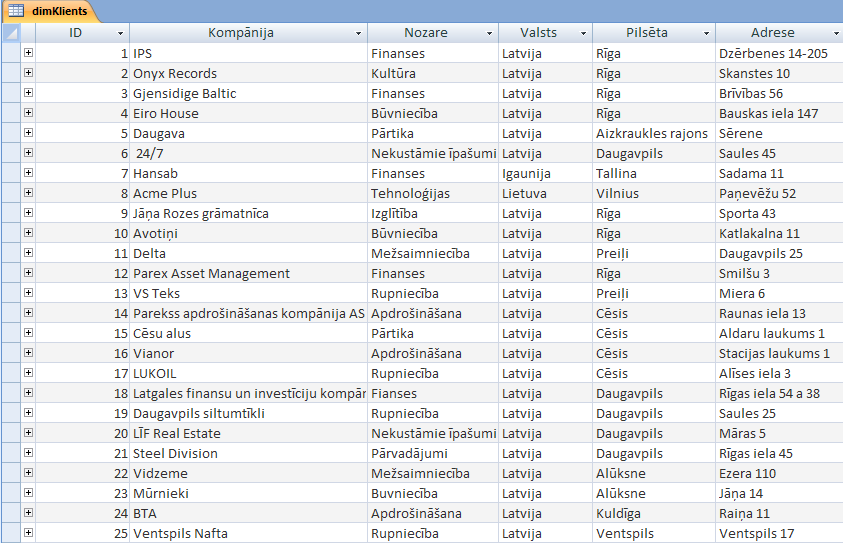
Uzdevumā nostādne ir paredzēta Latvijas tirgus izpētei, šoreiz IT pakalpojumu sfērā. Analizējot informāciju no vairākiem datu avotiem, tika atlasītas vairākas vadošās IT kompānijas un šo kompāniju īss apraksts. Bez tām tika noskaidroti pašlaik viss pieprasītākie IT pakalpojumi, kurus atbalsta un piedāvā šīs kompānijas. Tika nolemts izveidot Access datu bāzi, kura uzglabātu apkopoto informāciju pārskata periodā no 2007. pa 2008. gadam pa ceturkšņiem, šajā laika periodā tika nolemts apskatīt informāciju tikai par doto kompāniju viss izdevīgākajiem projektiem un papildus apkopot informāciju arī par klient kompānijām, kurām esošais projekts tika izstrādāts. Projekta ietvaros ir jāparedz detalizētu informācijas pārskatīšanu un analīzi, par šīs problēmas atrisinājumu tika nolemts izveidot OLAP datu kubu, ar kura palīdzību tiktu izveidoti vairāki Excel pārskati, kuros lietotājs var organizēt datus tā ka viņam uz doto brīdi ir nepieciešams, lai izanalizētu tirgus patreizējo stāvoklī.

Uzdevuma izpildes pirmais etaps bija izveidot Access datu bāzi ar nosaukumu *Services* (pakalpojumi) un tajā paredzēt IT kompāniju (*dimIT\_kompanija*), klient kompāniju (*dimKlients*), pakalpojumu (*dimPakalpojums*) un periodu (*dimPeriods*) tabulās, kuras turpmāk kalpos par OLAP kuba dimensijām un vienu galveno (centralizēto) faktu tabulu *Fakti*, kura saistīs šīs četras dimensiju tabulas. Tabulu veidošanai tika izmantota rīkjosla *Create*,kura piedāvā veidot gan tabulas, gan pārskatus, gan formas un vaicājumus Access 2007 vidē. Tabulas tika izveidotas *Design View* skatā, lietojot maskas, uzradot laukus kuri nevar būt tukši u.tml.



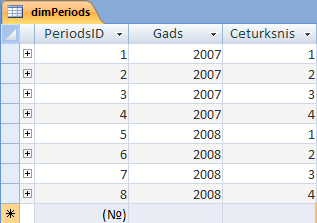
1. att. Tabula *dimIT\_kompanija*.

Tabula *dimIT\_kompanija* satur informāciju par vadošākajām IT kompānijām Latvijas tirgū minētāja laikā periodā.



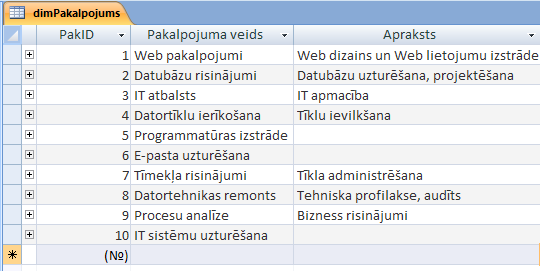
1. att. Tabula *dimKlients*.

Tabula *dimIT\_kompanija* uzglāba datus par klientu kompānijām, kuras ir vai nu bija vadošo IT kompāniju klienti.



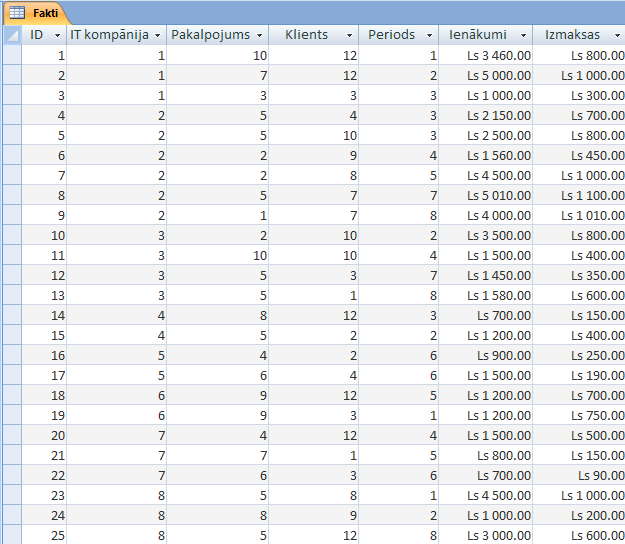
1. att. Tabula *dimPeriods*.

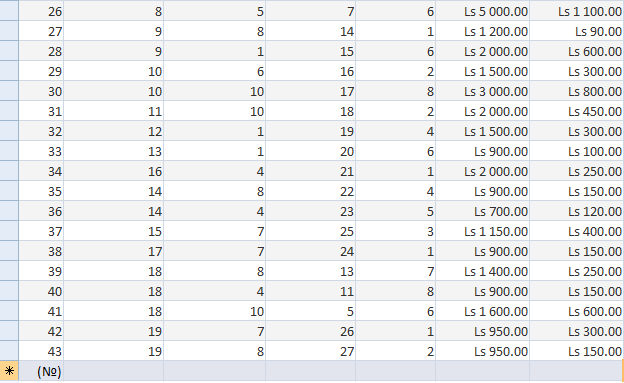
Tabula *dimPeriods* paredzēta pārskata perioda attēlošanai, t.i., tabulas raksti gads un ceturksnis uzrādīs viss izdevīgāk realizētu pakalpojumu šajā laikā periodā starp kādu vadošo IT kompāniju un klient kompāniju.



1. att. Tabula *dimPakalpojums*.

Tabula *dimPakalpojums* parāda vispieprasītākos IT pakalpojumus, kurus nodrošina vadošas IT kompānijas.

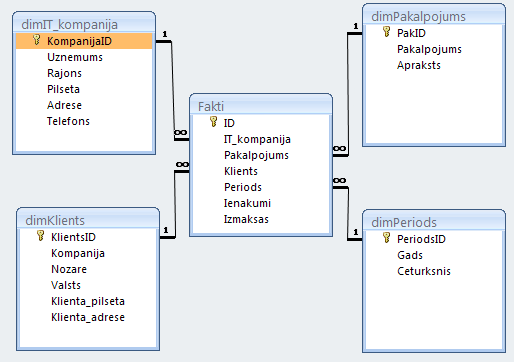




1. att. Tabula *Fakti*.

Tabula *Fakti* saista četras dimensijas tabulas un ir galvenā tabula, ka jau var pamanīt tā apkopo informāciju no visām tabulām, tāda veida tiek norādīts kādas vadošas IT kompānijas viss izdevīgais pakalpojums (projekts), kas tika realizēts kādai klient programmai. Bez tam tabula satur vēl divus papildus atribūtus – ienākumi un izmaksas, šajās kolonnas tiek uzradītas konkrēta projekta iegūtie ienākumi un izmaksas (patērēta degviela, programmas pārdošanas licenzēs pirkšana u.t.t). Dotiem papildus laukiem nebija nepieciešamības veidot atsevišķas tabulas, tāpēc tie ir izveidoti galvenajā tabulā.

Tabulas tika sasaistītas „sniegpārslas” struktūrā, izmantojot *Relationship* konstruktoru (skat. att. 6).

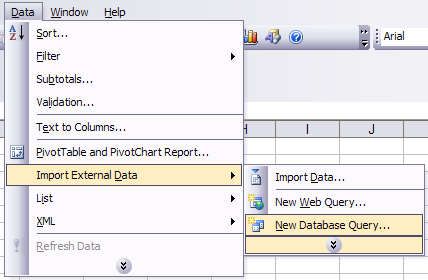


1. att. „Sniegpārslas” struktūra.

# OLAP KUBA VEIDOŠANA

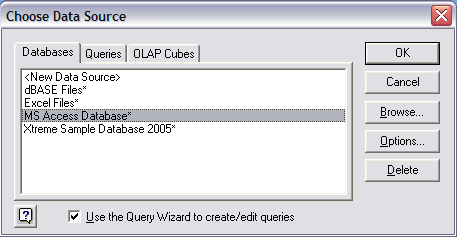
Tagad, kad mums ir datu bāze ar atbilstošam dimensijām un faktiem, varam konstruēt OLAP kubu, uz kurā pamata, ka jau iepriekš tika minēts, veidosim *PivotTable* pārskatus MS Excel. Doti pārskati ļaus analītiķim izanalizēt viss pelnošāku kompāniju, viss izdevīgāko un pieprasītāka pakalpojumu, klientu skaitu ārvalstis u.tml.

OLAP kuba izveidošanai, atveram MS Excel 2003 ( jaunāka Office 2007 versijā OLAP datu kubs vairs netiek publiski atbalstīts – neeksistē OLAP kuba veidošanas veidnis, kaut gan objektmodulis ļauj *PivotTable* pārskatus piesaistīt OLAP kubam). Šeit atveram *Data* izvēlni, no izvēlnes sarakstā izvēlāmies *Import External Data*, datu importēšanai no kāda datu avota. Atvērsies vēl viena izvēlne ar *Import Data* ( kāda speciāla datu avota savienojuma noradīšanai), *New Web Query* ( kādas informācijas importēšanai no Web avota), *New Database Query* ( jauna datu avota veidošana ) komandām, izvēlāmies trešo pēdējo (skat. att. ).



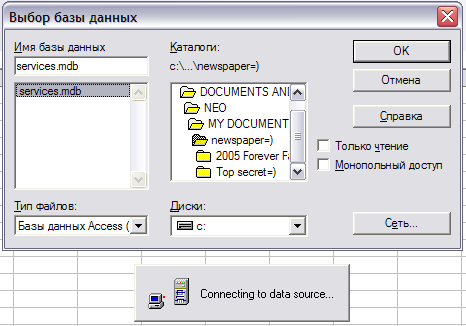
1. att. *New Database Query* komanda.

Atvērtajā *Choose Data Source* logā par datu avotu ir jānorāda MS Access datubāzi (skat. att. 8).



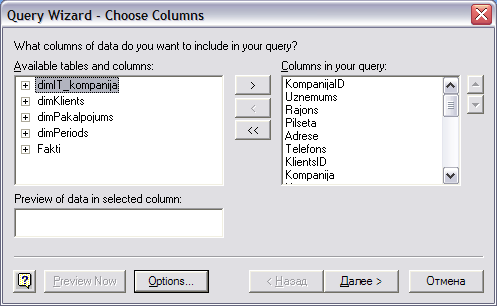
1. att. *Choose Data source* logs.

Pēc komandpogas *OK* nospiešanas atvērās pārlūks, kurā ir jānorāda mūsu izveidota Access datubāze. Sameklējot datu avotu, pārlūkā ir iezīme to (skat. att. 9 services.mdb) un jānospiež pogu *OK*, pēc kuras nospiešanas tiks izveidots savienojums.



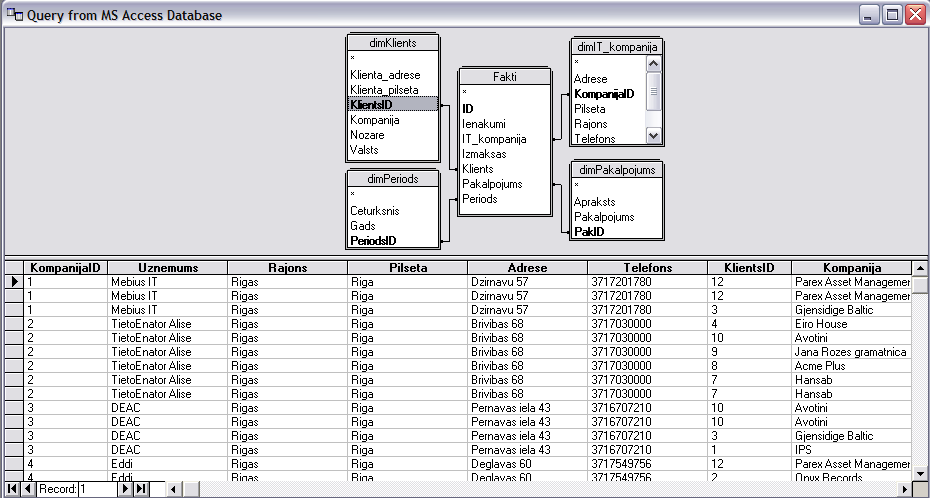
1. att. *Select DataBase* pārlūks.

Kad ir izveidots savienojums, atvērās *Query Wizard* forma. Formas kriesajā puse ir uzradītas datubāze pieejamas tabulas un iekš tabulām kolonu jeb tabulas atribūtu nosaukumu saraksts (bez tabulām forma vēl piedāvātu skatus – uz vaicājuma pamatā izveidotas tabulas, ka arī slēptas sistēmas tabulas – logā *Choose Data Source Options* formā bija jānorāda, lai tas tiktu rādītas ). Savukārt formas labajā pusē ir redzams *Colums in your query* teksta lauks, uz kuru ir jāpārvieto tabulu kolonnas, izmantojot pogu ar attiecīgo norādi „>”. Pilnvērtība vaicājuma izveidošanai, kurām ir jāimitē mūsu Access datu bāze, pārvietosim visas tabulas un kolonnas.



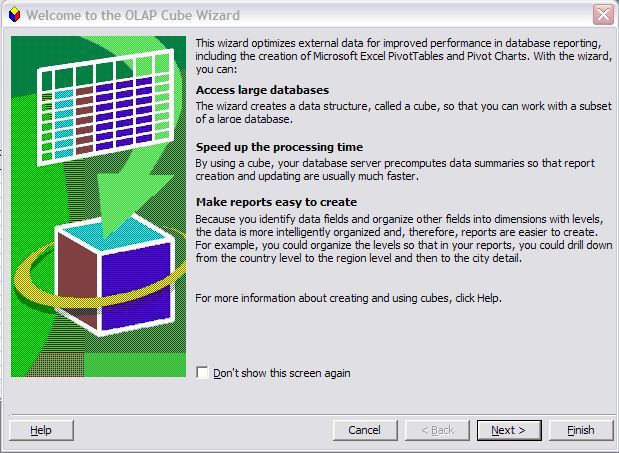
1. att. *Query Wizard* forma.

Nospiežot pogu *Next*, tiek izveidots vaicājums uz mūsu datubāzes pamatā (trūkums, *Query Wizard* nespēj saglabāt saites starp tabulām, tāpēc tas ir jāuzrāda atkārtoti). Tātad atvērās *Microsoft Query* formā, kurā atrodas *Query from MS Access Database* apakšforma (skat. att. 11). Apakšformā ir izvietotās mūsu faktu un dimensiju tabulas, pēc tabulu atkārtotas sasaistes, zemāk var redzēt tabulu ierakstu korektu atbilstību (pamatojoties uz *Fakti* tabulu).



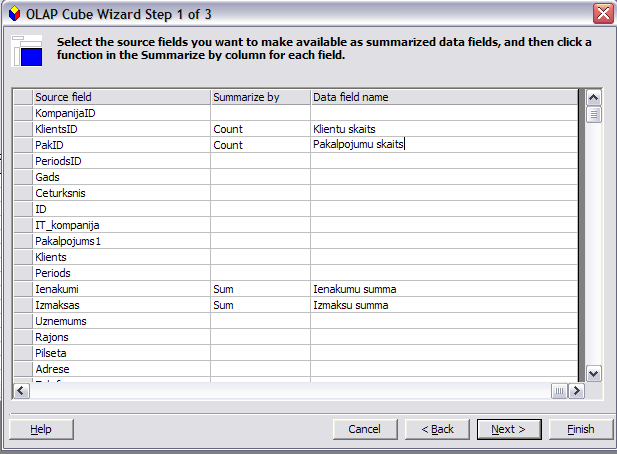
1. att. *Query from MS Access* apakšforma.

Izpildot šādus priekšnosacījumus, t.i., importējot MS Access datubāzi ieguvām vaicājumu, uz kurā pamata tagad konstruēsim daudzlīmeņu OLAP datu kubu. Formā *Microsoft* *Query* izvēlāmies *File* izvēlni un no turienes *Create OLAP Cube* komandu. Rezultāta tiks atvērts OLAP datu kuba veidošanas veidnis *OLAP Cube Wizard* (skat. att. 11).



1. att. *OLAP Cube Wizard* veidnis.

Pēc iepazīšanos ar veidni, spiežam pogu *Next*. Var pamanīt, ka veidnis mums piedāvās iziet trīs OLAP kuba projektēšanas soļus. Pirmajā solī ir jāizvēlas tabulu lauki, kuriem tiks pielietotas agregācijas funkcijas ( OLAP kubs paredzēts pirmkārt tieši integrēto datu iegūšanai un izvadei, kas tiek iegūti agregējot noteikta kuba apakšgabala datus). Dotie fakti jeb mēri (izvēlētie datu elementi) tiks izveidoti kuba krustpunktā starp dimensijām. Mainot kuba dimensiju vērtības, mainīsies arī fakti, attēlojot mums nepieciešamo informāciju par kādu mūs interesējošo objektu.



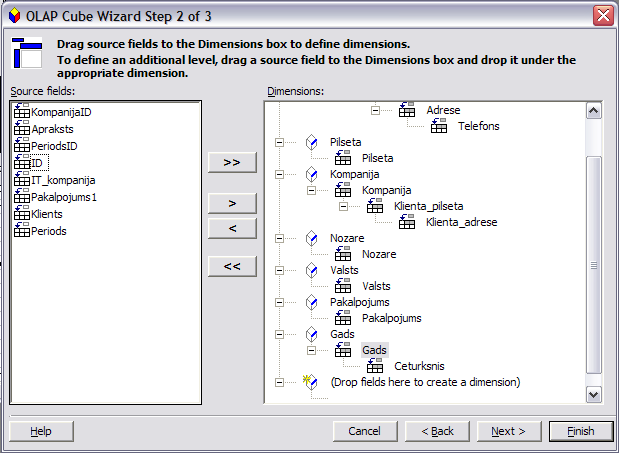
1. att. Faktunoradīšana.

Atbilstoši uzdevuma nostādnei, par faktiem tika izvēlēti *KlientsID* (klientu kompāniju uzskaitei), *PakID* (pakalpojumu uzskaitei), ienākumu un izmaksu lauki (kopējo ienākumu un izmaksu noteikšanai) *Source field* kolonnā. Blakus kolonnā *Summarize By* ir jānorāda agregācijas funkcijas, kuras tiks pielietotās atbilstošiem laukiem un trešā *Data field name* kolonnā ir jānorāda lauku nosaukumi.

Nākamajā otrajā solī ir jādefinē kuba dimensijās, kuras sastādīs kuba šķautnes. Atkarībā no izvēlētām dimensijām, fakti iegūs tās vai citas vērtības. Definēt dimensijas var no sarakstā *Source fields* pārvietojot attiecīgo tabulas kolonnu uz tekstā lauku *Dimensions* zem *Parent* elementa *Drop fields here to create a dimension*. Pēc pārvietošanas pārvietota kolonnā iegūst *Parent* elementa vērtību. Dimensijas var paredzēt, ka daudzlīmeņu, tās izdarāms ievietojot vēl vienu dimensiju zem *Parent* dimensijas (hierarhija).

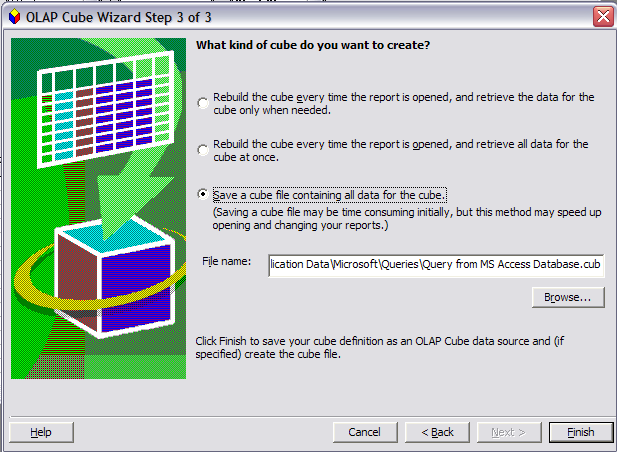
Tika izveidotas sekojošas dimensijas un līmeņi:

* IT kompānija: Uzņēmums🡪Adrese🡪Telefons (paredzēsim informācijas analīzei);
* Pilsēta: Pilsēta (paredzēsim datu filtrēšanai pēc pilsētām);
* Klient kompānija: Kompānija🡪Klienta pilsēta🡪Klienta adrese (datu analīzei);
* Nozare (datu filtrēšanai pēc nozares);
* Valsts (datu filtrēšanai pēc valstīm);
* Pakalpojums (datu analīzei);
* Periods: Gads🡪Ceturksnis (datu filtrēšanai pēc pārskata periodā).



1. att. Dimensiju definēšana.

Trešajā solī ir jānorāda kuba veids, kuru mēs vēlāmies izveidot. Mūsu gadījuma vispiemērotākais kuba veids būs tāds, kurš uzreiz sevi glabās visu informāciju (skat. att. 15). Šāds kubs tiks izveidots, ka atsevišķs fails ar paplašinājumu \*.cub un \*.ogy.

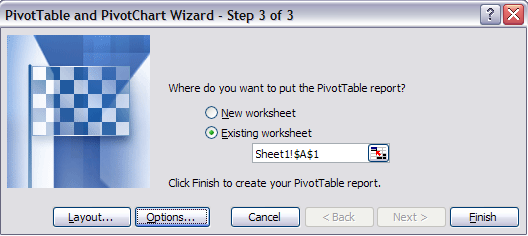


1. att. OLAP kuba veidošana un saglabāšana.

# PIVOTTABLE PĀRSKATU VEIDOŠANA

Pēc OLAP datu kuba izveidošanas un saglabāšanas *Microsoft Query* forma piedāvā atgriezties Excel dokumentā un izveidot pārskatus, izmantojot tikko izstrādātu OLAP daudz dimensiju kubu.

Tātad pēc noklusēšanas uzreiz tiek piedāvāts *Pivot Table* pārskatu veidošanas veidnis, kurš piedāvāta noradīt pārskata izvietojumu dokumentā – sakot no kādas noteiktas Excel šūnas vai nu ievietot pārskatu jaunajā Excel darba lapā, un satur divas svarīgas komandpogas *Layout*  un *Options* (skat. att. 16).



1. att. *Pivot Table* pārskatu veidošanas veidnis.

Pēc pogas *Layout* nospiešanas atveras veidņa *Layout* forma, kura piedāvā lietotājam izvietot faktus un dimensijas, tāda veida kāds ļautu organizēt nepieciešamo datu analīzi un pārskatīšanu. Tātad formas labajā puse tiek izveidotas dimensijas un fakti, kuros ir jāpārvieto uz attiecīgiem laukiem – ROW, DATA, COLUMN un PAGE. Šāds lauku izvietojums ir balstīts uz OLAP kuba pamatā – vidū fakti un šķautnes veido dimensijas.

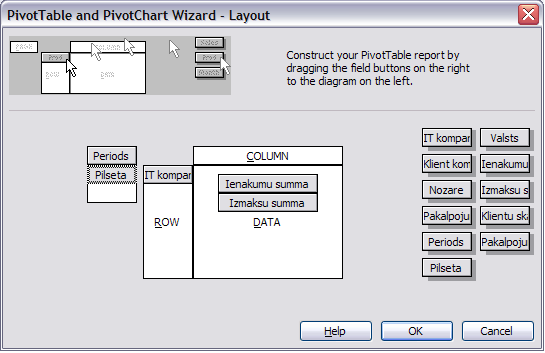
Lauks PAGE ir paredzēts datu atlasīšanai pēc ievietotas dimensijas vai vairāku dimensiju vērtībām. Jā dimensija ir daudzlīmeņu, tad var paredzēt datu detalizētāku atlasīšanu. Šajā laukā var ievietot tikai dimensijas.

Lauks ROW paredzēts dimensiju ievietošanai, pēc kuru vērtībām turpmāk tiek organizētas faktu vērtības. Parasti ievieto tas dimensijas, kurām tiek veikta analizē. Lauks ROW paredz grupēšanu pa rindām.

Lauks COLUMN analogi laukam ROW ievieto dimensijas, kurām tiek veikta datu analizē. Bet šoreiz dimensiju lauku vērtības attēlosies pārskata augšēja malā – kolonā un grupēšana tiek veikta pēc kolonam.

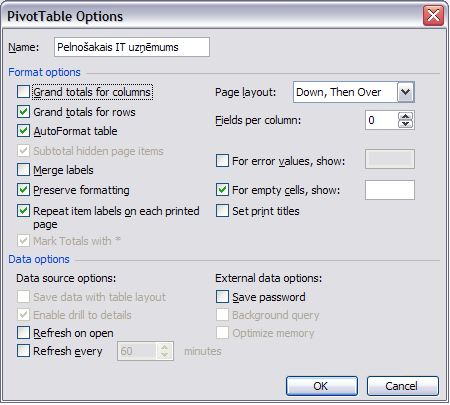
Lauks DATA paredzēts vienīgi faktu ievietošanai. Kā jau minēts fakti ir agregējamie lauki.

Izveidosim pārskatu, kurš ļautu izanalizēt, kurā no kompānijām ir viss pelnošāka un kurai ir viss lielākas izmaksas par visu apkopoto pārskata periodu. Tātad DATAlaukā ievietosim divus faktus - Ienākumu un Izdevumu summa un ROW laukā ievietosim vadošas IT kompānijas (IT kompānijas dimensiju). Datu uzskatāmībai PAGE laukā ievietosim Periods un Pilsēta dimensijas. Lauku COLUMN Pagaidām neaiztiksim (skat. 16. att.). Rezultāta jāsanāk pārskatām, kurš parāda IT kompāniju sarakstu un blakus kompānijām to kopējas izmaksas un ienākumus, pārskata augšēja daļā ir jābūt izvēlnei ar datu atlases iespējamību pa gadiem un ceturkšņiem un pilsētām. Izkārtojot dimensijas un faktus, spiežam OK.



1. att. *Pivot Table* pārskatu veidošanas veidnis.

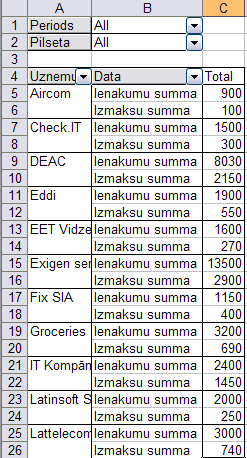
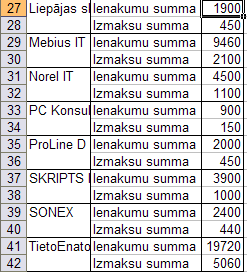
Bez lauku pārvietošanas, noradīsim arī papildus opcijas mūsu veidotājam *PivotTable* pārskatam, kas ir izdarāms *PivotTable Options formā,* kas tiek inicializēta nospiežot uz pogu *Options PivotTable* pārskatu veidošanas veidnī.



1. att. *Pivot Table* *Options* forma.

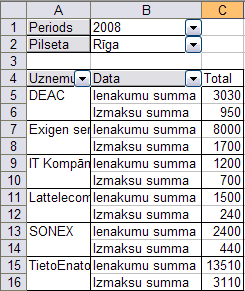
Forma piedāvā uzstādīt formatējuma opcijas un datu opcijas. Datu opcijas paredz pārskata atjaunošanu pēc kāda laika momenta – vai tiks atjaunots uzreiz pēc dokumenta atvēršanas (*Refresh on Open*) vai arī ik pēc 60 minūtēm. Formatējuma opcijas paredz pārskata izveidojuma noradīšanu, vai radīt kolonnu vērtību kopsummu apakšā zem laukiem (*Grand totals for column*) un rindu vērtību kopsummu (*Grand totals for rows*) u.t.t. Formas *Name* teksta laukā var noradīt pārskata nosaukumu. Veiksim dažas izmaiņas ( izmaiņas paradītas 18. att.) – nosauksim pārskatu un atslēgsim *Grand totals for column* opciju un nospiedīsim *OK*.

Pēc visu darbību izpildes spiežam pogu *Finish*, ka rezultātā Excel dokumentā tiek izveidots sekojošs *PivotTable* pārskats:



1. att. Pārskats Pelnošākais IT uzņēmums.

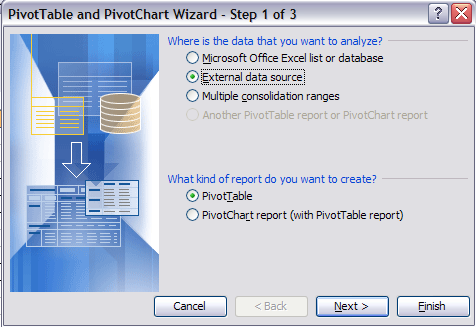
Vadoties pēc pārskata pelnošākais uzņēmums ir TietoEnator un šīm uzņēmumam arī ir viss lielākas izmaksas. Lai apskatītu Rīgas vadošo IT uzņēmumu izmaksas un ienākumus uz 2008 gadu, pārskata periodā laukā norādām vērtību 2008 un pilsētu Rīgu. Veicot šādu datu atlasi tika iegūts sekojošs rezultāts:



1. att. Rīgas IT uzņēmumu ienākumi un izmaksas par 2008. gadu.

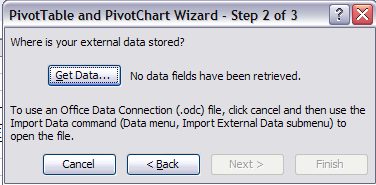
Vadoties pēc pārskata pelnošākais uzņēmums arī ir *TietoEnator IT Alise*.

Lai uzzinātu par viss pieprasītākajiem IT pakalpojumiem, piemēram, kāda noteiktajā nozare, izveidosim vēl vienu pārskatu. Atvērām *Data* izvēlni un izvēlāmies *PrivotTable and PivotChart Report* komandu, kā rezultāta atvērsis ar tādu pašu nosaukumu veidnis. Dotajā veidni ir jāizvēlas 2. opcija *External data source* un jānospiež pogu *Next* (skat. 21. att.).



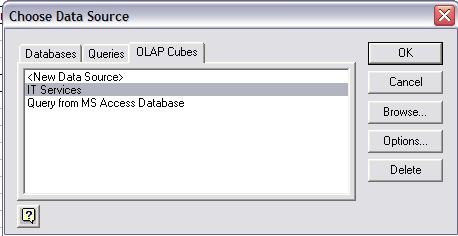
1. att. *PivotTable* *and* *PivotChart Report* *Wizard* veidnis.

Nākamajā otrajā solī ir jānorāda datu avots (mūsu izveidotājs OLAP kubs), kas ir izdarāms nospiežot uz pogas *Get Data* ( skat. att. 22.).



1. att. Poga *Get Data*.

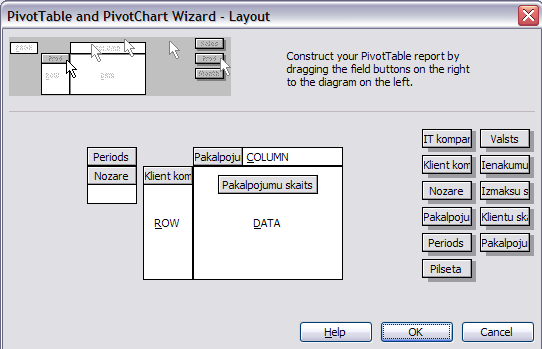
Pēc nospiešanas paradās *Choose Data Source* logs, kurā ir jāizvēlas *OLAP Cubes* ieliktnis un sarakstā jāatrod mūsu izveidotais datu kubs (skat. att. 23).



1. att. *Choose DataSource* logs.

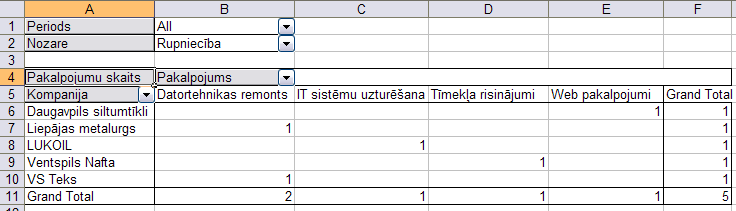
Izpildot iepriekš nosacītas darbības, mēs esam izveidojuši savienojumu ar OLAP datu kubu un trešajā soli parādās atkal jau pazīstamais un apskatītais *PivotTable* *and* *PivotChart Report* *Wizard* veidnis, kurā izvēlāmies pogu *Layout* (skat. att. 16).

Šoreiz PAGE laukā klāt Periods dimensijai ievietojām Nozares dimensiju, jo atlasīsim datus par nozarēm. ROW laukā ievietosim Klient kompāniju dimensiju un COLUMN laukā Pakalpojumi dimensiju, savukārt, DATA laukā ievietosim Pakalpojumu skaits faktu.



1. att. Pārskata par pieprasītākajiem pakalpojumiem konstruēšana.

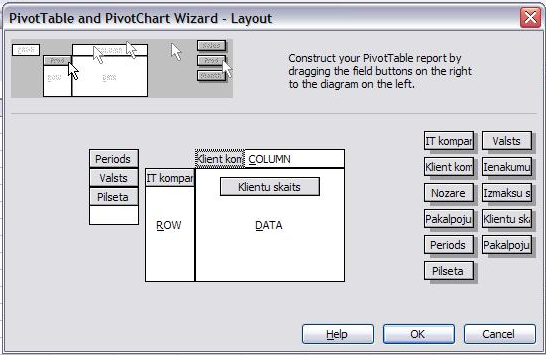
Rezultāta iegūstām pārskatu, kurš ļauj izanalizēt kādas nozares kādi IT pakalpojumi ir dominējoši. Piemēram, atlasīsim informāciju par kompānijām, kuru darbību sfēra ir rūpniecību nozare (skat. 25. att.).



1. att. Pārskats par pieprasītākajiem pakalpojumiem nozarēs.

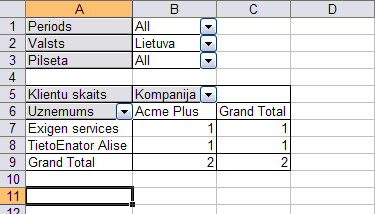
Spriežot pēc pārskata laikā periodā no 2007. un 2008. gadam tika pieprasīti Web pakalpojumi, tīmekļu risinājumi, IT sistēmu uzturēšana un datortehnikas remonts.

Izveidosim vēl vienu *PivotTable* pārskatu, kas ļaus noteikt klientu skaitu, kurus apkalpoja vai joprojām apkalpo kāda konkrēta IT kompānija noteiktajā laikā momentā. Bez tām paredzēsim detalizētas informācijas apskatīšanu, atlasot klientus pēc valsts un pilsētas. Tātad atlasīšanu veiksim pēc Perioda, Valsts un Pilsētas dimensijas, savukārt, klientu skaitu uzskaitīs fakts Klientu skaits. ROW laukā ievietosim IT kompānijas dimensiju un COLUMN laukā Klient kompānijas dimensiju (skat. att. 26).



1. att. Pārskata par klientu skaitu konstruēšanu.

Noteiksim cik klientu no Lietuvas apkalpo Latvijas IT kompānijas. Atlasot nepieciešamos datus tika konstatēts, ka Lietuva ir tikai viena kompānija, kas sadarbojās gan ar Exigen Services, gan TietoEnator kompāniju (skat. 27 att.).



1. att. Pārskats par klientu skaitu Lietuvā, kurus apkalpo IT kompānijas.

# SECINĀJUMI

Tika izstrādāts praktiskais darbs, kurā uzdevums bija izveidot datu vitrīnu, izmantojot OLAP datu kubu. Tika iepazītas jaunas DBVS iespējas, t.i., kā veidot OLAP analītiskas sistēmas un ar MS Excel palīdzību realizēt vairākas *PivotTable* pārskatu tabulas, kas ļauj lietotājam detalizēti izanalizēt apkopoto informāciju un pieņemt lēmuma kādas problēmas risināšanai.

Iepazinos ar jaunu veidu, ka ir iespējams saistīt vairākas tabulas, uzdevuma gadījumā ar „sniegpārslas” struktūru. Darbs bija radošs, jo darba gaitā tika apkopota informācija no vairākiem datu avotiem, līdz ar to bija interesanti organizēt un sistematizēt *PivotTable* pārskata tabulas datus vēlama rezultātā iegūšanai. OLAP datu kubs ir paredzēts datu apkopošanai, lietojot vairākas agregācijas funkcijas, šajā ziņa tas dod lielisko iespēju lielo datu apjomu analīzei.

# IZMANTOTA LITERATŪRA

1. V. Vinogradova, Lekciju konspekts priekšmetam ”Datu bāzes vadības sistēmas”, Rīga, RTU, 2008.