Latvijas Universitāte

Datorikas fakultāte

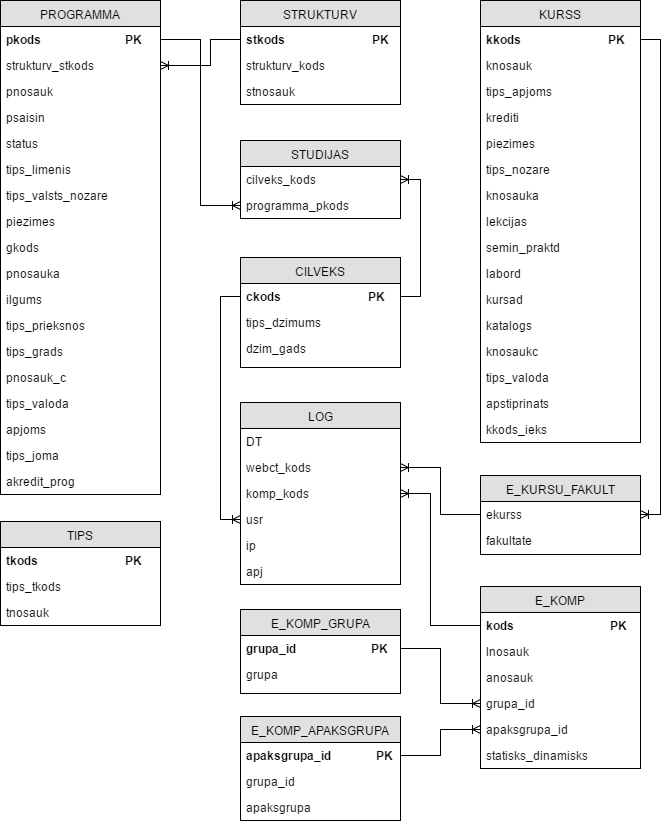
**Praktiskais darbs**

Kursā „DatZ6054: Datu noliktavu izvēlētas nodaļas”

**Autors:** Dāvis Sporāns DS16042

# Avots

Datu avotam par pamatu ņemti LUIS, WebCT, kā arī log faili atbilstoši praktiskā darba vajadzībām. Datu avots aprakstīts ar datubāžu relāciju shēmu (1. att.).

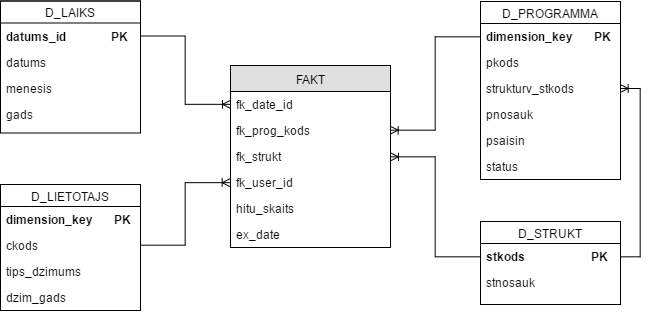


. 1. att. Datu avota relāciju shēma

# Zvaigznes shēma

Praktiskā darba pildīšanai uzdots uzdevums savākt datus:  
*Aktīvākie studenti pēc kopēja hitu skaita pa programmām.*

Uzdevuma risnāšanai tika izveidota zvaigznes shēma, tā aprakstīta ar relāciju diagrammu (2. att.).



2. att. Relāciju diagramma zvaigznes shēmai

# Zvaigznes shēmas hierarhijas

Dimensija “**D\_LAIKS**”

Gads 🡨 Mēnesis 🡨 Diena

Dimensija “**D\_PROGRAMMA**”

Fakultāte (Struktūrvienība) 🡨 Programma

# Dimensiju un faktu tabulu apraksts

1. tabula Dimensija “D\_ LIETOTAJS”

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mērķa tabula | Mērķa kolonna | Datu tips mērķa kolonnā | Avota tabula | Avota kolonna | Izteiksme |
| D\_LIETOTAJS | dimension\_key | NUMBER(10) | - | - | Automātiski ģenerēta sekvence |
|  | ckods | NUMBER(10) | Cilveks | ckods |  |
|  | tips\_dzimums | VARCHAR2(10) | Cilveks | tips\_dzimums |  |
|  | dzim\_gads | NUMBER(10) | Cilveks | dzim\_gads |  |

. tabula Dimensija “D\_ PROGRAMMA”

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mērķa tabula | Mērķa kolonna | Datu tips mērķa kolonnā | Avota tabula | Avota kolonna | Izteiksme |
| D\_PROGRAMMA | dimension\_key | NUMBER(10) | - | - | Automātiski ģenerēta sekvence |
|  | pkods | VARCHAR2(5) | Programma | pkods |  |
|  | strukturv\_stkods | VARCHAR2(12) | Programma | strukturv\_stkods |  |
|  | pnosauk | VARCHAR2(400) | Programma | pnosauk |  |
|  | psaisin | VARCHAR2(8) | Programma | psaisin |  |
|  | status | VARCHAR2(4) | Programma | status |  |

. tabula Dimensija “D\_ STRUKT”

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mērķa tabula | Mērķa kolonna | Datu tips mērķa kolonnā | Avota tabula | Avota kolonna | Izteiksme |
| D\_STRUKT | stkods | VARCHAR2(20) | Strukturv | stkods |  |
|  | stnosauk | VARCHAR2(100) | Strukturv | stnosauk |  |

4. tabula Dimensija “D\_ LAIKS”

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mērķa tabula | Mērķa kolonna | Datu tips mērķa kolonnā | Avota tabula | Avota kolonna | Izteiksme |
| D\_LAIKS | datums\_id | NUMBER(5) | - | - | Automātiski ģenerēta sekvence |
|  | datums | DATE | - | - | Automātiski ģenerēts |
|  | menesis | NUMBER(5) | - | - | Automātiski ģenerēts |
|  | gads | NUMBER(5) | - | - | Automātiski ģenerēts |

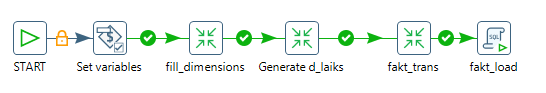
. tabula Fakts " FAKT"

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mērķa tabula | Mērķa kolonna | Datu tips mērķa kolonnā | Avota tabula | Avota kolonna | Izteiksme |
| FAKT | fk\_date\_id | NUMBER(5) | LOG | DT | Referncējas uz D\_LAIKS tabulu |
|  | fk\_prog\_kods | VARCHAR2(5) | Studijas | PROGRAMMA\_KODS | Inner join LOG  Where ckods=usr |
|  | fk\_strukt | VARCHAR2(12) | Strukturv | STKODS | Balstoties uz programmas kodu |
|  | fk\_user\_id | NUMBER(10) | LOG,  CILVEKS | USR, CKODS | Tiek savākts tikai tad, ja ir MAX(hitu\_skaits) programmā |
|  | hitu\_skaits | NUMBER(10) | LOG | - | Max(count(\*)) from LOG |
|  | ex\_date | DATE | - | - | Trunc(SYSDATE) |

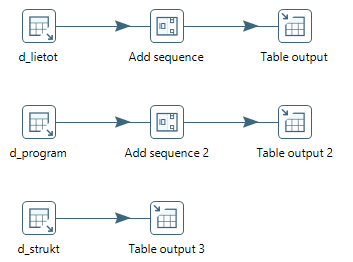
# Transformāciju apraksts

Lai, efektīvi savāktu apstrādātu un ievadītu datus datu noliktavā, tika izmantots ETL rīks “Pentaho”. Rīkā tika izveidotas 3 transformācijas un *load* darbs, izmantojot uzdevumā dotos ievades sākuma un beigu parametrus, kā arī WebCt kodu un student skaitu.

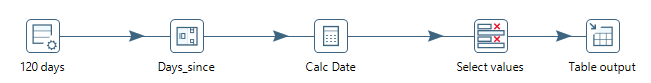
Visu transformāciju procesu pārvalda galvenais darbs “dwh\_prakt”:



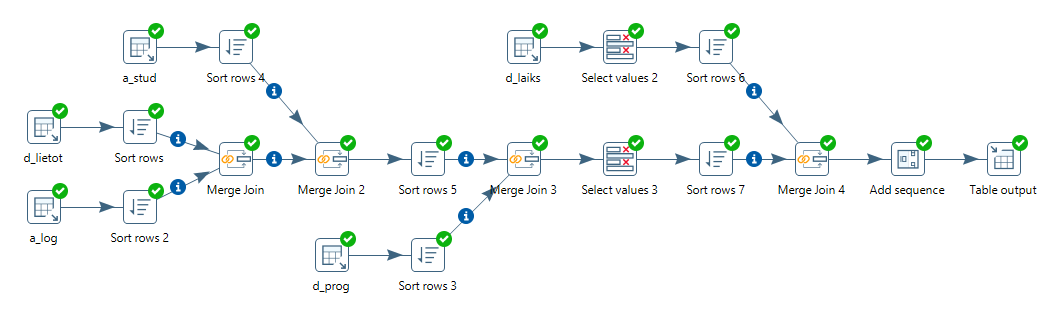
Transformācija “fill\_dimensions” aizpilda dimensiju tabulas (izņemot D\_LAIKS):



Transformācija “Generate d\_laiks” aizpilda laika dimensijas “D\_LAIKS” tabulu:



Transformācija “fakt\_trans” savāc datus faktu tabulai un veicot nepieciešamās transformācijas un noglabā tās temporārajā tabulā “tf\_fakt”.



Visbeidzot dati ar “Load” darbu tiek ievietoti “Fakt” tabulā. Load darba SQL komanda:

insert into fakt

(select fk\_date\_id, fk\_prog\_kods, fk\_strukt, fk\_user\_id,skaits,trunc(sysdate) from

(select fk\_date\_id,fk\_strukt,fk\_prog\_kods,fk\_user\_id,skaits,

rank() over (partition by fk\_date\_id,fk\_strukt,fk\_prog\_kods order by skaits desc) skaits\_rank

from

(select fk\_date\_id,fk\_strukt, fk\_prog\_kods, fk\_user\_id,count(f\_id\_seq) skaits

from tf\_fakt

group by fk\_date\_id,fk\_strukt, fk\_prog\_kods, fk\_user\_id))

where skaits\_rank <=${st\_skaits})

;