Datu publicēšanas platformas pilnveidošana saskaņā ar tehnisko specifikāciju

V. 1.0.1

VRAA.DPP.VDL.ADP.1.0.1.2019.08.19

Dokumenta autortiesības

Šo dokumentu Valsts reģionālās attīstības aģentūras (VRAA) uzdevumā ir izstrādājis A/S "RIX Technologies" projekta „”, Nr. 2.2.1.1/17/I/009, ietvaros. Dokuments ir izmantojams saskaņā ar 2018. gada 4. septembra līguma ID Nr. „ saskaņā ar tehnisko specifikāciju” nosacījumiem.

Izmaiņu vēsture

|  |  |  |  |
| --- | --- | --- | --- |
| Datums | Versija | Apraksts | Autors |
| 09.08.2019 | 1.0.0 | Dokumenta bāzes versija. | A/S RIX Technologies |
| 19.08.2019 | 1.0.1 | Papildinājumi, atbilstoši Pasūtītāju komentāriem. | A/S RIX Technologies |

Saturs

[1 Ievads 6](#_Toc17099618)

[1.1 Dokumenta nolūks 6](#_Toc17099619)

[1.2 Saistītie dokumenti 6](#_Toc17099620)

[2 Dbeaver 7](#_Toc17099621)

[2.1 Dbeaver lejuplāde 7](#_Toc17099622)

[2.2 Dbeaver uzstādīšana Windows 10 7](#_Toc17099623)

[2.3 Dbeaver uzstādīšana uz Linux Ubuntu 11](#_Toc17099624)

[2.4 Darbs ar Dbeaver rīku 12](#_Toc17099625)

[2.4.1 Jauna projekta izveidošana Dbeaver rīkā 12](#_Toc17099626)

[2.4.2 Jauna pieslēguma izveidošana 15](#_Toc17099627)

[2.4.3 Izejas datu pievienošana datu saglabāšanai mašīnlasāmā CSV failu formāta 19](#_Toc17099628)

[3 Libreoffice 23](#_Toc17099629)

[3.1 Libreoffice lejuplāde 23](#_Toc17099630)

[3.2 Libreoffice uzstādīšana Windows 10 23](#_Toc17099631)

[3.3 LibreOffice uzstādīšana uz Linux Ubuntu 25](#_Toc17099632)

[3.4 Darbs ar LibreOffice 26](#_Toc17099633)

[3.4.1 Excel datnes atvēršana ar LibreOffice Calc 26](#_Toc17099634)

[3.4.2 Izejas datu pievienošana datu saglabāšanai mašīnlasāmā CSV failu formāta 27](#_Toc17099635)

[4 Pentaho Data Integration 29](#_Toc17099636)

[4.1 Datu noliktavu piepildīšanai ar datiem.Pentaho Data Integration lejuplāde 30](#_Toc17099637)

[4.2 Pentaho Data Integration uzstādīšana Windows 10 30](#_Toc17099638)

[4.3 Pentaho Data Integration uzstādīšana Ubuntu 30](#_Toc17099639)

[4.4 Darbs ar Pentaho Data Integration 30](#_Toc17099640)

[4.4.1 Jauna transformāciju projekta izveidošana datu atlases rīkā 30](#_Toc17099641)

[4.4.2 MySql datu bāzes pievienošana 31](#_Toc17099642)

[4.4.3 Ieejas datu pievienošana transformāciju projektam 32](#_Toc17099643)

[4.4.4 Datu transformāciju pievienošana transformāciju projektā 39](#_Toc17099644)

[4.4.5 Izejas datu pievienošana transformāciju projektam 42](#_Toc17099645)

[4.4.6 Transformāciju izpildīšana 44](#_Toc17099646)

[4.4.7 Izejas datu pievienošana transformāciju projektam 47](#_Toc17099647)

[5 CSV datnes augšuplāde CKAN, izmantojot komandrindas rīku 47](#_Toc17099648)

[5.1 Ruby uzstādīšana 47](#_Toc17099649)

[5.1.1 Ruby uzstādīšana uz Windows darbstacijas 47](#_Toc17099650)

[5.2 Komandrindas rīka uzstādīšana 49](#_Toc17099651)

[5.3 Komandrindas rīka lietošana 49](#_Toc17099652)

[5.3.1 Komandrindas rīka konfigurācijas datne 49](#_Toc17099653)

[5.3.2 Komandrindas rīka izsaukšanas piemērs 50](#_Toc17099654)

# Ievads

## Dokumenta nolūks

Ieskats publiski pieejamo rīku Dbeaver, Libreoficce, Pentaho Data Integration lietošanā un uzstādīšanā.

## Saistītie dokumenti

Dokuments ir lasāms kopā ar šādiem dokumentiem:

1. Tehniskā specifikācija. Tehniskā specifikācija pakalpojumu iegādei no EIS kataloga pozīcijas CI110P.18.1. „Drupal pakalpojumi” „Datu publicēšanas platformas pilnveidošana”. VARAM un SIA “Datu Sistēmas”, 2018 (04.09.2018. Līguma Nr. 13-7/18/233 “Līgums par DARBA UZDEVUMA “Datu publicēšanas platformas pilnveidošana saskaņā ar tehnisko specifikāciju” izpildi”, Pielikums Nr.1).

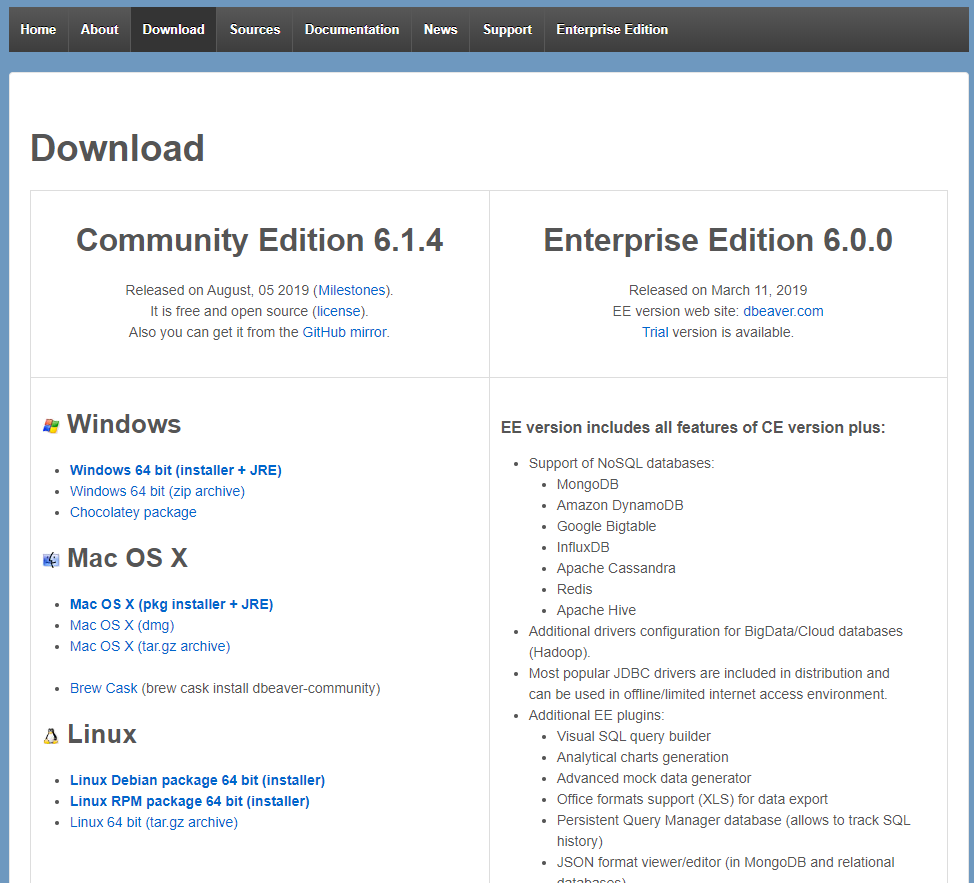
# Dbeaver

DBeaver ir bezmaksas atvērtā koda grafisks datu bāzes pārvaldības rīks datu bāzes izstrādātājiem un administratoriem. DBeaver var izmantot, lai pārvaldītu plašu datu bāzu vadības sistēmu (DBVS) klāstu.

Dbeaver programmatūra ir pieejama saitē “<https://dbeaver.io/download/>”.

## Dbeaver lejuplāde

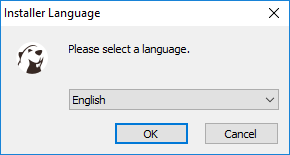
Saitē <https://dbeaver.io/download/>, atkarībā no lietotās operētājsistēmas, izvēlieties aktuālo Dbeaver versiju lejuplādei uzklikšķiniet uz saites. Lietotājam, kas izpilda instalāciju, jābūt administratora tiesībām.



Ilustrācija 1: Izvēlaties programmatūru atbilstoši platformai

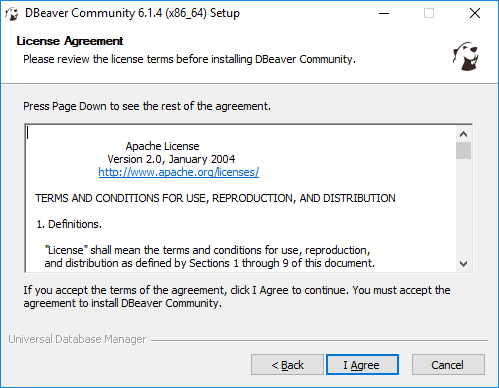
## Dbeaver uzstādīšana Windows 10

Windows 10 darbstacijā izpildiet (ar dubultklikšķi) lejuplādēto instalācijas pakotni. Pirmajā logā pēc instalācijas palaišanas izvēlieties kādu no instalācijas valodām (piemēram, angļu) un nospiediet pogu “OK”. Nākamajā logā nospiediet pogu “Next”.



Ilustrācija 2:Instalācijas valodas izvēle

Logā, kur jāapstiprina licences noteikumi, izlasiet Licences noteikumus un ar pogu “I Agree” apstipriniet tos.

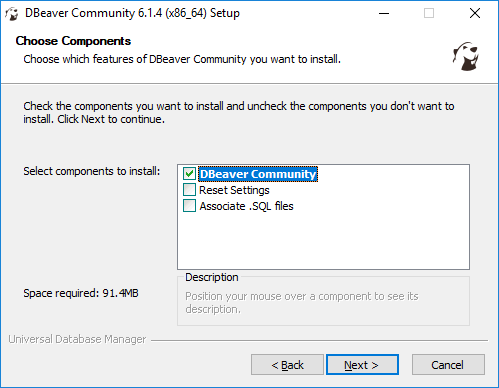


Ilustrācija 3:Licences noteikumi

Nākamajā logā nospiediet “Next”.

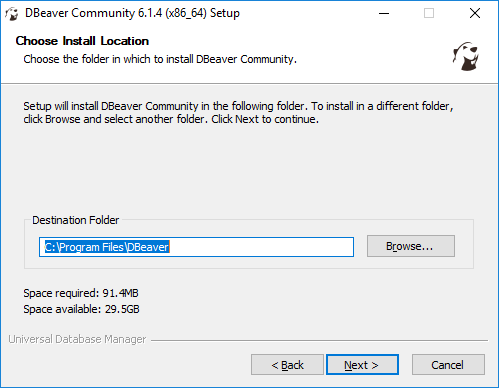
Ja parādās operētājsistēmas brīdinājumus, ka programmatūra tiks instalēta uz datora cietā diska, apstipriniet to.

Pēc apstiprināšanas atkārtoti izvēlieties instalēšanas valodu, nākamajā logā nospiediet “Next” un tad apstipriniet licences nosacījumus.



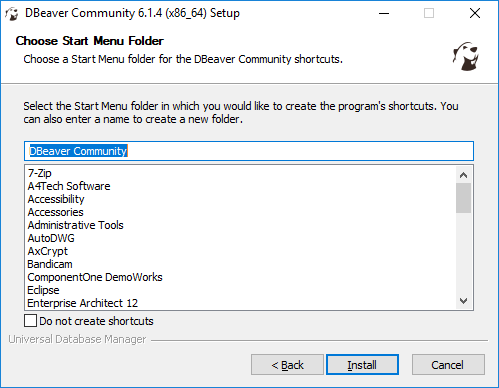
Ilustrācija 4:DBeaver komponenšu izvēle

Dbeaver komponenšu izvēles logā atstājiet noklusētās komponentes un tad nospiediet “Next”. Nakamajā logā izvēlieties vietu uz cietā diska, kur tiks instalēta programmatūra un nospiediet “Next”.



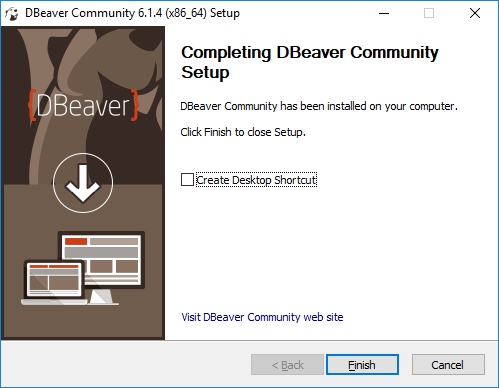
Ilustrācija 5:Programmatūras instalācijas vietas izvēle

Nākamajā logā nospiediet pogu “Install”.

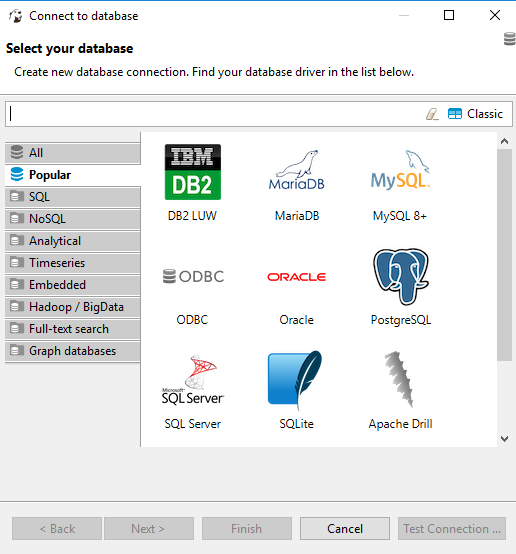


Ilustrācija 6: Programmatūras instalēšanas iniciēšana

Kad instalācija pabeigta, izvēlieties, vai uz darbvirsmas nepieciešama ikonas izveide, un pabeidziet instalēšanu ar pogu “Finish”.



Ilustrācija 7:Instalēšanas pabeigšana



Ilustrācija 8:Dbeaver datu bāzes izvēle Windows (pirmais logs pēc piestartēšanas)

## Dbeaver uzstādīšana uz Linux Ubuntu

Dbeaver darbībai nepieciešama Java. Pirms instalēšana to nepieciešams uzstādīt (pirms JAVA uzstādīšanas dažos gadījumos var būt nepieciešams atjaunot Ubuntu paku indeksu ar komandu sudo apt update). Viens no veidiem, kā uzstādīt JAVA, ir lietot komandu “sudo apt install default-jre”.

Pēc Java instalēšanas nepieciešams lejupielādēt jaunāko DBeaver RPM pakotni. Viens no veidiem, kā to izdarīt, ir lietot wget rīku. Ja uz operētājsistēmas tas nav uzstādīts, tad to no komandrindu rīka var izdarīt ar komandu “sudo apt-get install wget”.

Atveriet jaunu komandrindas termināli un izpildiet komandu:

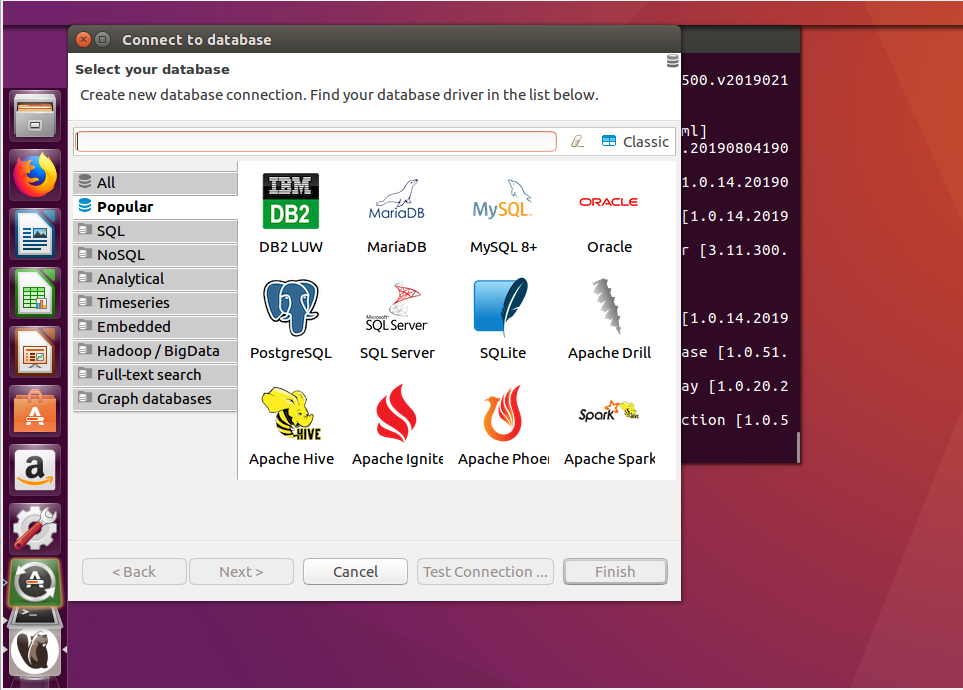
wget https://dbeaver.io/files/dbeaver-ce-latest-stable.x86\_64.rpm

Kad programatūras instalācijas pakotne ir lejuplādēta, var uzsākt instalēšanu ar rpm / dnf, alien vai yum pakotņu pārvaldnieku. Ja pakotņu pārvaldieki nav pieejami, tos var uzstādīt, piemēram, ar komandu “sudo apt-get install alien”. Instalēšanas piemērs, lietojot alien pakotņu pārvaldnieku:

**sudo alien -i ./dbeaver-ce-latest-stable.x86\_64.rpm**

Pēc veiksmīgas dbeaver uzstādīšanas, to no termināla var aktivizēt ar komandu:

$ dbeaver

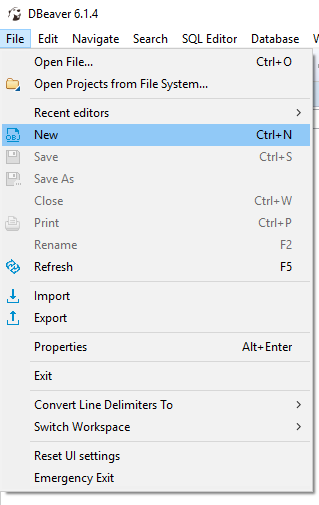


Ilustrācija 9:Dbeaver datu bāzes izvēle Ubuntu (pirmais logs pēc piestartēšanas)

## Darbs ar Dbeaver rīku

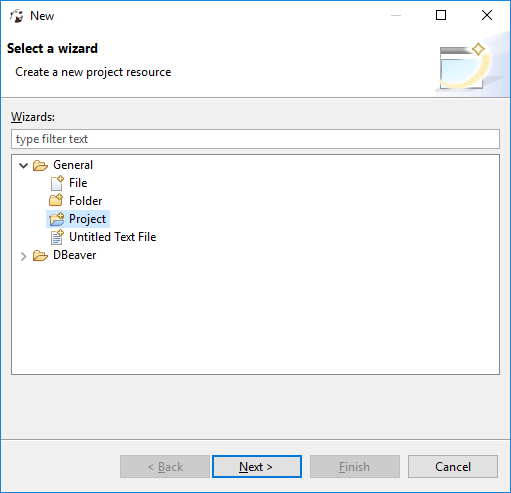
### Jauna projekta izveidošana Dbeaver rīkā

Lai izveidotu jaunu projektu, kopējā izvēļņu blokā uzklikšķina uz izvēlnes File->New:



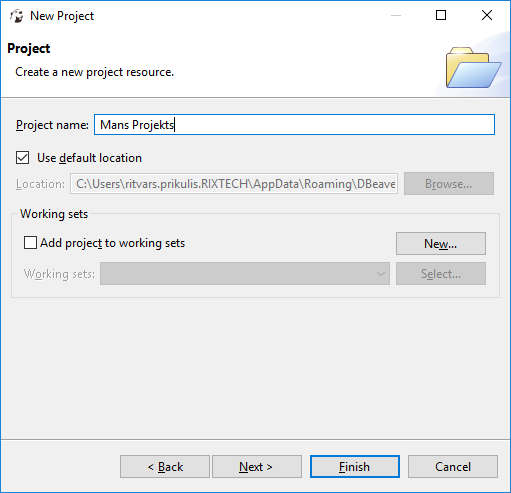
Ilustrācija 10: Jauna Dbeaver projekta izveide

Iznirstošajā logā izvēlas “General->Project” un nospiež pogu “Next”:



Ilustrācija 11:Dbeaver projekta izvēlne

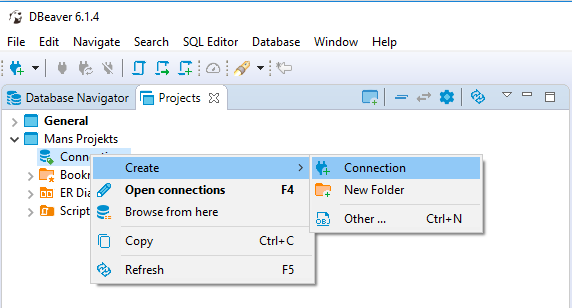
Nākamajā logā ieraksta projekta nosaukumu, piemēram, “Mans projekts” un nospiež “Finish” pogu:



Ilustrācija 12:Projekta nosaukuma ievade

### Jauna pieslēguma izveidošana

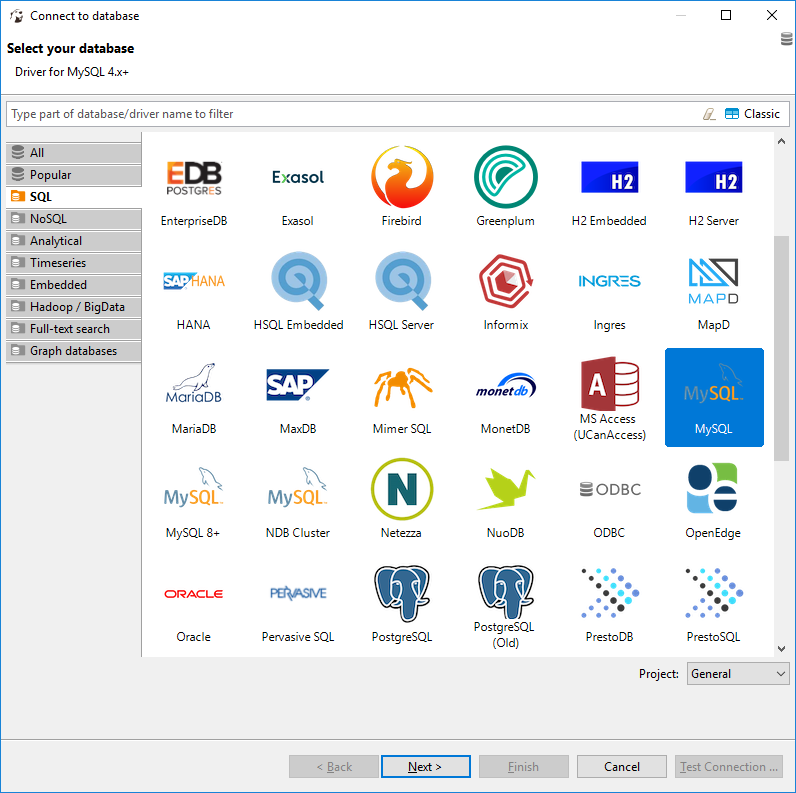
Dbeaver “Projects” navigācijas blokā (ja ekrānā nav redzams, to var iespējot “Window” izvēlnē), zem “Mans Projekts” (Projekta nosaukums, kuru izveidojāt) atrod pogu (izvēlni) “Connections” un, pozicionējot uz pogas peles kursoru, nospiež peles labo taustiņu. Tad izvēlas “Create->Connection”:



Ilustrācija 13:1.4.2 Jauna datubāzes pieslēguma izveidošana

#### Jauna Mysql datubāzes pieslēguma izveidošana

Pēc nodaļā “2.4.2 Jauna pieslēguma izveidošana” aprakstītā pieslēguma izveidošanas izvēlas pieslēguma veidu. Šajā piemērā kā datu avotu izvēlēsimies MySQL. Pēc datu avota izvēles nospiež pogu “Next”:



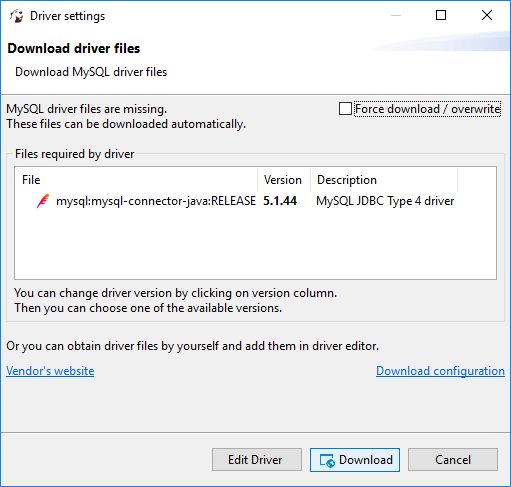
Ilustrācija 14:Mysql datubāzes pieslēguma izveidošana

Nākamajā logā jānorāda Datubāzes pieslēgšanās informācija. Ja datu bāze uzstādīta uz tā paša servera un root lietotājam nav uzstādīta parole, tad logu var aizpildīt šādi un nospiest pogu “Test Connection”:



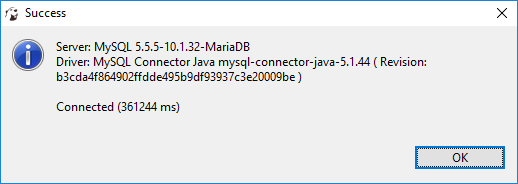
Ilustrācija 15: Datu bāzes pieslēguma informācijas norādīšana

Ja uz darbstacijas nebūs uzstādīti pieslēguma izveidei nepieciešamie draiveri, tiks attēlots Draiveru uzstādīšanas logs. Draiverus uzstāda, nospiežot pogu “Download”:



Ilustrācija 16:Mysql Draiveru uzstādīšana

Pēc draiveru uzstādīšanas vai ja tie jau bija uzstādīti tiks parādīts paziņojums par veiksmīgu pieslēgšanos:



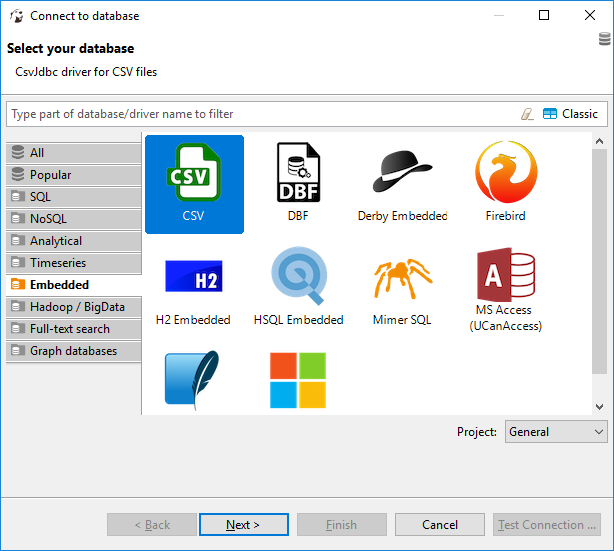
Ilustrācija 17: Veiskmīga pieslēgšanaš pie MySql datu bāzes

Ja pieslēgšanās ir veiksmīga, to apstiprina ar pogas “Ok” nospiešanu un datu bāzes pieslēguma izveides logā nospiež “Finish”.

#### Jauna Excel(CSV) pieslēguma izveidošana

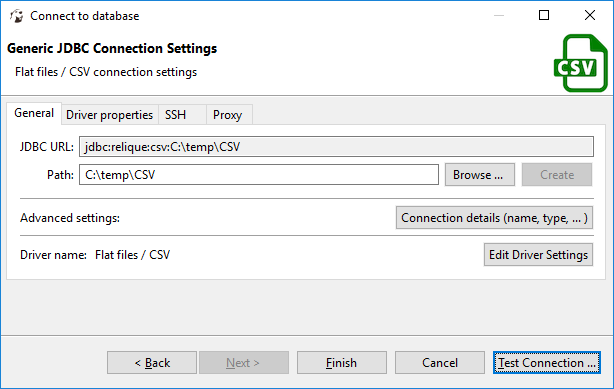
Dbeaver tiešā veidā nenodrošina datu apstrādi ar MS Excel. Bet nepieciešamības gadījumā iespējams Excel datni saglabāt kā .csv datni un izveidot CSV pieslēgumu.

Pēc nodaļā “2.4.2 Jauna pieslēguma izveidošana” aprakstītā pieslēguma izveidošanas jāizvēlas pieslēguma veids. Šajā piemērā izvēlēsimies CSV kā datu avotu. Pēc datu avota izvēles jānospiež “Next”:



Ilustrācija 18: CSV pieslēguma izveidošana

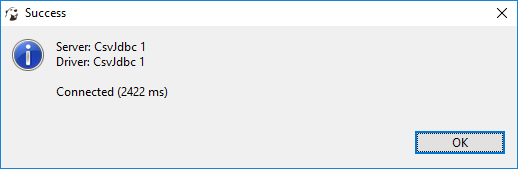
Nākamajā logā norāda ceļu (Path) līdz mapei, kurā atrodas CSV datnes, un nospiež pogu “Test Connection”:



Ilustrācija 19: CSV pieslēguma informācijas norādīšana

Ja uz darbstacijas nebūs uzstādīti pieslēguma izveidei nepieciešamie draiveri, tad tiks attēlots draiveru uzstādīšanas logs. Draiverus uzstāda, nospiežot pogu “Download”.

Pēc draiveru uzstādīšanas vai ja tie jau bija uzstādīti, tiks attēlots logs par veiksmīgu pieslēgšanos:



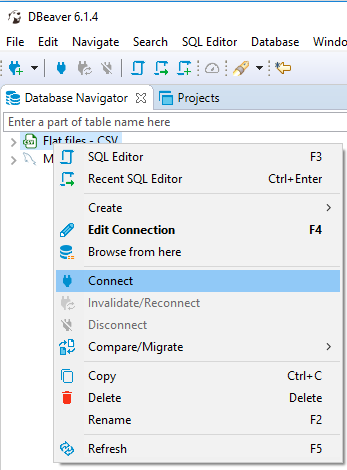
Ja pieslēgšanās ir veiksmīga, to apstiprina, nospiežot pogu “Ok”. Datu bāzes pieslēguma izveides logā nospiež “Finish”.

### Izejas datu pievienošana datu saglabāšanai mašīnlasāmā CSV failu formāta

Lai sagatavotu datus saglabāšanai mašīnlasāmā formātā, sākotnēji nepieciešams izveidot pieslēgumus datu avotiem (skatīt nodaļu “2.4.2 Jauna pieslēguma izveidošana”).

#### EXCEL (CSV) izejas datu sagatavošana saglabāšanai

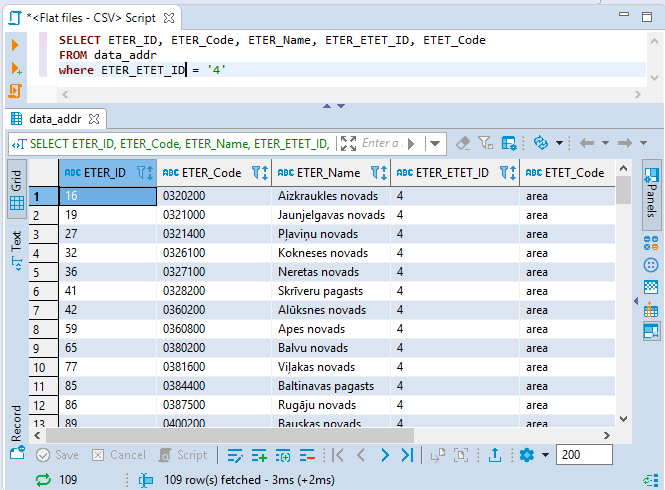
Uzsākot darbu, nepieciešams pieslēgties CSV datu avotam. Dbeaver “Database Navigator” navigācijas blokā (ja ekrānā netiek attēlots, to var iespējot “Window” izvēlnē), jāizvēlas nepieciešamais datu avots, pozicionējot uz tā peles kursoru un nospiežot peles labo taustiņu. Tad jānospiež “Connect”:



Ilustrācija 20:Pieslēgšanās CSV datu avotam

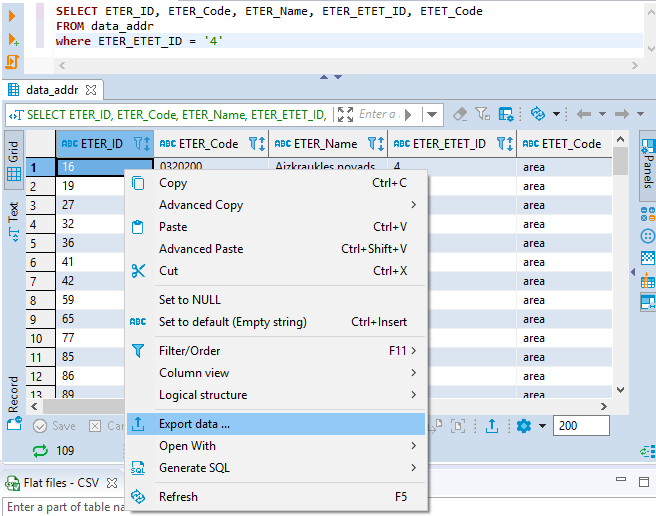
Tālāk atver “Dbeaver SQL Editor” rīku, piemēram, nospiežot klaviatūras taustiņu “F3”.

“Dbeaver SQL Editor” rīkā ievada SQL vaicājumu datu atlasei un to izpilda, piemēram, ar taustiņu kombināciju “Ctrl + Enter”:



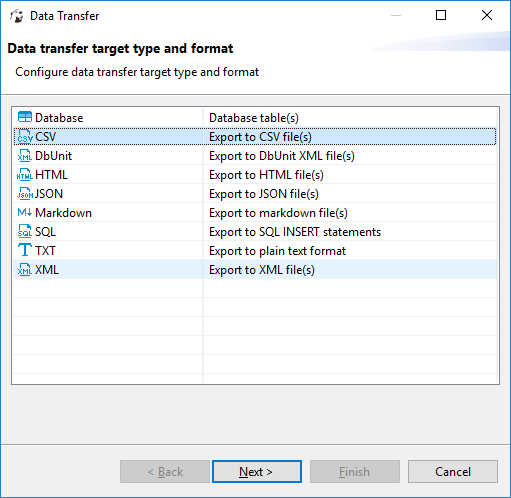
Ilustrācija 21:Eksportējamo CSV datu atlase

Kad eksportam nepieciešamie dati atlasīti, ar peles labo taustiņu uzklikšķina atlases rezultātam un nospiež pogu “Export data”:



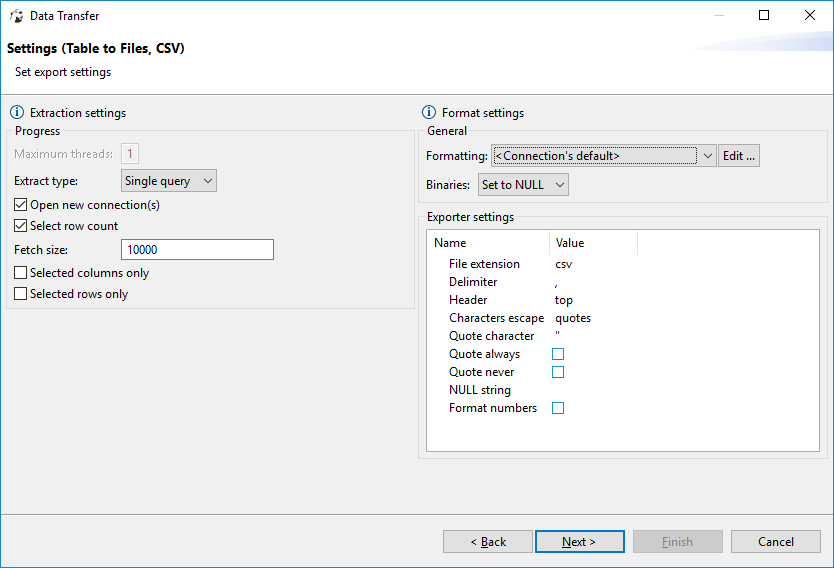
Ilustrācija 22: CSV datu eksportēšanas iniciēšana

Datu nodošanas logā izvēlas “CSV” un nospiež “Next”:



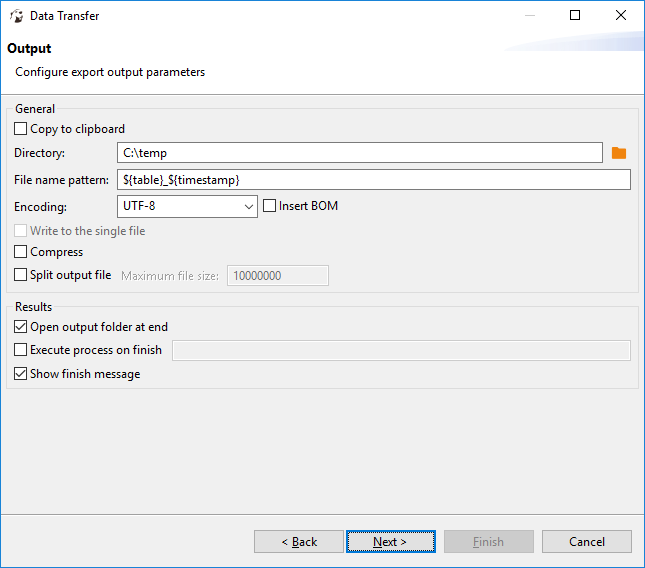
Ilustrācija 23: Eksporta formāta izvēle

Pārbauda, vai noklusētie Dbeaver eksporta parametri atbilst RFC4180 standartam (nepieciešamības gadījumā veic korekcijas) un nospiež pogu “Next”:



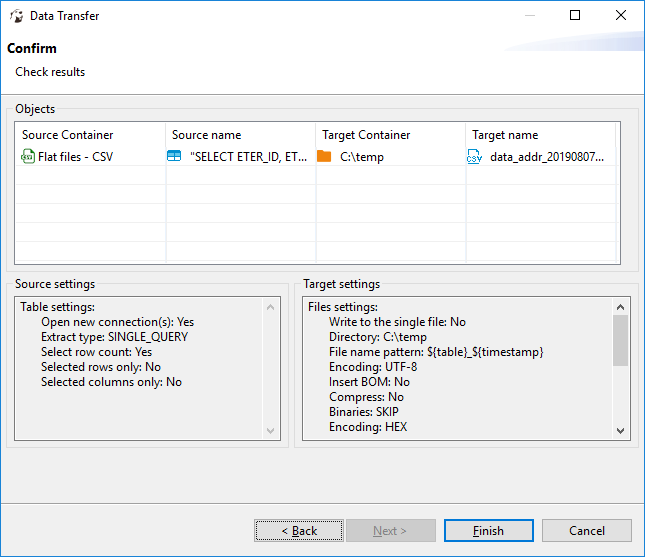
Ilustrācija 24:Eksporta parametru aizpilde

Norāda mapi, uz kuru eksportēt datni, un nospiež pogu “Next”:



Ilustrācija 25:Eksporta mapes norādīšana

Apstiprināšanas logā nospiež “Finish”:



Ilustrācija 26:Eksporta pabeigšana

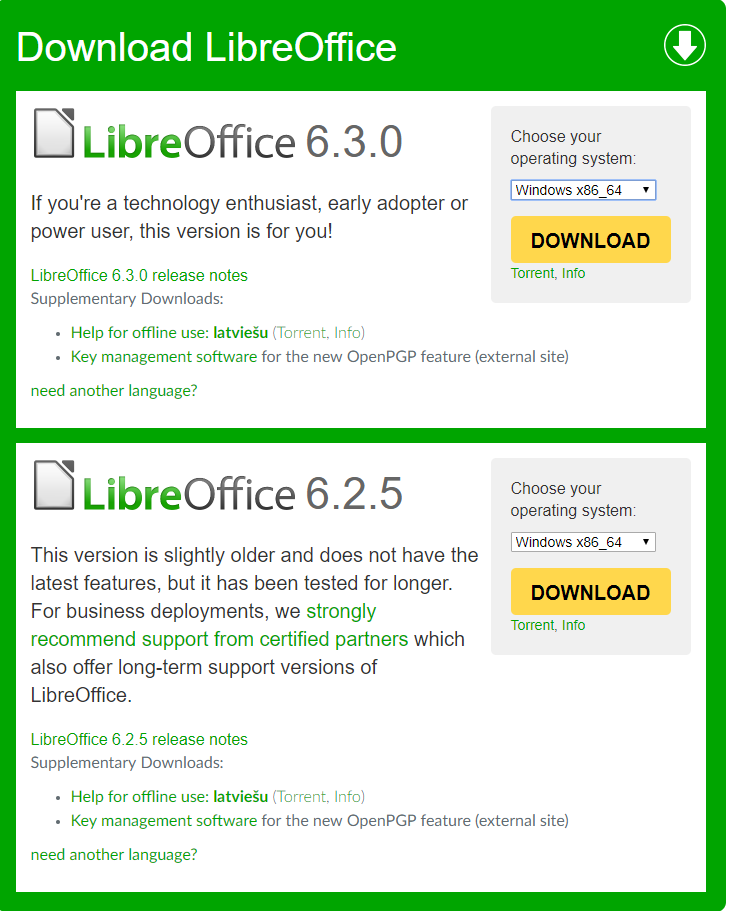
# Libreoffice

LibreOffice ir jaudīgs un bezmaksas biroja komplekts, ko izmanto miljoniem cilvēku visā pasaulē. LibreOffice interfeiss un funkcijām bagātie rīki palīdz atbrīvot jūsu radošumu un paaugstina produktivitāti ikdienas darbā. LibreOffice ietver vairākas lietojumprogrammas, kas padara to par universālāko bezmaksas un atvērtā koda biroja komplektu tirgū: Writer (tekstapstrāde), Calc (izklājlapas), Impress (prezentācijas), Draw (vektorgrafika un blokshēmas), Base (datu bāzes) un Math (formulu rediģēšana).

LibreOffice pieejams saitē <https://www.libreoffice.org/download/download/><https://dbeaver.io/download/>.

## Libreoffice lejuplāde

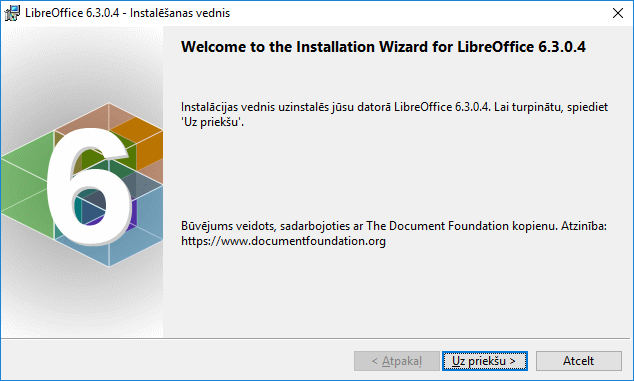
Saitē <https://www.libreoffice.org/download/download/><https://dbeaver.io/download/> atkarībā no lietotās operētājsistēmas izvēlas aktuālo Libreoffice versiju lejuplādei, uzklikšķinot uz saites. Lietotājam, kas izpilda instalāciju, jābūt administratora tiesībām.



Ilustrācija 27: Izvēlaties programmatūru atbilstoši platformai

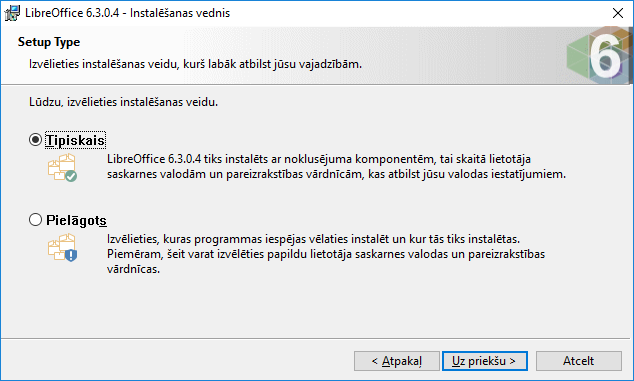
## Libreoffice uzstādīšana Windows 10

Windows 10 darbstacijā izpilda (ar dubultklikšķi) lejuplādēto instalācijas pakotni. Pirmajā logā pēc instalācijas palaišanas nospiež pogu “Uz priekšu”:



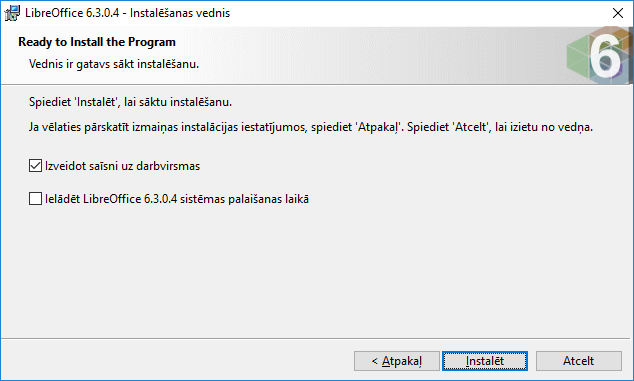
Ilustrācija 28:LibreOffice uzstādīšanas uzsākšana

Otrajā logā izvēlas instalēšanas viedu “Tipiskais” un nospiež pogu “Uz priekšu”:



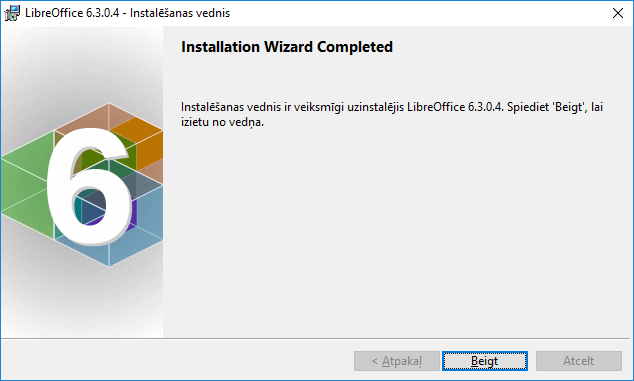
Ilustrācija 29:Instalēšanas veida izvēle

Nākamajā logā nospiež pogu “Instalēt”:



Ilustrācija 30:Instalēšanas iniciēšana

Ja operētājsistēma prasa apstiprināt programmatūras uzstādīšanu, tad apstiprina to. Kad programmatūra uzstādīta, instalēšanu pabeidz ar pogas “Beigt” nospiešanu:



Ilustrācija 31:Instalēšanas pabeigšana

Nepieciešamības gadījumā jārestartē darbstacija.

## LibreOffice uzstādīšana uz Linux Ubuntu

Terminālī aktivizē komandu: snap install libreoffice.

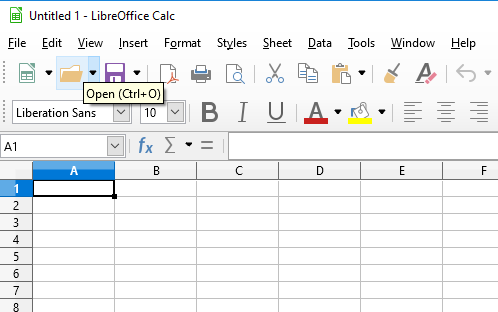
Ja tiek prasīts apstiprināt lietotāja paroli, izdara to. Lietotājam jābūt administratora tiesībām.

## Darbs ar LibreOffice

LibreOffice līdzīgi kā Microsoft Office lieliski noder tekstuālu datņu apstrādei. Datu kopu izveidei un uzturēšana lietderīgi izmanto LibreOffice Calc (MS Excel analogs).

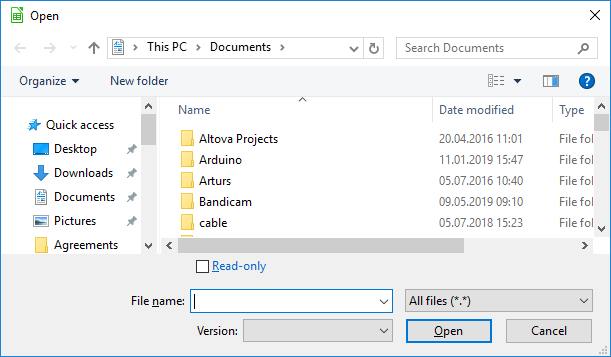
### Excel datnes atvēršana ar LibreOffice Calc

Lai atvērtu datnes ar LibreOffice Calc, jāuzspiež uz atvērtas mapes ikonas vai jānospiež taustiņu kombinācija “Ctrl + O”:



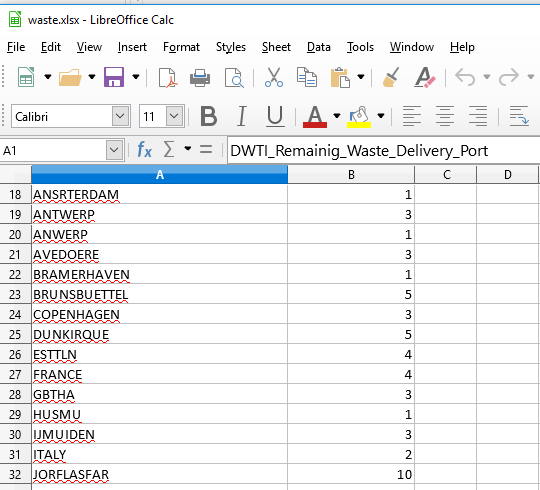
Ilustrācija 32:Datnes atvēršana

Datņu atvēršanas logā jāatrod Excel datne, ar peli tai jāuzklikšķina un jānospiež poga “Open”:



Ilustrācija 33: Datnes meklēšana

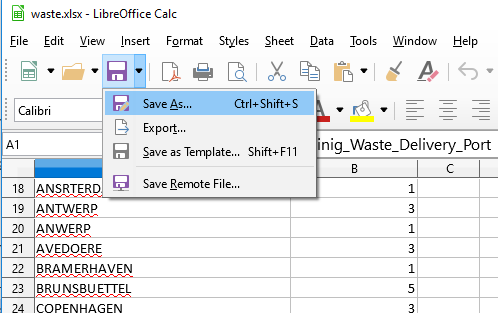
Tiks atvērta Excel datne:



Ilustrācija 34:Atvērta Excel datne

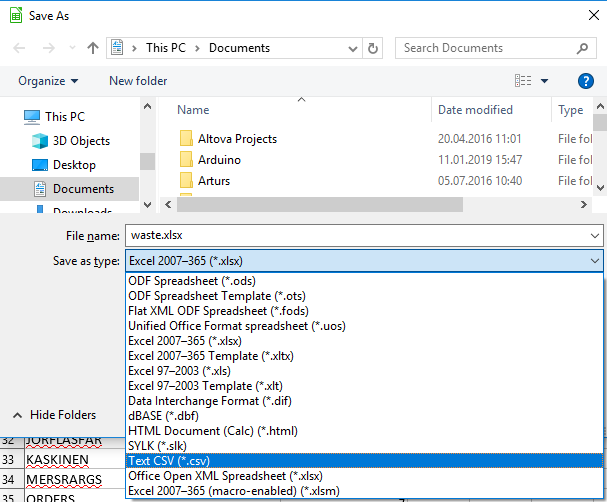
### Izejas datu pievienošana datu saglabāšanai mašīnlasāmā CSV failu formāta

Jebkurus datus, kas sagatavoti LibreOffice Calc rīkā, var saglabāt kā mašīnlasāmus datus. Lai to izdarītu, jānospiež poga “Save As” vai taustiņu kombinācija “Ctrl + Shift + S”:



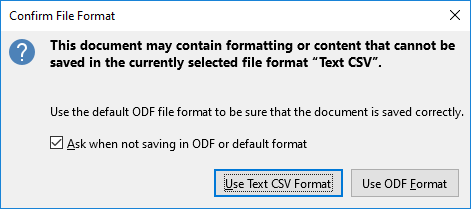
Ilustrācija 35:Datu saglabāšana

Datu saglabāšanas logā izvēlas datu formātu, kādā saglabāt datus – izvēlas “CSV” un nospiež pogu “Save”:



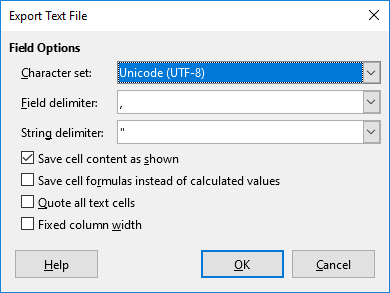
Ilustrācija 36:Datu saglabāšana csv formātā

Ja parādās logs par datnes formāta apstiprināšanu, nospiež pogu “Use Text CSV Format”:



Ilustrācija 37:Datnes formāta apstiprināšana

Nakamajā solī nepieciešams norādīt eksporta parametrus. Ja tie neatbilst RFC4180 standartam, tad nepieciešams veikt korekcijas un nospiest “Ok”:



Ilustrācija 38:Eksporta parametru norādīšana

# Pentaho Data Integration

Pentaho Data Integration rīks nodrošina datu izvilkšanas, transformēšanas un ielādes (ETL) iespējas.

Pentaho Data Integration parasti tiek izmantots:

* Datu migrācijai starp dažādām datu bāzēm un lietojumprogrammām;
* Lielu datu kopu ielādei datu bāzēs;
* Datu tīrīšanai pa soļiem, sākot no ļoti vienkāršām līdz ļoti sarežģītām transformācijām;
* Datu integrācijai.

## Datu noliktavu piepildīšanai ar datiem. Pentaho Data Integration lejuplāde

Saitē <https://sourceforge.net/projects/pentaho/files/Data%20Integration/7.1/> ir pieejama Pentaho data Integration programmatūra.

## Pentaho Data Integration uzstādīšana Windows 10

Pentaho Data Integration nav jāinstalē uz darbstacijas. Iepriekšējā nodaļā lejuplādēto programmatūru nepieciešams atarhivēt darbstacijā. Lai palaistu Pentaho Data Integration, nepieciešams atvērt datni “Spoon.bat”.

## Pentaho Data Integration uzstādīšana Ubuntu

Lai lietotu Pentaho Data Integration uz Linux operētājsistēmas, jābūt iepriekšuzstādītai JAVA programatūrai (vismaz JAVA 8 versija).

Lai lietotu Pentaho Data Integration, uz servera nepieciešams uzinstalēt papildus 3. pušu programmatūru sudo apt-get install libwebkitgtk-1.0-0.

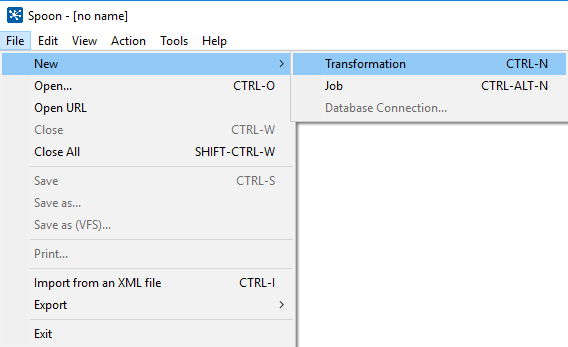
Iepriekš lejuplādēto programmatūru nepieciešams atarhivēt darbstacijā. Pirms Pentaho Data Integration aktivizēšanas nepieciešams atarhivētajā mapē aktivizēt komandu: chmod +x \*.sh.

Kad tas izdarīts, programmatūru var aktivizēt ar komandu ./spoon.sh.

## Darbs ar Pentaho Data Integration

### Jauna transformāciju projekta izveidošana datu atlases rīkā

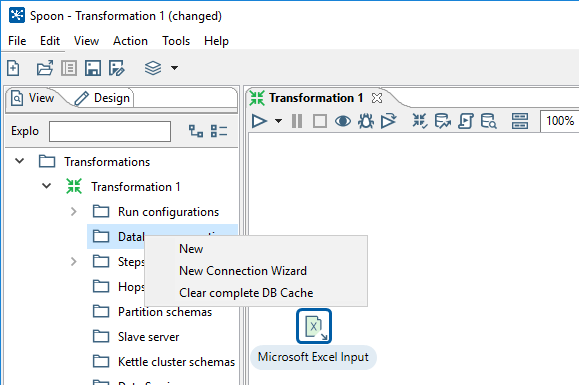
Lai izveidotu jaunu transformācijas projektu, izvēlnē norāda File->New->Transformation vai izmanto taustiņu kombināciju “Ctrl + N”:



Ilustrācija 39:Jauna transformāciju projekta izveidošana

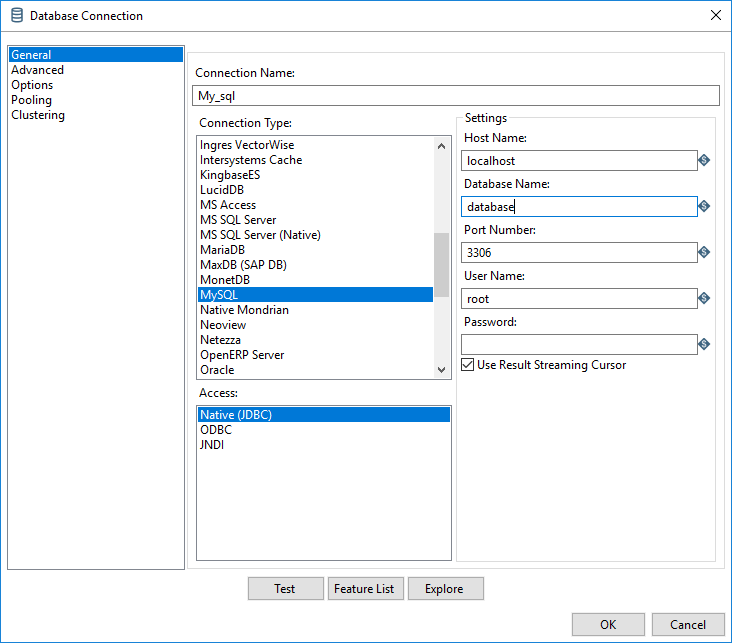
### MySql datu bāzes pievienošana

Pentaho Data Integration “View” logā sadaļā “Transformations” jāatrod transformācija, kas tika izveidota nodaļā “4.4.1 Jauna transformāciju projekta izveidošana datu atlases rīkā”. Zem tā koka izvēlnē atrod “Database connections” un, nospiežot peles labo taustiņu, izvēlnē izvēlas “New”:



Ilustrācija 40:Jauna datu bāzes pieslēguma izveidošana

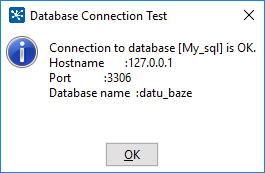
Logā “Database Connection” nokonfigurē izvēlētās datubāzes pieslēgumu, piemēram, “Mysql”:



Ilustrācija 41: Datu bāzes pieslēguma konfigurēšana

Pēc pieslēguma nokonfigurēšanas nospiež pogu “Test”. Ja tiek atgriezta kļūda, tad biežākais iemesls ir neaktuāli JDBC draiveri. Tos var lejuplādēt saitē “<https://dev.mysql.com/downloads/connector/j/5.1.html>”. Jāizvēlas platformu neatkarīgs konektors, kur .zip datnē esošā JAR datne jāiekopē Pentaho Data Integration mapes apakšmapē “lib”.

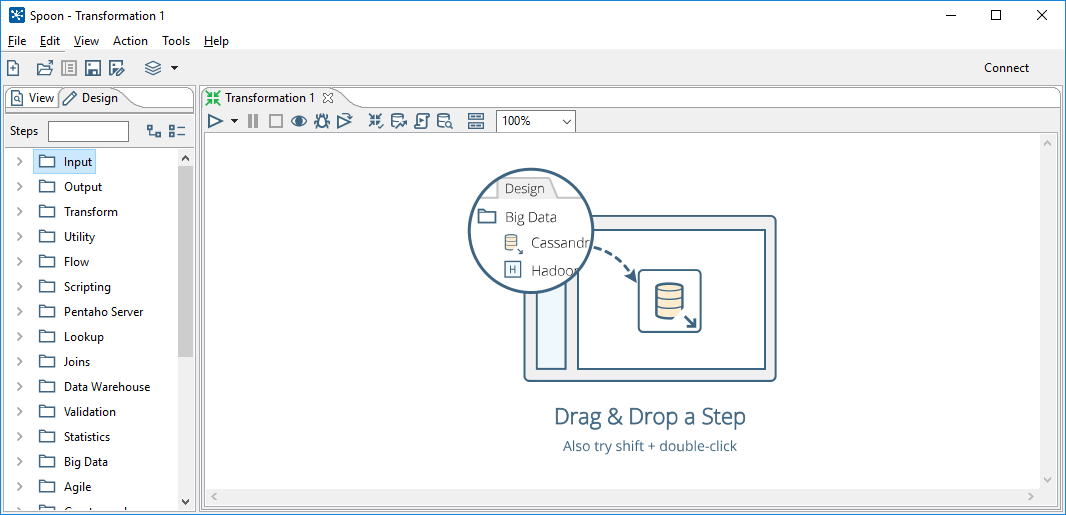
Ja konfigurācija ir veiksmīga, tiek attēlots šāds paziņojums:



Ilustrācija 42:Paziņojums par veiksmīgu pieslēgšanos datu bāzei

### Ieejas datu pievienošana transformāciju projektam

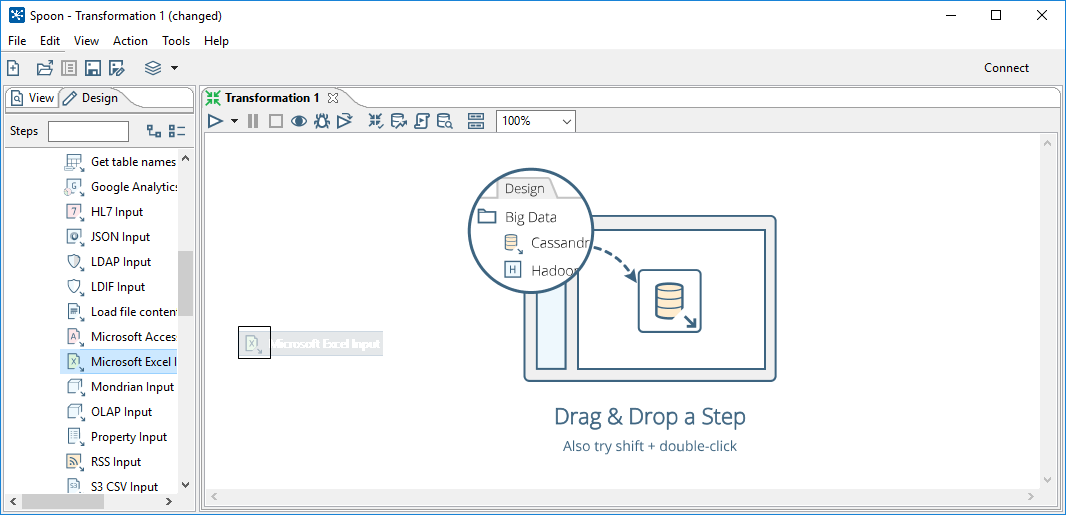
Jebkuram Transformāciju projektam var pievienot dažāda veida ieejas datus. Tas darāms Transformācijas Design logā:



Ilustrācija 43:Ieejas datu pievienošana

#### Excel datnes pievienošana

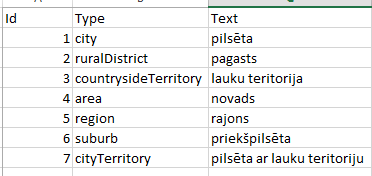
Pentaho Data Integration “Design” logā sadaļā “Input” jāatrod “Microsoft Excel Input” un jāpārvelk uz transformācijas logu:



Ilustrācija 44:Excel pievienošana transformāciju projektam

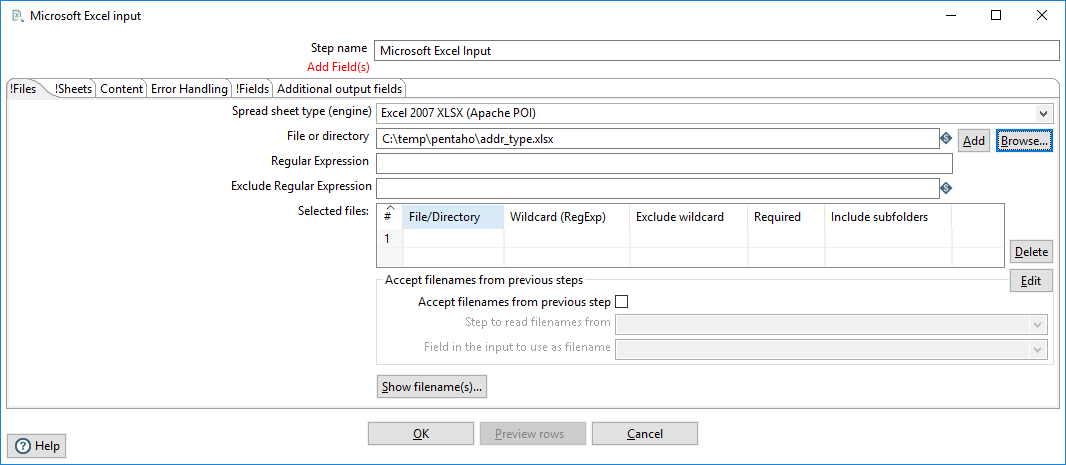
#### Excel ieejas datu konfigurēšana

Piemērā tiks veikta šādas Excel datnes konfigurēšana:



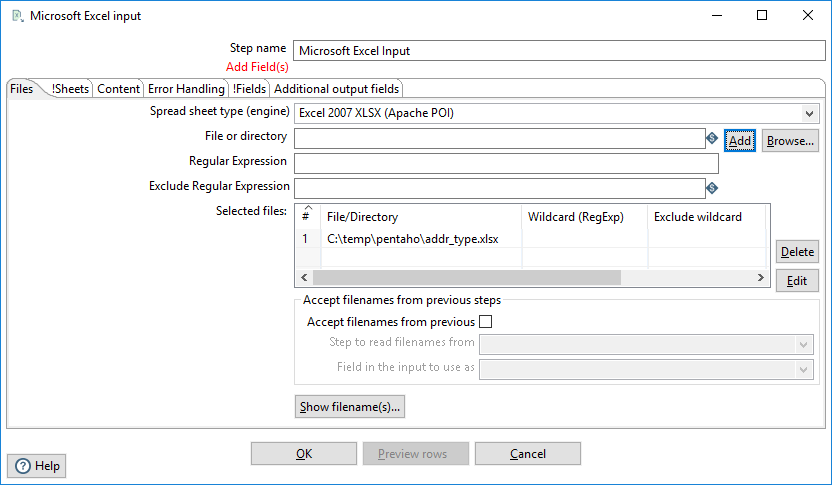
Ilustrācija 45:Excel datnes piemērs

Lai rīks saprastu, kādi dati jāapstrādā, nepieciešama Excel ieejas datu objekta konfigurēšana. Ar dubultklikšķi uzklikšķinot uz Excel ieejas datu objekta, tiek parādīts konfigurācijas logs:



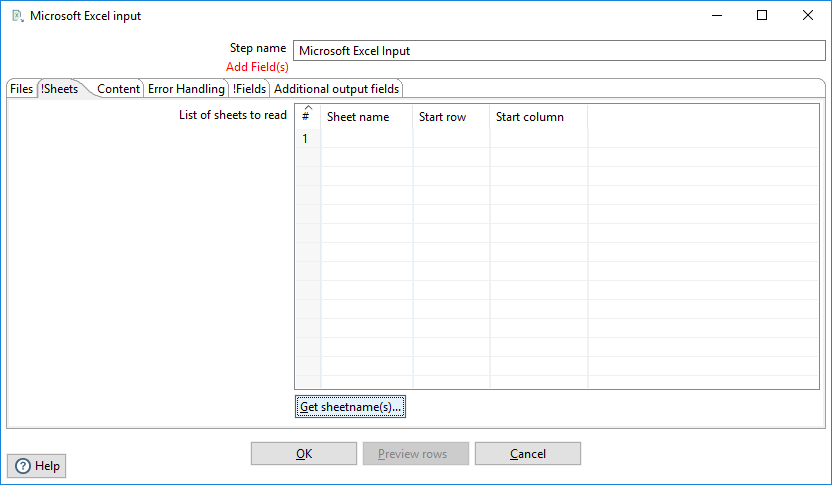
Ilustrācija 46:Excel ieejas datu konfigurēšana

Konfigurācijas logā sākotnēji nepieciešams pievienot Excel datnes vai mapi, kur atrodas apstrādājamās datnes. Datņu pievienošana notiek ar pogu “Browse”, ko nospiežot atveras operētājsistēmas datņu meklētājs. Pēc nepieciešamās datnes vai mapes atrašanas, Konfigurācijas logā nospiež pogu “Add”. Izvēlētās datnes redzamas konfigurācijas logā sadaļā “Selected files”:



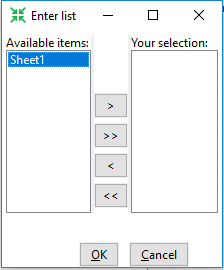
Ilustrācija 47: Konfigurācijas logs - izvēlētas datnes

Pēc datņu izvēles “Tab” izvēlnē uzklikšķina uz “Sheets”, kur nospiež pogu “Get sheetname(s)”:



Ilustrācija 48: Izklājlapas izvēle

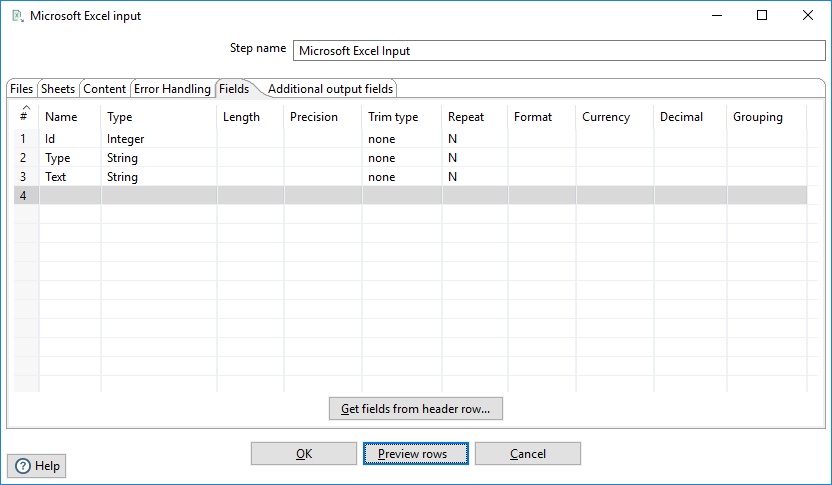
Iznirstošajā logā no saraksta izvēlas pieejamās izklājlapas:



Ilustrācija 49:Izklājlapu izvēle

Ar bultām pārnes nepieciešamās lapas no “Available items” saraksta uz “Your selection”, pēc tam apstiprinot ar pogu “Ok”.

Tālāk “Tab” izvēlnē uzklikšķina uz “Fields”, kur nospiež pogu “Get fields from header row…”, lai lauki no Excel datnes tiktu ielasīti rīkā apstrādei:

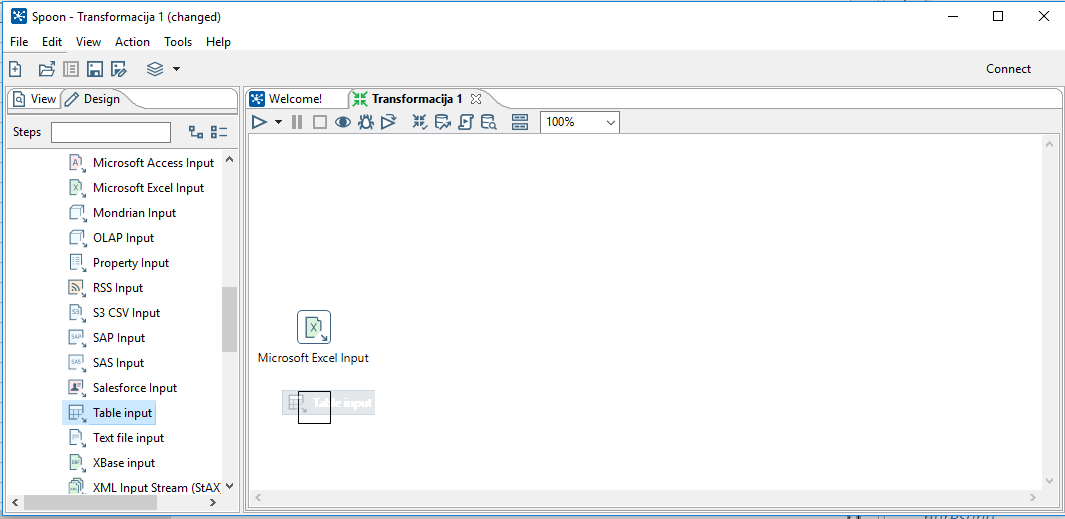


Ilustrācija 50: Datu lauku izvēle

Pēc lauku ielasīšanas rīkā jāsakonfigurē Datu tipi (Type) un citi nepieciešamie parametri. Kad konfigurēšana pabeigta, nospiež pogu “Ok”. Excel datnes pievienošana transformāciju projektam līdz ar to ir pabeigta.

#### MySql tabulas pievienošana

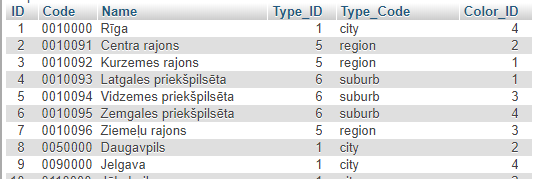
Pentaho Data Integration “Design” logā sadaļā “Input” jāatrod “Table Input” un jāpārvelk uz transformācijas logu:



Ilustrācija 51:MySql tabulas pievienošana transformāciju projektam

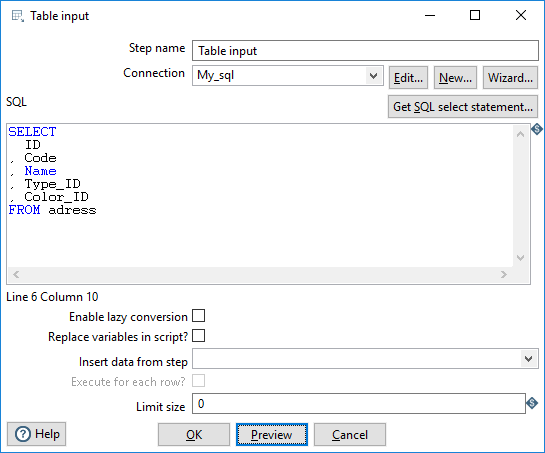
#### MySql tabulas ieejas datu konfigurēšana

Piemērā tiks veikta šādas MySql tabulas konfigurēšana:



Ilustrācija 52:MySQL tabulas piemērs

Lai rīks saprastu, kādi dati jāapstrādā, nepieciešama Tabulas ieejas datu objekta konfigurēšana. Ar dubultklikšķi uzklikšķinot uz Tabulas ieejas datu objekta, tiks parādīts konfigurācijas logs (piemērā jau ar aizpildītām vērtībām):

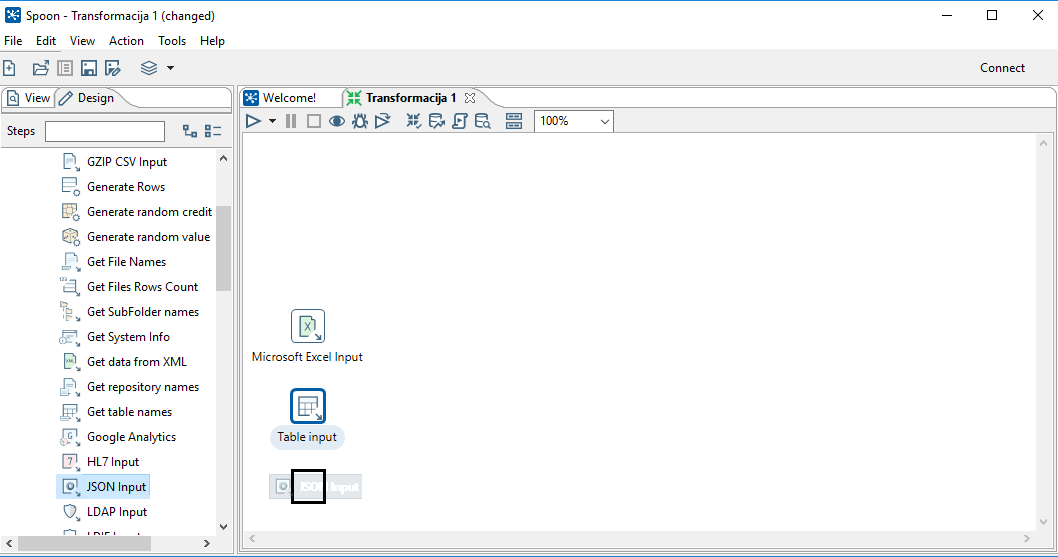


Ilustrācija 53:Tabulas konfigurācijas logs

Primāri nepieciešams aizpildīt “Connection” lauku ar pieslēgumu, kas tika izveidots nodaļā 4.4.2 MySql datu bāzes pievienošana. Tālāk konfigurācijas logā iespējams ierakstīt vaicājumu laukā SQL vai arī, nospiežot pogu “Get SQL select statement”, to uzģenerēt. Kad aizpildīta minimālā konfigurācija, ar pogu “Preview” var pārbaudīt rezultātu. Veiksmīgas izpildes gadījumā jānospiež “OK”.

#### JSON datnes pievienošana

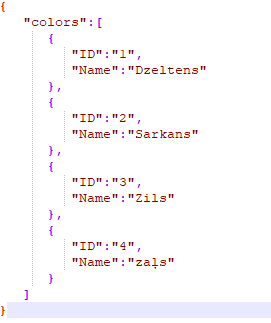
Pentaho Data Integration “Design” logā sadaļā “Input” atrod “JSON Input” un pārvelk uz transformācijas logu:



Ilustrācija 54:JSON datnes pievienošana transformāciju projektam

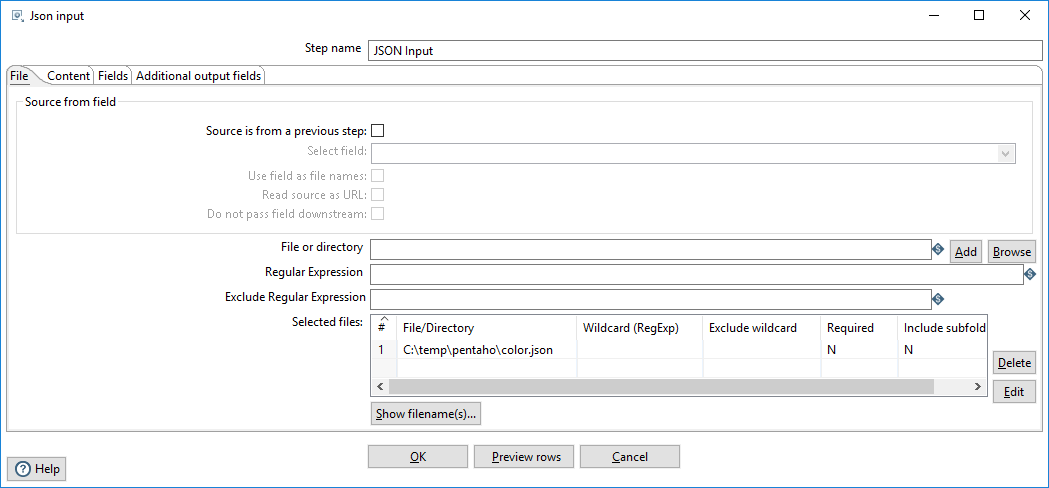
#### JSON datnes ieejas datu konfigurēšana

Piemērā tiks veikta šādas JSON datnes konfigurēšana:



Ilustrācija 55:JSON datnes piemērs

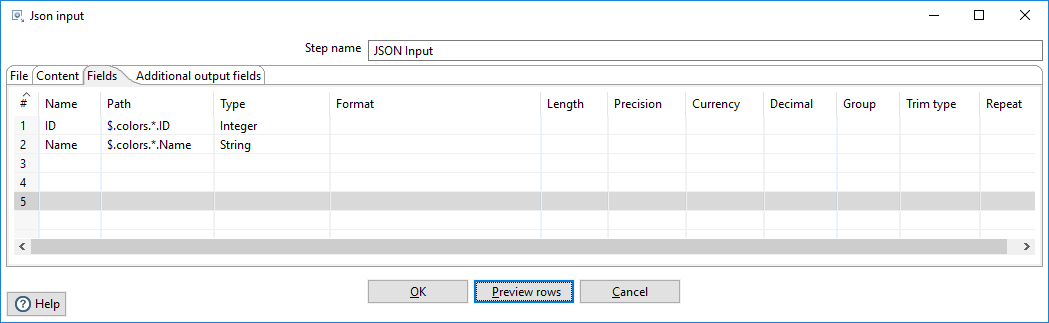
Lai rīks saprastu, kādi dati jāapstrādā, nepieciešama JSON ieejas datu objekta konfigurēšana. Ar dubultklikšķi uzklikšķinot uz JSON ieejas datu objekta, tiek attēlots konfigurācijas logs:



Ilustrācija 56: JSON ieejas konfigurēšanas logs

Konfigurācijas logā sākotnēji nepieciešams pievienot JSON datnes vai mapi, kur atrodas apstrādājamās datnes. Datņu pievienošana notiek ar pogu “Browse”, ko nospiežot atveras operētājsistēmas datņu meklētājs. Pēc nepieciešamās datnes vai mapes atrašanas, Konfigurācijas logā jānospiež poga “Add”. Izvēlētās datnes parādīsies konfigurācijas logā sadaļā “Selected files”.

Pēc datņu izvēles “Tab” izvēlnē uzklikšķina uz “Fields”, kur nepieciešams aprakstīt katru no apstrādājamiem laukiem. Piemērā “ID” un “Name”:



Ilustrācija 57:JSON lauku definēšana

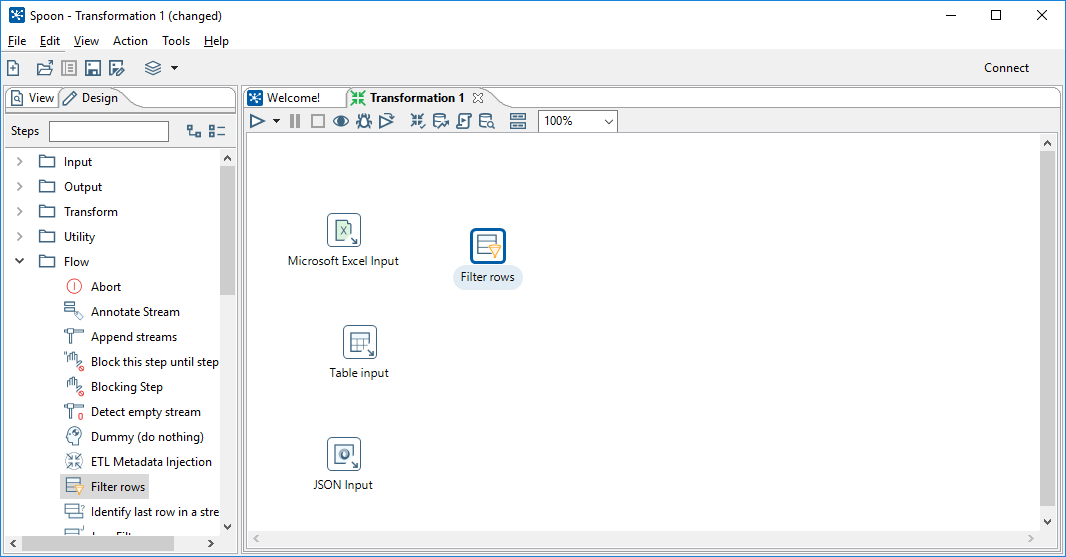
Obligāti jāaizpilda kolonnu “Name” un “Path” vērtības, kā arī vēlams aizpildīt kolons “Type” vērtības. Kad konfigurēšana pabeigta, nospiež pogu “Ok”. JSON datnes pievienošana transformāciju projektam līdz ar to ir pabeigta.

### Datu transformāciju pievienošana transformāciju projektā

Piemēros tiks veidotas uzmeklēšanas (lookup) transformācijas. MySql tabulai tiks pievienotas klasifikatoru vērtības no JSON un Excel datiem.

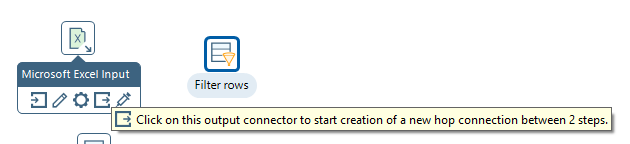
#### Teksta ieejas datu tukšo rindu filtrēšana

Lai minimizētu riskus datu apstrādē, tiek rekomendēts pirms transformāciju pievienošanas izfiltrēt no teksta datnēm visus ierakstus, kas ir tukši (it sevišķi Excel datnei, jo rīki mēdz ielasīt arī tukšas rindiņas). Pentaho Data Integration “Design” logā sadaļā “Flow” jāatrod “Filter rows” un jāpārvelk uz transformācijas logu:



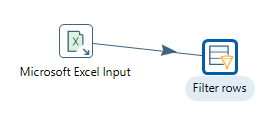
Ilustrācija 58:Ieejas datu filtrēšana

Lai norādītu, kurš objekts ir jāfiltrē, peles kursoru nepieciešams novietot uz filtrējamā objekta, piemēram, Microsoft Excel Input:



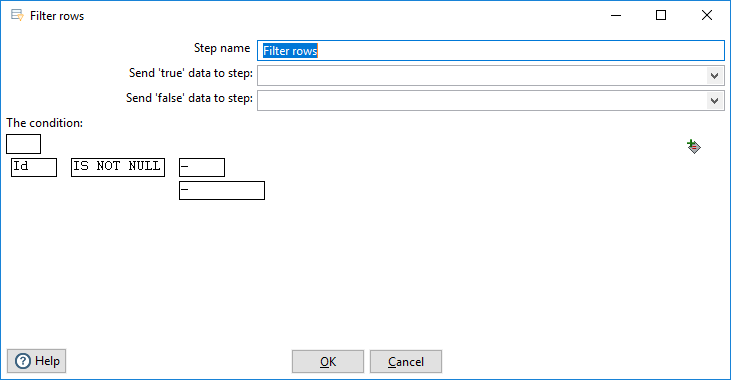
Ilustrācija 59:Objektu savienojuma pievienošana

Jāizvēlas nepieciešamais savienojuma vieds, piemēram, “output”. Tālāk rīkā ar peli uzklikšķina uz objekta, ar ko jāveic savienošana, piemērā uz “Filter rows”. Rezultātā tiek iegūts šāds skats:



Ilustrācija 60:Savienojums starp diviem objektiem

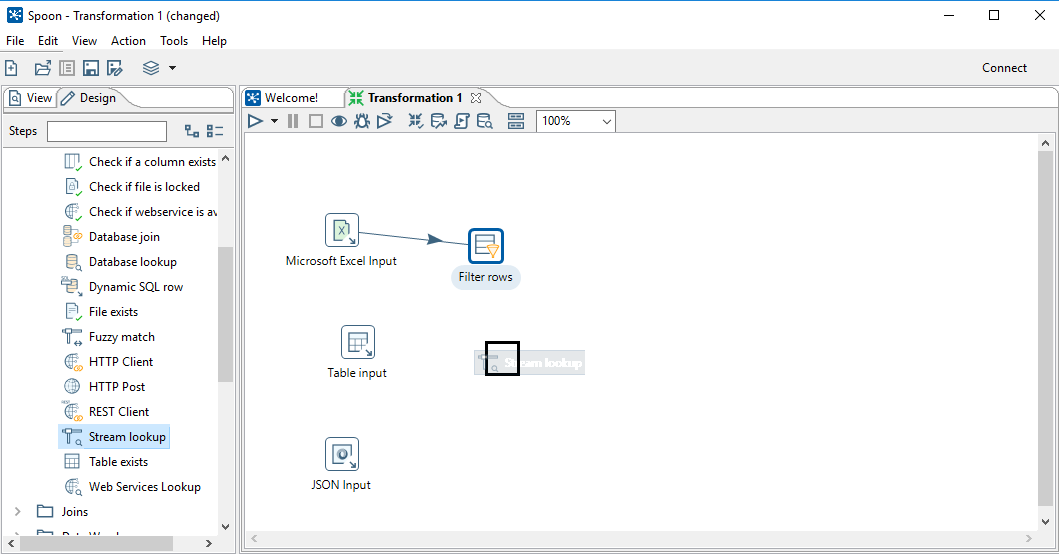
Pēc šāda principa jāveido visas nepieciešamās plūsmas. Lai izfiltrētu visas tukšās rindas, ar dubultklikšķi jāizvēlas filtrēšanas objekts un jāieliek filtrēšanas kritēriji. Piemēram, šādi:



Ilustrācija 61:Tukšu rindu filtrēšana

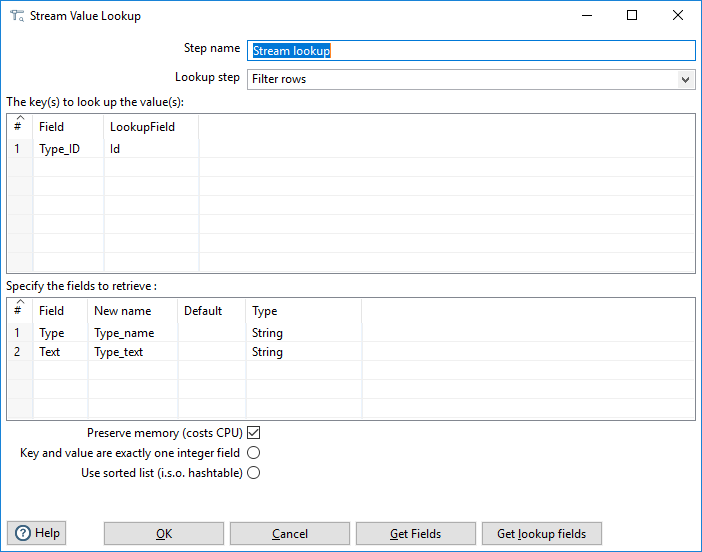
#### Uzmeklēšanas transformācijas pievienošana

Pentaho Data Integration “Design” logā sadaļā “Lookup” atrod “Stream Lookup” un pārvelk uz transformācijas logu:



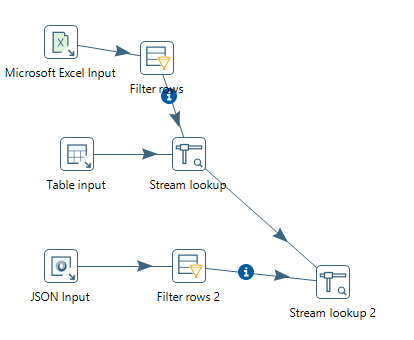
Ilustrācija 62:Uzmeklēšanas transformācijas pievienošana

Lai konfigurētu Uzmeklēšanas transformāciju, nepieciešams izveidot savienojumus starp “Filter rows” objektu un no jauna izveidoto “Lookup” objektu, kā arī starp “Table input” objektu un “Lookup” objektu (ar plūsmas virzienu uz “Lookup” objektu). Kad tas izdarīts, jāveic “Lookup” objekta konfigurēšana. Ar dubultklikšķi izvēlas “Lookup” objektu, lai tiktu atvērts konfigurācijas logs. Tajā jāveic konfigurācija atbilstoši datiem, kādi ir ieejas datos. Attēlā piemērs, kas balstīts uz vadlīnijās izmantotajiem datiem:



Ilustrācija 63:Uzmeklēšanas transformācijas konfigurācijas piemērs

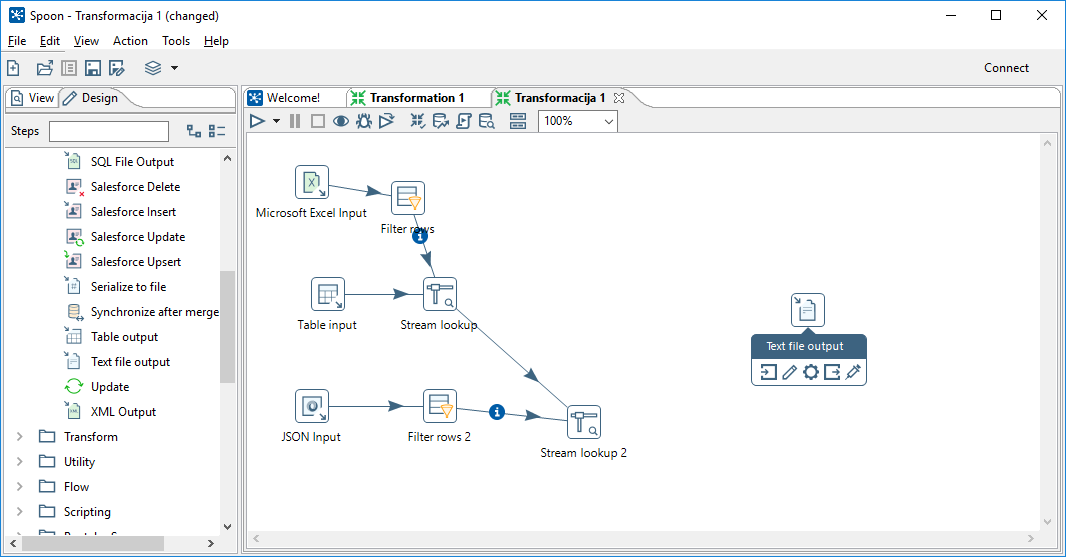
Pēc identiska principa jāizveido arī JSON un Tabulas datu transformācija, iegūstot šādu transformāciju:



Ilustrācija 64:Uzmeklēšanas transformācijas

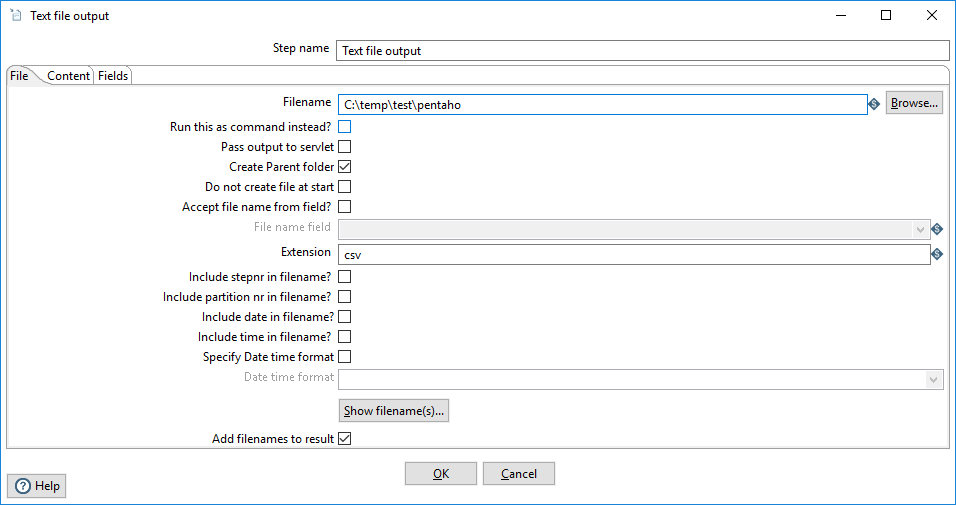
### Izejas datu pievienošana transformāciju projektam

Pentaho Data Integration “Design” logā sadaļā “Output” atrod “Text file output” un pārvelk uz transformācijas logu:



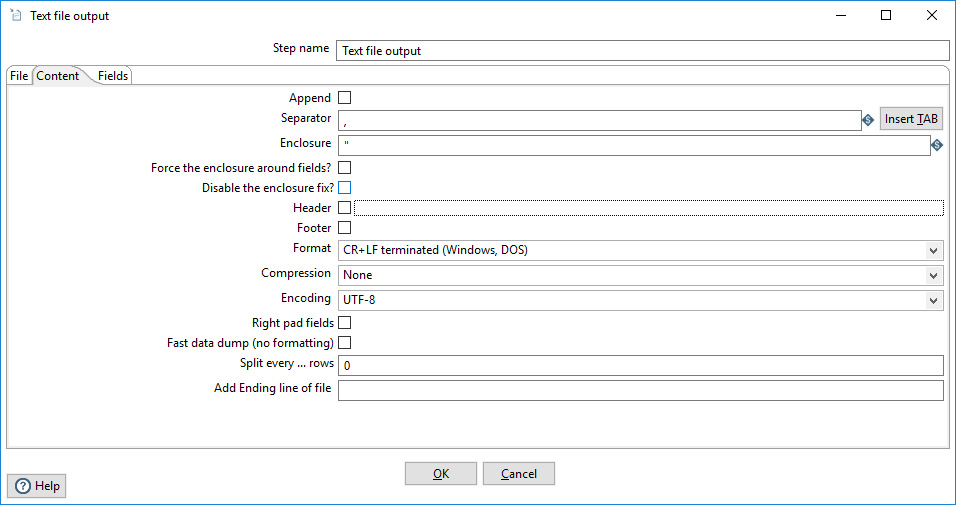
Ilustrācija 65:Izejas datu pievienošana

No “Stream lookup 2” objekta izveido savienojumu ar “Text file output”, pēc kā ar dubultklikšķi var veikt jaunā objekta konfigurēšanu. Logā, kas tiks atvērts, jānorāda datnes nosaukums (filename) un paplašinājums (extention):



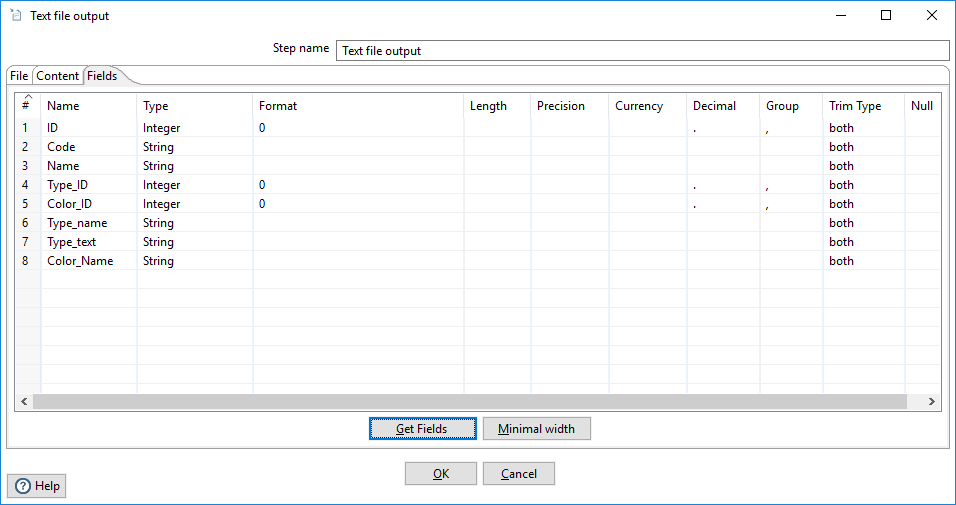
Ilustrācija 66:Izejas datnes norādīšana

Pēc datņu izvēles “Tab” izvēlnē uzklikšķina uz “Content”, kur nepieciešams precizēt eksporta parametrus atbilstoši RFC4180 standartam:



Ilustrācija 67:Eksporta parametru norādīšana

Pēc eksporta parametru norādīšanas “Tab” izvēlnē uzklikšķina uz “Fields”, kur nepieciešams uzskaitīt laukus, kuri jāizvada:



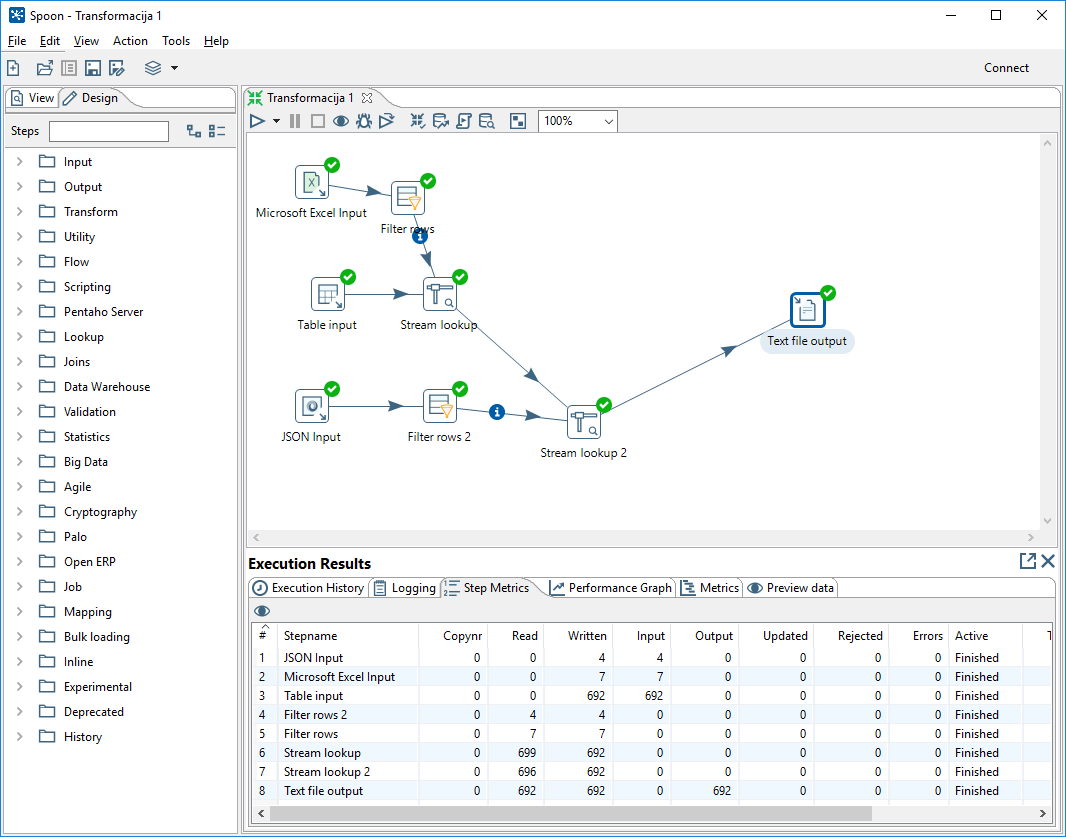
Ilustrācija 68:Izejas datu lauku norādīšana

### Transformāciju izpildīšana

Jebkuru transformāciju var izpildīt transformāciju logā no GUI vai arī, veidojot speciālus JOB, var veikt secīgu dažādu transformāciju izpildi.

#### Transformāciju izpildīšana no GUI

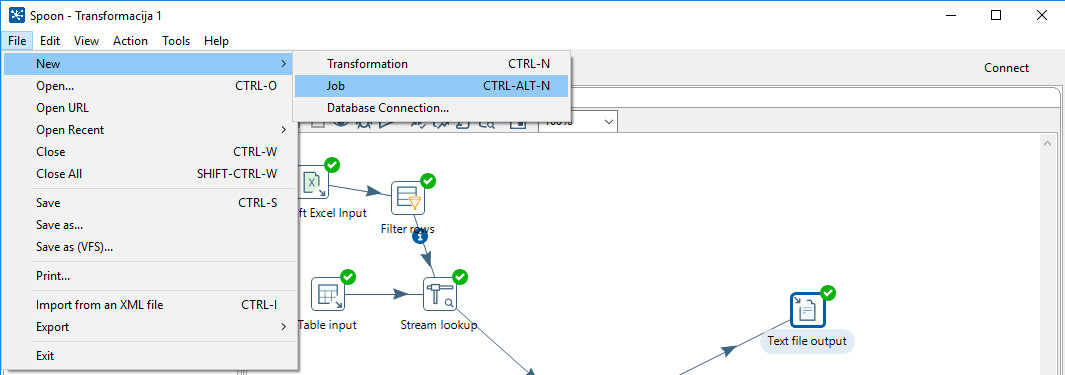
Lai izpildītu izveidoto transformāciju, nepieciešams nospiest “F9” vai “Play” ikonu un iznirstošajā logā nospiest pogu “run”. Piemērā izveidotās transformācijas izpilde:



Ilustrācija 69:Transformācijas izpilde

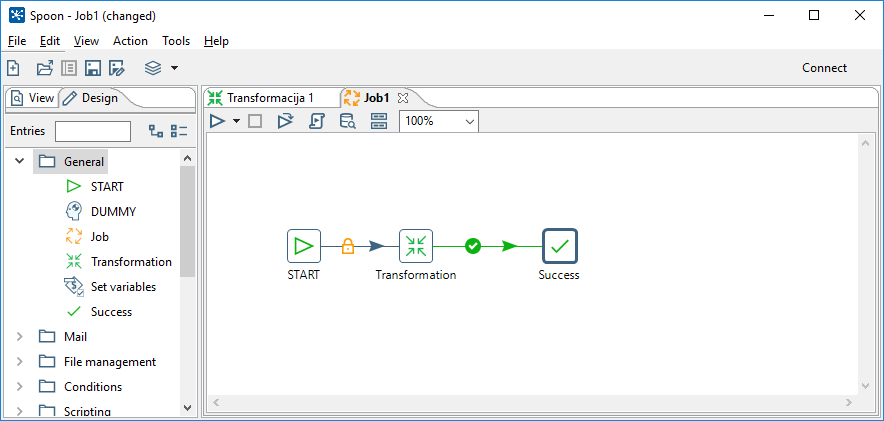
#### Transformāciju izpildīšana ar JOB

Lai transformācijas varētu izpildīt no JOB, nepieciešams izveidots Pentaho Data Integration JOB:



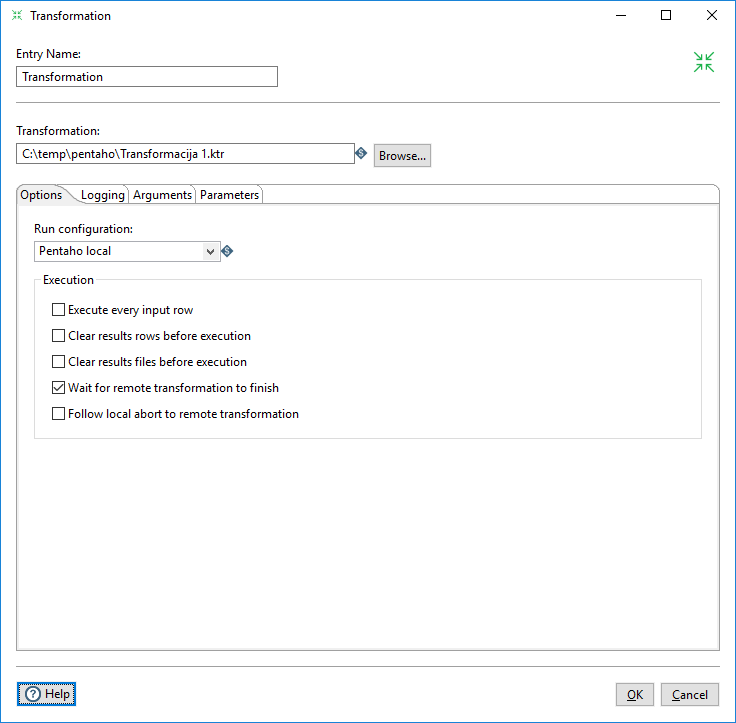
Ilustrācija 70:Jauna Job izveide

JOB uzturēšanas formā Pentaho Data Integration “Design” logā sadaļā “General” jāatrod “Start”, “Transformation”, “Success” un visi trīs jāpārvelk uz transformācijas logu, kā arī jāsavieno ar savienojumiem:



Ilustrācija 71:Vienkārša Job izveide

Ar dubultklikšķi uz “Transformation” objekta jāatver transformācijas uzturēšanas forma, kur jānorāda izpildāmā transformācija:



Ilustrācija 72: Izpildāmās transformācijas norādīšana

Izveidoto Job jāsaglabā uz darbstacijas cietā diska, lai to būtu iespējams izpildīt arī no komandrindas.

Lai to izdarītu windows sistēmā, nepieciešams izveidot Job.bat datni

kitchen.bat /file:C:\temp\pentaho\Job1.kjb /level:Basic

un Job2.bat datni

ckancli.rb upload -d C:\temp\test -c config.json -r resource.json -v schema.json

Lai to izdarītu linux sistēmā, nepieciešams izveidot Job.sh datni:

kitchen.sh /file:/tmp/pentaho/Job1.kjb /level:Basic

un Job2.sh datni:

ckancli.rb upload -d /tmp/tests/ -c config.json -r resource.json -v schema.json

Operētājsistēmā pieejamā regulāro programmu izpildes rīkā (sheduler, crontab) jāiekonfigurē, ka faili izpildāmi secīgi.

Job1.kjb ir Pentaho Data Integration jobs, kas tika izveidots šajā nodaļā.

### Izejas datu pievienošana transformāciju projektam

# CSV datnes augšuplāde CKAN, izmantojot komandrindas rīku

## Ruby uzstādīšana

Lai darbinātu komandrindas rīku, nepieciešams, lai uz darbstacijas būtu uzstādīts Ruby. Nepieciešama vismaz Ruby 2.0.0 versija.

### Ruby uzstādīšana uz Windows darbstacijas

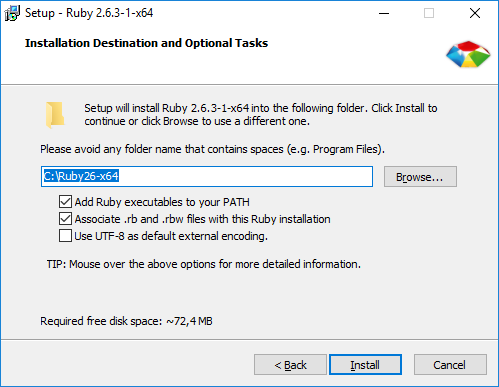
Ruby instalācijas atrodamas saitē <https://rubyinstaller.org/downloads/>. Lejuplādējiet kādu no tām un veiciet instalācijas izpildi ar lietotāju, kuram ir administratora tiesības uz darbstacijas.

Pirmajā logā nepieciešams apstiprināt licences nosacījumus un nospiest pogu “Next”:



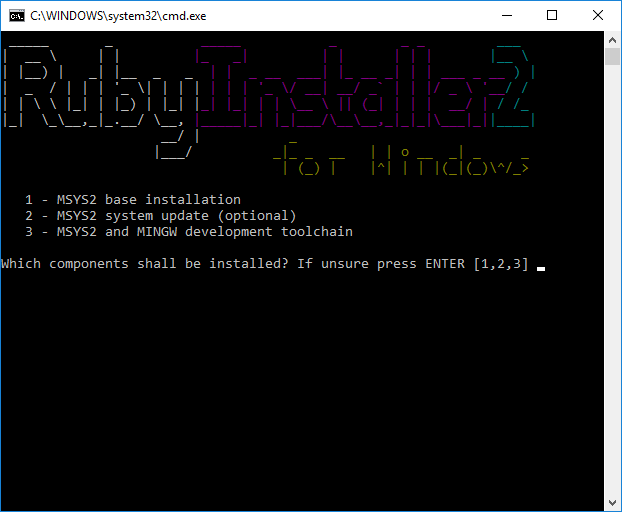
Ilustrācija 73:Licences nosacījumu apstiprināšana

Jāizvēlas mape, kurā instalēt Ruby, un jānospiež poga “Install”:



Ilustrācija 74:Ruby instalēšanas iniciēšana

Pēc instalēšanas pabeigšanas jānospiež poga “Finish”. Tiks atvērts Ruby termināla logs, kurā jāturpina instalēšana, nospiežot taustiņu “Enter”:



Ilustrācija 75:Ruby uzstādīšana

Ja uz darbstacijas nebūs uzstādīta MSYS2 komponente, tad būs nepieciešams to uzstādīt, MSYS2 logā nospiežot “Next” 3 reizes. Kad instalēšana pabeigta, nospiež “Finish”.

## Komandrindas rīka uzstādīšana

Nepieciešams atvērt komandrindas terminālu ar administratora tiesībām un izpildīt komandu “gem install ckan\_cli”.

Pēc komandas izpildes tiek uzstādīts komandrindas rīks, kuru par palaist ar komandu “ckancli.rb”.

## Komandrindas rīka lietošana

### Komandrindas rīka konfigurācijas datne

Komandrindas rīka izpildes gaitā nepieciešamo konfigurāciju var nodot ar –c komandu json formātā:



Ilustrācija 76:Konfigurācijas paraugs

Konfigurācijas paraugs lejuplādējams saitē –

<https://github.com/datagovlv/ckan_cli/blob/master/example_files/config.json>.

Konfigurācijā obligāti jānorāda ckan\_api sekcija. Pārējās sekcijas nav obligātas.

Komandrindas rīka konfigurācijas api\_key lauka vērtība atrodama CKAN lietotāja profilā, kura vārdā vēlaties iesūtīt datus – lauks API atslēga. Url vērtību nepieciešams uzzināt pie CKAN uzturētāja.

Pārējie konfigurācijas parametri ir e-pastu notifikāciju izsūtīšanai un aizpildāmi, ja vēlaties saņemt notifikācijas par apstrādes procesu kļūdām un/vai veiksmīgu apstrādi.

### Komandrindas rīka izsaukšanas piemērs

Komandrindas rīkam ir publicētas piemēra datnes, uz kurām balstās šajā dokumentā apskatītais piemērs. Datnes publicētas saitē –

<https://github.com/datagovlv/ckan_cli/tree/master/example_files>.

Visas datnes, kas publicētas, nepieciešams lejuplādēt un saglabāt uz darbstacijas diska. Piemēram, mapē c:\temp\tests vai /tmp/tests – atkarīgs no izmantojamās operētājsistēmas.

Atveriet jaunu termināla logu un kā aktīvo izvēlaties direktoriju c:\temp\tests vai /tmp/tests – atkarīgs no izmantojamās operētājsistēmas.

Izsauciet komandrindas rīku ar komandu:

ckancli.rb upload -d C:\temp\test -c config.json -r resource.json -v schema.json

vai

ckancli.rb upload -d /tmp/tests -c config.json -r resource.json -v schema.json