

Tallinna Ülikool  
Informaatika Instituut

Magistritöö

# Vabavara Talend- programmeerimisvahendi analüüs

Koostaja: Alex Aisting

Juhendaja: Jaagup Kippar

Autor: ..... „..... „2011

Juhendaja: ..... „..... „2011

Instituudi direktor: ..... „..... „2011

Tallinn 2011

## **Autorideklaratsioon**

Deklareerin, et käesolev magistritöö on minu töö tulemus ja seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on viidatud.

.....

(kuupäev)

.....

(autor)

## Sisukord

Autorideklaratsioon.....	2
Sisukord.....	3
Sissejuhatus.....	5
1. Magistritöö olemus.....	6
1.1. Teema aktuaalsus.....	6
1.2. Uurimisprobleem ja töö eesmärk.....	7
1.3. Töö tulemus.....	7
1.4. Magistritöö struktuur.....	8
2. ETL- programmeerimisvahendid.....	8
2.1. ETL protsessi olemus.....	8
2.1.1. ETL mõiste.....	8
2.1.2. ETL protsessi kasutamine andmelaos.....	10
2.1.3. Programmipõhine versus käsitsi kirjutatud ETL protsess.....	11
2.2. ETL tarkvara vahendite võrdlus.....	13
2.2.1. Kaitstud omandiõigusega ja avatud lähtekoodiga vahendid.....	13
2.2.2. ETL tarkvara vahendite monitooring.....	14
2.2.3. Momentum indexi võrdlus.....	15
2.2.4. ETL turu suurus ja pakkujate turujaotus.....	15
2.2.5. ETL tarkvarafirmade poolt koostatud võrdlused konkurentidega.....	16
2.3. ETL tarkvara turuliidrid.....	16
2.3.1. Talend.....	16
2.3.2. Pentaho (Kettle).....	16
2.3.3. InformaticaPowerCenter.....	17
2.3.4. OracleWarehouseBuilder (OWB).....	17
3. Talend programmeerimisvahendi ülevaade.....	18
3.1. Talend teoreetiline osa.....	18
3.1.1. Talendi põhimõte.....	18
3.1.2. Talendi partnerid.....	19
3.1.3. Talendi plussid ja miinused.....	20

3.2.	Näidisülesande lahendamine Talend tarkvaraga.....	21
3.3.	Lahenduskaik Talend avatud lähtekoodiga vahendiga.....	22
3.3.1.	Loo uus projekt ( <i>Create a new local project</i> ).....	23
3.3.2.	Ärimudel ( <i>Business Model</i> ) .....	23
3.3.3.	Metaandmed ( <i>Metadata</i> ).....	25
3.3.4.	Andmebaasi ühendus ( <i>Db Connections</i> ).....	25
3.3.5.	Töö disainija ( <i>Job Designer</i> ) .....	26
3.3.6.	Päringute- koostaja ( <i>SqlBuilder</i> ).....	27
3.3.7.	Kogu töö teostamine töölaual .....	30
4.	Talendi populaarsus .....	32
4.1.	Küsitluse kokkuvõte.....	32
4.2.	Autori ja teiste kasutajate kogemused Talendi kasutamisel.....	33
4.2.1.	Autori kogemused.....	33
4.2.2.	Ühe Talendi kasutaja arvamus .....	34
4.2.3.	Teise Talendi kasutaja praktilise ülesande teostus.....	35
4.3.	Kümme põhjust, miks valida avatud lähtekoodiga programmeerimisvahend Talend 37	
	KOKKUVÕTE .....	42
	Summary .....	43
	Kasutatud allikad .....	44
	Lisad.....	48
Lisa 1	Koodinäide MySQL-i ühenduste metaandmetest.....	49
Lisa 2	Näide MySQL-i ühenduse omadustest .....	53
Lisa 3	Küsimustik.....	54
Lisa 4	ETL pakujate võrdlus .....	57

## Sissejuhatus

Tänapäeval on IT muutunud ettevõtte äritegevuse lahutamatuks osaks, IT toetab ärieesmärkide saavutamist. Äri nõuab IT-lt teostada ülesannet kvaliteetselt, lühikese ajaga ja võimalikult odavalt. Äri jaoks on esmatähtis raha, aeg ja korralik tulemus.

Ettevõtte üks osakond töötab Oracle aplikatsioonis, rahandus töötab SAP aplikatsioonis, mingit osa protsessi ketis toetavad suurarvutid (*Mainframe*), osad andmed on salvestatud Oracle andmebaasi, suurarvutitesse ja mõned andmed VSM failina jne. Kui ühel päeval oleks sealt vaja välja võtta aruandeid, siis selleks on kaks võimalust:

- käsitsi genereerida erinevatest süsteemidest erinevad aruanded ja integreerida need omavahel
- tuua kõik andmed erinevatest süsteemidest/rakendustest kokku ja teha nendest andmeladu ning koostada sealt raportid vastavalt esitatud nõuetele

Teine võimalus on enamasti parem lahendus. (Techtiks Informatica Group, kp. puudub)

Ning selle teostamiseks, kui äritegevuse nõuded on paika pandud, analüüsitud lähteallikate süsteeme ja välja arendatud andmelao mudel, tuleb leida millega viia läbi ETL protsess. Kuna see on otsustava tähtsusega andmelao loomise projekti edus.

ETL protsessi kasutatakse andmeladude moodustamise ning äriettevõtetes andmete töötlemise, analüüsimise stsenaariumides andmete kättesaamiseks algallika rakendustest ja laadimiseks andmelattu. ETL käigus järkjärgult teostatakse andmete liikumist, korrektuuri, sünkroniseerimist ja muundamist, samuti pidevat andmete importi ja eksporti. (Extract Transform Load, 2010)

ETL projekteerimise protsess on kõige aeganõudvam etapp andmelao loomise projektis. See on tihti nii, et üle 50% ajast, mis on pühendatud andmelao projektile, kulub ETL protsesside kavandamisele ja arendamisele. (An introduction to Data Warehouses and the Data Warehousing., 2007) Seetõttu on õige ja sobiva ETL programmeerimisvahendi valimine kogu projekti õnnestumiseks väga oluline.

Tuleb ette, et tehtud IT kulutused ei ole seotud IT efektiivse rakendamisega. Seega on võimalik IT panna efektiivselt äri toetama ja seda väiksemate kuludega, kuid see eeldab „õigete asjade” tegemist/valimist.

Magistritöö annab ülevaate ETL protsessi olemusest, erinevatest ETL vahenditest ning magistritöö käigus tutvustatakse lähemalt Talend programmeerimisvahendit.

# 1. Magistritöö olemus

Tehakse ülevaade töö aktuaalsusest, püstitatakse uurimisprobleem ja koostatakse töö eesmärgid, selgitatakse töö tulemust ja esitatakse töö struktuur.

## 1.1. Teema aktuaalsus

Mitte väga kauges minevikus ei olnud ETL (andmelao kasutamise protsess ja eelkõige andmete käsitlemine) tarkvara veel olemas ja selleks, et luua andmeladu tuli käsitsi kirjutada koodi, et liigutada andmeid ühest allikast teise. Seal oli palju koodi - palju korduvaid koodiridu. Pärast koodi valmis kirjutamist tuli seda aga tihti käsitsi täiendada. Iga kord, kui olemasolevat süsteemi muudeti, kui eesmärgi määratlus muutus, kui lõppkasutaja tahtis midagi uut, tuli käsitsi täiendada/muuta koodi. Seetõttu oli eelnevatel aegadel andmelao haldamine tõeline katsumus. (Inmon, ETL: Extract, Transform and Load for Data Warehousing, 2008)

1995 a. tulid laiemal üldsuse jaoks välja esimesed ETL tarkvarad. Juhtpositsioonil oli Prisma Solutions, millele järgnes Informatica, DataStage ja paljud teised ETL tehnoloogiaid. Suure maksumuse tõttu olid nad suunatud vaid suurettevõtetele (telekommunikatsiooni-, kindlustusfirmad, pangad). Aastal 2000 tuli turule uut tüüpi ETL tarkvara, mis oli mõeldud keskklassi ja väiksematele ettevõtetele. Liider sellel turul on Talend. Talend on funktsionaalne ETL tööriistade komplekt, kuid avatud lähtekoodiga, seetõttu on taskukohane ka keskklassi maailmas. Talend pakub oma põhituuma tasuta, mida saab alla laadida internetist. Talend põhituuma peale saab lisada ka muid funktsioone ja teenuseid. Talend sobitub turule hea funktsionaalsuse ning märkimisväärselt väiksema hinnaga kui ükski teine konkurent. See on hea võimalus keskmise suurusega ettevõtetele, kes vajavad ETL kuid kes ei vaja täishinnaga ja suurt ETL paketti, mida pakutakse ja kasutatakse suurtes ettevõtetes. (Inmon, A Brief History of ETL, 2008)

Üks põhjus, miks hinnata, edendada ja tutvustada ETL tööriistu on, et paljud organisatsioonid loovad veel tänapäeval oma andmeladu käsitsi, kirjutavad keerulised SQL-id ja salvestatud protseduurid. Paljud organisatsioonid ikka veel elavad justkui "kiviajas". Andmelaod arendatakse siiani käsitsi, kasutades kas SQL või PL/SQL, kuigi arendaja tootlikkust võiks asjakohase ETL tööriista kasutamisel tõsta 3 kuni 5 korda. Lisaks on andmelao ehitamine ETL tööriista kasutades usaldusväärsem ja stabiilsem, sest rohkem aspekte saab üksteise

suhtes kontrollida ja jälgida, metaandmed on selles osas hea näide. Tuleb siiski märkida, et üksnes ETL tööriista kasutamine ei taga automaatselt edu. (Background of the ETL Tool Survey, 2009)

## **1.2. Uurimisprobleem ja töö eesmärk**

Töötades logistikaga tegelevas ettevõttes IT-tarkvara projekteerijana tekkis magistritöö autoril vajadus teostada kiirelt ja kvaliteetselt liidestamist erinevat liiki andmebaaside vahel, arvestades olemasolevate infosüsteemide keerukust.

Seni kasutatud meetodid (käsitsi koodi kirjutamine) osutusid ebaefektiivseks, kuna ajakulu oleks olnud liiga suur. Magistritöö autoril puudus ülevaade, millist tarkvara kasutada, saavutamaks optimaalset tulemust. Üks töökaaslane soovitas aga kasutada Talend vahendit.

Sellest lähtuvalt sai püstitatud ka magistritöö uurimisprobleem:

- milliseid tarkvaralahendusi eksisteerib andmete käsitlemiseks erinevate andmebaaside vahel?
- kas Talend on sobiv mulle esitatud ülesande lahendamiseks?

Uurimisprobleemist lähtudes said püstitatud järgmised eesmärgid:

- tutvuda ETL- protsessi olemusega
- teostada erinevate ETL - tarkvaratoodete võrdlus
- tutvuda põhjalikumalt Talend tarkvaraga
- lahendada näidisülesanne Talend tarkvara kasutades
- viia läbi küsitlus tarkvarafirmades ETL- tarkvara ja Talendi kasutamise osas
- kajastada teiste kasutajate kogemusi Talend tarkvaraga töötamisel

## **1.3. Töö tulemus**

Töö tulemusena antakse ülevaade andmete ühest allikast teise liigutamise olemusest, tutvustatakse selleks kasutatavaid ETL – vahendeid ning antakse põhjalikum ülevaade Talend vahendist (teoreetiline pool kui ka praktiline näidiskonstruktsiooni teostamine). Kajastatakse autori ja teiste kasutajate kogemusi ja arvamusi Talend vahendiga töötamisel.

Magistritöö tulemusena valmib tutvustav ülevaade Talend tarkvarast ning selle kasutamise võimalustest.

## 1.4. Magistritöö struktuur

Magistritöö koosneb sissejuhatausest, neljast peatükist, kokkuvõttest, inglise keelsest resümeest, kasutatud kirjanduse loetelust ja lisadest.

Esimeses peatükis tuuakse välja magistritöö aktuaalsus, uurimisprobleem ja püstitatud eesmärgid.

Teises peatükis tutvustatakse üldist ETL – protsessi olemust ja tehakse ülevaade erinevatest ETL vahenditest ja nende võrdlus.

Kolmandas peatükis tuuakse välja Talendi põhiomadused, eelised ja puudused, realiseeritakse näidisülesanne kasutades Talend vahendit.

Neljandas peatükis analüüsitakse Talendi kasutamise efektiivsust ja tehakse järeldused.

## 2. ETL- programmeerimisvahendid

Selles peatükis antakse ülevaade ETL protsessi olemusest, tutvustatakse ja võrreldakse ETL tarkvara vahendeid.

### 2.1. ETL protsessi olemus

#### 2.1.1. ETL mõiste

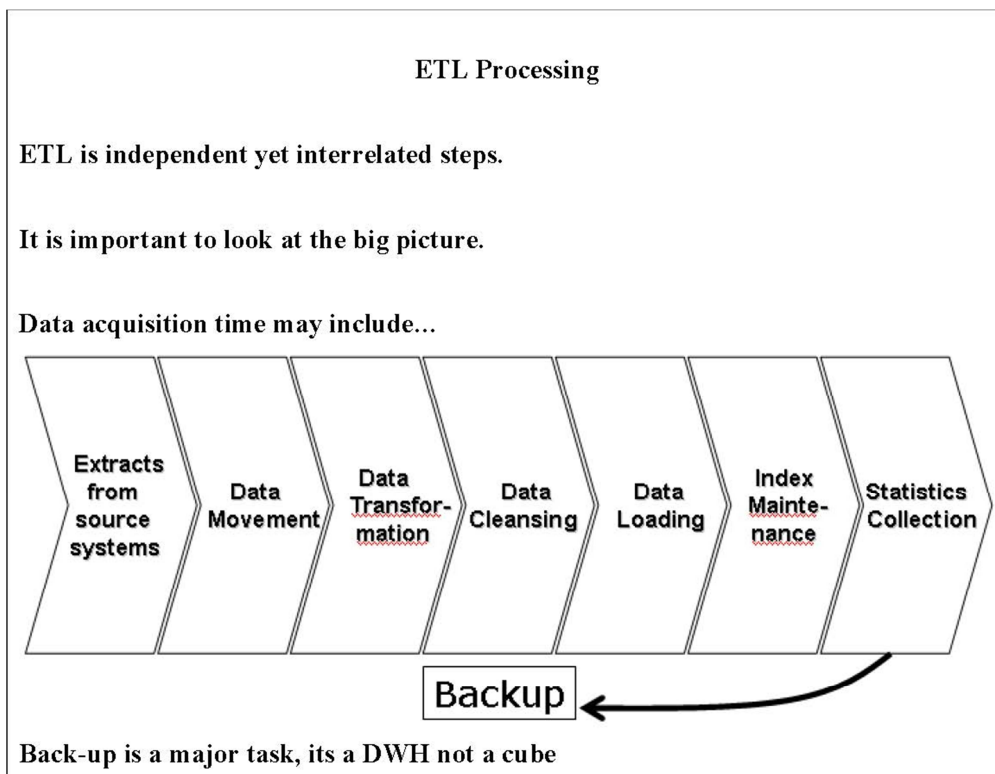
**ETL** on lühend kolmest sõnast *Extract*, *Transform* and *Load*. ETL protsess on andmete **väljavõtmine** (*Extract*), peamiselt erinevat liiki süsteemidest, **muundamine** (*Transform*) sellisesse struktuuri, mis on sobivam aruandluseks ja analüüsiks ning lõpuks andmete **laadimine** (*Load*) andmebaasi. (Joonis 1) allpool kuvab neid ETL samme.

- *Extract* - väljavõtte allikast Andmete väljavõtmine erinevatest sisemistest ja välistest ning struktureeritud ja/või struktureerimata allikatest. Liht päringud saadetakse algallikasse, kasutades natiiv ühendusi, teksti päringuid, ODBC või OLE-DB vahetarkvara. Andmed pannakse niinimetatud töötlemisalale (*Staging Area*), tavaliselt samas struktuuris nagu algallikas. Mõnel juhul me tahame ainult andmeid,



mis on uued või mida on muudetud, päringutele kuvatakse ainult muutused. Mõned ETL tööriistad teevad seda automaatselt, pakkudes muutunud andmete hõivamise mehhanismi.

- *Transform* – muundamine. Kui andmed on saadaval töötlemisalal, moodustavad nad kõik ühe platvormi ja ühe andmebaasi. Nii me saame lihtsalt ühendada ja liita (*union*) tabeleid, filtreerida ja sorteerida andmeid, kasutades spetsiifilisi atribuute, muundada teise struktuuri ja teha ärilisi arvutusi. Selles ETL protsessi etapis, saame kontrollida andmete kvaliteeti ja puhastada andmeid, kui vaja.
- *Load* – andmebaasi sisse laadimine. Lõpuks andmete sisestamine andmelattu, tavaliselt fakti ja dimensiooni põhistes tabelitesse. Sealt saab andmeid ühendada, liita ja laadida mujale kui peetakse vajalikuks.

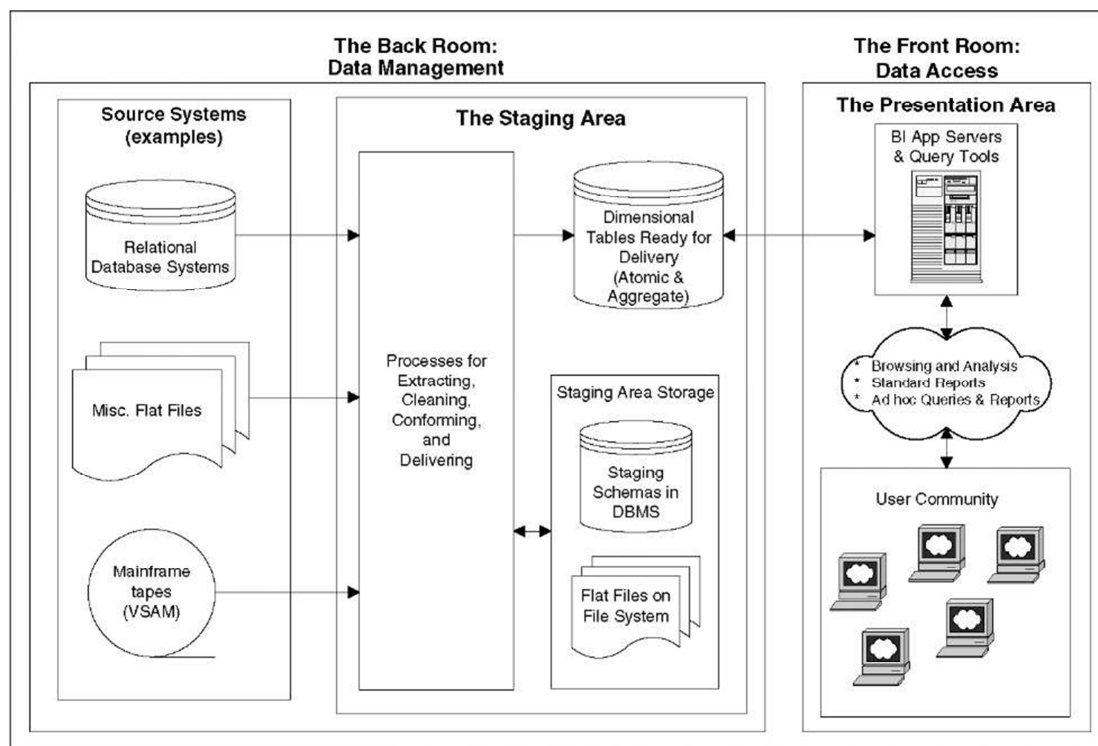


Joonis 1 ETL protsessi arhitektuur (*Extract Transform Load (ETL) | Data Warehousing, kp. puudub*).

ETL - andmelao kasutamise protsess ja eelkõige andmete käsitlemine. Tänapäeval on ETL lisaks eeltoodule andmete profileerimine, kvaliteedi kontroll, monitoring ja korrastamine, reaalajas ja nõudmisel olevate andmete käsitlemine ja integreerimine (*Service Oriented Architecture SOA*) ning metaandmete haldamine. (What is ETL, Extract Transform and Load?, kp. puudub)

## 2.1.2. ETL protsessi kasutamine andmelaos

ETL protsessi kasutatakse andmelaos andmete käsitlemiseks. Andmelaos võib tinglikult jagada kaheks: tagaruum ja eesruum, need on füüsiliselt, loogiliselt ja administratiivselt üksteisest eraldatud. Teisisõnu, enamasti tagaruum ja eesruum on erinevates seadmetes, sõltuvad erinevatest andmestruktuuridest ja neid haldab erinev IT-personal.



Joonis 2 Andmelaos eesruum ja tagaruum (Kimball, 2004, lk 16)

Andmete ettevalmistamine, mida sageli nimetatakse andmehalduseks, hõlmab andmete väljavõtmist ja muundamist ning selle informatsiooni toomist andmelaos eesruumi. Eesruum on mõeldud päringute koostamiseks tagaruumis ettevalmistatud andmetest.

Seda ülesehitust on võrreldud restoraniga, kus tagaruumi moodustavad laod, kus hoitavaid toiduaineid võrreldakse andmetega. Köök vastab siis töötlemisalale (*Staging area*), kus toimub ümberkujundamine, muundamine. Ning söögisaal vastab eesruumile, kuhu jõuab tulemus vastavalt esitatud soovile.

Töötlemisala, mida võib nimetada andmelaos köögiks, on koht kuhu pääsevad ligi ainult kogunud andmete integratsiooni spetsialistid. See on rajatis, kus asuvad algallikast välja võetud, puhastatud ja muundatud andmed, mis on ette valmistatud laadimiseks andmelattu. (Kimball, 2004, lk 16-17)

### 2.1.3. Programmpõhine versus käsitsi kirjutatud ETL protsess

Kumba eelistada? On situatsioone kus käsitsi kirjutatud ETL kood tuleb odavam ja põhjendatum aga teises olukorras jällegi on ETL vahendit kasutada mõistlikum. Kõik sõltub projekti ulatusest ja nõudmistest. Valiku osas otsuse tegemisel tuleb järgida tavalist projekti planeerimist nagu kõikide arendusprojektide tehnoloogiliste otsuste puhul. Arendusmeeskond saab teha otsuse peale põhjalikku projekti mahu, esitatud nõudmiste ja teostava personali olemasolu hindamist.

Programmpõhine versus käsitsi kirjutatud ETL protsess.

#### Programmpõhise ETL eelised:

- Lihtsam, kiirem, odavam arendus. Tööriista maksumuse teeb projekt ise tasa, kui see on piisavalt suur või keeruline. Tsitaat ETL tööriista müüjalt: "Väärtusliku vahendi eesmärk ei ole lahendada triviaalseid probleeme, vaid muuta võimatud probleemid võimalikuks."
- Paljud ETL vahendid omavad integreeritud metaandmete reposiitorit, mis võimaldab sünkroniseerida metaandmed algallikas, sihtmärgi andmebaasis. Tehniliste oskustega äri inimesed, kes ei ole programmeerijad, saavad seetõttu ise efektiivselt kasutada ETL vahendeid.
- Enamik ETL tööriistu genereerivad automaatselt metaandmed igal protsessi sammul juurutades järjepidevat metaandmetel põhinevat metoodikat, mida kõik arendajad peavad järgima.
- Andmevoo ja andmete järjestuse ajaloo dokumentatsiooni saab tihti genereerida automaatselt metaandmete reposiitorist. ETL vahend hõlmab enamasti veakäsitlust ja monitooringut, juhaks kui midagi läheb valesti.
- ETL tööriistadel on sisseehitatud ühilduvused paljude allika ja sihtmärgi süsteemide jaoks, mis on kasulik, kui tuleb tööd teha paljude erinevate allikate või sihtmärkidega. ETL vahend peaks suutma teostada kõiki sorte kompleks andmetüüpide konversiooni. ETL tööriistad tavaliselt pakuvad ka käsu rea peal krüpteerimist ja pakkimist.
- Enamik ETL tööriistu omavad head jõudlust isegi väga suure andmekogumi korral. Tuleks mõelda programmpõhisele ETL valikule, kui teie ETL andmete maht on väga suur või kui see toimub paari aasta jooksul.

- Mõned ETL tööriistad omavad muutuse mõju analüüsi, mis aitab ette näha muutmise tagajärge edasises töös.

#### Käsitsi kirjutatud ETL eelised:

- Automatiseeritud üksuse katsetamine vahendid on saadaval käsitsi kodeerimise süsteemis, kuid mitte vahendipõhise lähenemisega. Näiteks *JUnit library* on kõrgelt hinnatud ja hästi toetatud vahend üksuse katsetamiseks Java programmis. On sarnaseid pakette ka teistele keeltele. Võite kasutada ka skriptikeelt, nagu *Tcl* või *Python*, luua katseandmed, käivitada ETL protsess ja kontrollida tulemusi. Testimisprotsessi automatiseerimine läbi ühe nendest meetoditest, oluliselt parandab tootlikkust ja tulemuse kvaliteeti.
- Objekt-orienteeritud programmeerimise tehnika aitab teha kõik muundamised ühtmoodi tõrketeavituses, valideerimises ja metaandmete uuenduses. Ei pruugi saavutada sellisel tasemel kontrolli nagu ETL-vahendi puhul.
- Käsitsi kirjutatud süsteemis saab palju otsesemalt käsitleda metaandmeid. Tegelikult metaandmeid juhtivad käsitsi kirjutatud süsteemid on palju paindlikumad kui mis tahes alternatiiv.
- Lühike ETL süsteemi nõuete analüüs kiiresti suunab edasi failipõhisele töötlemisele, mitte andmebaasi salvestatud protseduuridele. Failipõhised protsessid on palju selgemad. Nad on lihtsalt kodeeritud, lihtsalt testitud ja hästi mõistetavad.
- Tööriistal põhinev lähenemine võib piirata kasutajat vahendi müüja võimekuse ning kordumatu skriptikeele kaudu. Aga käsitsi võib arendada süsteemi ühtses ja hästi tuntud keeles.
- Käsi-kodeeritud ETL annab piiramatu paindlikkuse, juhul kui see on tõepoolest see mida vajatakse. Võid sõna otseses mõttes teha seda mida tahad. Paljudel juhtudel võib olla suureks eeliseks ainulaadne lähenemine või teises keeles teostamine. (Kimball Ralph, 2003) ja (Kimball, 2004, lk 10-12)

ETL on andmelao koostamise protsessi kõige aeganõudvam ja kallim osa. 15 aastat tagasi enamus ETL süsteeme olid käsitsi kirjutatud. ETL tarkvara turg on aga viimasel kümnendil jõudsalt kasvanud ning nüüdseks kasutatakse enamuses programmipõhist ETL protsessi. (Mundy, 2008)

## 2.2. ETL tarkvara vahendite võrdlus

Tutvustatakse erinevaid ETL tarkvara võrdluseid, kuna mahukamate ja üksikasjalisemate uuringute tulemused on kõik tasulised, siis magistritöös antakse neist vaid üldistav ülevaade.

### 2.2.1. Kaitstud omandiõigusega ja avatud lähtekoodiga vahendid

ETL vahendeid/programme leiame üsnagi palju: tasulised, tasuta, prooviversioone.

Jagunevad üldjoontes kaheks:

- kaitstud omandiõigusega
- avatud lähtekoodiga

#### Kaitstud omandiõigusega ETL vahendite tootjad

IBM DataStage ja Informatica on näited kaitstud omandiõigusega ETL vahenditest, mis omavad ulatuslikku kogemust ja tausta ETL vahendite turul. Omavad suutlikkust käsitleda väga suuri ja keerulisi andmemahte ebaühtlastes keskkondades. Need tooted pakuvad palju erinevaid funktsioone ja seetõttu nõuavad põhjalikku koolitust tõhusaks kasutamiseks. Muud lahendused selles kategoorias on Oracle Warehouse Builder, SAS Integration Server ja Business Objects Data Integration. Arvestades kulude ja jõupingutustega nende lahenduste rakendamisel, sobivad need tavaliselt kõige paremini suurte, keeruliste ja kogu ettevõtte ärianalüüsi sisaldavate stsenaariumitega. (Business Intelligence - Data warehousing - ETL, kp. puudub)

#### Avatud lähtekoodiga ETL vahendite tootjad

Turu teise spektri moodustavad avatud lähtekoodiga vahendite müüjad kes pakuvad ETL lahendusi mis on nüüd sisse viimas elujõulisi tehnoloogilisi alternatiive. Java raamistiku peal töötavad Talend, Pentaho, CloverETL, Enhydra Octopus ja Apatar<sup>1</sup> on näited selles kategoorias olevatest lahendustest. Nemad pakuvad tavaliselt oma tarkvara versioone tasuta, avatud lähtekoodi litsentsiga. Samas tavaliselt pakuvad ka majanduslikult litsentsitud versiooni, mis on võimaluste rohkem ning hõlmab toetust ja garantiid. (Adeptia, kp. puudub) Kui tegu on väga suurte andmemahtudega või laia andmeallikate hulgaga võib osutada vajalikuks kaitstud omandiõigusega ETL vahendite kasutamine, mis on erinevate võimaluste

---

<sup>1</sup> <http://cycloppedia.com/view/cloveretl.html>

poolest rikkalikumad. Kuid avatud lähtekoodiga vahendid on suurepärane võimalus andmete integreerimiseks olukordades, mis on rohkem keskendunud tootele, mis vajab andmete integratsiooni. (Pentaho Data Integration (PDI), Talend and IBM WebSphere DataStage, kp. puudub)

### **2.2.2. ETL tarkvara vahendite monitooring**

2003 a. alates on Hollandi päritolu sõltumatu konsultatsiooni ja analüüsi kompanii Passioned teostanud ETL vahendite turu monitooringut - *ETL Tool Survey*. Kahjuks paljud ettevõtte on arvamusel, et valides turuliidri, on nad automaatselt saanud õige lahenduse oma ettevõtte jaoks. Monitooring annab põhjaliku võrdluse erinevatest ETL vahenditest, millega tutvumine enne ETL vahendi valikut aitaks leida just täpselt ettevõtte vajadustele vastava tarkvaralahenduse. Jaanuaris 2011 a. hõlmas *ETL Tool Survey* uuring 19 populaarseimat ETL vahendit ja neid võrreldi 90 kriteeriumi alusel. (Täis uuringu tulemused saab osta interneti vahendusel hinnaga 750.- EUR.) (ETLtool, kp. puudub)

*ETL Tool Survey* 2009 monitooring maksab 425.- EUR aga Vincent McBurney on selle kohta teinud tasuta lühiülevaate, ta toob välja, et on olemas vaid üksikud ETL vahendite võrdlused, kuna neid on raske koostada ning ajakohastada.

- Omavahel võrreldi 19 erinevat ETL vahendit. Mille hulgas kallid tarkvarad nagu DataStage ja Informatica PowerCenter, odavad nagu näiteks Pervasive ja IKAN ning avatud lähtekoodiga vahendid nagu Pentaho, Clover ETL ja Talend.
- Clover ETL ja Talend paigutuvad kõige enam paranenud toodete nimekirja. Selles raportis omavad Talend ja Clover ETL enamuses võrdluskategooriates paremaid skooore kui Pentaho Data Integration (Kettle project).
- Kui on kaalumisel kas võtta kasutusele avatud lähtekoodiga või kaitstud omandiõigusega ETL vahend, siis see on hea võrdlusraport hindamaks tuleviku võimaluste ja kasutatavuse skooore, andes ideid millele tähelepanu pöörata vahendi valikul. (McBurney, My Review of the ETL Vendor Comparison Report from WWW.ETLTOOL.COM, 2009)

### **2.2.3. Momentum indexi võrdlus**

Selle võrdluse käigus on 96 avatud lähtekoodiga erinevate valdkondade tarkvara firmat järjestatud vastavalt Momentum Index skoorile. Võrdlus on läbi viidud 9 märts 2011 a.

Talend ja Pentaho paigutusid esikolmikusse.

1. Talend 89.5
2. Kaltura 76.2
3. Pentaho 74,3

(Momentum Index, 2011)

### **2.2.4. ETL turu suurus ja pakkujate turujaotus**

2007 a. võrdles Forrester juhtivaid kaitstud omandiõigusega ETL pakkuvaid ettevõtteid 68 kriteeriumi alusel ja leidis, et IBM ning Informatica omavad liiderpositsiooni kaitstud omandiõigusega ETL pakkujate seas, tänu võimele teostada operatiivset andmete integreerimist keerulistes keskkondades ning säilitada jõuline andmehalduse võimekus. Business Objects ja Oracle tuuakse ka välja liidritena, just tänu märkimisväärsetele kasutatavuse ja mastaapsuse võimalustele, kuid peamiselt on need kasutusel andmete ladustamise stsenaariumides ning vähem leiavad rakendust andmete operatiivsel integreerimisel. SAS ning Ab Initio kuuluvad turu tugevate mõjutajate hulka, sinna kategooriasse liigitatakse ka iWay Software, Microsoft, Pervasive ja Sunopsis, mis on küll suunatud rohkem andmete integreerimise professionaalidele. Uue tulijana on turul Sybase.

Analüüs on avaldatud 2.mail 2007 “The Forrester Wave™: Enterprise ETL, Q2 2007 (maksumus US \$2495). (McBurney, Tooling Around in the IBM InfoSphere , 2004 ) , (Rob Karel, 2007)

ETL tarkvara 2008 a. turujaotuse ja järgneva viie aasta tuleviku suundade kohta on tehtud aruanne - Worldwide Data Integration and Access Software 2008 Vendor and Segment Analysis 2009–2013 Forecast. (maksumus \$3,500) (Olofson, 2009)

### **2.2.5. ETL tarkvarafirmade poolt koostatud võrdlused konkurentidega**

ETL tarkvarafirmade poolt on koostatud arvukalt võrdlevaid analüüse konkurentidega. Näitena toon Adeptia ettevõtte poolt koostatud võrdleva aruande, mis on toodud Lisa 4 all. (Adeptia, kp. puudub)

## **2.3. ETL tarkvara turuliidrid**

Tuginedes eelnevale ETL tarkvara võrdlusele on välja toodud avatud lähtekoodiga ja kaitstud omandiõigusega turuliidrid järgmises nimekirjas:

- Talend
- PentahoKettle
- InformaticaPowerCenter
- OracleWarehouseBuilder

### **2.3.1. Talend**

- Talend on avatud lähtekoodiga andmete integreerimise töörist (vahend). Toode omab GNU General PublicLicense, version 2
- Kasutab koodi genereerimist ja kasutab GUI (GraphicalUserInterface) (rakendatakse-sarnaselt Eclipse RC)
- Ettevõtte alustas oma tegevusega 2005a.
- Omab tunduvalt vähemat kogukonda kui Pentaho, kuid toetatud 2 finantsettevõtte poolt.
- Genereerib Java või Perl koodi mida hiljem saab kasutada, rakendada serveri peal.
- Mugav graafiline liides, mis võimaldab kohandada ka SQL lauset
- Omab mugavaid monitooringu ja logimise süsteeme.

### **2.3.2. Pentaho (Kettle)**

- Pentaho on kommerts avatud lähtekoodiga integratsioonivahend tuntud veel nimega Kettle



- Kasutab uuenduslikku ja innovatiivset meta lähenemisviisi ja omab tugevat ning lihtsalt kasutatavat graafilist liidest
- Ettevõtte alustas oma tegevust 2001 aastal
- Omab tugevat kogukonda ja üle 13500 registreeritud kasutaja
- Kasutab iseseisvat Java generaatorit
- Liigutab andmeid paljude andmebaaside ja failide hulgas
- Pentaho pakub ettevõtete jaoks aruandluse, analüüsi, monitooringu ja integratsiooni võimeid äritegevuses

### **2.3.3. InformaticaPowerCenter**

- InformaticaPowerCenter omab väga head kaitstud omandiõigusega andmete integratsiooni kompleksi
- Ettevõtte asutati juba 1993
- On turuosa liider andmete integreerimisel
- Omab üle 2600 kliendi. Üle 100 nendest on valitsuse organisatsioonid
- Ettevõtte on orienteeritud andmete integratsioonile
- Omab üsna suurt paketti vahendeid ettevõtte süsteemide integreerimise jaoks nii uuemate kui vanemate süsteemide puhul

### **2.3.4. OracleWarehouseBuilder (OWB)**

- Oracle Warehouse Builder (OWB) ja nende integreerimise vahend on loodud Oracle organisatsiooniga. Oracle asendas Warehouse Builder (OWB) nende uue integratsiooni tootega Oracle Data Integrator, mis pakub sisseehitatud graafilist liidest realiseerimaks integratsiooni protsesse, käsitlemaks andmeid kliendi andmetes.
- On orienteeritud suurtele ettevõtetele
- Omab suurt kogukonda, häid tugiteenuseid ja koolitusi üle kogu maailma

### **3. Talend programmeerimisvahendi ülevaade**

Antakse põhjalikum ülevaade Talend vahendist nii teoreetilise kui praktilise poole pealt.

#### **3.1. Talend teoreetiline osa**

Talend tarkvara kuulub niinimetatud ETL vahendite hulka.

Antud tarkvaras kõik projektiga seotud tegevused on kooskõlastatud ja koordineeritud, lihtsasti muudetavad. Kergesti saab luua uusi versioone oma projekti/töösse, muuta programmeerimise tsüklit (arendamine--> testimine --> valmis projekt). Peale olulise komponendi või parameetri muutmist pakub programm võimaluse automaatselt uuendada kogu projekti sisu.

##### **3.1.1. Talendi põhimõte**

Antud tarkvara kasutades saab realiseerida nii lihtsaid ülesandeid kui ka suuremahulisi projekte. On võimalik ühenduse loomine praktiliselt kõikide tuntud andmebaasidega, failidega opereerimine, FTP, HTTP, Servlet ja muud. Pakub andmete sünkroniseerimist, liidestamist, koopiade tegemist ja muud.

Talend hõlmab kolme peamist valdkonda:

- andmete integreerimine
- andmete kvaliteet
- põhiantmete haldamine infosüsteemis

Peamised faktid:

- suurepärased tugiteenused
- suur hulk ühilduvusi
- avatud lähtekoodiga vahend
- lihtne õppida ja kasutada
- mugavad kasutusjuhendid ja iseseisvaks õppimiseks mõeldud manuaalid
- kaasaegne foorum, kust leiab suurema hulga vastuseid oma küsimustele
- võimalus laadida alla lisakomponendid

Mida saab juurde osta:

- toote koolitusi
- professionaalset tuge
- konsultatsioone

### 3.1.2. Talendi partnerid

Partnerid võib jagada gruppidesse:

- konsultatsiooni partnerid
- tehnoloogia partnerid

Tarkvara tugineb tugevale kogukonnale ja Talend on rajatud ulatusliku partnerite võrgustiku peale, Talendi Alliance programm, mis on Talend ökosüsteemi lahutamatu osa. Et mahutada erinevaid ettevõtteid ja vajadusi, see programm pakub mitmeid võimalusi teha koostööd Talend-i tehnoloogia, tugi ja müügi meeskondadega. Talend Alliance programm pakub juurdepääsu lisa tööriistadele, pakkudes pühendatud tuge ja koolitust, et tagada optimaalne koostöö Talend ja partnerite meeskonna vahel. Talend Alliance partnerid on organisatsioonid, mis pakuvad täiendavaid tooteid või teenuseid Talend kogukonnas. Talend jätkab oma partnerite võrgustiku loomisega ning värbab dünaamilisi ja asjatundlikke partnereid. Nende hulka kuuluvad süsteemiintegraatorid, koolituskeskused, sõltumatu tarkvara müüjad jne.

Talendi poolt on toetatud OEM programm, mis võimaldab kolmanda osapoole tarkvara müüjal, kas avatud lähtekoodiga või kaitstud omandiõigusega, lülitada Talendi tehnoloogia nende lahendustesse. (EOLAS, 2010 ) ja (Actuate and Talend Collaborate to Deliver Open Source BI Solution | Business Wire, 2011 )

Tehnoloogia partnerite hulka kuulub üle 30 ettevõtte, nendest on mulle rohkem tuntud:

- Jaspersoft<sup>2</sup>
- MySQL<sup>3</sup>
- SugarCRM<sup>4</sup>
- Ingres<sup>5</sup>

Lähemalt olen kokku puutunud Jaspersoft ja MySQL toodetega.

---

<sup>2</sup>[www.jasperforge.org](http://www.jasperforge.org)

<sup>3</sup><http://www.mysql.com>

<sup>4</sup><http://www.sugarcrm.com>

<sup>5</sup><http://www.ingres.com/>

### 3.1.3. Talendi plussid ja miinused

Plussiks võiks nimetada:

- Antud tarkvara töötab praktiliselt igas operatsiooni süsteemis. See tähendab et tugi ja käivitus on võimalik teostada: Windows, Linux, Mac, Solaris operatsiooni süsteemides. Lisaks veel valikuks kas 32 või 64 bit süsteem.
- Väga populaarne ning teadmiste/vastuste rikas foorum. Antud foorumis leiab vastused suuremale hulgale oma küsimustele. Registreeritud kasutajad saavad edukalt püstitada oma küsimuse ning saada neile vastuseid.
- Webinarid<sup>6</sup>– üle interneti on võimalik (eelnevalt registreerides) ja vebinaari ajal olles õigel veebisaidil (saadetakse link emaili teel peale registreerimist) osaleda online koolitusel. Üsna mugav vaadata kodus, interneti kohvikus, või tööl. Antakse ülevaade teooriast ja seejärel ka näitlik ülesanne. Iga online koolituse lõpus saab esitada ja saada kohe vastused oma küsimustele. Antud teemal probleemiga või ettepanekuga saab võtta ühendust ka emaili teel. Hiljem on võimalik antud salvestus leida ka online koolituste arhiivis.
- Versioonide vahetus toimub väga pehmelt. Näiteks kui olete teinud oma projekti eelnevas versioonis ja nüüdseks olete hakanud kasutama uuemat Talendi versiooni - ei tulene sellest kasutajale probleeme

Tarkvara põhimiinusteks võiks nimetada:

- Juurdepääs andmete allikate juurde nõuab JDBC draiverit
- Jätkuv Talend-i areng on enamasti tingitud emettevõttest, mida rahastab investeerimisfirma. On oht, et ettevõtte ei suuda osutada põhiteenuseid, mis on ainuke tulutoov allikas, ja seetõttu kaob toetav rahastaja. See vähendaks tugiteenuseid ja uuenduste loomist ning tekib risk, et ühilduvus uute andmebaasidega ei toimi. Kuna see firma on rohkem kui 5 aastat vana ja omab suurt hulka kasutajaid, on antud risk praegu juba üsna madal.
- Puuduvad täiendavad profileerimise, andmete kvaliteedi või metaandmete haldamise tooted. Need rollid võib täita kolmanda osapoole tooted, kuid vähe on avatud lähtekoodiga valikuid hetkel saadaval.

---

<sup>6</sup><http://www.talendforge.org/forum/>

- Toode võib jagada ressursse üle võrgu, kuid puudub automaatika andmete eraldamisel ja partitsioneerimisel ja see võib kaasa tuua kitsaskohti töötamisel. (Solutions - BI - Extract, Transform, Load, kp. puudub), (Talend - MIKE2.0, the open source methodology for Information Development)

### 3.2. Näidisülesande lahendamine Talend tarkvaraga

Näidisülesanne:

- saada kätte andmed personalitarkvarast
- edastada siseveebi haldajale, siseadministraatorile, domain-i administraatorile
- andmete liikumine toimuks 1 kord päevas.

Projektijuht palus magistritöö autoril teostada lühikese ajaga ülesande, mille sisuks oli tuua välja personali tarkvara andmebaasist (PostgreSQL) töötavate töötajate nimekiri koos nende andmetega (Nimi, Perenimi, email, ametikoht, ülemus, tööle asumise aeg, sünnipäev). Need andmed on vajalikud intraneti jaoks. Magistritöö autor võttis ühendust siseveebi haldajaga, täpsustas faili formaadi ja koha. Koostas vajaliku päringu, lisas vajalikud atribuudid ning ajastas programmi käivitamist oma masinas. Fail vajalike andmetega tekkis, tegin läbi vajalikud testid ning järgnes edukas rakendamine LIVE-i. Projektijuhi poolt antud ülesanne sai teostatud. Süsteemi administraator ja domaini siseadministraator samuti tundsid huvi nende andmete vastu, kuna nende töö hõlmab õiguste jagamist ning töötajate kontode avamist ja sulgemist. Tuli uuesti projekti hakata muutma, täiendama, et ka nemad saaksid neid andmeid kasutada.

Probleemi olemus:

- andmed peavad liikuma erinevate andmebaaside vahel: Postgresql, Oracle, MySQL
- lisaks saata emaili peale teavitused

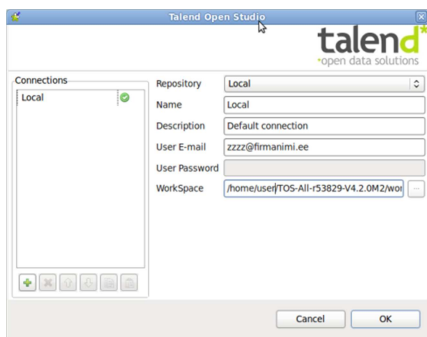
Postgresql-i andmebaasile tuli lisaks luua ühendus Oracle (kus antakse kasutajatele eriõigused) ja MySQL-i baasiga, kus saab veebihaldaja kergesti andmed kätte. Lisaks tuli saata iga kord muudatustest info emailile (domaini administraatorile), kes soovis näha tööle tulnud ja töölt lahkunud inimeste nimekirja. Talendi abil sai viie minutiga loodud MySQL andmebaasiga ühendus, hangitud vajalikud andmed, koostatud vajalik raport ja kantud üle teise andmebaasi. Talend osutus väga tõhusaks ja vajalikuks vahendiks.

### 3.3.Lahenduskäik Talend avatud lähtekoodiga vahendiga

Veebsaidil<sup>7</sup> tuleb valida alla laadimiseks omale sobiv versioon. Antud töö kirjutamisel olid saadaval järgmised programmi versioonid: TOS-V4.2.0RC1(minestone) ja TOS-V4.1.2 (General Availability). Otsustasin V4.1.2 kasuks, kuna 4.2 oli alles äsja valminud.

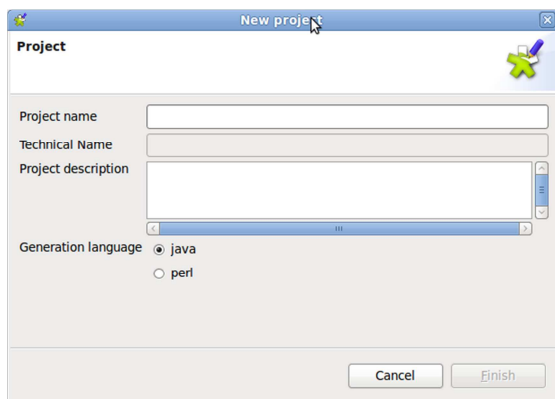
Järgmise sammuga tuleks valida kas kasutajale on vajalik Windows või siis Unix, Linux (GTK based) versioon. Olen 64 bit Linuxi kasutaja ja laen alla: TOS-All-V4.1.2. Pakin lahti ja käivitan käsurea peal: `#!/TalendOpenStudio-linux-gtk-x86_64`

Peale esimest käivitamist programm palub sisestada projekti põhiandmed.



*Joonis 3 Talendi käivitamise aken*

Peale OK nupu vajutamist tuleb järgmine aken, kus on oluline esmase käivitamise puhul määrata Action all: Create a new local project, Import demo project, Import existingproject(s) aslocal või Deletelocalproject(s). Näiteülesande puhul loome uue projekti, seda illustreerib Joonis 4. Oluline määrata Java või Perl kompileerimise programmeerimise keel.



*Joonis 4 Uue projekti loomise aken*

<sup>7</sup><http://www.talend.com/download.php>

### 3.3.1. Loo uus projekt (*Create a new local project*)

Valime *Create a new local project*, peale *Go* nupu vajutamist tuleb ette uue projekti seadistamise aken. Kus tuleb määrata projekti nimi ja *Generation language*: *java* või *perl* (millest edaspidi genereeritakse/kompileeritakse koodi). *Project description* alla saab lisada memo projekti kohta. Peale kohustuslike väljade täitmist vajutame *Finish* nuppu. Uus projekt on tekkinud, korraga saab tegeleda vaid ühe projekti täitmisega, teise projekti käivitamisel suletakse eelmine projekt.

### 3.3.2. Ärimudel (*Business Model*)

Mõistlik on alustada oma projekti ärimudelist (*Business Model*).

Talend Open Studio võimaldab projekteerimise süsteeme, ühendusi, protsesside ja nõuete standardiseeritud töövoogu märke läbi lihtsa graafilise raamatukogu.

Ärimudel võimaldab andmete integratsiooni projektiga seotutel graafiliselt esitada nende vajadused, olenemata tehnilise rakendamise nõuetest. Ärimudel aitab IT töötajal mõista neid väljendatud vajadusi ja üle kanda need tehnilisse protsessi (töökohta). Nad tavaliselt sisaldavad nii süsteeme ja protsesse mis juba toimivad ettevõttes aga samuti neid, mida läheb vaja tulevikus.

Ärimudeli disainimine on osa ettevõtete "parimatest tavadest", mille organisatsioonid peaksid väga varajases staadiumis lisama andmete integratsiooni projekti, et tagada selle edu. Sest ärimudel tavaliselt aitab avastada ja lahendada kiiresti projekti kitsaskohti ning nõrku külgi, nad aitavad piirata eelarve ületamist ja/või vähendada esmast investeeringut. Siis projekti rakendamise ajal ja pärast saab ärimudelit üle vaadata ja parandada, et peegeldada kõik nõutavaid muutusi.

Ärimudel on vähem tehniline vaade äri töövoogu vajadustest.

Üldiselt tüüpiline ärimudel sisaldab juba endas strateegilisi süsteeme või protsesse, mis töötavad ettevõttes samuti kui uued vajadused. Saab sümboliseerida need süsteemid, protsessid ja vajadused kasutades erinevaid kujundeid ning luua seoseid nende vahel. Tõenäoliselt kõiki neist saab kergesti kirjeldada, kasutades reposiitoreid ja formateerivaid vahendeid.

Talend Open Studio projekteerimise tööalas saate kasutada mitmeid vahendeid, et:

- juhtida teie ettevõtte vajadusi
- luua ja määrata arvukalt repositoreid oma mudeli objektidele
- määratleda oma mudeli objekti ärimudeli omadused

Seal saame defineerida kõik nõudmised ja ülesande vajadused. Paremalt pool saame kasutada tööriistade riba „palette”. Vaadake kui lihtsasti sain kajastatud sinna oma ülesande diagrammi.

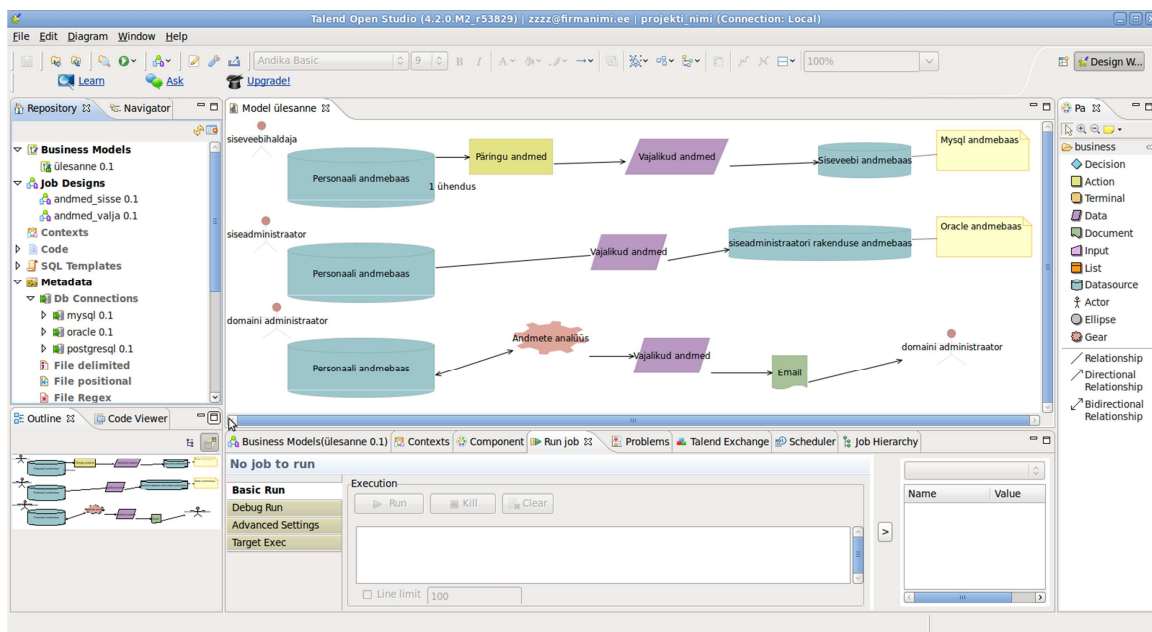
Ärimudeli vahend (*Modeler*) on tehtud järgnevatest paneelidest:

Talend Open Studio projekteerimise tööruum palett, kujundite ja joontega on spetsiifiline äriliseks modelleerimiseks ärimudeli paneel näitab täpset teavet kõigi või osade mudeliosade kohta. (Talend, kp. puudub)

Ärimudeli komponendid saab ühendada omavahel:

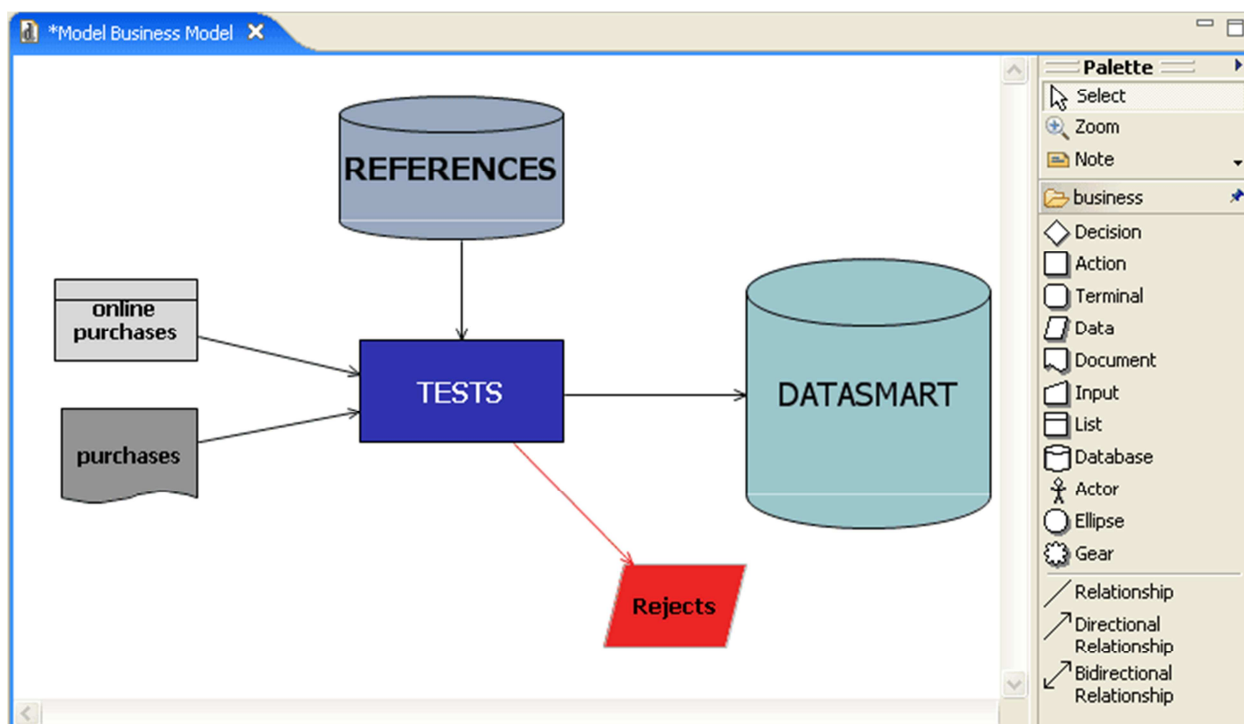
- ühesuunaliselt
- kahesuunaliselt
- seosnooltega

Juurde saab kirjutada memo tekstid ja lisainfo, noolte ja joonte paksust saab muuta, samamoodi saab muuta fondid ja värvid tekstiboksid.



Joonis 5 Näiteülesande ärimudeli näide





Joonis 6 Ärimudeli näide internetist (datawarehouse4u (Joonis), kp. puudub)

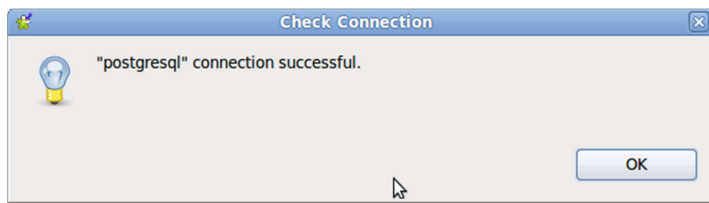
### 3.3.3. Metaandmed (Metadata)

Metaandmeteks nimetatakse andmeid andmete kohta, Talend-is on objektide metaandmed oluline osa - kirjeldavad dokumente. Kirjeldades oma ühendusi, vajaminevaid faile, emailid jne. Peale seda saame väga hõlpsasti realiseerida oma projekti töö disainija (*JobDesigner*) abil. Kirjeldan siis kõik ühendused, mida mul läheb vaja ning seadistan emaili.

### 3.3.4. Andmebaasi ühendus (Db Connections)

Ühenduse loomisel valin Metadata all oleva: DbConnections. Parema hiireklõpsuga valime menüüst Create connection, seejärel avaneb aken kus on oluline defineerida ühenduse nimi (Name) teised väljad on valikulised. Järgmisena vajutame Next, teise sammuga peame valime andmebaasi tüübi. Nimekiri ületab 30 ühenduse liiki, mõned andmebaasid on kirjeldatud mitme reaga. Tuleks valida omaale sobiv ühendus.

Valin PostgreSQL, peale kasutaja, parooli tunnust, serveri nime, andmebaasi nime sisestamist vajutame nupule Check. Juhul kui andmed olid õigesti sisestatud ja ühenduse loomine õnnestus kuvatakse vastav teade, mida illustreerib Joonis 7.



### *Joonis 7 Eduka ühenduse loomise teade*

Teen sama protseduuri ka teiste andmebaasidega. Mu valikusse kuulub Oracle ja MySQL andmebaasidega ühendamine. Siinkohal pööran tähelepanu asjaolule, et Oracle ühendamisel on mitu erinevat võimalust.

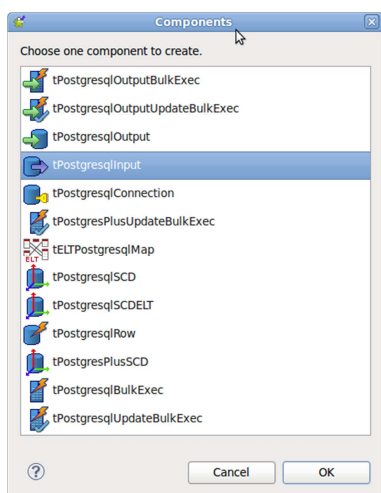
Väike nipp: Juhul kui olete korra juba teinud ühe nimega baasi ja olete ta kustutanud, siis ei lase programm teha uuesti sama nimega metaandmeid enne, kui samanimeline objekt on kustutatud programmi prügikastist (*RecycleBin*)-is.

Peale ühenduste defineerimist, saab hakata kasutama töö disainijat, millega saab ära teostada põhiülesande.

#### **3.3.5. Töö disainija (*Job Designer*)**

Parema hiireklahvi klõpsuga kutsume välja menüü ning valime esimese variandi - tekitame töö (*Create Job*).

Peale nime andmist ülejäänud ei ole kohustuslikud väljad ja neid ei pea sisestama. Nüüd valime metaandmete (*Metadata*) alt andmebaasi ühendused (*DbConnections*) ja sealt meie poolt tehtud ühenduse nimega postgresql, tirime selle meie poolt äsja tehtud töö nimega: „andmed\_valja” kohale. Programm pakub automaatselt valikud, kus on välja toodud põhitegevused mida saab teha meie poolt valitud metaandmetest. Kuna meie kirjeldatud ülesandes esimene soov on saada kätte andmed personali andmebaasist, siis valime: tPostgresqlInput. All valikust peame valima komponendid (*Components*) või teha topeltkliki ilmunud töö (*Job*) paneelile „postgresql” nimelisele komponendile mida illustreerib Joonis 8.



*Joonis 8 Näiteülesande Postgresql ühenduse viisid*

Vaatame lähemalt komponendi menüüd. Seal on juba defineeritud metaandmetes host, port, andmebaas, kasutaja, kasutaja salasõna. Lisaks peab defineerima päringu. Päringut saab defineerida programmi sisseehitatuna või eraldi kirjeldatud konfiguratsiooni failis. Vaatame hetkel tavalist sisseehitatud meetodit. Kui on täpselt teada päring saab selle kohe kirjutada jutumärkide vahele. Vastasel juhul tuleks valida päringu lahtri kõrval asuvat kolme punkti. Antud link viib päringute koostaja (*Sql Builder*) juurde. Enne seda tuleks saada kätte andmebaasi metaandmed teha (*Retrieve schema*).

### **3.3.6. Päringute- koostaja (*SqlBuilder*)**

Talendis on olemas sisseehitatud sql päringute jaoks suhteliselt primitiivne ja lihtne vahend: *SqlBuilder*. Selle abil saab edukalt ülesse ehitada sql päringut, kontrollida ja siluda tulemust. Enam ei ole vajalik eraldiseisvat SQL klienti.

Vaatame nüüd seda lähemalt: paremale lahtrisse saab kirjutada oma päringu, programm pakub/abistab kirjutamist, välja pakkudes muutujad, andmebaasi, tabeli või lahtri nimetusi. Vasakule alla tulevad päringu tulemused, saab vaadata mitu päringu tulemust – lahterdatud tabulaatoriga.

Oletame meil andmed puuduvad – saab samuti kergesti need sinna importida.

Kirjutame päringut – selle ülesande lõigu puhul näeb nii välja:

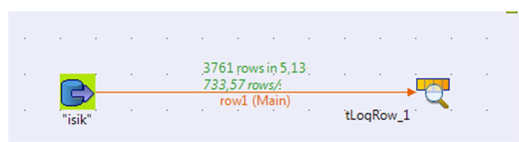
```
"select public.tbl_demo.eesnimi
        ,public.tbl_demo.nimi
        ,public.tbl_demo.email
        ,public.tbl_demo.ametikood
        ,public.tbl_demo.ylem_kood
        ,public.tbl_demo.tlalgus
        ,public.tbl_demo.synniaeg
from      public.tbl_demo;"
```

Peale päringu sobivust vajutame OK nupule.

Selle päringu tulemusi saab kergesti näha, monitoorida ja logida, kui panna alguses komponent rea lugeja (*tLogRow*). Samas konsool näitab päringu edenemist ja täitmist. Logi hõlmab

ühete veateadet, samuti alguse ja ka lõpu sõnumeid. Samuti näitab töö tulemust rea lugeja (*tLogRow*) komponent, mida kasutatakse töökoha disainil. Peale tulemuse ülevaatamist kustutame selle ära meie tööst. Antud päring jääb programmi sisse, kui me soovime seda salvestada konfiguratsiooni peab valima komponentide alt päringu tüüp. Ja muuta seda repositoori (*Repository*) peale. Rea lugeja komponendi leiame paremal pool komponentide paneelist (*palette*), otsimiseks võib edukalt kasutada filtrit. Kui antud komponent on leitud tirime seda töölehele (*JobDesigns*), seejärel ühendame komponent postgresql ja rea lugeja (*tLogRow*) omavahel. Selleks parema klahviga klikime postgresql ja valime rea (*Row*)--> Põhi (*Main*) suunamine noole rea lugeja (*tLogRow*) peale.

Peale seda sammu on komponendid omavahel ühendatud, käivita (*Run*) (*Job*) abil käivitame protseduuri.

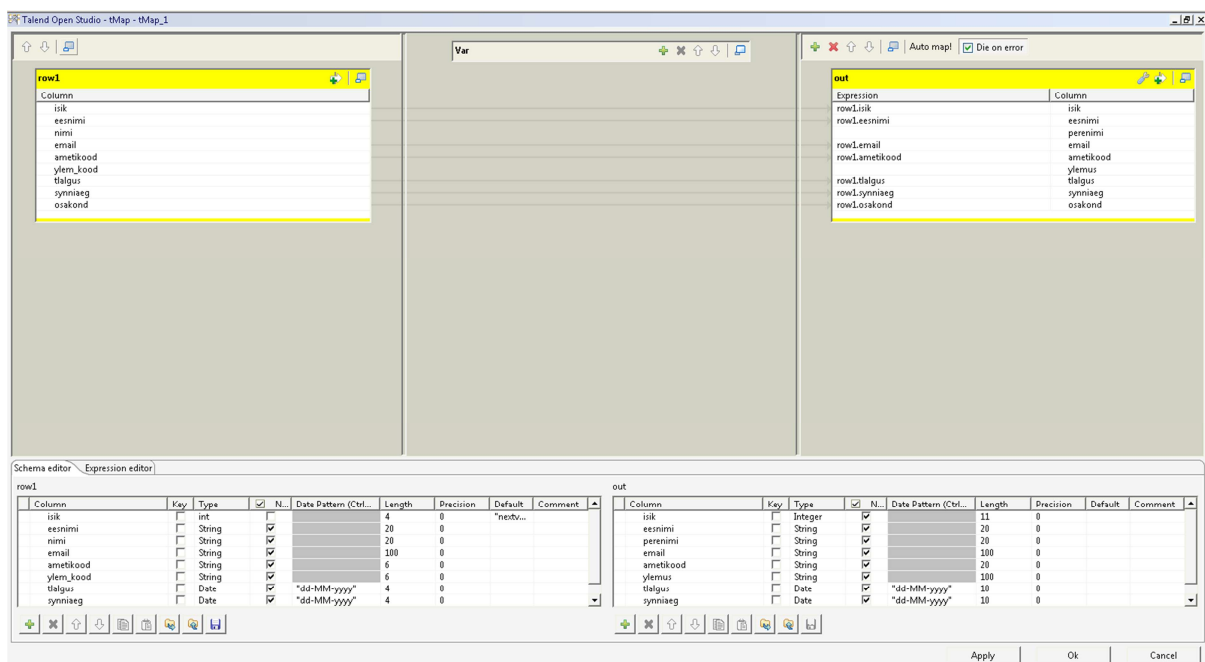


*Joonis 9 Monitooringu ja logi aken*

Protseduuri käivitamisel näeme päringu tulemusi, päringu kestvust. Nüüd kui on näha, et andmed liiguvad andmebaasist välja ja päringu tulemused on näha teostamise (*Execution*) juures (Joonis 9), võtame ära rea lugeja (*tLogRow*) komponendi ning lisame 1 alamülesande komponendi, mida varem oleme nimetanud MySQL.

Enne seda tuleb pärida skeem (*schema*), teeme parema hiire klõpsu MySQL peal ning (*Retrieve schema*) andmete kätte saamiseks. Otsime meid huvitava tabeli/tabeleid ning vajutame NEXT ning OK nupule. Nüüd meil on tekkinud MySQL alla skeemina kirjeldatud

tabel. Tirime selle tabeli meie töö (*Job*) peale, järgmise sammuga otsime komponentide seast mapingu vahendi (tMap). Selle komponendi abil saame edukalt teha mapingu andmete vahel. Parema hiire klõpsuga seome postgresql komponendi, valime põhi (*Main*) ja seome noolega mapingu (*tMap*) komponendiga. Järgmise sammuga seome mapingu (*tMap*) komponendi MySQL komponendiga ja anname talle nime out. Programm viisakalt pakub hankida infot sihtmärgi (*target*) kohta, vastame Yes. Vajutame mapingu (*tMap*) peale kaks korda, avab mapingu programm kus saame ära ühendada väljad. Kiirem viis ühendamiseks - valime üleval automaatne mapping (*autoMap*). Kuid programm suudab leida ja ära siduda samanimelisi väärtusi, ülejäänud mapingu teeme ise käsitsi. (Joonis 10) Kui kõik seosed on loodud vajutame OK nuppu.



Joonis 10 Näidisülesande mapping

Esimene ülesande lõik on teostatud, nüüd saame alustada ülesande teise lõiguga. Selleks metaandmetest valime uuesti postgresql nimelises tabelis skeemi nimega „tbl\_demo“ ja tirime selle töölauda. Avame alt komponendi kirjelduse ja sisestame komponendi kirjelduses (*Component*) päringu (*Query*):

```
SELECT
tbl_demo.isik,
tbl_demo.eesnimi,
tbl_demo.nimi,
tbl_demo.email,
```

tbl\_demo.ametikood,

tbl\_demo.osakond

FROM tbl\_demo

Antud päringuga saame kätte vajalikud andmed teise ülesande löigu teostamisel.

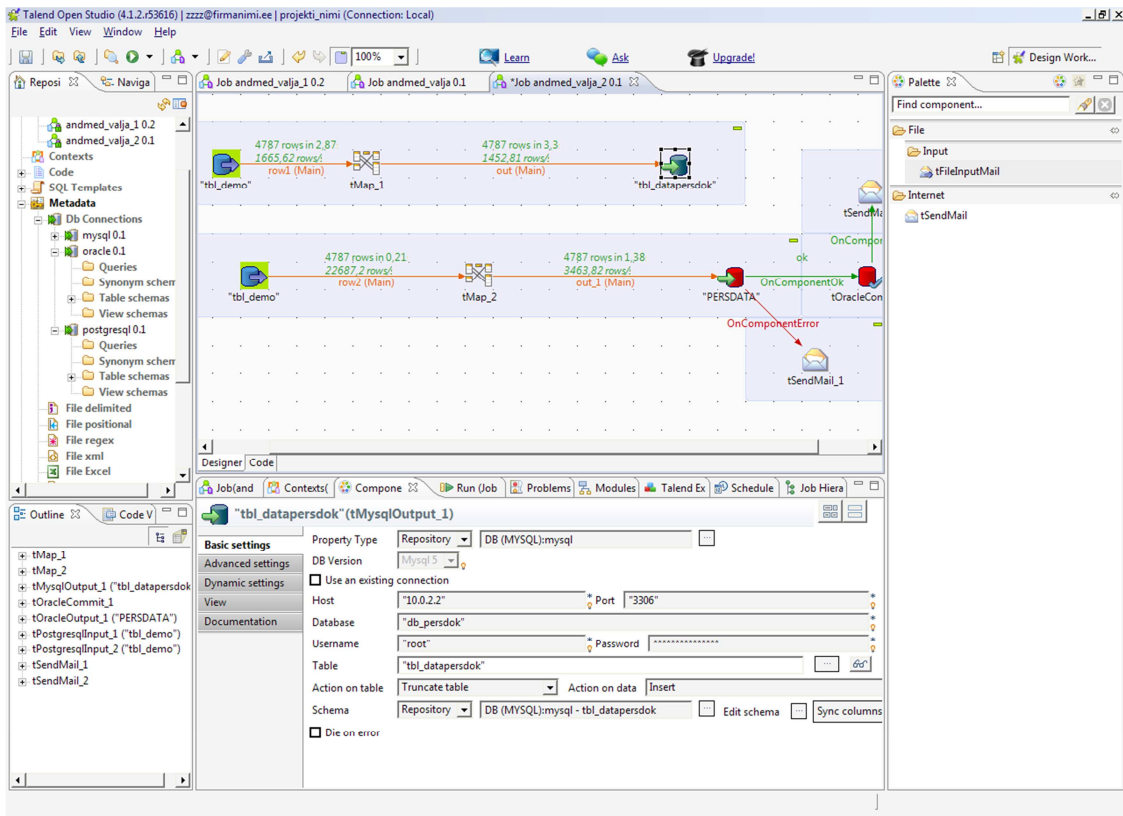
### 3.3.7. Kogu töö teostamine töölaual

Samuti avame Oracle nimelise andmebaasi ühenduse ning sealt valime tabeli skeemi „persdata“. Lisame komponentide alt (*tMap*), selle abil uuesti ühendame andmed omavahel. Pööran tähelepanu, et Oracle komponentide kirjelduses pole vaja tabeli tegevuse (*Action on table*) peal midagi valida. Otsime komponentide hulgast (*tSendMail*) ja (*tOracleCommit*) komponendid. Antud komponentide abil saab andmete puhul kinnitada ja saata emaili teel info kasutajale. Parema hiireklahvi klõpsuga seome postgresql komponendi, valime põhi (*Main*) ja seome noolega (*tMap*) komponendiga. Järgmise sammuga seome (*tMap*) komponendi Oracle komponendiga ja anname talle nime „out1“. Programm pakub hankida infot sihtmärgi (*target*) kohta, vastame Yes. Vajutame (*tMap*) peale kaks korda, avab mappingu programm, kus saame ära ühendada väljad. Peale väljade ühendamist vajutame OK peale. Andmed on seotud postgresql – (*tMap*) ja (*tOracleOutput*) komponendiga, järgmise sammuga ühendame (*OracleOutput*) ja (*tOracleCommit*), selleks teeme parem hiire klõps ja valime päästiku (*trigger*), alammenüüst valime korras tegevuse peale (*OnComponentOk*) ja tõmbame (*tOracleCommit*) peale. Viimase sammuga võiks siduda tõrke puhul (*OracleOutput*) (*tSendMail*) komponendiga vastava sisuga. Viimane saadab veateate juhul kui päringu tulemused mingil juhul pole ilmunud või on toonud veateate. Viimase sammuga käivitame (*Run*) käsu (*Run Job*) peal.

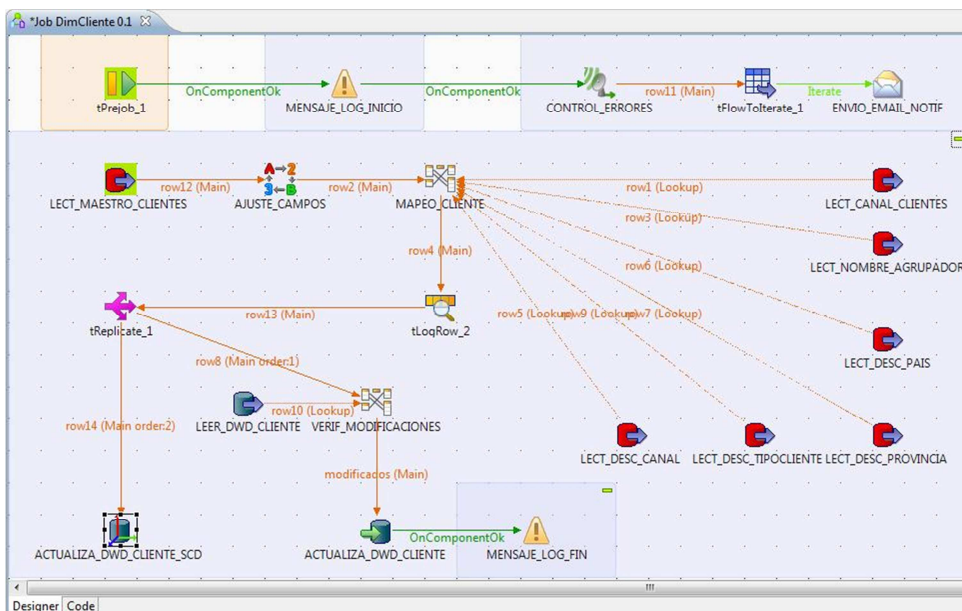
Peale edukat käivitamist näeme, kuidas andmed liiguvad komponentide vahel, jälgime statistikat, algus ja lõpp aega. Põhi käivitamise (*Basic Run*) kõrval olemas ka (*Debug Run*) tabulaator, selle abil saab (*Tracers Debug*).

Kui soovite muuta versiooni või projekti staatust (arendus, test või valmis toodang), selleks tuleb üleval menüüst valida: File→Edit project properties, Project Settings alt valime Version Management.

Kui töö on valmis ja vigu ei tule võime selle projekti arendusosa lugeda lõppenuks (Joonis 11) ja selle välja eksportida serverisse ülesse panemiseks. Selleks üleval menüüst valime File→ Export, alammenüüst valime JAR, määrame ekspordi jaoks kataloogi ja vajutame FINISH.



Joonis 11 Näidisülesande lõppfaas



Joonis 12 Arendusprojekti lõpp faas - näide internetist (Job Documentation Generated by Talend Open Studio (Joonis), 2010)

## 4. Talendi populaarsus

### 4.1. Küsitluse kokkuvõte

Küsimustik oli koostatud eesmärgiga, saada ülevaade Eesti ettevõtetes kasutatavatest IT integratsiooni vahenditest ning nende huvist uue tarkvara vastu.

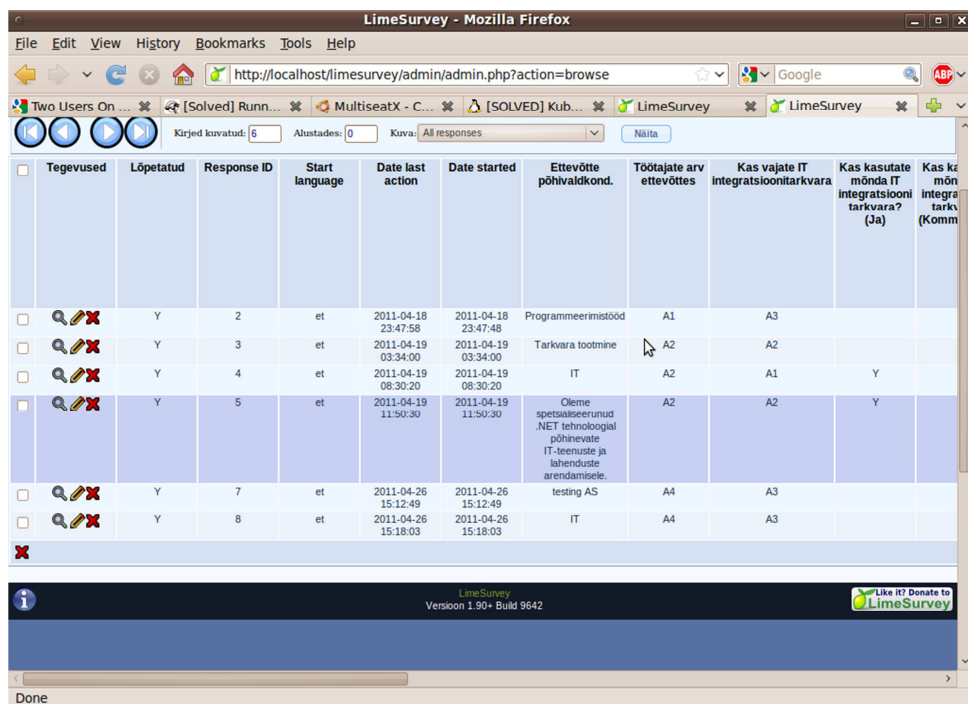
Küsimustikule vastamine oli anonüümne, vastamiseks oli nädal aega 18 aprill – 26 aprill 2011. Kokku saatsin elektroonilise küsimustiku 60-le IT lahendustega tegelevale ettevõttele. Küsimustikule vastas kõigest 6 ettevõtet ehk 10%. (Joonis 13) Neist kahel on olnud kokkupuudet andmete integreerimise vahenditega.

Kuna vastajaid oli vähe siis saadud andmete alusel põhjalikku analüüsi Talend vahendi kasutamise kohta Eesti ettevõtetes teha ei saa.

Magistritöö autorile on kogemuste ja tähelepanekute alusel jäänud mulje, et Talend vahendi kasutamine Eestis ei ole väga populaarne.

Põhjuseid võib olla mitmeid:

- teadmatus ETL tarkvarast ja vähesed teadmised antud valdkonnas
- teadmatus taskukohastest avatud lähtekoodiga ETL vahendite olemasolust
- puudub ülevaade ja kokkupuuted, kuidas Talend vahend saab ettevõttele kasulik olla



The screenshot shows the LimeSurvey administration interface in Mozilla Firefox. The browser address bar displays 'http://localhost/limesurvey/admin/admin.php?action=browse'. The interface includes a navigation menu, a search bar, and a table of survey responses. The table has columns for 'Tegevused', 'Lõpetatud', 'Response ID', 'Start language', 'Date last action', 'Date started', 'Ettevõtte põhivaldkond.', 'Töötajate arv ettevõttes', 'Kas vajate IT integratsioontarkvara', 'Kas kasutate mõnda IT integratsiooni tarkvara? (Ja)', and 'Kas kehtib mõn integratarkv (Komm)'. The table contains 8 rows of data, with the first row having a 'Y' in the 'Lõpetatud' column and a '2' in the 'Response ID' column.

Tegevused	Lõpetatud	Response ID	Start language	Date last action	Date started	Ettevõtte põhivaldkond.	Töötajate arv ettevõttes	Kas vajate IT integratsioontarkvara	Kas kasutate mõnda IT integratsiooni tarkvara? (Ja)	Kas kehtib mõn integratarkv (Komm)
	Y	2	et	2011-04-18 23:47:58	2011-04-18 23:47:48	Programmeerimistööd	A1	A3		
	Y	3	et	2011-04-19 03:34:00	2011-04-19 03:34:00	Tarkvara tootmine	A2	A2		
	Y	4	et	2011-04-19 08:30:20	2011-04-19 08:30:20	IT	A2	A1	Y	
	Y	5	et	2011-04-19 11:50:30	2011-04-19 11:50:30	Oleme spetsiaalseerunud .NET tehnoloogiat põhinevate IT-teenuste ja lahenduste arendamisele.	A2	A2	Y	
	Y	7	et	2011-04-26 15:12:49	2011-04-26 15:12:49	testing AS	A4	A3		
	Y	8	et	2011-04-26 15:18:03	2011-04-26 15:18:03	IT	A4	A3		

Joonis 13 Küsimustikule vastused



## 4.2. Autori ja teiste kasutajate kogemused Talendi kasutamisel

Talend võiks kasulik olla infosüsteemide arendamise ja integratsiooniga tegelevatele firmadele. Samuti võib abiks olla näiteks tarkvaraarendajale, testijale, siseadministraatorile. Magistritöös on pikemalt välja toodud kahe Talend-i kasutaja arvamus.

### 4.2.1. Autori kogemused

Esimest korda kuulsin Talendi tarkvarast oma töökaaslaselt 2008 a., kes soovitas mul kasutada Talendit ühe konkreetse projekti teostamisel. Ise olen sellest ajast alates Talendit kasutades realiseerinud üle 20 suurema projekti ning hulgaliselt väiksemaid tööülesandeid, väljavõtteid, andmete ekspordi või impordi infosüsteemi, aruannete koostamist ja emaili peale saatmise automatiseerimist.

Mainin siin mõned projektid/ülesanded:

- andmete korjamine, filtreerimine arvutite paremeetrise kohta (genereeritud aida32 programmiga) ja nende edastamine MySQL-i baasi, tulevikus kasutamise ja analüüsi jaoks
- kampaaniate tabelite analüüs (valideerimine), nende importimine põhiaandmebaasi
- tootepuu tootmine ja toodete edastamine e-pood-i, veebi (MsSQL andmebaasist)
- magistritöös kirjeldatud näidisülesanne, kus oli vaja personalitöötaja poolt sisestatud andmed saata mitmesse infosüsteemi

Kas ja mille poolest on Talendiga ülesannete lahendamine efektiivsem?

Ühendus baasiga luuakse hetkega, metaandmete säilimine, nende käsitlemine on kiire ja lihtne. Ühe projekti raames on kogu info jooksvalt kättesaadav ja muudetav, lihtne seadistada konfiguratsiooni faili, ei pea olema eraldi päringute programm, kuna see on sisse ehitatud. Kõik need väljatoodud omadused hoiavad palju aega kokku. Kui peaks käsitsi looma ühendused läheks tunduvalt kauem aega. Lihtne andmebaasi import või eksport võtaks ilma spetsiaalsete vahenditeta oluliselt rohkem aega.

Talendit kasutamine on majanduslikult jõukohane paljudele, kuna Talend Open Studio on vabavara (freeware) ja avatud lähtekoodiga (open source), operatsioonisüsteemist sõltumatu. Ei pea Talendit installerima – lihtsalt tõmbad internetis ära, pakid lahti ja töötab (eeldusel et olemas JDK). Talendit nõudmine on Java installimine, serveri peal see tähendab, et veel üks teenus mida peab üleval ja silma peal hoidma.

Algajale kasutajale soovitaks tulevikus Talendiga projekti teostamisel vaadata arhiveeritud veebinare, nendest leiab palju kasu, teiseks avada manuaal ja asuda oma projekti teostama. Talendi ja andmebaasi disaini vahendi võrdlus.

Kui palju kuulub aega, et importida andmebaasi sisse Exceli tabelist andmeid?

Kõik oleneb:

- vahendist
- andmetest
- pädevusest

Igal juhul aeg ja kvaliteet on vägagi tähtsad selle juures. Vaatame mõningaid vahendeid millega saaks importida näiteks Oracle baasi Excelist. Nendest olulised toon välja:

- DBDesigner - on MySQL andmebaasi disainimisvahend, mis pakub tuge ka teistele andmebaasidele, nende hulgas: SQL Server, Oracle, SQL Lite and ODBC.<sup>8</sup> (selle juures pean mainima, et see on parim lahendus mida olen siiani leidnud, esiteks: omandab GPL litsentsi, teiseks toetab Windows ja ka Linux operatsioonisüsteeme).
- Toad – suurepärase vahend, Toad Productivity Software for Database Development and Administration, Toad on kuulus kasvava DBA, arenduse ja andmete analüüsi produktiivsuse osas.<sup>9</sup>

Kui tuleks andmed lihtsalt importida – siis selleks on tõepoolest need vahendid ideaalsed. Nende vahenditega teeks selle samuti kiiresti, kuid kui vähemalt natukene tuleks andmeid muuta, täiendada – jääb nende vahenditega jänni. Selle eest Talendis tehtud töö võimaldaks vajadusel andmetega ka manipuleerida.

#### 4.2.2. Ühe Talendi kasutaja arvamus

Suures, üle 100 töötajaga, logistika ettevõttes töötava 15 aastase töökogemusega tarkvara arendaja ülevaade Talendi kasutamisest oma töös:

Esimest korda kuulsin, et on olemas selline kaasaegne integratsioonivahend nagu Talend, oma ettevõtte IT arendusjuhilt. See võis olla umbes 2008 aastal. Esimene ülesanne, mille ma lahendasin Talendi abiga oli järgmine:

---

<sup>8</sup> <http://www.fabforce.net/dbdesigner4/>

<sup>9</sup> <http://www.quest.com/toad/>

- laohaldus tarkvara genereerib igal öösel tekstifaili, milles asuvad tooted, mis peavad olema interneti teel tellitavad
- talendi abiga kannan andmed sellest tekstifailist MySQL andmebaasi tabelisse, mis on internetitellimise süsteemi põhjaks
- kui tabelis juba on antud toote kirje, siis värskendan andmed, kui sellist toodet tabelis ei ole, siis lisan antud toote tabelisse
- kui tabelis on tooteid, mida tekstifailis ei ole, siis need eemaldan tabelist.

Varem olin selle ülesande realiseerinud C Borland 4 ja seal selle ülesande realiseerimise maht oli 150 koodirida, millele kulus aega umbes 8 tundi. Nüüd oli vaja programmi sisse viia muudatused aga programmi kood ei olnud enam säilinud, otsustasin ülesande uuesti lahendamisel kasutada Talend tarkvara. Selleks kulus mul siis 4 tundi, tänapäeval teeksin seda poole tunniga.

Kui oskuslikult Talendi kasutada, siis on ta väga efektiivne, kuna väga paljud alamülesanded saab lihtsalt graafiliselt hiirega objekte lohistades ja omavahel ühendades ära teha ja Talend ise genereerib selle pealt Java koodi. Seda koodi saab ise veel täiendada aga lihtsamate ülesannete puhul pole seda vaja teha. Talendi plussideks loengi seda, et praktiliselt polegi vaja programmeerimisoskust ja lihtsamad integratsiooni ülesanded saab suhteliselt ruttu valmis. Talendi miinuseks võiks tuua selle, et ta on suhteliselt ressursinõudlik (Java VM), samuti võibolla selle, et alguses võib Talend tunduda veidi keeruline.

Algajale soovitaks Talendi põhitõed selgeks õppida, sest siis saab paljud integratsiooni ülesanded lahendada väga kiiresti ilma erilise programmeerimiseta.

#### **4.2.3. Teise Talendi kasutaja praktilise ülesande teostus**

Minu kokkupuudet Talend Open Studioga võib pidada pinnapealseks, arvestades kõiki võimalusi mida tööriistaga saab teha. Toode avaldab muljet suure hulga komponentidega, mis võimaldavad ligipääsu andmete allikatele ja sihtmärkidele ning manipuleerida ja integreerida andmeid mitmel erineval moel. Koodi genereerimine hõlmab Java või Perli koodi, on üsna laiendatav. Ma testisin Talend Open Studio Version 3.1, masin töötab 64-bit Ubuntu Linux 9.04. TOS on saadaval ka versioonid PowerPC ja x86 versioonide Linux, Solaris, 32- ja 64-bitiste versioonidega Windows ja Apple masinad OS X. Alustasin katsetamist TOS programmi käivitamisega ja uue projekti loomisega. Sel hetkel kui ma võiks valida Java-või Perl-põhise projekti loomise ning minu katsetused jooksid ummikusse Java-

põhise projektiga. Ma soovisin importida firma finantsandmed MySQL andmebaasi, selleks tuli luua kõigepealt metaandmete elemendid komadega eraldatud väärtused tekstifaili, kus kogutakse andmed, millega tahan edasi töötada. TOS esitas mulle väga lihtsa viisardi, mille käigus selgitati välja minu andmefail ja identifitseeriti kõik andmed separaatori ja tähemärkide vahel. Nagu ka teiste vahendite puhul, mida olen kasutanud, Talend sisaldab eelvaate akent, kus on näha ootuspäraselt töödeldud fail.

Kasutasin ka seda viisardit, et anda minu tulba mõistetele nimed ja andmetüübid. Aja informatsiooni lisamiseks andmefaili, määratlesin õige kuu-päev-aasta formaadi. Kui ma salvestasin metaandmete definitsiooni, mille viisard aitas mul luua, ma võin kohaldada selle määratluse teistesse sarnaselt vormindatud andmefailidesse. Nii näiteks, kord minu defineeritud metaandmetele põhinedes, milleks olid 2010 valimiste tsükkel, võiksin kasutada sama definitsiooni importides eelmiste valimiste tsüklid. Olen loonud eraldi metaandmeelemendi minu MySQL andmebaasi. See samm on seotud ühenduse informatsiooni välja täitmisega minu andmebaasi jaoks, just nagu ma oleks seadistanud andmebaasipäringu vahendi ühenduse. Ma võiks valida 31 erinevat liiki andmebaasi ühendust, mis sisaldab kõiki andmebaasi tooteid kui ka ODBC ja JDBC ühendusi.

Minu allikas ja eesmärgi elemendid on valmis, ma tirisin igat elementi oma töölauale, mille ma olin äsja loonud. Ma panin tähele, et minu piiritletud fail oleks sisend element ja et mu andmebaas aktsepteerib väljundit. Siis ma muutsin mõned andmebaasi väljundi elemendi seaded, näiteks, et TOS peaks looma uue tabeli ja saama andmeid. Ma käivitasin oma uue töö ja vaatasin kuidas Talendi read, read-sekundis ja möödunud-aja indikaatori näitajad märgivad minu katseandmete voolu MySQL andmebaasi. Need andmed olid olemas ja ma seadistasin uue töölaua - seekord on juba minu configureeritud andmebaasi element nagu sisendi allikas, samuti ka uus element, SugarCRM mooduli jaoks. Mida ma olen testinud väljundi aktsepteerimiseks. SugarCRM elemendi seadistamine oli sarnane minu andmebaasi ühenduse loomisega: võttes aluseks SugarCRM veebiteenuste URL ja minu autentimise info ning valisin tabeli mida tahtsin kasutada Sugar süsteemis. Ma lisasin ka kolmanda elemendi töö disainimiseks -tMap element, mis võimaldas mul mappida tulbad minu MySQL allikast minu valitud SugarCRM tabelisse, samuti muuta tulba väärtusi liikudes kahe vahel. Ma kasutasin Talend's expression builder, näiteks, et välja võtta perekonna nimi täisnime tulbast minu lähtetabelist, kasutades toote pakutavat funktsiooni. (Brooks, 2009)

### 4.3.Kümme põhjust, miks valida avatud lähtekoodiga programmeerimisvahend Talend

- Turu jaoks valmis tooted, mitte arendatavad versioonid

Talend Open Studio on terviklik toode hõlmates mitmeid funktsioone ja laia hulka ühilduvusi. Talend on lipulaev toode, see on kõige avatum, uuenduslikum ja võimsam andmete integratsiooni lahendus, mis on täna olemas. Talend Open Studio ei ole "kergekujuline toode" või katsetamise tarkvara, see sisaldab kõiki vajalikke funktsioone võimsa andmete integratsiooni protsessi teostamiseks ning on vabalt allalaaditav ja kasutatav GPL v2 litsentsi alusel.

Talend Integration Suite on täiustatud versioon, pakkudes täiendusi, suurettevõtte tasandi funktsionaalsust (koostöö, automatiseeritud kasutuselevõtt, laadimise tasakaalustus, monitooring) suurettevõtte klassi projektide jaoks. Lahendus hõlmab kõrgetasemelist tehnilist abi, et vastata ettevõtte küsimustele ja õiguslikku garantiid intellektuaalse omandi kaitseks (*IP indemnification*).

Talend Open Studio ja Talend Integration Suite funktsioonide võrdluse maatriks annab suurepärase ülevaate erinevate toodete omadustest. Maatriks on saadaval Talendi kodulehel.<sup>10</sup>

- Toote omandamisel ei teki takistusi

Talend Open Studio rakendamine on lihtne ja kiire - lihtsalt tuleb alla laadida uusim versioon Talend kodulehel ja see installeerida. Kui tekib mingeid probleeme (või lisateabe soov), siis kogukonna foorumid peaksid suutma selles osas aidata. Toode on tasuta, mis tähendab, et ei ole vaja selle kasutamist õigustada juhtkonnale või alustada formaalset omandamise protsessi, enne töö alustamist.

Ei kulu aega administratiivsetele ülesannetele ja tarkvara müüjaga kohtumisele. Toodet võib kasutada piiramatul viisil ja ajal. Samuti on tasuta toote uuendused.

---

<sup>10</sup> <http://www.talend.com/products-data-integration/Talend-Data-Integration-Features-Comparison-Matrix.pdf>

- Kiiresti õpitav

Talend vahend on kasutajasõbralik ja lihtne käsitseda. Graafiline kasutajaliides on intuiitiivne ja ei vaja formaalset koolitust.

Talendi töö disainer (*Job Designer*) pakub nii graafilist kui funktsionaalset vaadet aktuaalsele integratsiooni protsessile, kasutades komponentide kogu (*Component Library*) - komponentide ja ühilduvuste graafilist paletti. Integratsiooni protsess toimub lihtsalt komponentide ja ühilduvuste lohistamisel ja paigutamisel töölehel, joonistades ühendused ja seosed nende vahele ja pannes paika nende omadused (enamik omadusi pärinevad metaandmetest).

Talendi äri modelleerija (*Business Modeler*) tasakaalustab ülalt-alla lähenemist, mis võimaldab ärist väljaspool asuvatel sidusrühmadel osaleda integratsiooniprotsesside disainimisel ning jälgida arengu edusamme.

- Stabiilne ja prognoositav hinnamudel

Kaitstud omandiõigusega pakujate puhul kehtib andmete maks („*data tax*”), mis suurendab kulusid töödeldes täiendavaid andmeid sisaldavaid servereid, andmeallikaid / sihtmärke või segi üleminekul mitme tuumaga protsessorile vajab täiendavate litsentside ostmist. Seega, infrastruktuuri kulud ei ole prognoositavad ja ettevõtted võivad ettenägematutest kuludest tingituna ületada oma võimalused.

Talend lahenduse maksumus põhineb andmete integratsiooni protsessi arendajate arvil. Võite kasutada uusi andmeid nii nagu vaja. Näiteks kui luua uus aplikatsioon või omandada uusi äri – olukordi on mõnikord raske eelnevalt ette ennustada - ei ole vaja osta täiendavaid litsentse. Veelgi enam, kui firma liigub arendus valdkonnast hoopis hoolduspoolele, ei pea säilitama kõiki litsentse.

- Tugev kogukond

Kui viidatakse avatud lähtekoodile on seal ka kogukond. Avatud lähtekoodiga tarkvara kasutajatele toob selle kogukond kasu nii toe kui tootearenduse osas. Talendi kogukond pakub palju ressursse, et hõlbustada Talend lahenduse rakendamist ja haldamist -foorumid, viki, õpetused, blogid, vigade haldaja jne. Lisaks valdaja julgustab kasutajad, kes arendavad

ühilduvusi, funktsioone või muudavad lähtekoodi, andes sellest tagasisidet suuremale kogukonnale.

*Talend Babili* tõlgib kasutajaliidese abil. Neli kuud pärast *Talend-Babili* käivitamist-esimene kogukonna tõlke aplikatsioon- Talend võttis kasutusele üheksa keelepaketti, mille alusel sai lokaliseerida enamiku Talendi kasutajaid. Talend vahend on ainuke andmete integratsiooni lahendus tänasel turul, kes pakub lokaliseerimise võimalusi kasutajatele ja arendajatele, laiendades Talendi ulatust üle maailma.

- Ettevõtte poolsed tugiteenused

Vastupidiselt levinud arvamusele pakuvad avatud lähtekoodiga müüjad samuti tõelist tugiteenust, samalaadse kvaliteediga kui kaitstud omandiõigusega pakkujad. Sellises olukorras on ka Talend, kelle teenuste eesmärk on lihtsustada meeskonnatööd ja suurendada tootlikkust. Need teenused on juurutatud Talend ekspertide või Talend- sertifitseeritud partnerite poolt ning pakuvad samu teenuseid mis suured kaitstud omandiõigusega pakkujad.

Talend teab oma tooteid kõige paremini. Kasutajad kes allkirjastavad tehnilise toe lepingu, saavad endale meelerahu. Kuna Talend pakub väga head professionaalset tuge „*Talend 's Professional Support Services*“.

- Laialdane ühilduvus

Rohkem kui 400 ühilduvust, Talend lahendused annavad praktiliselt piiramatu ühilduvuse ettevõtte süsteemidega, andmebaasidega, tarkvara pakettidega, suurarvutitega, failidega, veebiteenustega jne.

Ükski teine tänapäeval turul kättesaadav lahendus ei pakub nii palju ühilduvusi. See on tingitud suurest hulgast Talendi kogukonna kasutajatest, kes teevad oma avastused üldsusele kättesaadavaks. Analoogselt Talendi professionaalse toe osakond jagab oma IT arenguid kogukonnaga võit-võit situatsiooni käsitlusel.

- Projektide universaalsus ja mitmekülgsus

Talend andmete integratsiooni lahendus ei ole piiratud standard ETL funktsionaalsusega äriandmete käsitlemisel (*Business Intelligence*), vaid saab kasutada ka operatiivsete andmete integratsiooni projektides. Tavaliselt tehakse neid siiani käsitsi, kuid integratsiooni vahendi näol saaks seda tõhustada.

Kõik organisatsioonid, sõltumata nende suurusest, seisavad silmitsi andmete haldamise mahu plahvatusliku suurenemisega. Hinnanguliselt igal aastal andmemaht kahekordistub, mida põhjustab tootmisliinide digitaliseerimine ja andmekiipide (RFID) kasutamine, automatiseeritud ERP ja CRM protsessid, kontori ja suhtlustarkvara, äriandmete käsitlemise tulemuslikkuse analüüsi vahendid, uued sidetehnoloogiad (email, VoIP), jne.

Andmete integratsiooni projektid on üsna erinevad. Talend lahendusega saab kasutada olemasolevaid arendusi uutes projektides, parandades uue projekti turule jõudmise aega ja tootlikkust.

- Funktsionaalsus, jõudlus, töökindlus

Kaugel stereotüübist, et avatud lähtekoodiga lahendusi ümbritseb vajakajäämisi professionaalsuses, Talend lahendus on tõeline äri tööriist. Pakkudes funktsionaalsuse taset, mis on võrdne kaitstud omandiõigusega pakkujatega. Ja avatud lähtekoodiga ei tähenda, et lahendus on välja töötatud vabatahtlike poolt oma vabast ajast. Talendil on oma R & D meeskond ja, nagu eespool arutletud, tema lahendused on rikastatud kogukonna panusega.

Kuigi Talend kontrollib toote tegevuskava, firma pidevalt kuulab kogukonna ja selle klientide arvamusi ja arvestab vajadustega, et aidata muutustega toimida. Talend on pühendunud pakuma võimsaid ja usaldusväärseid lahendusi, et olla atraktiivne paljude kasutajate jaoks.

- Optimeeritud kestus ja maksumus

Talend lahendused on 50% kuni 80% odavam kui samaväärsed kaitstud omandiõigusega turul pakutavad lahendused. Talend on ka odavam juurutada, hallata ja toetada.

Lisaks aitavad nad kaasa kiiremale arengule ja produktiivsuse tõusule võrreldes omandiõigusega kaitstud vahendite ja käsitsi kodeerimisega.



Näiteks ärimudeli vahend (*Business Modeler*) ühtlustab suhtlemise sidet ärikasutajate ja arendusmeeskonna vahel, vähendades aega, mis kulub ettevõttel tavaliselt vajaduste kindlaks tegemiseks ning uue tootearengu käivitamiseks, andes kasutajatele asjakohast ärilist teavet.

Lisaks saab andmete integratsiooni protsessi monitooringu hõlpsasti integreerida ülemaailmsesse järelevalve konsooli, mis tagab kasutajatele kiirema ja vastutustundlikuma laienemise protseduuri, vähendades seisakute kulukust ja tagades andmete hetke oleku. (Talend, 2009)

## KOKKUVÕTE

Magistritöö eesmärk on anda ülevaade Talend programmeerimisvahendist, et tutvustada arendajatele ja ettevõtetele Talend vahendi võimalusi andmete käsitlemiseks ning töötlemiseks.

ETL protsessil on andmeladu haldavas ja äriandmeid analüüsisivas (Business Intelligence) ettevõttes väga tähtis osa. Andmelao kavandamise ja arendamise protsessis kulub pool ajast ETL protsessi kujundamisele. Sobiva ja õige ETL protsessi valik tagab edaspidise eduka ja sujuva andmelao töö. (ETL - Extract Transform Load, kp. puudub)

Talend oli esimene avatud lähtekoodiga andmete integratsiooni tarkvara pakkuja ning sellega põhjustas Talend viimasel aastakümnel maailmas pöörde, kui lasi välja esimese versiooni Talend Open Studio 2006. Erinevalt patenteeritud, suletud lähtekoodiga lahendustest, mida saavad endale lubada vaid suured organisatsioonid, teeb Talend äriklassi andmete integratsioonivõimalused taskukohaseks ka väiksematele ettevõtetele. Kuna avatud lähtekoodil põhinevad ETL vahendid on kättesaadavad alles mõned aastad, siis ettevõtetele kes teostavad oma ETL protsessi ikkagi veel käsitsi kodeerituna annab magistritöö teadmised ETL programmipõhise teostamise võimalustest ja eelistest. (Talend, 2007 )

Eesmärgi saavutamiseks tuli tutvuda ETL protsessi teoreetilise poole olemusega, mille teostamisel toetuti peamiselt Ralph Kimball raamatule „Datawarehouse ETL toolkit“ (2004), Bill Inmon, keda on nimetatud ka andmete ladustamise isaks, avaldatud artiklitele ning erinevatele interneti lehekülgedele. Autori poolt teostati näidisrakendus Talend vahendiga, kajastati ka teiste kasutajate kogemusi Talendiga töötamisel.

Magistritöö käigus viidi läbi küsitlus andmete integratsioonivahendite ja Talend vahendi kasutamise kohta Eesti ettevõtetes. Uuringus paluti osaleda 60 ettevõttel, kellest realselt vastas kuus ning neist kahel on olnud kokkupuudet andmete integreerimise vahenditega. Kuna vastajaid oli vähe siis saadud andmete alusel põhjalikku analüüsi teha ei saa.

Loodan, et töö põhjal saavad lugejad hea ülevaate Talend vahendist, suudavad enam näha ja leida võimalusi selle programmeerimisvahendi kasutamiseks.

## Summary

The title of the Master's thesis is „Talend freeware – developer tool analyze“, the main objectives of this master thesis are to share and analyze the main of open source ETL and integration tool powered by Talend Corporation.

Main goal is to explain what is the ETL tools, to give explanation why is better to use in data warehouse such kind of tool, to give overview nowadays existing ETL tools and describe how to work with them.

The main Keywords: ETL, integration tool, data management, data integration, data integration systems, open source integration software, data quality, application integration, data migration, data synchronization.

The length of the thesis is 60 pages with 13 figures. The list of references contains 31 items of literature and Web resources. The thesis is written in Estonian.

In master thesis author gives overview of all popular ETL tool comparisons to help select the best ETL tool to fit your organization. The author of this master thesis recommends Talend – open source, what can solve many needs in your Data Warehouse to bring into strong data quality, monitoring etc.

Implementing Talend Open Studio in your company is convenient and easy—just to download the latest version from Talend’s website and install it. No license, no trial version is needed, just get one and start to use.

In the end of the master thesis, author brings out ten reasons for choosing Open Source Data Integration tool Talend.

The author has conducted in master's thesis with Open Source tool a demo project, a new user of Talend can take it for example in project.

Author hopes that sharing this powerful great tool with other Estonia developers will help company’s to choose right tool to manage their needs in Data Warehousing.

## **Kasutatud allikad**

- An introduction to Data Warehouses and the Data Warehousing.* (2007). Kasutamise kuupäev: 17. Aprill 2011. a., allikas ETL - Extract Transform Load:  
<http://www.sqlbook.com/Data-Warehousing/ETL---Extract-Transform-Load-6.aspx>
- Background of the ETL Tool Survey.* (2009). Kasutamise kuupäev: 11. Aprill 2011. a., allikas ETLtool: <http://www.etltool.com/etlsurveybackground.htm>
- Extract Transform Load.* (25. Oktoober 2010. a.). Kasutamise kuupäev: 26. Märts 2011. a., allikas Extract, Transform, Load:  
[http://it.toolbox.com/wiki/index.php/Extract,\\_Transform,\\_Load](http://it.toolbox.com/wiki/index.php/Extract,_Transform,_Load)
- Job Documentation Generated by Talend Open Studio (Joonis).* (2010). Kasutamise kuupäev: 6. Märts 2011. a., allikas roberto-espinoza: [http://roberto-espinoza.es/talend/DimCliente/DimCliente\\_0.1.html](http://roberto-espinoza.es/talend/DimCliente/DimCliente_0.1.html)
- Actuate and Talend Collaborate to Deliver Open Source BI Solution | Business Wire.* (14. Veebruar 2011 . a.). Kasutamise kuupäev: 24. Märts 2011. a., allikas Actuate and Talend Collaborate to Deliver Open Source BI Solution | Business Wire:  
<http://www.businesswire.com/news/home/20110214005172/en/Actuate-Talend-Collaborate-Deliver-Open-Source-BI>
- Adeptia. (kp. puudub). *ETL Vendors Comparison.* Kasutamise kuupäev: 17. Aprill 2011. a., allikas ETL Vendors Comparison - Adeptia ETL Suite, Informatica Powercenter, DataStage ETL, SSIS, Pervasive Integrator, Talend, Pentaho Kettle, CloverETL::  
[http://www.adeptia.com/products/etl\\_vendor\\_comparison.html](http://www.adeptia.com/products/etl_vendor_comparison.html)
- Brooks, J. (8. September 2009. a.). *REVIEW: Talend Open Studio Makes Quick ETL Work of Large Data Sets.* Kasutamise kuupäev: 4. Märts 2011. a., allikas REVIEW: Talend Open Studio Makes Quick ETL Work of Large Data Sets - Database - News & Reviews - eWeek.com: <http://www.eweek.com/c/a/Database/REVIEW-Talend-Open-Studio-Makes-Quick-ETL-Work-of-Large-Data-Sets-281473/>
- EOLAS. (4. Veebruar 2010 . a.). *TALEND AND BUSINESS & DECISION CREATE PARTNERSHIP TO MANAGE INFORMATION CAPITAL - EIM.* Kasutamise kuupäev: 17. Märts 2011. a., allikas TALEND AND BUSINESS & DECISION CREATE PARTNERSHIP TO MANAGE INFORMATION CAPITAL - Business & Decision Group: <http://www.businessdecision.com/Actualite/755/talend-and->

business-decision-create-partnership-to-manage-information-capital/eim/1238-news.htm

ETLtool. (kp. puudub). *Background of the ETL Tool Survey*. Kasutamise kuupäev: 24. Märts 2011. a., allikas ETLtool: <http://www.etltool.com/etlsurveybackground.htm>

Inmon, B. (8. Mai 2008. a.). *A Brief History of ETL*. Kasutamise kuupäev: 3. Märts 2011. a., allikas A Brief History of ETL by Bill Inmon - BeyeNETWORK: <http://www.b-eye-network.com/view/7040>

Inmon, B. (01. Jaanuar 2008. a.). *ETL: Extract, Transform and Load for Data Warehousing*. Kasutamise kuupäev: 22. Veebruar 2011. a., allikas <http://www.b-eye-network.com/view/6680>

Kimball Ralph, N. G. (31. Mai 2003. a.). Is Hand-Coded ETL the Way to Go? *Intelligent Enterprise*.

Kimball, R. (2004). *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data* (1 tr.). Wiley.

*Business Intelligence - Data warehousing - ETL*. (kp. puudub). Kasutamise kuupäev: 21. Veebruar 2011. a., allikas Definition and concepts of the ETL process: [http://etl-tools.info/en/bi/etl\\_process.htm](http://etl-tools.info/en/bi/etl_process.htm)

*datawarehouse4u (Joonis)*. (kp. puudub). Kasutamise kuupäev: 3. Aprill 2011. a., allikas <http://datawarehouse4u.info/Talend-Open-Studio.html>

*ETL - Extract Transform Load*. (kp. puudub). Kasutamise kuupäev: 25. Aprill 2011. a., allikas The Extract Transform Load (ETL) design process is perhaps the most time consuming stage of the Data Warehouse project. This article outlines the main ideas and processed that are undertaken by ETL.: <http://www.sqlbook.com/Data-Warehousing/ETL---Extract-Transform-Load-6.aspx>

*Extract Transform Load (ETL) | Data Warehousing*. (kp. puudub). Kasutamise kuupäev: 27. Aprill 2011. a., allikas Extract Transform Load (ETL) | Data Warehousing: <http://free-books-online.org/computers/data-warehousing/extract-transform-load-etl/>

*Solutions - BI - Extract, Transform, Load*. (kp. puudub). Kasutamise kuupäev: 20. Märts 2011. a., allikas Solutions - BI - Extract, Transform, Load: <http://www.tripodsoftcs.com/solution/bi/ETL.htm>

*What is ETL, Extract Transform and Load?* (kp. puudub). Kasutamise kuupäev: 4. Aprill 2011. a., allikas What is ETL, Extract Transform and Load?: <http://www.etltool.com/what-is-etl.htm>

- McBurney, V. (28. Oktoober 2004 . a.). *Tooling Around in the IBM InfoSphere* . Kasutamise kuupäev: 28. Aprill 2011. a., allikas Toolbox Today - Bill Strikes Again:  
<http://it.toolbox.com/blogs/infosphere/forrester-etl-wave-from-2004-to-2007>
- McBurney, V. (6. August 2009. a.). *My Review of the ETL Vendor Comparison Report from WWW.ETLTOOL.COM*. Kasutamise kuupäev: 24. Märts 2011. a., allikas Tooling Around in the IBM InfoSphere : <http://it.toolbox.com/blogs/infosphere/my-review-of-the-etl-vendor-comparison-report-from-wwwetltoolcom-33310>
- Momentum Index. (9. Märts 2011. a.). *96 Open Source Companies Ranked by Momentum Index - Momentum Index*. Kasutamise kuupäev: 8. Aprill 2011. a., allikas Momentum Index: [http://momentumindex.com/96-open-source-start-ups-ranked-by-momentum-index-3/#disqus\\_thread](http://momentumindex.com/96-open-source-start-ups-ranked-by-momentum-index-3/#disqus_thread)
- Mundy, J. (6. Aprill 2008. a.). *Kimball University: Should You Use An ETL Tool?*  
 Kasutamise kuupäev: 23. Märts 2011. a., allikas Intelligent Enterprise: Should You Use An ETL Tool?:  
<http://www.intelligententerprise.com/showArticle.jhtml;jsessionid=DNBKSH5WLWCBEQSNLPCCKHSCJUNN2JVN?articleID=207602398>  
 ([http://www.kimballgroup.com/html/articles\\_search/articles2008/0804E.html](http://www.kimballgroup.com/html/articles_search/articles2008/0804E.html))
- Olofson, C. W. (November 2009. a.). *Worldwide Data Integration and Access Software 2008 Vendor and Segment Analysis*. Kasutamise kuupäev: 13. Märts 2011. a., allikas IDC - Document at a Glance - 220559: <http://www.idc.com/getdoc.jsp?containerId=220559>
- Pentaho Data Integration (PDI), Talend and IBM WebSphere DataStage. (kp. puudub). *ETL*. Kasutamise kuupäev: 23. Märts 2011. a., allikas Solutions - BI - Extract, Transform, Load: <http://www.tripodsoftcs.com/solution/bi/ETL.htm>
- Rob Karel, C. M. (2. Mai 2007. a.). *The Forrester Wave™: Enterprise ETL, Q2 2007*. Kasutamise kuupäev: 21. Aprill 2011. a., allikas The Forrester Wave™: Enterprise ETL, Q2 2007 - Forrester Research:  
[http://www.forrester.com/rb/Research/wave%26trade%3B\\_enterprise\\_etl%2C\\_q2\\_2007/q/id/39637/t/2](http://www.forrester.com/rb/Research/wave%26trade%3B_enterprise_etl%2C_q2_2007/q/id/39637/t/2)
- Talend - MIKE2.0, the open source methodology for Information Development*. (kuupäev puudub). Kasutamise kuupäev: 17. märts 2011. a., allikas <http://mike2.openmethodology.org/wiki/Talend>
- Talend. (22 . Octoober 2007 . a.). *Talend Upgrades the Industry's First Open Source Data Integration Solution*. Kasutamise kuupäev: 4. Aprill 2011. a., allikas Talend Upgrades the Industry's First Open Source Data Integration Solution | Business Integration /

Middleware:: <http://www.mcpressonline.com/software-news/business-integration-middleware/talend-upgrades-the-industrys-first-open-source-data-integration-solution.html>

Talend. (16. September 2009. a.). *The Top 10 Reasons for Choosing Open Source Data Integration - A Talend White Paper*. Kasutamise kuupäev: 7. Aprill 2011. a., allikas A Talend White Paper: [http://www.talend.com/news/newsletter16\\_en.php](http://www.talend.com/news/newsletter16_en.php)

Talend. (kp. puudub). *User Guide (UG) of Talend Open Studio: Provides general use information*. Kasutamise kuupäev: 14. Märts 2011. a., allikas Open Source Integration Software and Tools - Talend: <http://www.talend.com/downloads/download.php?version=user-guide-41b-en&src=HomePage&kid=&srv=eu1>

Techtiks Informatica Group. (kp. puudub). *Introduction to ETL - Business Case*. Kasutamise kuupäev: 23. Aprill 2011. a., allikas Introduction to ETL - Business Case: [http://informatica.techtiks.com/etl\\_tools.html](http://informatica.techtiks.com/etl_tools.html)

**Lisad**



## Lisa 1 Koodinäide MySQL-i ühenduste metaandmetest

```
<?xml version="1.0" encoding="UTF-8"?>
<xmi:XMI          xmi:version="2.0"          xmlns:xmi="http://www.omg.org/XMI"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:TalendMetadata="http://www.talend.org/metadata/connection/2010"
xmlns:orgomg.cwm.resource.relational="http://orgomg/cwm/resource/relational.ecore"
xmlns:relational="http://www.talend.org/cwm/resource/relational/2010">
  <TalendMetadata:DatabaseConnection          xmi:id="_dSGi8GAIEeCiQs63fc893Q"
name="MySQL" dataPackage="_g30ZwGR1EeC7VKQONn2ekg" DatabaseType="MySQL"
DriverClass="org.gjt.mm.MySQL.Driver"
URL="jdbc:mysql://10.0.2.2:3306/db_persdok?noDatetimeStringSync=true"
dbVersionString="MYSQL_5"          Port="3306"          Username="root"
Password="0YVKc0eMnRSv9MrLxRHIkg==" ServerName="10.0.2.2" DatasourceName=""
FileFieldName="" SID="db_persdok" SqlSynthax="SQL Syntax" DbmsId="MySQL_id"
ProductId="MYSQL" AdditionalParams="noDatetimeStringSync=true" StandardSQL="true"
SQLMode="false" UiSchema=""/>
  <orgomg.cwm.resource.relational:Catalog          xmi:id="_g30ZwGR1EeC7VKQONn2ekg"
name="db_persdok" dataManager="_dSGi8GAIEeCiQs63fc893Q">
    <ownedElement          xsi:type="relational:TdTable"
xmi:id="_g30ZwWR1EeC7VKQONn2ekg"          name="tbl_datapersdok"
id="_f8INKGR1EeC7VKQONn2ekg" label="tbl_datapersdok" tableType="TABLE">
      <taggedValue xmi:id="_g30ZwmR1EeC7VKQONn2ekg" tag="Comment" value=""/>
      <feature          xsi:type="relational:TdColumn"          xmi:id="_g30Zw2R1EeC7VKQONn2ekg"
name="isik" label="isik" length="11" sourceType="INT" talendType="id_Integer">
        <taggedValue xmi:id="_g30ZxGR1EeC7VKQONn2ekg" tag="Comment" value=""/>
        <initialValue          xsi:type="relational:TdExpression"
xmi:id="_g30ZxWR1EeC7VKQONn2ekg" language="COLUMN_DEF"/>
        <sqlDataType          xmi:id="_g30ZxmR1EeC7VKQONn2ekg"          name="INT"
numericPrecisionRadix="10" javaDataType="4" nullable="columnNullable"/>
      </feature>
```

```

    <feature xsi:type="relational:TdColumn" xmi:id="_g30Zx2R1EeC7VKQONn2ekg"
name="eesnimi" label="eesnimi" length="20" sourceType="VARCHAR"
talendType="id_String">
    <taggedValue xmi:id="_g30ZyGR1EeC7VKQONn2ekg" tag="Comment" value=""/>
    <initialValue
xsi:type="relational:TdExpression"
xmi:id="_g30ZyWR1EeC7VKQONn2ekg" language="COLUMN_DEF"/>
    <sqlDataType xmi:id="_g30ZymR1EeC7VKQONn2ekg" name="VARCHAR"
numericPrecisionRadix="10" javaDataType="12" nullable="columnNullable"/>
</feature>
    <feature xsi:type="relational:TdColumn" xmi:id="_g30Zy2R1EeC7VKQONn2ekg"
name="perenimi" label="perenimi" length="20" sourceType="VARCHAR"
talendType="id_String">
    <taggedValue xmi:id="_g30ZzGR1EeC7VKQONn2ekg" tag="Comment" value=""/>
    <initialValue
xsi:type="relational:TdExpression"
xmi:id="_g30ZzWR1EeC7VKQONn2ekg" language="COLUMN_DEF"/>
    <sqlDataType xmi:id="_g30ZzmR1EeC7VKQONn2ekg" name="VARCHAR"
numericPrecisionRadix="10" javaDataType="12" nullable="columnNullable"/>
</feature>
    <feature xsi:type="relational:TdColumn" xmi:id="_g30Zz2R1EeC7VKQONn2ekg"
name="email" label="email" length="100" sourceType="VARCHAR"
talendType="id_String">
    <taggedValue xmi:id="_g30Z0GR1EeC7VKQONn2ekg" tag="Comment" value=""/>
    <initialValue
xsi:type="relational:TdExpression"
xmi:id="_g30Z0WR1EeC7VKQONn2ekg" language="COLUMN_DEF"/>
    <sqlDataType xmi:id="_g30Z0mR1EeC7VKQONn2ekg" name="VARCHAR"
numericPrecisionRadix="10" javaDataType="12" nullable="columnNullable"/>
</feature>
    <feature xsi:type="relational:TdColumn" xmi:id="_g30Z02R1EeC7VKQONn2ekg"
name="ametikood" label="ametikood" length="20" sourceType="VARCHAR"
talendType="id_String">
    <taggedValue xmi:id="_g30Z1GR1EeC7VKQONn2ekg" tag="Comment" value=""/>
    <initialValue
xsi:type="relational:TdExpression"
xmi:id="_g30Z1WR1EeC7VKQONn2ekg" language="COLUMN_DEF"/>

```

```

    <sqlDataType xmi:id="_g30Z1mR1EeC7VKQONn2ekg" name="VARCHAR"
numericPrecisionRadix="10" javaDataType="12" nullable="columnNullable"/>
  </feature>
  <feature xsi:type="relational:TdColumn" xmi:id="_g30Z12R1EeC7VKQONn2ekg"
name="ylemus" label="ylemus" length="100" sourceType="VARCHAR"
talendType="id_String">
    <taggedValue xmi:id="_g30Z2GR1EeC7VKQONn2ekg" tag="Comment" value=""/>
    <initialValue
xsi:type="relational:TdExpression"
xmi:id="_g30Z2WR1EeC7VKQONn2ekg" language="COLUMN_DEF"/>
    <sqlDataType xmi:id="_g30Z2mR1EeC7VKQONn2ekg" name="VARCHAR"
numericPrecisionRadix="10" javaDataType="12" nullable="columnNullable"/>
  </feature>
  <feature xsi:type="relational:TdColumn" xmi:id="_g30Z22R1EeC7VKQONn2ekg"
name="tlaligus" label="tlaligus" length="10" sourceType="DATETIME"
talendType="id_Date" pattern="&quot;dd-MM-yyyy&quot;;">
    <taggedValue xmi:id="_g30Z3GR1EeC7VKQONn2ekg" tag="Comment" value=""/>
    <initialValue
xsi:type="relational:TdExpression"
xmi:id="_g30Z3WR1EeC7VKQONn2ekg" language="COLUMN_DEF"/>
    <sqlDataType xmi:id="_g30Z3mR1EeC7VKQONn2ekg" name="DATE"
numericPrecisionRadix="10" javaDataType="91" nullable="columnNullable"/>
  </feature>
  <feature xsi:type="relational:TdColumn" xmi:id="_g30Z32R1EeC7VKQONn2ekg"
name="synniaeg" label="synniaeg" length="10" sourceType="DATETIME"
talendType="id_Date" pattern="&quot;dd-MM-yyyy&quot;;">
    <taggedValue xmi:id="_g30Z4GR1EeC7VKQONn2ekg" tag="Comment" value=""/>
    <initialValue
xsi:type="relational:TdExpression"
xmi:id="_g30Z4WR1EeC7VKQONn2ekg" language="COLUMN_DEF"/>
    <sqlDataType xmi:id="_g30Z4mR1EeC7VKQONn2ekg" name="DATE"
numericPrecisionRadix="10" javaDataType="91" nullable="columnNullable"/>
  </feature>
  <feature xsi:type="relational:TdColumn" xmi:id="_g30Z42R1EeC7VKQONn2ekg"
name="osakond" label="osakond" length="100" sourceType="VARCHAR"
talendType="id_String">
    <taggedValue xmi:id="_g30Z5GR1EeC7VKQONn2ekg" tag="Comment" value=""/>

```

```
<initialValue                                xsi:type="relational:TdExpression"
xmi:id="_g30Z5WR1EeC7VKQONn2ekg" language="COLUMN_DEF"/>
  <sqlDataType    xmi:id="_g30Z5mR1EeC7VKQONn2ekg"    name="VARCHAR"
numericPrecisionRadix="10" javaDataType="12" nullable="columnNullable"/>
  </feature>
</ownedElement>
</orgomg.cwm.resource.relational:Catalog>
</xmi:XMI>
```

## Lisa 2 Näide MySQL-i ühenduse omadustest

```
<?xml version="1.0" encoding="UTF-8"?>
<xmi:XMI          xmi:version="2.0"          xmlns:xmi="http://www.omg.org/XMI"
xmlns:TalendProperties="http://www.talend.org/properties">
  <TalendProperties:Property          xmi:id="_dSGi8WAIEeCiQs63fc893Q"
id="_dSF74GAIEeCiQs63fc893Q"          label="MySQL"          creationDate="2011-04-
06T04:41:21.631-0430" modificationDate="2011-04-11T15:54:30.108-0430" version="0.1"
statusCode="DEV" item="_dSGi82AIEeCiQs63fc893Q">
    <author href="../../talend.project#_LCg_YF9cEeCditfL8-0axw"/>
  </TalendProperties:Property>
  <TalendProperties:ItemState xmi:id="_dSGi8mAlEeCiQs63fc893Q" path=""/>
  <TalendProperties:DatabaseConnectionItem          xmi:id="_dSGi82AIEeCiQs63fc893Q"
property="_dSGi8WAIEeCiQs63fc893Q" state="_dSGi8mAlEeCiQs63fc893Q">
    <connection href="MySQL_0.1.item#_dSGi8GAIEeCiQs63fc893Q"/>
  </TalendProperties:DatabaseConnectionItem>
</xmi:XMI>
```

## Lisa 3 Küsimustik

Olen koostanud järgmise küsimustiku, kasutades:“ LimeSurvey – avatud lähtekoodiga uuringu aplikatsioon (lihtne ja vaba) “, mis annaks parema ülevaate integratsioonivahendite vajadusest, nõudlusest ja olulisusest.

Pöördun Teie poole palvega osaleda uuringus, mis on kokku pandud Tallinna Ülikooli magistrandi Alex Aisting magistritöö „Vabavara Talend- programmeerimisvahendi analüüs " raames. Töö eesmärk on tutvustada ja välja selgitada ettevõtete vajadus, kogemused antud tarkvara osas.

Küsimustikule laekunud vastused:

### 1. Ettevõtte põhivaldkond (põhitegevus)

Vastaja 1	Programmeerimistööd
Vastaja 2	Tarkvara tootmine
Vastaja 3	IT
Vastaja 4	Oleme spetsialiseerunud .NET tehnoloogial põhinevate IT-teenuste ja lahenduste arendamisele.
Vastaja 5	testing AS
Vastaja 6	IT

### 2. Töötajate arv ettevõttes

Vastaja 1	1-10
Vastaja 2	11-50
Vastaja 3	51-100
Vastaja 4	51-100
Vastaja 5	51-100
Vastaja 6	51-100

### 3. Kas vajate IT integratsioonitarkvara?

Vastaja 1	Pole kindel
Vastaja 2	Ei
Vastaja 3	Ja
Vastaja 4	Ei
Vastaja 5	Pole kindel
Vastaja 6	Pole kindel

### 4. Kas kasutate mõnda IT integratsiooni tarkvara?

Vastaja 1	Ei
Vastaja 2	Ei
Vastaja 3	Ja
Vastaja 4	Ja (SaaS)
Vastaja 5	Ei
Vastaja 6	EI

### 5. Kas ettevõttes on kasutatud, kokkupuudet olnud Talend tarkvaraga [www.talend.com](http://www.talend.com)?

Vastaja 1	Ei
Vastaja 2	Ei
Vastaja 3	Ei
Vastaja 4	Ei
Vastaja 5	Ei
Vastaja 6	EI (pole olnud kokkupuudet)

6. Kui Ja, siis kuidas olete rahule jäänud?

Vastaja 1	Ei
Vastaja 2	Ei
Vastaja 3	Ei
Vastaja 4	Ei
Vastaja 5	Ei
Vastaja 6	Ei (pole olnud kokkupuudet)

7. Kas soovite rohkem teada sellest tarkvarast?

Vastaja 1	Ei
Vastaja 2	Ei
Vastaja 3	Ei
Vastaja 4	Ei
Vastaja 5	Ja
Vastaja 6	Ja



## Lisa 4 ETL pakkujate võrdlus

Detailne võrdlusmaatriks

Feature	Adeptia ETL Suite	Informatica Powercenter	IBM DataStage	Microsoft SSIS	Pervasive Data Integrator	Talend Integration Suite	Pentaho Kettle Enterprise	CloverETL
Easy to Use	Yes	No	No	Yes	OK	Yes	Yes	Yes
Web-based UI	Yes	No	No	No	No	No	No	No
Multi-role Team Collaboration	Yes	Yes	Yes	No	No	Yes	No	No
Process Centric Approach	Yes	No	No	No	No	No	No	No
Enables SOA	Yes	No	No	No	No	No	No	No
Reusable Services Repository	Yes	No	No	No	No	No	No	No
Single-install, Light Footprint	Yes	No	No	No	Yes	Yes	No	Yes
Large Volume Performance	Yes	Yes	Yes	OK	OK	OK	OK	Yes
Clustering and Job Distribution	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Data Partitioning	Yes	Yes	Yes	No	Yes	OK	No	Yes
Automatic Recovery of Flows	Yes	Yes	Yes	No	No	No	No	No

XA-Transaction Rollbacks	Yes	Yes	Yes	No	No	No	No	No
Meta-Driven Approach vs.Code	Yes	Yes	No	No	No	No	Yes	Yes
Built-in Scheduler	Yes	Yes	Yes	No	No	Yes	No	No
Real-time Triggers	Yes	Yes	Yes	No	No	No	No	No
Non-RDBMS Connections	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Web Services Client	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Publish Flows as Web Services	Yes	Yes	No	No	No	No	No	No
Join Multiple Sources	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Split Data Streams	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Graphical Data Mapper	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Complex Transformations	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Data Validations	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Preview Source Data In-Design	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes

Run Mapping Rules In-Design	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Library of New Mapping Functions	Yes	Yes	Yes	No	No	No	No	No
Complex Lookups	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Automatic Documentation	Yes	Yes	Yes	No	No	No	No	No
Human Workflow for Error Handling	Yes	No	No	No	No	No	No	No
Plug-in External Programs	Yes	Yes	Yes	No	Yes	No	No	No
Alerts and Notifications	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Version Control	No	Yes	Yes	No	No	Yes	No	No
Deploy in Cloud Option	Yes	No	No	No	Yes	No	Yes	No
Commerical or Open Source	C	C	C	C	C	OS	OS	OS
Price	\$	\$\$\$\$	\$\$\$	\$\$	\$\$	\$\$	\$	\$

**Märkus:**

- OK tähendab limiteeritud funktsionaalsust või et funktsionaalsus on saadaval läbi kolmanda poole
- Selle võrdleva maatriksi sisu on koostatud avalikult kättesaadava informatsiooni ülevaatamisel
- Litsentsite hinnakiri kajastab esimese kasutusaasta kulusid

\$\$\$\$	\$150,000+
\$\$\$	\$60,000 to \$150,000
\$\$	\$20,000 to \$60,000
\$	Less than \$20,000