

KARELIA-AMMATTIKORKEAKOULU
Tietojenkäsittelyn koulutusohjelma

Timo Jalkanen

TIETOVARASTOINTI- JA BUSINESS INTELLIGENCE
-YMPÄRISTÖJEN VERTAILU

Opinnäytetyö
Helmikuu 2016



OPINNÄYTETYÖ
Helmikuu 2016
Tietojenkäsittelyn koulutusohjelma

Karjalankatu 3
80200 JOENSUU
013 260 600

Tekijä(t)
Timo Jalkanen

Nimeke
Tietovarastointi- ja Business Intelligence -ympäristöjen vertailu

Toimeksiantaja
Karelia-ammattikorkeakoulu

Tiivistelmä

Opinnäytetyön tavoitteena oli vertailla erilaisia tietovarastointi- ja business intelligence -ympäristöjä keskenään ja selvittää, millaisille organisaatioille eri valmistajien ympäristöt sopivat. Opinnäytetyössä käydään läpi tietovarastointi- ja business intelligence -aiheeseen liittyviä keskeisiä käsitteitä ja tekniikoita. Lisäksi tarkastellaan eri ympäristöistä tiedonsiirto- ja raportointi -ominaisuuksia ja ympäristöjen visuaalista ilmettä.

Työssä vertailtiin viiden eri valmistajan, Microsoftin, Jaspersoftin, Talendin, Pentahon ja Oraclen, ympäristöjä. Ympäristöistä kaksi on kaupallisia ja kolme ilmaisohjelmistoja. Ympäristöjen ominaisuuden selvitettiin kokeilemalla niiden toimintaa käytännössä.

Tuloksena saatiin selville, minkälaisille organisaatioille eri ympäristöt sopivat. Microsoftin ympäristö sopii parhaiten koulutukseen, pienille ja keskisuurille organisaatioille. Jaspersoftin ja Talendin ympäristöt soveltuvat pääasiassa pienille yrityksille. Pentahon ympäristöt sopivat parhaiten koulutukseen, ja pienille yrityksille. Oraclen ympäristöt sopivat pääasiassa suurille organisaatioille niiden suurien hankintakustannuksien vuoksi.

Kieli
suomi

Sivuja 57

Asiasanat
tietovarastointi, business intelligence, ETL-prosessi, dataintegraatio



THESIS
February 2016
Degree Programme in Business
Information Technology

Karjalankatu 3
80200 JOENSUU
FINLAND
013 260 600

Author (s)
Timo Jalkanen

Title
Comparison of Data Warehousing and Business Intelligence Environments

Commissioned by
Karelia University of Applied Sciences

Abstract

The purpose of the thesis was to compare different data warehousing and business intelligence environments, and find out which of the environments are suitable for certain kinds of organizations. The thesis investigates the key concepts and techniques related to data warehousing and business intelligence. In addition, data transmission and reporting features were examined as well as the visual image of the environments as a whole.

Five different manufacturers environments were compared; Microsofts, Jaspersofts, Talends, Pentahos and Oracles. Two of the environments are commercial and three of them are open source, free to download software. The features included in the environments were examined by testing them in practice.

As a result it was found out which of the environments are suitable for certain kinds of organizations. Microsoft's environment is best suited for educational purposes, and for small and medium sized organizations. Jaspersoft's and Talend's environments are primarily suitable for small organizations. Pentaho's environments are best suited for educational purposes, and for small organizations. Oracle's environments are primarily aimed for large organizations due to their high acquisition costs.

Language
Finnish

Pages 57

Keywords

Data Warehousing, Business Intelligence, ETL-process, Data Integration

Sisältö

1	Johdanto	6
2	Tietovarastoinnin ja business intelligenen perusteet	7
2.1	Tietovarasto	7
2.2	Paikallisvarasto	9
2.3	ETL-prosessi.....	11
2.4	Dataintegraatio	15
2.5	Business Intelligence	18
3	DW/BI-järjestelmät ja ominaisuudet	20
3.1	Microsoft SQL Server	20
3.2	Talend Open Studio for Data Integration	23
3.3	Jaspersoft ETL Community Edition.....	24
3.4	JasperReports Server Community Edition	26
3.5	Pentaho Kettle	28
3.6	Pentaho Report Designer	29
3.7	Oracle Data Integrator	31
3.8	Oracle Business Intelligence.....	33
4	Järjestelmien vertailu	35
4.1	Ominaisuuksien vertailu.....	35
4.2	Käyttöliittymien vertailu	40
5	Vertailun tulokset ja johtopäätökset	49
5.1	Microsoft SQL Server	49
5.2	Jaspersoft ETL & JasperReports	50
5.3	Talend Open Studio for Data Integration	51
5.4	Pentaho Kettle & Report Designer.....	51
5.5	Oracle Data Integrator & Business Intelligence	52
6	Pohdinta.....	52
	Lähteet.....	55

Lyhenteet

BI	Business Intelligence, suomeksi liiketoimintatiedon hallinta. Organisaation tekemää järjestelmällistä liiketoimintaan liittyvien tietojen hankintaa, tallennusta ja analysointia. (Wikipedia 2015a.)
CRM	Customer Relationship Management, suomeksi asiakkuudenhallinta. CRM sisältää asiakaslähtöisen ajattelutavan organisaatiossa sekä siihen liittyvät tietojärjestelmät. (Wikipedia 2015b.)
DW	Data Warehouse, suomeksi tietovarasto. Organisaation keskeinen paikka tiedon säilyttämistä varten. (Wikipedia 2015c.)
ERP	Enterprise Resource Planning, suomeksi toiminnanohjausjärjestelmä. Organisaation tietojärjestelmä, joka integroi eri osastojen toimintoja, kuten tuotantoon, toimitukseen ja varastonhallintaan liittyviä toimintoja, sekä laskutukseen ja kirjanpitoon liittyviä toimintoja. (Wikipedia 2015d.)
ETL	Extract-Transform-Load. Prosessi, jonka aikana tietoa kerätään, muokataan ja siirretään. (Wikipedia 2015e.)
SQL	Structured Query Language. Standardoitu kyselykieli, jolla relaatiotietokantaan voi tehdä erilaisia hakuja, muutoksia ja lisäyksiä. (Wikipedia 2015f.)

1 Johdanto

Tämän opinnäytetyön tarkoitus on ensisijaisesti vertailla eri organisaatioiden tarjoamia tietovarastointi- (DW) ja business intelligence (BI) -ympäristöjä, jotta kyseisiä ohjelmistoja hankkivat tahot voisivat punnita vaihtoehtojaan ja saada näkemystä hinta-laatusuhteesta eri ympäristöjen välillä. Tämän vertailun lisäksi kerron yleisellä tasolla näiden ohjelmistojen ominaisuuksista ja kyseisten ympäristöjen yhteydessä käytettävistä termeistä ja tekniikoista.

Työssä tutustuttiin moniin eri tietovarastointi- ja business intelligence -ympäristöihin sekä teoreettisella tasolla että käytännössä kokeilemalla. Teoreettinen osuus pohjautuu pääasiassa verkosta saatavaan materiaaliin. Käytännön kokeilu toteutetaan lataamalla ilmaisia avoimen lähdekoodin (open source) ohjelmistoja sekä lisäksi hyödyntämällä Karelia-ammattikorkeakoulun oppimisympäristöjä vastaaviin kaupallisiin ohjelmistoihin. Toimeksiantajalle opinnäytetyö tarjoaa tietoa päätöksenteon tueksi erilaisia tietovarastointi- ja business intelligence -ratkaisuja pohdittaessa ja oppimisympäristöjä kehitettäessä.

Johdanto-luvun jälkeen toisessa luvussa käsitellään tietovarastointi- ja business intelligence -ympäristöihin kuuluvia tekniikoita sekä perustietoa niiden toiminnasta. Kolmannessa luvussa esitellään eri yritysten tietovarastointi ja business intelligence -järjestelmiä ja niiden ominaisuuksia. Neljännessä luvussa tehdään varsinainen vertailu näiden järjestelmien ja niiden ominaisuuksien välillä. Viidennessä luvussa esitellään vertailun tuloksia ja näiden tulosten perusteella tehtyjä johtopäätöksiä. Lopuksi pohditaan vielä opinnäytetyön prosessia ja tulevaisuuden näkymiä.

Opinnäytetyön tarkoituksena on vastata seuraaviin kysymyksiin:

- Mitä tietovarastointi- ja business intelligence -ympäristöt ovat?
- Millaisia tekniikoita näihin ympäristöihin sisältyy?
- Millaisia ominaisuuksia näistä ympäristöistä löytyy?
- Millaisia eroja eri yritysten DW/BI -ohjelmistoissa on?

2 Tietovarastoinnin ja business intelligencen perusteet

2.1 Tietovarasto

Tietovarasto (eng. Data Warehouse) on yksinkertaisimmillaan kokoelma dataa useista eri lähteistä, jotka on suunniteltu tukemaan sekä strategista että taktista päätöksentekoa. Tietovaraston päätarkoitus on luoda selkeä ja johdonmukainen kuva liiketoiminnan tilanteesta tietyinä ajankohtana. Käyttäjät voivat suorittaa verkkokyselyitä ja samalla tehdä tiedonlouhintaa (eng. Data mining) hyödyntämällä erilaisia tietovarastointityökaluja. (Data Warehouses 2015a.)

Yrityksillä ja eri organisaatioilla on käytössään paljon erilaista tietoa hajautettuna eri tietokantoihin. Niillä on myös tarve saada nämä eri lähteiden tiedot kootua yhteen paikkaan, josta ne ovat saatavilla nopeasti ja tehokkaasti yrityksen eri osastojen ja henkilöiden käyttöön. Yrityksen johtoportaalille ja sen päätöksentekoa varten tämä tieto on hyvin tärkeää. Kaikki tarpeellinen tieto yrityksen tai muun organisaation sisällä pitäisi antaa johdon käyttöön, jotta se pystyisi tekemään mahdollisimman tarkkoja ja liiketoimintaa hyödyttäviä päätöksiä. Kaikki tämä tieto täytyy olla mahdollisimman helposti luettavissa ja analysoitavissa, jotta päätöksenteko olisi tehokasta ja tilanteeseen nähden optimaalista. Faktoihin perustuvat ja optimaaliset päätökset ovat kriittisiä organisaation toiminnan kannalta, kun toimitaan kilpailevassa kvartaalitaloudessa. Yrityksen johtoporras käyttää kerättyä tietoa monessa eri asiassa: esimerkiksi toiminnan ennustamisessa ja suunnittelussa, budjettien analysoimisessa ja resurssien ohjaamisessa. (Honkavaara 2013, 4.)

Tarkemmin sanottuna tietovarasto on paikka, minne kerätään yrityksen kaikkien eri tietojärjestelmien tiedot yhteen ”pakettiin”. Kun kaikki tieto sijaitsee yhdessä paikassa, eli tietovarastossa, yrityksen johtajat pääsevät siihen käsiksi helposti. Tietovarastossa olevaan dataan voidaan käyttää analysointityökaluja ja muita toimintoja, mikä auttaa johtajia päätöksien tekemisessä. Mitä paremmin liiketoiminnan tiedot on järjestetty ja mitä virheettömämpiä ja ajantasaisempia ne

ovat, sitä tehokkaammin ja paremmin nämä tiedot ovat myös analysoitavissa. Lisäksi tietovarastosta löytyvää dataa voidaan hyödyntää raportoinnissa sekä tietojen yhdistelyssä. Tietoa voidaan viedä eheyttynä tietovarastosta organisaation työntekijöiden tietojärjestelmiin. Eheyttelyllä tiedolla tarkoitetaan esimerkiksi osoitetietojen päivittämistä yritysrekisteriin sekä lakkautettujen yritysten poistamista tietokannoista. Toisin sanoen eheytetty tieto on ajantasaista ja liiketoiminnan kannalta tärkeää. (Honkavaara 2013, 4.)

Useimmissa tapauksissa tietovarasto sijaitsee SQL-kielellä toteutetussa tietokannassa. Tietokanta koostuu tauluista (eng. tables), ja nämä taas sisältävät kenttiä (eng. fields). Näiden taulujen koko on riippuvainen niiden sisältämien kenttien määrästä. Yksittäinen taulu voisi esimerkiksi sisältää kaikki perustiedot yrityksestä. Tietovarastossa olevat tiedot voisivat koostua yritysmaailmassa tunnetuista objekteista kuten asiakas, osoite, lasku, tuote, määrä yms. (Honkavaara 2013, 4–5.)

Voidaan sanoa, että tärkein tietovarastosta löytyvä tieto on aika, varsinkin kun käsitellään tiedon analysointia. Analysointivaiheessa aikaleima määrittelee, mitä on asiakas on tehnyt mihinkin aikaan, ja kuinka asiakkuus on syventynyt tai etääntynyt ajan saatossa. Kun vertaillaan tapahtumia keskenään aikatietojen avulla, hyvin todennäköisesti saadaan tietoa siitä, miksi asiakkuus on muuttunut. Hyvin suunniteltu tietovarasto pitääkin kirjaa kaikista muuttuneista tiedoista, ja kaikilla muutoksilla on omat avaimensa ja rivinsä tietovarastossa. (Honkavaara 2013, 6.)

Ajan kuluessa tietovarastoon kertyy enemmän ja enemmän tietoa. Tietovaraston koko kasvaakin räjähdysmäisesti, etenkin suurissa yrityksissä, joissa on monia henkilöitä syöttämässä ja muokkaamassa attribuutteja kymmeniä, ellei satoja kertoja päivässä. Tietovarastoon siis viedään uutta tietoa joka päivä, kaikista asiakkuuksista, tuotteista, ongelmista yms. Kaikki tämä syötetty tieto kasvattaa tietokannan kokoa, ja tämän vuoksi tietovarasto vaatii tarkkaan suunniteltua tiedon järjestelyä ja massiivista laskentatehoa. Kaikki tieto pitääkin ottaa talteen ja varastoida tietovarastoon ja järjestellä siellä, jotta tietoa voidaan analysoida ja hakea nopeasti ja tehokkaasti. Periaatteessa tietovarasto toimii hyvin

samalla tavalla kuin aivan tavallinen, fyysinen varasto. Jos tavarat eivät ole järjestyksessä ja kirjattuna oikein, tavaroita ei löydetä helposti varaston uumenista. (Honkavaara 2013, 6.)

Yrityksessä on siis erilaisia tietojärjestelmiä, joiden sisältämiä tietoja kerätään yhteen, yhtenäiseen paikkaan (eli tietovarastoon), josta ne voidaan sitten muuntaa tarvittaessa yhteensopivaksi eri tietojärjestelmien kanssa käytettäväksi. Tällainen tietojärjestelmä on esimerkiksi asiakkuudenhallintajärjestelmä (CRM), jonka kautta myyntitehtävissä toimivat henkilöt voivat hallinnoida asiakassuhteita ja tehdä myyntityötä. Toinen tällainen tietojärjestelmä on toiminnanohjausjärjestelmä (ERP), jossa ovat kaikki asiakkaisiin liittyvät myyntisaatavat sekä tilushistoriat. Lisäksi on muita tietojärjestelmiä, joita eri yrityksen osastot hyödyntävät. Esimerkiksi markkinointiosastolla on tietoa markkinointiaktiviteeteista, joita on kohdistettu tiettyihin yrityksen asiakkaisiin. Vastaavasti tuotannolla on tietoa tuotantokannasta tai siitä, mitä valmiita tuotteita varastossa on juuri kyseisenä ajankohtana. Yrityksen asiakaspalvelulla taas on tiedot asiakkaiden kyselyistä sekä avoimista tai ratkaistuista ongelmista. (Honkavaara 2013, 6–7.) Tämän perusteella voidaan todeta, että organisaation kaikki osastot ja osastojen työntekijät tuottavat jatkuvasti tietoa, jota analysoimalla saadaan aiemmin mainittu selkeä ja johdonmukainen kuva liiketoiminnan tilanteesta juuri kyseisellä hetkellä.

2.2 Paikallisvarasto

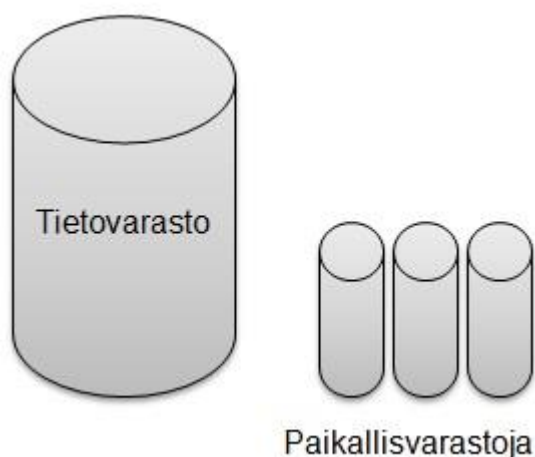
Paikallisvarasto (eng. Data Mart) on tietovaraston alaryhmä, joka on yleensä suunnattu tietyille liiketoiminnan osa-alueelle. Paikallisvarastot ovatkin kuin pieniä siivuja tietovarastosta. Toisin sanoen tietovarastot ovat koko liiketoiminnan laajuisia, eli niissä on kaikkien osastojen tiedot, kun taas paikallisvarastot kuuluvat yksittäiselle yrityksen osastolle, eli esimerkiksi myyntiosaston tiedot ovat paikallisvarastossa. Joissain suuremmissa yrityksissä jokainen osasto tai liiketoiminnan yksikkö on vastuussa omasta paikallisvarastostaan sekä niiden laitteistosta, ohjelmistosta ja tiedoista. Tämä mahdollistaa jokaisen osaston itsenäisen tiedonkulun; osastot voivat käyttää, manipuloida ja kehittää tietojaan

millä tavalla itse haluavat, mutta eivät kuitenkaan pääse muuttamaan tai rikko-
maan muiden osastojen paikallisvarastoja tai itse yrityksen tietovarastoa. (Wiki-
pedia 2015g.)

Organisaatiot rakentavat tietovarastoja ja paikallisvarastoja siksi, että tietokan-
noissa olevia tietoja ei ole järjestelty siten, että niistä löytäisi nopeasti tiedot, joi-
ta liiketoiminta tarvitsee juuri sillä hetkellä. Myös monimutkaiset kyselyt voivat
viedä paljon aikaa, ennen kuin käyttäjät saavat tarvitsemansa tiedot ulos. Tämä
johtuu siitä, että tietokantajärjestelmät on suunniteltu prosessoimaan miljoonia
muutoksia joka päivä. Toisin kuin liiketoiminnalliset tietokannat, jotka on suunni-
teltu jatkuvasti päivitettäväksi, tietovarastot ja paikallisvarastot ovat kirjoitus-
suojattuja, eli niiden tietoja pystyy vain lukemaan. (Wikipedia 2015g.)

Paikallisvarasto on siis periaatteessa tiivistetty ja keskitetympi versio tietovaras-
tosta, mikä heijastuu organisaation jokaisen osaston liiketoimintaprosesseihin.
Jokainen paikallisvarasto on omistettu tietylle yrityksen toiminnolle tai toiminta-
alueelle. On yleistä, että useita paikallisvarastoja käytetään samanaikaisesti,
jotta pystytään vastaamaan jokaisen, erillisen liiketoimintayksikön (esim. kirjan-
pidon, markkinoinnin, myynnin yms.) tarpeisiin.

Tiivistetysti voidaan siis sanoa, että tietovarastossa on useiden eri lähteiden ja
osastojen tietoja. Tietovarastossa on erittäin yksityiskohtaista tietoa, ja se integ-
roi kaikkien lähteiden tiedot. Paikallisvarastossa taas on yleensä vain yhden
osaston, esimerkiksi myyntiosaston, tiedot. Siinä on vähemmän yksityiskohtais-
ta tietoa ja se keskittyy tiedon integrointiin vain yhdestä kohteesta. Graafisesti
tietovaraston ja paikallisvaraston välinen ero kaikessa yksinkertaisuudessaan
voisi olla yksi suuri tietokanta ja monta pientä tietokantaa. (Kuva 1.)



Kuva 1. Tietovarasto vs. paikallisvarasto.

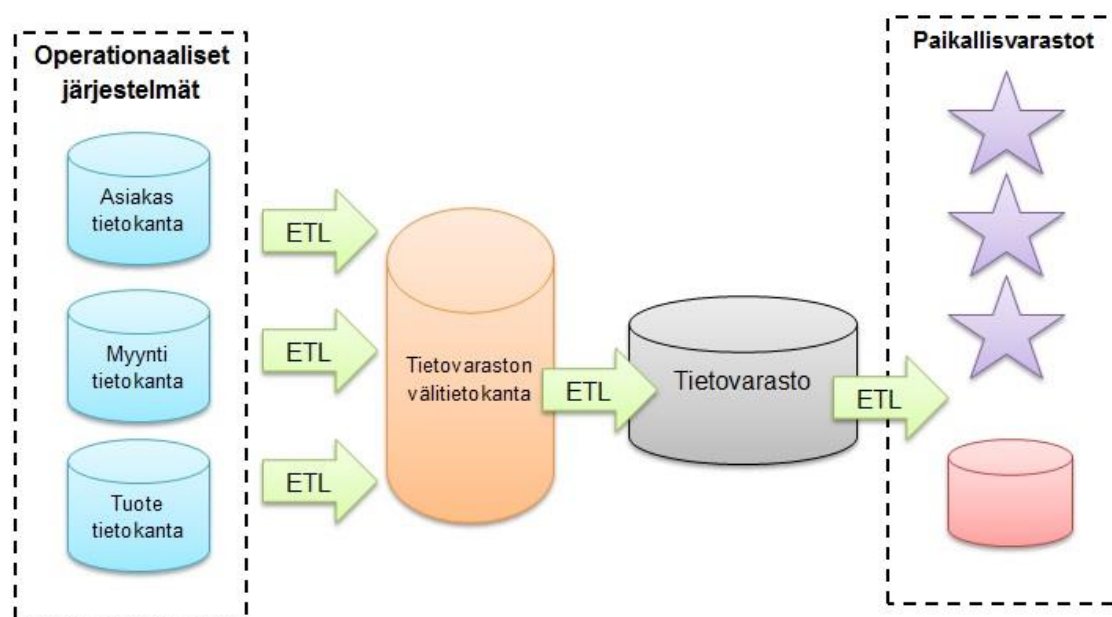
2.3 ETL-prosessi

ETL-prosessi (Extract – Transform – Load) on merkittävä osa tietovarastojen sekä muiden suurien tietokantojen luomisessa. ETL on prosessi, jossa lähdetietokannasta kerätään (eng. Extract) halutut tiedot, ja sitten muokataan (eng. Transform) nämä kerätyt tiedot sopivaan muotoon ja lopuksi viedään (eng. Load) kyseiset tiedot kohdetietokantaan. ETL-prosessi suoritetaan joko yhtenä suurena pakettina kerran yössä, jos tietomäärät ovat erittäin suuria, tai jos tietomäärät ovat pienempiä, niin prosessi voidaan suorittaa pienemmissä osissa useita kertoja päivässä. Joka tapauksessa siirrettävä tietomäärä on niin suuri, että se vaikuttaa työntekijöiden työskentelyyn, koska prosessin suoritus aiheuttaa hidastelua tietojärjestelmissä, joista tietoa luetaan. ETL-prosessin avulla tehdään tietovarasto, ja sen avulla tietovarastoa myös päivitetään jatkuvasti. Kuten muidenkin projektien kanssa, myöskään ETL-prosessin kehittäminen ei lopu kun tietovarastoprojekti saadaan päätökseen, vaan sitä kehitetään paremmaksi jatkuvasti sekä ylläpidetään kuin mitä tahansa muutakin ohjelmistoprojektia. (Honkavaara 2013, 15.)

ETL-prosessista puhutaan yleensä vain tietovarastoinnin yhteydessä. Tämä tekniikka on kuitenkin myös hyödynnettävissä muissakin asioissa, kuten esimerkiksi kahden operatiivisen tietojärjestelmän välisessä tiedonsiirrossa. Kyseisessä tapauksessa ETL-mallin vaiheet kuitenkin suoritetaan eri järjestyksessä, ja puhutaankin jo ELT-mallista. Kyseisessä muokatussa mallissa tiedot kerä-

tään kohteena olevan järjestelmän tietokannan välitauluihin (eng. Staging Table), minkä jälkeen kerätyt tiedot muunnetaan tarvittavaan muotoihin välitaulujen sisällä, ennen kuin tiedot siirretään oikeisiin paikkoihin tietokannassa. Vastavasti normaalissa ETL-prosessissa on erillinen välitietokanta (eng. Staging Area), jossa tiedot muunnetaan tarvittavaan muotoon, ennen kuin ne ladataan kohdejärjestelmän tietokantaan. ETL-prosessi toimii aina omana prosessinaan, joten käsiteltävien tietojärjestelmien käyttäjät eivät voi vaikuttaa tiedonsiirron kulkuun millään tavalla. Toisin sanoen tietojärjestelmiä käyttävät työntekijät eivät voi useinkaan samanaikaisesti käsitellä tiedonsiirron aikana niitä tietoja, joita ETL-prosessi käsittelee. (Kallio 2012, 9–10.)

Graafisesti kuvattuna koko tietovarastointiympäristö voisi näyttää siltä, että toisella puolella ovat kaikki tietokannat, joista halutaan kerätä tietoa. Tämän jälkeen nämä tietokannat siirretään ETL-prosessin avulla välitietokantaan. Välitietokannassa taas halutut tiedot muunnetaan sopivaan muotoon, josta ne sitten vielä siirretään tietovarastoon. Lopuksi tietovarastosta voidaan tehdä omia pienempiä siivujaan, jotka ovat sitten paikallisvarastoja. (Kuva 2.)



Kuva 2. Tyypillinen tietovarastointiympäristö (Mukaien Data Warehouses 2015b).

ETL-prosessissa "Extract" tarkoittaa tiedon poimintaa kyseisen organisaation tietokannoista tai muista lähteistä, joista haluttua tietoa löytyy. Tätä dataa voi-

daan hakea kaikista lähteistä, jotka ovat organisaation tietoverkon sisällä ja joihin on asetettu lukuoikeudet. Myös internetistä saatavia lähteitä on mahdollista käyttää. Erilaisia lähteitä voivat olla esimerkiksi relaatiotietokannat, taulukkolaskentaohjelmat (kuten Excel), henkilörekisterit ja eri internetsivustot. Tämä tiedonpoimintavaihe on tärkeä, sillä se vaikuttaa myöhemmin tuleviin vaiheisiin konkreettisesti. On tärkeää, että tiedot poimitaan ihan alusta asti oikein ja että kaikki haluttu tieto on saatavilla (ei rikkonaisia linkkejä tai tiedostoja). Jos tämä ensimmäinen, kriittinen vaihe epäonnistuu, myös muut, sen jälkeen tulevat vaiheet epäonnistuvat, koska tätä virhetilannetta ei pystytä korjaamaan jälkikäteen. (Honkavaara 2013, 16–17.)

Seuraavaksi on vuorossa ”Transform”-vaihe, jossa suurin osa työstä prosessin sisällä tapahtuu. Tässä vaiheessa käydään läpi kaikki tiedot, jotka aiempi Extract-vaihe poimi halutuista lähteistä. Tietojen läpikälyprosessi aloitetaan siivouksella, eli kaikista tiedoista pitää poistaa esimerkiksi Null-arvot, tehdä tarvittavat tyyppimuunnokset, muuttaa halutut tiedot lähdetietokannan muodosta sellaiseen muotoon, että kohdetietokantakin sen ymmärtää. Lopuksi vielä suoritetaan tiedon täydentäminen. Transform-vaiheessa on myös mahdollista verrata muistissa olevaa tietoa ja tarkistaa, onko kyseistä tietoa jo olemassa ja tarvitseeko sitä täydentää lähteestä. Transform-vaiheessa tehdään myös tarvittavat merkkijonon muutokset tai mahdollisesti suoritetaan aritmeettisia laskuoperaatioita. Kaikkea dataa voidaan sitten järjestellä ja yhdistellä sekä tallentaa kaikki muutokset muistiin. (Honkavaara 2013, 17.)

Prosessin aikana tietovirrasta poistetaan esimerkiksi duplikaattiarvot, eli jos lähdetietokannassa on samaa objektia kaksi (tai useampaa) kappaletta, nämä ylimääräiset arvot poistetaan. Duplikaatti voi olla tietokannassa jopa tismalleen samoilla tiedoilla tai hiukan muokattuina (esim. vanhentunut versio arvosta). Näitä duplikaattiarvoja syntyy lähdetietokantaan siinä tapauksessa, jos tietoja ei tarkisteta ennen uuden arvon luontia manuaalisesti tai automaattisesti tietokannan toimesta. Näitä kaksoisarvoja voi syntyä myös aiemmin tehdyissä tietokantojen yhdistelyissä tai muiden ETL-prosessien kautta. (Honkavaara 2013, 17.)

Transform-vaiheen aikana hallinnoidaan myös niin sanottu ”likainen data”. Se on ”tietokannasta löytyvää väärää, epäjohdonmukaista ja ei haluttua tietoa”. Tällainen tieto tulee lähdejärjestelmistä Extract-vaiheen aikana ETL-prosessiin. Transform-vaiheessa siivotaan dataa, jotta se olisi tarpeeksi hyvälaatuista ja mahdollisimman virheetöntä varastoitavaksi tietovarastoon. (Honkavaara 2013, 17.)

Transform-vaihe on erittäin tehokas ja ETL-työkaluista riippuen prosessin aikana voidaan tehdä halutuille tiedoille paljon erilaisia muokkauksia. Etukäteen suunnittelu on kuitenkin tärkeää, koska mahdolliset virheet todennäköisesti pidentävät tiedon muuttamiseen vaadittavaa aikaa moninkertaiseksi. Jos tässä prosessin vaiheessa tehdään virheitä, ne vaikuttavat seuraavan vaiheen eli siirtämisen läpimenoon sekä tiedon eheyteen, kun sitä luetaan kohdetietokannasta eli tässä tapauksessa tietovarastosta. (Honkavaara 2013, 17.)

Viimeisenä prosessissa on ”Load”-vaihe, joka tarkoittaa aikaisemmassa vaiheessa muunnettujen tietojen siirtämistä kohdetietokantaan. Tämä kohdetietokanta voi olla minkä tahansa tyyppinen tietokanta, mutta usein se on relaatiotietokanta, kuten tietovarastot yleensä ovat. Yksinkertaisissa tapauksissa tämä siirtäminen toteutuu hyvinkin suoraviivaisesti, jolloin prosessiin määritellään, mitkä tiedot sijoitetaan kohteessa haluttuun kenttään. Prosessin alussa tehdään määrittely, jonka mukaan tiedetään onko kyseessä tietojen ylikirjoittaminen vai halutaanko säästää tietoa siitä, mikä arvo kentässä on ennen kuin muutoksia tehdään. Jos halutaan tehdä vain ylikirjoitus, niin se ei vaadi mitään ylimääräisiä toimenpiteitä, vaan data vedetään vanhan päälle. Usein kuitenkin halutaan pitää historiatietoja tallessa. Tämän tekeminen vaatii historiataulun lisäämistä tietokantaan ja tiedoille omat avaimet, joiden avulla tiedetään, mikä arvo kentässä on ollut missäkin ajankohdassa. Tässä tapauksessa jokaiseen muutokseen täytyy laittaa aikaleima, jotta pystytään seuraamaan muutoksia. (Honkavaara 2013, 17–18.)

Tehokkaan ja luotettavan ETL-prosessin toteuttamisessa on lukuisia haasteita, koska tietoyhteiskunta on nykyään jo niin moniulotteinen sekä nopeasti muuttuva tietoyhteiskunta. Yksi näistä haasteista on yksinkertaisesti tietomäärät Big

Datan vuoksi. Tietomäärät kasvavat eksponentiaalisesti koko ajan, mikä pakottaa ETL-prosessit skaalautumaan suurelle määrälle erilaista, monimutkaista dataa. Toinen ongelma on tiedon nopeus. Nykyaikaiset yritykset tarvitsevat reaaliaikaista tietoa jatkuvasti, jotta ne voivat tehdä hyviä ja toimivia päätöksiä. Tämän reaaliaikaisen tiedon tarpeen vuoksi tietovarastoja ja paikallisvarastoja täytyy päivittää jatkuvasti ja latausaikojen pitäisi samalla olla nopeampia. Tämä vaatii reaaliaikaista käsittelyä sekä eräajoja. (Astera 2015.)

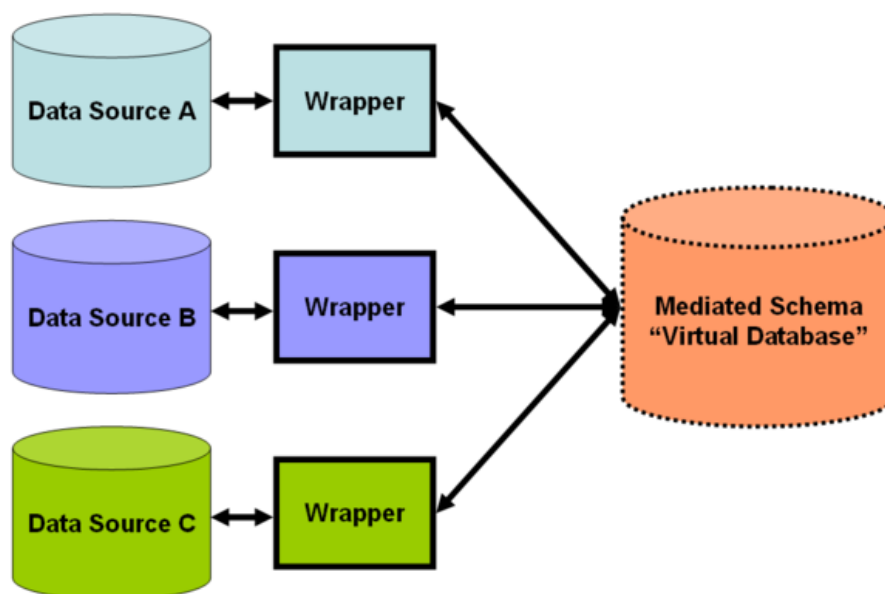
Kolmantena haasteena ovat erilaiset lähteet – tietojärjestelmät monimutkaistuvat, ja niiden on saatava tietoa yhä useammista eri lähteistä. ETL-prosessien täytyy siis pystyä keräämään tietoa hyvinkin erilaisista järjestelmistä, tietokannoista ja tiedostoista. Vastaavasti monipuoliset kohteet ovat myös ongelma. Business intelligence -järjestelmät, tietovarastot ja paikallisvarastot ovat kaikki erilaisia rakenteeltaan, joten tarvitaan todella monipuoliset datan muunnosominaisuudet. ETL-prosesseissa olevat muunnos (Transform) -vaiheet voivat olla hyvinkin monimutkaisia. Dataa täytyy koota yhteen, jäsentää, laskelmoida, prosessoida tilastollisesti jne. Usein dataintegraatio projekteissa käytetään useita eri data lähteitä, ja tämän vuoksi tarvitaan myös useita eri avaimia, jotta voidaan yhdistellä tietoa järkevästi. (Astera 2015.)

2.4 Dataintegraatio

Dataintegraatio on yksinkertaisimmillaan tiedon muuntamista ja kuljettamista tietojärjestelmästä toiseen. Data on tallennettu erilaisiin muotoihin käyttäen erilaisia teknologioita, ja tarvitaan yhtenäinen näkemys kaikesta organisaation datasta. Dataintegraatiosta tulee erittäin tärkeää sellaisissa tilanteissa, kun yhdistetään kahden yrityksen tietojärjestelmien tiedot, tai yhden suuren yrityksen sisällä olevien tietojärjestelmien yhdistämisessä, jotta voidaan tarkastella koko yrityksen tietoja; tätä kutsutaankin tietovarastoksi, josta on kerrottu aikaisemmassa luvussa. (Dataintegration 2015.)

Kuvassa 3 on yksinkertainen luonnos dataintegraatio-ratkaisusta. Järjestelmäsuunnittelija rakentaa sovittimen, jonka avulla käyttäjät voivat suorittaa kyselyi-

tä. Virtuaalinen tietokanta liittyy lähdetietokantoihin wrapperin avulla tarvittaessa.



Kuva 3. Yksinkertainen luonnos dataintegraatio-ratkaisusta (Wikipedia 2015h).

Luultavasti tunnetuin dataintegraation toteutustapa on organisaation tietovaraston rakentaminen, mutta dataintegraatio-käsite sisältää kuitenkin useita muita alueita, kuten tiedonsiirron (eng. Data migration), yritysten sovellusten ja tiedon integraation sekä perustietojen hallinnan (eng. Master Data Management) (Dataintegration 2015).

Dataintegraatiossa on myös omat haasteensa. Ensisilmäyksellä voisi luulla, että suurin haaste on tekninen toteutus, eli kuinka voisi integroida tietoa monesta lähteestä jotka eivät sovi yhteen. Kuitenkin paljon suurempi haaste liittyy itse dataintegraation kokonaisuuteen; siinä täytyy olla erilaisia vaiheita. Suunnitteluvaiheessa täytyy huomioida, että tiedon integrointi hyödyttää koko liiketoimintaa eikä vain IT-osastoa. Yrityksessä pitäisi olla henkilö, joka ymmärtää tiedon merkityksen liiketoiminnan sisällä ja joka osaa johtaa keskustelua pitkän aikavälin dataintegraation suunnittelussa, jotta dataintegraatiosta tulisi mahdollisimman yhdenmukainen, onnistunut ja etenkin liiketoiminnan kannalta hyödyllinen. (Dataintegration 2015.)

Toteutusvaiheessa esitutkimuksen pitäisi olla jo valmis, jotta voidaan valita sopivat työkalut dataintegraatiojärjestelmän toteuttamiseen. Pienet tai aloittavat yritykset jotka ovat vasta aloittamassa tietovarastoinnin parissa joutuvat pohtimaan, mitkä työkalut toimisivat heidän tapauksessaan parhaiten. Suuret yritykset, jotka ovat jo aloittaneet dataintegraation projektejaan, ovat paremmassa asemassa, koska niillä on jo kokemusta asiasta, ja täten ne pystyvät laajentamaan jo olemassa olevaa järjestelmää ja hyödyntämään olemassa olevaa tietotaitoa järjestelmän toteutuksen parantamisessa. On tilanteita, jolloin uuden tai paremman teknologian käyttö tekee järjestelmästä paljon tehokkaamman verrattuna siihen, että pidettäisiin vanhat järjestelmät pystyssä. Tehokkuutta voidaan lisätä esimerkiksi etsimällä sopivampi työkalu, joka mahdollistaa paremman skaalautuvuuden tulevaa liiketoiminnan kasvua varten, tai ratkaisu, joka vähentää toteutuksen tai ylläpidon kuluja. (Dataintegration 2015.)

Kuten kaikessa muissakin tietojärjestelmäprojekteissa, myös dataintegraatiossa täytyy olla testausvaihe. Toteutuksen aikana ja sen jälkeen on tärkeitä tehdä testaus kunnolla, jotta voidaan varmistaa, että yhdistynyt data on oikeaa, virheetöntä ja ajantasaista. Sekä toteutusta tekevien teknisten henkilöiden, että liiketoiminnan loppukäyttäjien täytyy osallistua testaukseen, jotta saadaan varmuus, että tulokset ovat haluttuja. Siksi testaukseen pitäisi sisällyttää ainakin suorituskyvyn stressitesti (eng. Performance Stress test - PST), tekninen hyväksymistesti (eng. Technical Acceptance Testing – TAT) sekä käyttäjien hyväksymistesti. (eng. User Acceptance Testing – UAT) (Dataintegration 2015.)

On olemassa useita erilaisia dataintegraatiotekniikoita. Manuaalinen integraatio on vähiten automatisoitu tekniikka, jossa käyttäjät operoivat kaikkia tarvittavia tietoja erillisten lähdejärjestelmien tai verkkosivujen kautta. Yhtenäistä näkymää tiedoista ei ole. Hieman automatisoidumpi tekniikka on sovelluspohjainen integraatio, joka vaatii tiettyjä sovelluksia joiden avulla tehdään integroinnit. Tätä tekniikkaa voidaan käyttää vain hyvin rajallisessa määrässä sovelluksia. Kun taas lisätään automaatiota, tulee väliohjelmisto (eng. Middleware) dataintegraatio, jossa siirretään integraatiologiikka tietyistä sovelluksista uudelle väliohjelmistokerrokselle. Vaikka integraatiologiikkaa ei toteuteta sovelluksissa enää,

tarvitaan silti näitä sovelluksia osittain osallistumaan dataintegraation toteutukseen. Virtuaalinen integraatio jättää tiedot lähdejärjestelmiin, ja määrittelee näkymän jonka avulla kaikki liiketoiminnan osa-alueet pystyvät näkemään yhteisen näkymän esim. asiakkaasta. Pisimmälle automatisoitu tekniikka on fyysinen dataintegraatio, joka usein tarkoittaa uuden järjestelmän luomista, joka sitten varastoi lähdejärjestelmien tiedot ja pystyy hallinnoimaan tätä tietoa alkuperäisestä järjestelmästä riippumattomasti. Tunnetuin esimerkki tästä on aiemmin mainittu tietovarasto. (Dataintegration 2015.)

2.5 Business Intelligence

Business intelligence eli liiketoimintatiedon hallinta on järjestelmällistä organisaation harjoittamaa liiketoiminnan tietojen keräämistä, tallettamista ja tarkastelua. Business intelligencen tarkoituksena onkin esittää käytännöllistä tietoa, joka auttaa yritysjohtajia, myyntipäälliköitä ja muita käyttäjiä tekemään parempia ja osuvampia liiketoimintapäätöksiä. Business intelligence sisältää erilaisia työkaluja, sovelluksia ja menetelmiä, joiden avulla organisaatiot voivat kerätä tietoa sisäisistä järjestelmistä ja ulkoisista lähteistä, valmistella sitä analyysejä varten, kehittää ja ajaa kyselyitä tästä kyseisestä datasta. Tämän lisäksi voidaan luoda raportteja, mittareita ja tietojen visualisointia näiden tietojen perusteella, jotka voidaan sitten esitellä yrityksen päätöksentekijöille sekä operatiivisille työntekijöille. (Rouse 2015.)

Business intelligence -ohjelmistojen potentiaalisia hyötyjä ovat esimerkiksi päätöksenteon nopeutuminen ja tarkentuminen, sisäisten liiketoimintaprosessien optimointi, liiketoiminnan tehokkuuden lisääminen, uusien tulonlähteiden löytäminen ja kilpailuedun saaminen liiketoiminnan kilpailijoita vastaan. Business intelligence -järjestelmät voivat myös auttaa markkinatrendien tunnistamisessa sekä liiketoiminnan ongelmien havaitsemisessa. Business intelligence -data voi sisältää sekä aiemmin käsiteltyä, vanhaa liiketoiminnan tietoa, että uutta tietoa jota kerätään lähdejärjestelmistä sitä mukaa kun sitä luodaan. Tämän ansiosta business intelligence -analyysit tukevat sekä strategisia että taktisia päätöksentekoprosesseja. Alun perin business intelligence -työkaluja käyttivät vain data-

analyttikot ja muut IT-ammattilaiset, jotka suorittivat analyysseja ja tuottivat raportteja kyselyiden tuloksista päätöksentekijöille. Nykyään myös yritysjohtajat ja heidän työntekijänsä käyttävät business intelligence -sovelluksia, koska business intelligence -työkaluja on kehitetty käyttäjäystävällisempään suuntaan. (Rouse 2015.)

Business intelligence yhdistää laajan joukon data-analyysisovelluksia, kuten ad hoc -analyysit ja -kyselyt, liiketoiminnan raportit, OLAP-tekniikan ja erilaiset business intelligencen -sovellukset, kuten mobiili-, reaali-aikainen tai -operationaalinen BI. Business intelligence teknologia sisältää myös tiedon visualisointiohjelmistoja kaavioiden suunnittelua varten, sekä työkaluja business intelligence -mittaristojen ja -suorituskykyymittaristojen rakentamista varten. Nämä mittarit näyttävät graafisesti tietoa liiketoiminnan tilanteesta ja keskeisistä luvuista helposti ymmärrettävällä tavalla. (Rouse 2015.)

Business intelligence ohjelmistoihin voi sisältyä kehittyneen analytiikan työkaluja, kuten tiedonlouhinnan, ennakoivan analytiikan, text miningin, tilastollisen analyysin ja big data -analytiikan työkaluja. Kuitenkin monissa tapauksissa kehittyneen analytiikkaprojektin toteuttaa tiimi, jossa työskentelee datatutkijoita, tilastotieteilijöitä ja muita analytiikan ammattilaisia, kun taas vastaavat business intelligence -tiimit hoitavat yksinkertaisempia kyselyitä ja liiketoiminnan tiedon analysointia. Business intelligence -data tyypillisesti tallennetaan tietovarastoon tai pienempiin paikallisvarastoihin. Ennen kuin tietoa voidaan käyttää business intelligence -sovelluksissa, sitä täytyy siistiä dataintegraatiotyökalujen avulla, jotta voidaan varmistaa, että käyttäjät analysoivat tarkkaa ja yhdenmukaista tietoa. (Rouse 2015.)

Business intelligence -tiimit koostuvat yleensä business intelligence-päälliköstä, arkkitehteistä, kehittäjistä, liiketoimintatiedon analyttikoista ja tiedonhallinnan ammattilaisista. Näiden lisäksi tarvitaan myös liiketoiminnan puolen edustajia, jotka vaikuttavat prosessin kehittämiseen siten, että siitä olisi mahdollisimman paljon käytännön hyötyä yritykselle. Monet yritykset ovatkin korvaamassa perinteisen vesiputousmallin ketterällä business intelligence -kehityksellä ja tietovarastointi suunnitellaan siten, että se hyödyntää ketterän kehityksen mallia ja

tekniikoita, jotta business intelligence -projektit voitaisiin jakaa pienempiin osiin ja toimittaa loppukäyttäjille uusia toimintoja jatkuvasti. Ketterän kehityksen avulla yritykset voivat hyödyntää business intelligence -toimintoja nopeammin ja jaloistaa tai muokata kehityssuunnitelmaa sitä mukaa, kun liiketoiminnan tarpeet muuttuvat. (Rouse 2015.)

3 DW/BI-järjestelmät ja ominaisuudet

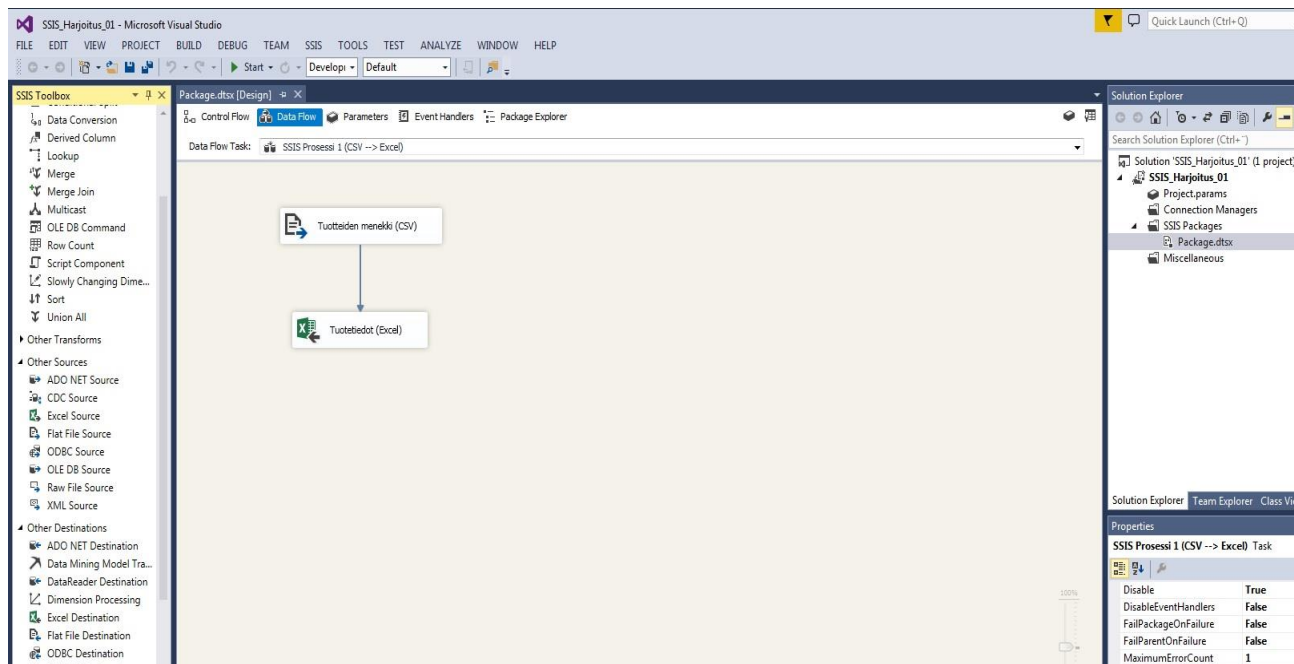
3.1 Microsoft SQL Server

Microsoft SQL Serverin ominaisuudet on jaettu eri osioiden (komponenttien) alle. Microsoftin SQL Server onkin tietokantaohjelmistopaketti, johon kuuluu monia erilaisia ominaisuuksia. Pakettiin kuuluu dataintegraatioon keskittyvä SSIS-osio, analyysipalveluihin keskittyvä SSAS-osio sekä raportointipalveluihin keskittyvä SSRS-osio. Näiden lisäksi SQL Serverissä on myös yleisesti käytetyt ”database engine”-tietokantapalvelut, jonka kautta muut osiot toimivat.

SQL Server Integration Services-osiota (SSIS) voidaan käyttää erilaisten tiedonsiirtoon liittyvien tehtävien suorittamiseen. SSIS sisältää nopeita ja joustavia tietovarastointityökaluja, joita käytetään tiedon keräämiseen, muuntoon ja siirtämiseen (eli aiemmin mainittuun ETL-prosessiin). Työkaluja voidaan myös käyttää automatisoituun SQL Serverin tietokantojen ylläpitoon ja moniuloitteisten datakuutioiden päivitykseen. (Wikipedia 2015i.)

Kuvassa 4 on Microsoft SQL Serverin SSIS-osion perusnäkyminen Visual Studiossa. Tässä kuvassa näkyvät erilaiset toiminnot vasemmalla olevassa työkaluvalikossa. Keskellä kuvassa näkyy esimerkkinä tehty dataintegraatio CSV-

tiedostosta excel-tiedostoon. Komponenttien välillä oleva nuoli ilmoittaa mistä tieto lähtee ja mihin se menee. Tämä myös tarkoittaa, että komponenttien välillä menee ETL-prosessi.

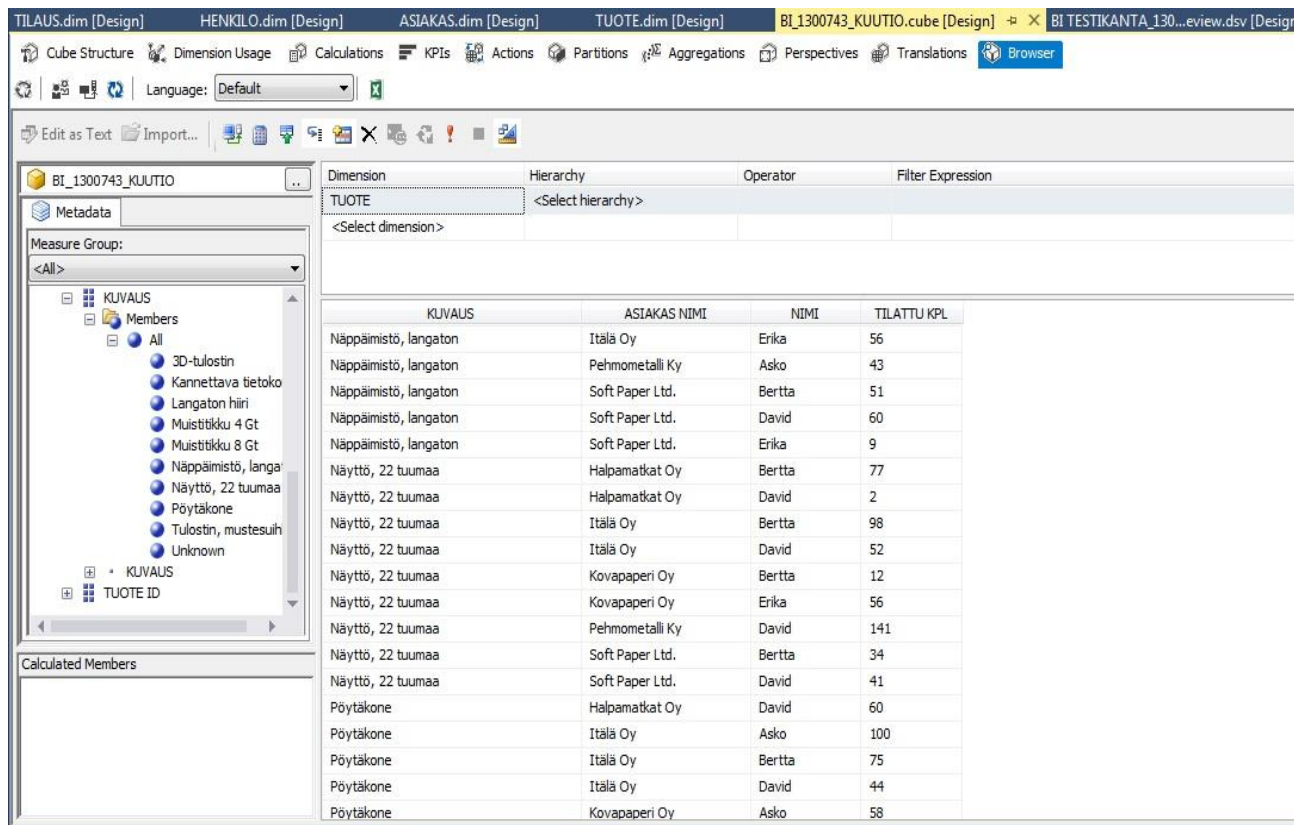


Kuva 4. Microsoft SQL Serverin SSIS-osan perusnäkyä Visual Studiossa (Kuvakaappaus)

Toinen Microsoftin SQL Serveristä löytyvä osio (tai komponentti) on SQL Server Analysis Services (SSAS). Tässä osiossa on online analytical processing (OLAP) ja tiedonlouhintatyökalut (data mining). SSAS-komponenttia käytetään tiedon analysointiin ja muokkaamiseen (esim. moniulotteisen kuution avulla) eri organisaatioissa. Tarkoituksena on siis saada käyttökelpoista tietoa monista eri tietokannoista tai tiedostoista sekä analysoida ja jalostaa sitä eteenpäin. SSAS-osiota käytetään myös tietojen rakenteiden suunnitteluun, luontiin ja hallintaan (Analyysikantojen toteutus SSRS projekteille). (Wikipedia 2015j.)

Kuvassa 5 on Microsoft SQL Serverin SSAS -ympäristöllä tehdyn yksinkertaisen harjoituksen lopputulos Visual Studiossa. Tässä kuvassa näkyvät erilaiset komponentit vasemmalla olevassa metadatatavalikossa. Keskellä kuvassa näkyy esimerkkinä tehty selaus tuotekuvauksesta, asiakkaan nimestä, henkilön nimestä sekä heidän tilaamastaan kappalemäärästä. Tätä näkymää ja tarkasteltavia

tauluja voisi muokata sellaiseksi kuin itse haluaa vetämällä komponentteja (drag & drop) metadatasta ikkunan keskellä olevaan laatikkoon.



Kuva 5. Microsoft SQL Serverin SSAS-osan perusharjoitus Visual Studiossa (Kuvakaappaus)

Pakettiin kuuluu kahden aiemmin mainittujen osioiden lisäksi vielä SQL Server Reporting Services (SSRS) komponentti. Se on palvelinpohjainen raportin luontijärjestelmä. SSRS:n avulla käyttäjät voivat nopeasti ja helposti luoda raportteja Microsoftin SQL Serverin käyttämistä tietokannoista. Tätä voitaisiin kutsua SQL Serverin business intelligence -osaksi. Osiosta löytyy laaja joukko työkaluja monipuolisten raporttien tuottamiseen. SSRS sisältää myös ominaisuuksia, jotka mahdollistavat visuaalisuuden hyödyntämisen. (Wikipedia 2015k.)

Kuvassa 6 on Microsoft SQL Serverin SSRS -ympäristöllä tehty testiraportti selaimessa. Tässä testissä on hyödynnetty ns. ”Drill down” -ominaisuutta, joka mahdollistaa tietojen tutkimisen tarkemmin. Tuotteet on ensin lajiteltu vain tilusmäärien mukaan, mutta voidaan ”porautua” syvemmälle ja lajitella tuotteet asiakkaiden ja kaupunkien mukaan, ja sitten vielä tarkennettuna eri myyjien mukaan.

Testiraportti_TimoJ				
Tuotteen kuvaus	Asiakkaan nimi	Kaupunki	Myyjä	Tilausmäärä
☐ Kannettava tietokone				457
	☐ Itälä Oy			127
		Jyväskylä	Bertta	51
		Jyväskylä	David	76
	☒ Kovapaperi Oy			276
	☒ Pehmometalli Ky			42
	☒ Soft Paper Ltd.			12
☒ Langaton hiiri				718
☒ Muistitikku 4 Gt				592
☒ Muistitikku 8 Gt				495
☒ Näppäimistö, langaton				694
☒ Näyttö, 22 tuumaa				513
☒ Pöytäkone				839
☒ Tulostin, mustesuihku				717

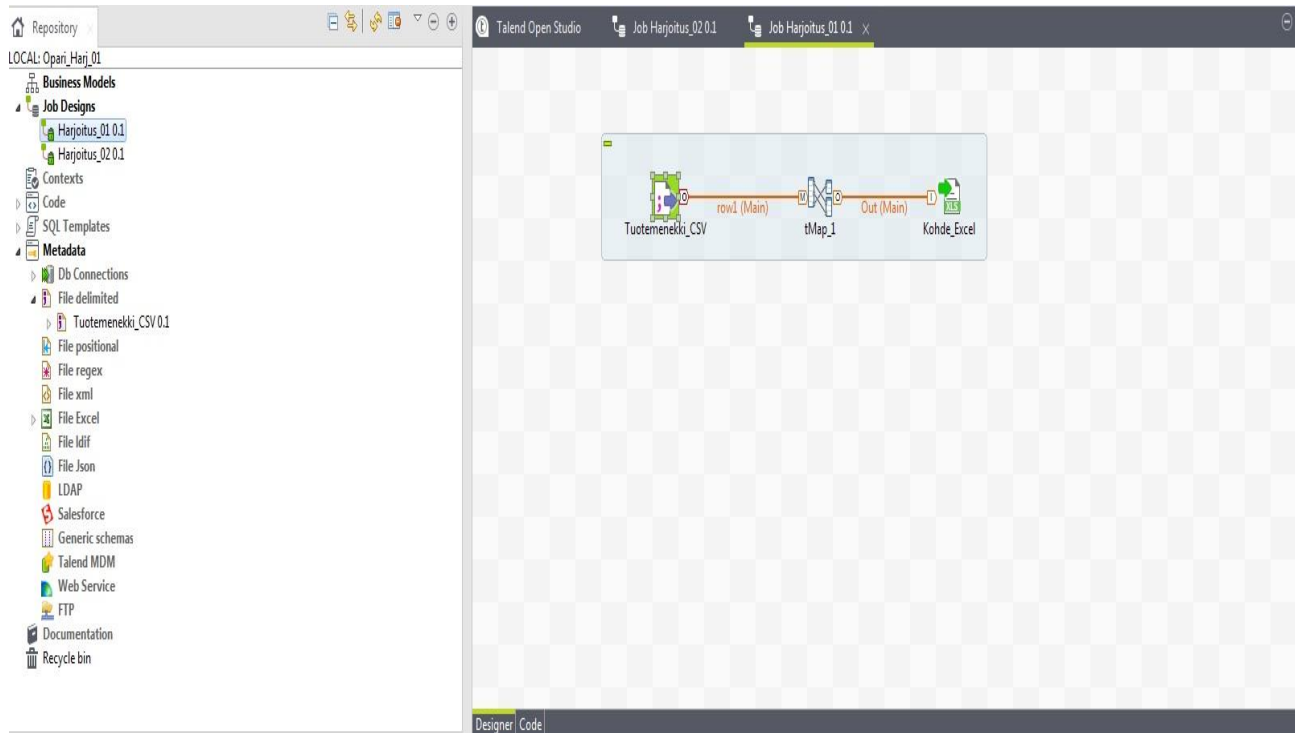
Kuva 6. Microsoft SQL Serverin SSRS-osan testiraportti (Kuvakaappaus)

3.2 Talend Open Studio for Data Integration

Talend Open Studio for Data Integration on Talendin kehittämä ilmainen dataintegroatioympäristö, joka vastaa aiemmin mainitun Microsoft SQL Serverin SSIS-osiota. Siinä on itsessään Visual Studion kaltainen kehitysympäristö, joka mahdollistaa graafisen käyttöliittymän ja erilaisten komponenttien vetämisen. Ominaisuuksiltaan tämä ympäristö on hyvin samankaltainen kuin Microsoft SQL Server, siinä on komponentteja, joita voidaan käyttää erilaisten tiedonsiirtoon liittyvien tehtävien suorittamiseen. Myös Talend Open Studio for Data Integration sisältää nopeita ja joustavia tietovarastointityökaluja, joita käytetään tiedon keräämiseen, muuntoon ja siirtämiseen (eli useaan kertaan mainittuun ETL-prosessiin). Aivan kuten vastaavassa Microsoftin tuotteessa, tällä ympäristöllä voidaan siirtää dataa erilaisten tiedostojen (CSV, Excel, XML) sekä tietokantojen välillä ja integroida näitä erilaisiin paketteihin (tai luoda / siirtää dataa tietovarastoihin). Talend Open Studio for Data Integration kuitenkin keskittyy vain dataintegroatioon, joten siinä ei ole niin monipuolisia toimintoja kuin Microsoft SQL Server –paketissa. (Talend 2016.)

Kuvassa 7 on Talend Open Studio for Data Integration -ympäristön perusnäkyvä. Kuten vastaavassa MS SQL Serverissä, kuvassa näkyvät erilaiset komponentit vasemmalla olevassa työkaluvalikossa. Keskellä kuvassa näkyy esi-

merkkinä tehty dataintegraatio CSV –tiedostosta Excel –tiedostoon. Komponenttien välillä oleva nuoli ilmoittaa, mistä tieto lähtee ja mihin se menee. Toisin kuin Microsoftin ympäristössä, Talendilla täytyy tehdä ”mapping” eli kartoitus komponenttien välille, eikä niitä määritellä suoraan komponentteihin itseensä.



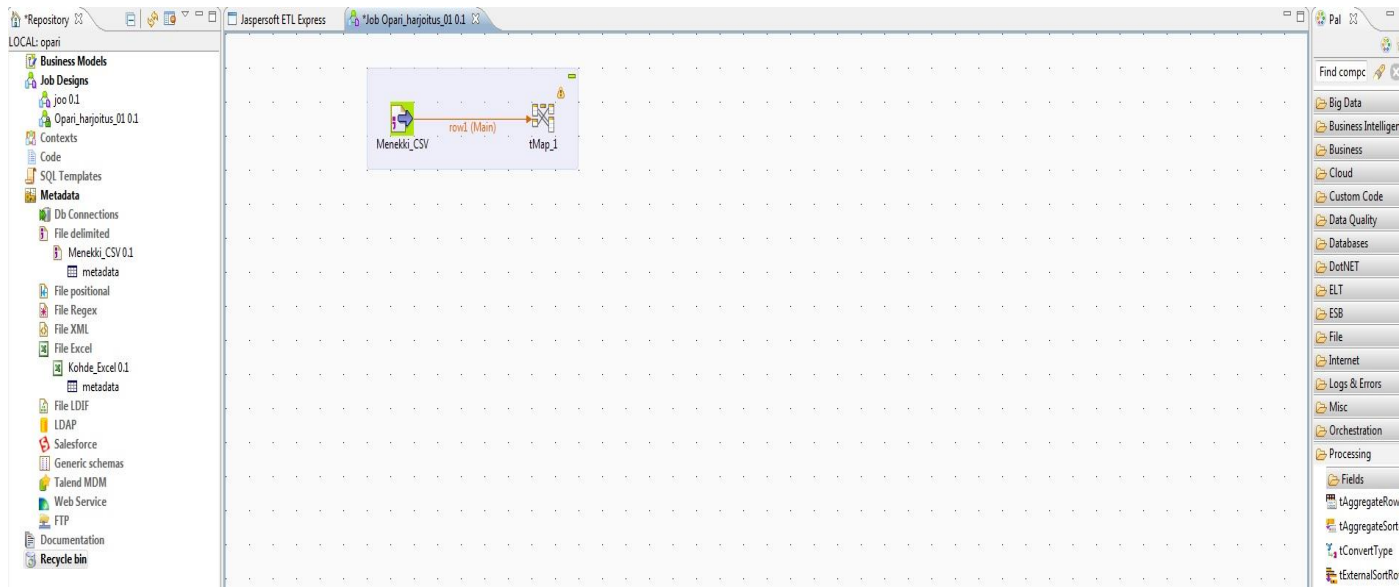
Kuva 7. Talend Open Studio for Data Integration perusnäky (Kuva-kaappaus)

3.3 Jaspersoft ETL Community Edition

Jaspersoft ETL Community Edition on TIBCO Jaspersoftin kehittämä ilmainen dataintegraatioympäristö, joka on lähes täysin identtinen Talendin Open Studio for Data Integrationin kanssa. Tämä ympäristö vastaa Microsoft SQL Serverin SSIS-osiota. Aivan kuten Talendin ympäristössä, siinä on itsessään Visual Studio kaltainen kehitysympäristö, joka mahdollistaa graafisen käyttöliittymän ja erilaisten komponenttien vetämisen. Nämä kaksi organisaatiota tekevät tiivistä yhteistyötä. Jaspersoftin omalla verkkosivustolla lukee näin: ”JasperETL is powered by Talend, the world leader in open source ETL and data integration technology. All of the Talend resources below apply to JasperETL.” (Jaspersoft 2015a.)

Toisin sanoen kaikki ohjeet, jotka soveltuvat Talendin dataintegraatio työkalulle, soveltuvat myös Jaspersoftin vastaavalle tuotteelle. Ominaisuuksiltaan myös tämä ympäristö on hyvin samankaltainen kuin Microsoft SQL Serverin SSIS-osoio, siinä on komponentteja, joita voidaan käyttää erilaisten tiedonsiirtoon liittyvien tehtävien suorittamiseen. Myös Jaspersoft ETL Community Edition sisältää nopeita ja joustavia tietovarastointityökaluja, joita käytetään ETL-prosessiin. Kuten vastaavassa Microsoftin tuotteessa, tällä ympäristöllä voidaan siirtää dataa erilaisten tiedostojen (CSV, Excel, XML) sekä tietokantojen välillä, ja integroida näitä erilaisiin paketteihin (tai luoda / siirtää dataa tietovarastoihin). Aivan kuten Talendin ympäristö, myös Jaspersoft ETL Community Edition keskittyy vain dataintegraatio -osaan, joten siinä ei ole niin monipuolisia toimintoja kuin Microsoft SQL Server -paketissa. (Jaspersoft 2015a.)

Kuvassa 8 on Jaspersoft ETL Community Edition -ympäristön perusnäkyvä. Vertaamalla tätä kuvaa Talendin ympäristöön voidaan havaita, että Jaspersoftin ja Talendin ympäristöt ovat lähes täysin identtisiä. Kuten vastaavassa Talendin ympäristössä, kuvassa näkyvät erilaiset komponentit vasemmalla olevassa työkaluvalikossa. Keskellä kuvassa näkyy esimerkki komponentteja. Komponenttien välillä oleva nuoli ilmoittaa mistä tieto lähtee ja mihin se menee. Aivan kuten Talendin ympäristössä, myös Jaspersoft ETL ympäristössä täytyy tehdä ”mapping” komponenttien välille, eikä niitä määritellä suoraan komponentteihin itseensä.



Kuva 8. Jaspersoft ETL Community Edition perusnäky (Kuvakaappaus)

3.4 JasperReports Server Community Edition

JasperReports Server Community Edition on TIBCO Jaspersoftin kehittämä ilmainen raporttien luontiympäristö, joka vastaa aiemmin mainitun Microsoft SQL Serverin SSRS-osiota. Jotta kyseistä ympäristöä voitaisiin käyttää, täytyy myös ladata TIBCO Jaspersoftin Studio, joka on hieman samanlainen kuin Visual Studio -kehitysympäristö. Tämä ohjelma mahdollistaa graafisen käyttöliittymän ja erilaisten komponenttien vetämisen. Ominaisuuksiltaan tämä ympäristö on melko samankaltainen kuin Microsoft SQL Server, siinä on komponentteja, joita voidaan käyttää erilaisten raporttien luontia varten. Vaikka toiminnallisuus on samanlainen kuin Microsoftin vastaavassa tuotteessa, käyttöliittymä on huomattavasti erilaisempi. Myös JasperReports on palvelinperustainen raportinluontijärjestelmä, jonka avulla käyttäjät voivat luoda raportteja erilaisista tietolähteistä. Tätä voitaisiin siis kutsua TIBCO Jaspersoftin business intelligence -ympäristöksi. Se ei ole yhtä helppo ja nopea kuin Microsoftin SQL Serverin SSRS-osa, mutta hyvin monipuolinen kyllä. JasperReportsissa on laaja joukko työkaluja monipuolisten raporttien tuottamiseen, tiedot voivat olla erilaisista lähteistä, kuten tietokannoista tai erilaisista tiedostoista, esimerkiksi CSV tai Excel tiedostoista. Myös JasperReports sisältää ominaisuuksia, jotka mahdollistavat visuaalisuuden hyödyntämisen, kuten erilaisten mittareiden käytön. (Jaspersoft 2015b).

Kuvassa 9 on TIBCO Jaspersoft Studio & JasperReports –ympäristöllä tehty testiraportti, joka on mukailtu heidän esimerkkiharjoituksestaan. Tässä testissä on käytetty testi-tietokantaa, josta tiedot on haettu raporttiin. Kaikki komponentit (otsikot, tiedot ym.) täytyy itse manuaalisesti vetää paikoilleen, eikä minkäänlaista automatisointia komponenttien asetteluun ole, toisin kuin Microsoft SQL Serverin SSRS-osiossa. Drill-down -ominaisuuden tukeminen on heikompaa tai pikemminkin se on vaikeampi toteuttaa. Parametrien ja hyperlinkkien avulla se on kuitenkin mahdollista, mutta kyseisessä harjoituksessa ei asiassa menty niin syväälle.



Coffee Title
Coffee SubTitle

>Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce auctor purus gravida arcu aliquam mattis. Donec et nulla libero, ut varius massa. Nulla sed turpis elit. Etiam aliquet mauris a ligula hendrerit in auctor leo lobortis.

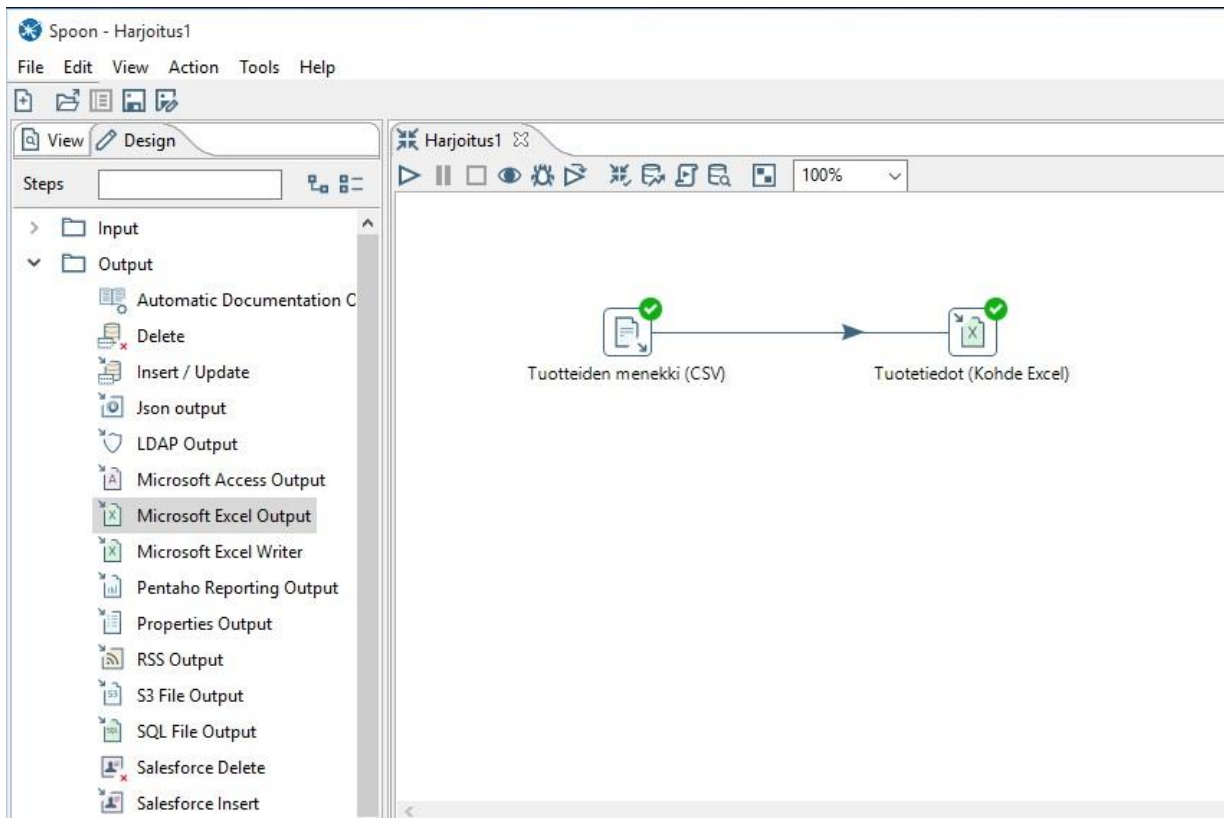
ORDERID	SHIPNAME	SHIPADDRESS	SHIPCITY	
10248	Vins et alcools Chevalier	59 rue de l'Abbaye	Reims	France
10249	Toms Spezialitäten	Luisenstr. 48	Münster	Germany
10250	Hanari Carnes	Rua do Paço, 67	Rio de Janeiro	Brazil
10251	Victualles en stock	2, rue du Commerce	Lyon	France
10252	Suprêmes délices	Boulevard Tirou, 255	Charleroi	Belgium
10253	Hanari Carnes	Rua do Paço, 67	Rio de Janeiro	Brazil
10254	Chop-suey Chinese	Hauptstr. 31	Bern	Switzerland
10255	Richter Supermarkt	Starenweg 5	Genève	Switzerland
10256	Wellington Importadora	Rua do Mercado, 12	Resende	Brazil
10257	HILARION-Abastos	Carrera 22 con Ave. San Cristóbal Carlos Soublette #8-35		Venezuela
10258	Ernst Handel	Kirchgasse 6	Graz	Austria
10259	Centro comercial Moctezuma	Sierras de Granada 9993	México D.F.	Mexico
10260	Ottiles Käseladen	Mehrheimerstr. 369	Köln	Germany
10261	Que Delicia	Rua da Panificadora, 12	Rio de Janeiro	Brazil
10262	Rattlesnake Canyon Grocery	2817 Milton Dr.	Albuquerque	USA
10263	Ernst Handel	Kirchgasse 6	Graz	Austria
10264	Folk och få HB	Åkergatan 24	Bräcke	Sweden
10265	Blondel père et fils	24, place Kléber	Strasbourg	France
10266	Wartian Herkku	Torikatu 38	Oulu	Finland
10267	Frankenversand	Berliner Platz 43	München	Germany
10268	GROSELLA-Restaurante	5ª Ave. Los Palos Grandes	Caracas	Venezuela
10269	White Clover	1029 - 12th Ave. S.	Seattle	USA

Kuva 9. JasperReports Server Community Editionilla tehty testiraportti (Kuva-kaappaus)

3.5 Pentaho Kettle

Pentaho Data Integration (myös Kettle -nimellä tunnettu) on Pentaho/Hitachin kehittämä ilmainen dataintegraatioympäristö, joka vastaa aiemmin mainitun Microsoft SQL Serverin SSIS-osiota. Tätä ympäristöä käytetään mukana tulevalla ”Spoon” -kehitysympäristöllä, joka on Visual Studion kaltainen kehitysympäristö. Myös Spoon mahdollistaa graafisen käyttöliittymän ja erilaisten komponenttien vetämisen. Kuten muutkin vertailussa olevat järjestelmät, Kettle on ominaisuuksiltaan hyvin samankaltainen kuin Microsoft SQL Server, joten siinä on komponentteja, joita voidaan käyttää erilaisten tiedonsiirtoon liittyvien tehtävien suorittamiseen. Kuten muissakin vastaavissa tuotteissa, myös Pentahon Kettle sisältää nopeita ja joustavia tietovarastointi työkaluja, joita käytetään tiedon keräämiseen, muuntoon ja siirtämiseen (eli ETL-prosessiin). Aivan kuten vastaavissa Dataintegraatio -ympäristöissä, myös Kettlen avulla voidaan siirtää dataa erilaisten tiedostojen (CSV, Excel, XML) sekä tietokantojen välillä, ja integroida näitä erilaisiin paketteihin (tai luoda / siirtää dataa tietovarastoihin). Kuten Jaspersoftin vastaava tuote, Pentahon Kettle keskittyy vain SSIS-osaan, eikä sitä ole integroitu isompaan pakettiin kuten Microsoftin SQL Serveriä. (Jaspersoft 2015a.) On myös mainitsemisen arvoinen asia, että Kettle vaatii JDK 7 (Java Development Kit) -kehitysympäristön, joka on jo hieman vanhentunut. (Jaspersoft 2015b).

Kuvassa 10 on Pentaho Kettle -ympäristön perusnäkyvä. Kuten vastaavassa Microsoftin SQL Serverissä, kuvassa näkyvät erilaiset komponentit vasemmalla olevassa työkaluvalikossa. Keskellä kuvassa näkyy esimerkkinä tehty data integraatio CSV-tiedostosta excel-tiedostoon. Komponenttien välillä oleva nuoli ilmoittaa mistä tieto lähtee ja mihin se menee. Käyttöliittymä on suoraviivaisempi kuin aiemmin tarkastellussa Jaspersoftin ratkaisussa.



Kuva 10. Pentaho Kettle -ympäristön perusnäky (Kuvakaappaus)

3.6 Pentaho Report Designer

Pentaho Report Designer on Pentaho/Hitachin kehittämä ilmainen raporttien luontiympäristö, joka vastaa aiemmin mainitun Microsoft SQL Serverin SSRS-osiota. Toisin kuin Jaspersoftin vastaavassa tuotteessa, Pentaho Report Designer ei vaadi mitään ylimääräisiä latauksia, vaan siinä tulee oma graafinen käyttöliittymä mukana. Ominaisuuksiltaan tämä ympäristö on melko samankaltainen kuin Microsoftin ja Jaspersoftin ratkaisut, siinä on komponentteja, joita voidaan käyttää erilaisten raporttien luontia varten. Toiminnallisuus on samanlainen kuin vastaavissa tuotteissa, ja käyttöliittymä on melko samanlainen kuin Jaspersoftin ympäristössä. Myös Pentaho Report Designer on palvelin pohjainen raportinluontijärjestelmä, jonka avulla käyttäjät voivat luoda raportteja erilaisista tietolähteistä. Tämä on siis Pentahon business intelligence -ympäristö. Se ei ole yhtä helppo ja nopea kuin Microsoftin SQL Serverin SSRS-osa, mutta vähemmän raskas kuin Jaspersoftin vastaava tuote. Kuten vastaavista ympäristöistä, myös Pentaho Report Designeristä on laaja joukko työkaluja monipuolisten ra-

porttien tuottamiseen, tiedot voivat olla erilaisista lähteistä, kuten tietokannoista tai erilaisista tiedostoista, esimerkiksi CSV tai Excel tiedostoista. Pentaho Report Designer pystyy tekemään raportteja monissa eri formaateissa yksinkertaisesti valitsemalla formaatin valikosta. Vaihtoehtoina ovat esimerkiksi HTML, Excel, CSV ja PDF -formaatit. Pentaho Report Designerissa visuaalisuuden hyödyntäminen on hieman heikompaa kuin vastaavissa tuotteissa.

Kuvassa 11 on Pentaho Report Designer -ympäristöllä tehty testiraportti, joka on mukailtu heidän tutoriaalistaan. Tässä testissä on käytetty testitietokantaa, josta tiedot on haettu raporttiin. Komponentit (otsikot, tiedot ym.) voidaan itse manuaalisesti vetää paikoilleen, mutta Pentahossa on myös Wizard, jonka avulla voidaan peruspalikat laittaa valmiiksi paikoilleen. Voidaankin sanoa, että Pentaho on kuin Microsoftin ja Jaspersoftin tuotteiden välimalli. Drill down -ominaisuuden toteuttaminen on vielä vaikeampaa kuin Jaspersoftin tuotteessa. Kuten Jaspersoftin vastaavassa ympäristössä, parametrien ja hyperlinkkien avulla se on kuitenkin mahdollista.

Pentaho Testiraportti				
Tuotteet ==> Myyjät ==> Varastosaldo/Hinta				
PRODUCTLINEClassic Cars				
PRODUCTNAME	PRODUCTVENDOR	PRODUCTDESCRI ..	QUANTITYINSTOCK	BUYPRICE
1968 Ford Mustang	Autoart Studio Design	Hood, doors and trunk all open to reveal highly detailed interior features. Steering wheel actually turns the front wheels. Color dark green.	68	95
1966 Shelby Cobra 427 S/C	Carousel DieCast Legends	This diecast model of the 1966 Shelby Cobra 427 S/C includes many authentic details and operating parts. The 1:24 scale model of this iconic lightweight sports car from the 1960s comes in silver and it's own display case.	8,197	29
1958 Chevy Corvette Limited Edition	Carousel DieCast Legends	The operating parts of this 1958 Chevy Corvette Limited Edition are particularly delicate due to their precise scale and require special care and attention. Features rotating wheels, working steering, opening doors and trunk. Color dark green.	2,542	16
1982 Camaro Z28	Carousel DieCast Legends	Features include opening and closing doors. Color: White. Measures approximately 9 1/2 Long.	6,934	47
1952 Alpine Renault 1300	Classic Metal Creations	Turnable front wheels; steering function; detailed interior; detailed engine; opening hood; opening trunk; opening doors; and detailed chassis.	7,305	99
1965 Aston Marlin DB5	Classic Metal Creations	Die-cast model of the silver 1965 Aston Martin DB5 in silver. This model includes full wire wheels and doors that open with fully detailed passenger compartment. In 1:18 scale, this model measures approximately 10 inches/20 cm long.	9,042	66

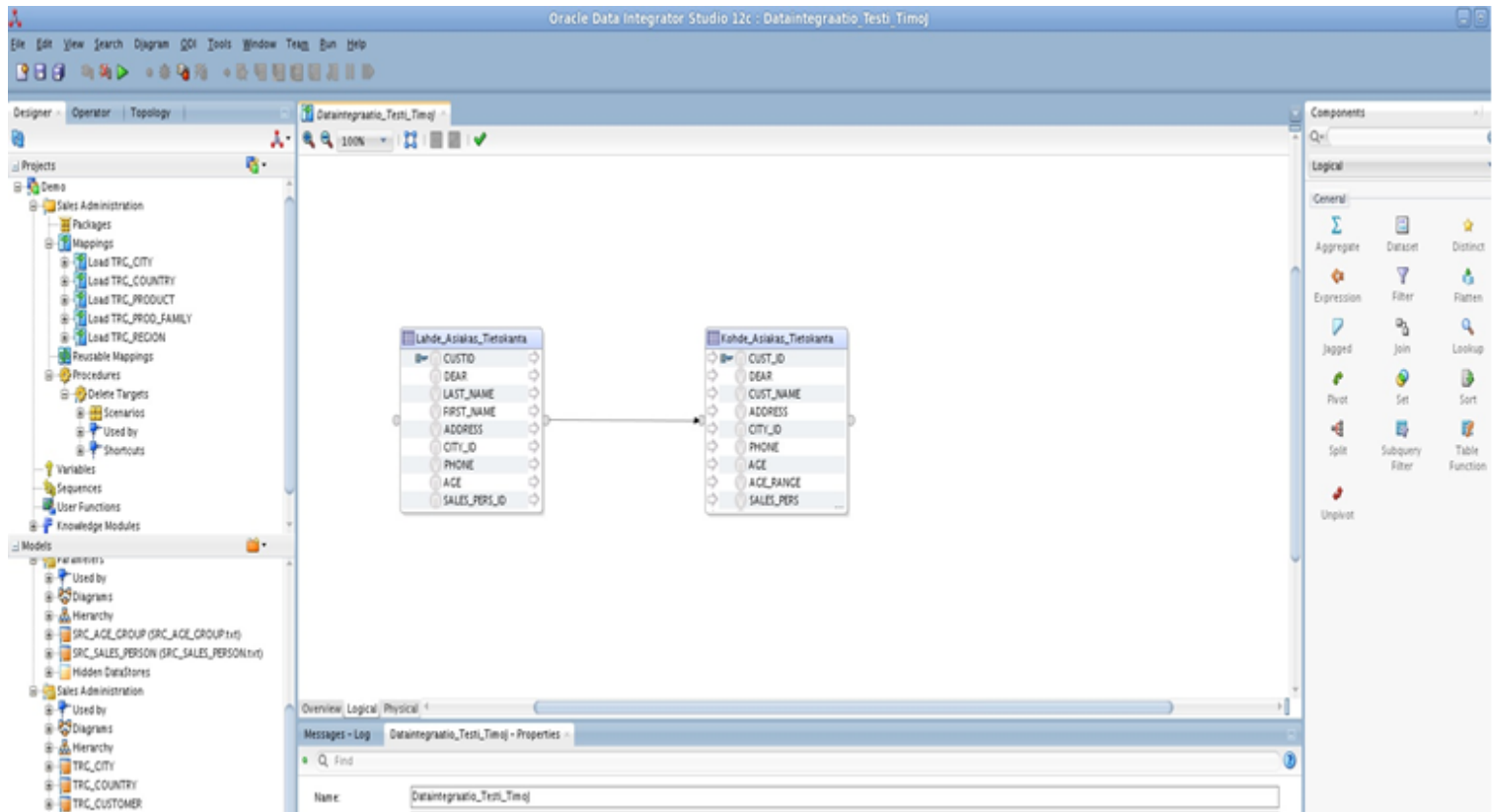
Kuva 11. Pentaho Report Designer -ympäristöllä tehty testiraportti (Kuvakaappaus)

3.7 Oracle Data Integrator

Oracle Data Integrator on Oraclen kehittämä kaupallinen dataintegraatioympäristö, joka vastaa Microsoft SQL Serverin SSIS-osiota. Maksullisen tuotteen vuoksi tässä opinnäytetyössä käytettiin vain kokeiluversiota ympäristöstä. Oraclen ympäristöjen asennus on huomattavasti työläämpää ja raskaampaa kuin kilpailijoiden ympäristöjen. Jos käyttäjät eivät ole IT-alan ammattilaisia, niin asennus ei välttämättä onnistu. Ennen kuin Oracle Data Integratoria voi edes käyttää, täytyy tehdä repositorio testiä varten, ja se luodaan aivan muualla kuin itse ohjelmassa. Tätä ympäristöä käytetään mukana tulevilla graafisella käyttöliittymällä, joka poikkeaa aika paljon vastaavista tuotteista. Oraclen Data Integrator on ominaisuuksiltaan samankaltainen kuin muut vertailussa olevat järjes-

telmät, joten siinä on komponentteja, joita voidaan käyttää erilaisten tiedonsiirtoon liittyvien tehtävien suorittamiseen. Dataintegraatio on kuitenkin hyvin vahvasti sidottuna tietokantoihin tässä ympäristössä, jopa käyttöliittymä muistuttaa enemmän tietokannanhallintaa kuin kilpailijoiden drag & drop -komponenttimenetelmät. Oraclen Data Integrator onkin tehokas (ja raskas) ympäristö, jossa on työkaluja ETL-prosessin laajamittaiseen suorittamiseen. Kuten monet muutkin kilpailijansa, Oraclen ympäristö keskittyy vain dataintegraatioon, eikä sitä ole integroitu isompaan pakettiin kuten Microsoftin SQL Serveriä. Oraclella on omat työkalut ja ympäristöt jokaiselle asiaan liittyvälle toiminnolle. Toisin kuin Pentahon Kettle, Oracle Data Integrator vaatii JDK 8 (Java Development Kit) -kehitysympäristön, joka on ajantasainen (ja Oraclen tuote sekin.) (Oracle 2015a).

Kuvassa 12 on Oracle Data Integrator -ympäristön perusnäkyminen. Tässä kuvassa näkyvät erilaiset komponentit vasemmalla olevassa työkaluvalikossa. Keskellä kuvassa näkyy esimerkkinä tehty dataintegraatio kahden tietokannan välillä, yksinkertaisesti lähdetietokannasta siirretään tietoa kohdetietokantaan. Komponenttien välillä oleva nuoli ilmoittaa mistä tieto lähtee ja mihin se menee. Toisin sanoen komponenttien välillä menee ETL-prosessi. Kuvan oikealla puolella on erilaisia työkaluja esimerkiksi tietojen yhdistelyä (Merge Join) varten.



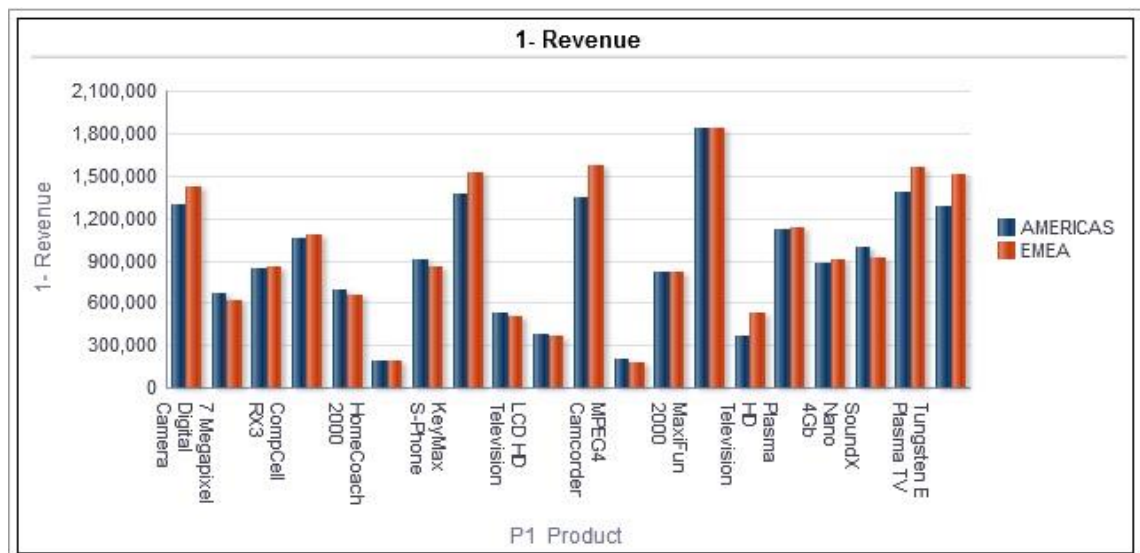
Kuva 12. Oracle Data Integrator -ympäristön perusnäky (Kuvakaappaus)

3.8 Oracle Business Intelligence

Oracle Business Intelligence on Oraclen kehittämä kaupallinen raporttien luontiympäristö, joka vastaa aiemmin mainitun Microsoft SQL Serverin SSRS-osiota. Kuten Oraclen Data Integrator -ympäristön kanssa, Myös Oracle Business Intelligencen täytyy asentaa työläästi ja raskaasti, ennen kuin ympäristöä voi käyttää. Oraclen ympäristö käyttää JDK 8 (Java Development Kit) -kehitysympäristöä, joka on ajantasalla. Oracle Business Intelligence vaatii kuitenkin oman logiikkapalvelininfrastruktuurinsa asennuksen ennen kuin itse ympäristön voi edes asentaa. Tämä on monivaiheinen prosessi, joka on huomattavasti työläämpää kuin vastaavilla tuotteilla. Ominaisuuksiltaan tämä ympäristö on kuitenkin hyvin samanlainen kuin Microsoftin ja Jaspersoftin ratkaisut, joten siinä on komponentteja, joita voidaan käyttää erilaisten raporttien luontia varten. Toiminnallisuus on samanlainen kuin vastaavissa tuotteissa, ja käyttöliittymä on hyvin selainpohjaista. Myös Oracle Business Intelligence on palvelin-pohjainen raportinluontijärjestelmä, jonka avulla käyttäjät voivat luoda raport-

teja erilaisista tietolähteistä. Tämä on mielestäni vertailussa olevista ympäristöistä kaikista vaikein. Kuten vastaavista ympäristöistä, myös Oracle Business Intelligencestä löytyy laaja joukko työkaluja monipuolisten raporttien tuottamiseen. Visuaalisuuden hyödyntäminen on melko samanlaista kuin Microsoftin vastaavassa tuotteessa (Ympäristöstä löytyy erilaisia mittareita ja indikaattoreita pelkkien kaavioiden lisäksi).

Kuvassa 13 on Oracle Business Intelligence -ympäristöllä tehty testiraportti, joka on mukailtu heidän esimerkistään. Myös tässä testissä on käytetty testitietokantaa, josta tiedot on haettu raporttiin. Komponentit voidaan itse manuaalisesti vetää paikoilleen. Tässä testissä on kuvattu perusnäkyä, jossa on vain testitiedoista tehty pylväskaavio. Drill down -ominaisuus on kuitenkin toteutettu hyvin ja yksinkertaisesti, kuten vastaavassa Microsoftin SQL Serverin SSRS-osassa.



Kuva 13. Oracle Business Intelligence -ympäristöllä tehty testiraportti (Kuva-kaappaus)

4 Järjestelmien vertailu

4.1 Ominaisuuksien vertailu

Seuraavassa vertaillaan eri järjestelmiä erilaisin vertailukriteerein, ja eroja on havainnollistettu taulukkojen avulla. Taulukossa 1 on perusvertailu, jossa kerrotaan eri ympäristöjen hankintakustannukset, niiden laitevaatimukset sekä tuetut käyttöjärjestelmät.

Taulukossa 1 on esitetty, että kaikissa ympäristöissä on varsin samanlaiset laiteistovaatimukset, jotka eivät ole nykyaikaisille tietokoneille oletettavasti ongelmallisia. Suurimmat erot ovat lähinnä levytilavaatimuksissa. Jos käytössä oleva laitteisto on kuitenkin hieman vanhentunut, Pentahon ympäristö on selvästi muita vertailussa olevia kevyempi. Tuetuissa käyttöjärjestelmissä on jo eroja, sillä Microsoftin ympäristö tukee vain Windowsia, kun taas muut järjestelmät tukevat vähintään Windowsia sekä Linuxia. Hinta on varmasti myös yksi ratkaiseva hankintakriteeri, ja niissä on merkittäviä eroja eri ympäristöjen välillä.

	Microsoft SQL Server	Jaspersoft ETL& JasperReports	Talend Data Integration	Pentaho Kettle & Report Designer	Oracle Data Integrator	Oracle Business Intelligence
Hinta	3000-4000€	0€ (Community Edition)	0€ (Open Studio Version)	0€ (Standard Edition)	5000 - 26000€	1290€ per käyttäjä
Suorittimen teho suositus	2 GHz	2,5 GHz	2,5 GHz	Tupla-ydin	2 GHz	1 GHz
Keskusmuistin suositus	4 Gt RAM	4-8 Gt RAM	4-8 Gt RAM	2 Gt RAM	3 Gt RAM	2 Gt RAM
Levytilavaimuus	6 Gt HDD	10 Gt HDD	900 Mt HDD	1,5 Gt HDD	2,5 Gt HDD	8,5 Gt HDD
Tuetut Käyttöjärjestelmät	Windows	Windows Mac Linux	Windows Mac Linux	Windows Mac Linux	Windows Linux UNIX	Windows Linux UNIX

Taulukko 1. Ympäristöjen laitevaatimukset ja hankintakustannukset.

Taulukossa 2 on esitetty mitä ominaisuuksia ympäristöt tukevat tietovarastoinnin osalta sekä mitä ominaisuuksia moniulotteisen tiedon käsittelyä varten löytyy. Toisin sanoen tarkastellaan, onko ETL-prosessille tuki ja minkälaisia tiedonsiirto rajapintoja ympäristöt tukevat. Moniulotteisen tiedon käsittelystä tarkastellaan, onko tukea OLAP-kuutioille ja tiedonlouhinnalle. Business intelligenen kannalta tärkeälle Drill down -ominaisuuden toteutukselle on myös omat pisteensä. Tämän lisäksi arvioidaan BI-käyttöliittymien muita ominaisuuksia yhdestä (1) kolmeen (3) tähdellä. 1 tähti on tyydyttävä, 2 tähteä on hyvä ja 3 tähteä on kiitettävä. Visuaalisuuden hyödyntämisellä tässä vertailussa tarkoitetaan erilaisten graafisten elementtien hyödyntämistä, kuten kaavioiden, mittareiden ja indikaattoreiden määrää tai laatua.

	Mic- rosoft SQL Server	Jaspersoft ETL & JasperRe- ports	Talend Data Integrati- on	Pentaho Kettle & Report Designer	Oracle Data Integra- tor	Oracle Business Intelligen- ce
ETL-prosessi tuki	Kyllä	Kyllä	Kyllä	Kyllä	Kyllä	Ei
Tiedonsiirto rajapinnat	CSV, Ex- cel, XML, Tietokan- nat	CSV, Excel, XML, Tietokan- nat	CSV, Excel, XML, Tieto- kannat	CSV, Excel, XML, Tieto- kannat	CSV, Excel, XML, Tieto- kannat	N / A
Moniulotteisen tiedon käsitte- ly: OLAP kuutiot	Kyllä	Ei Tarvitsee erillisen ympäristön	Ei	Ei Tarvitsee erillisen ympäristön	Ei Tarvitsee erillisen ympäristön	N / A
Moniulotteisen tiedon käsitte- ly: Data Mining	Kyllä	Ei	Ei	Erillinen ympäristö	Erillinen ympäristö	N / A
BI- käyttöliittymi- en ominai- suus: Raportit	Kyllä	Kyllä	Ei	Kyllä	Ei	Kyllä
BI- käyttöliittymi- en ominai- suus: Dashbo- ardit	Kyllä	Kyllä	Ei	Kyllä	Ei	Kyllä
BI- käyttöliittymi- en ominai- suus: Drill- down	***	**	*	*	*	***
Visuaalisuu- den hyödyn- täminen	***	**	*	*	**	**

Taulukko 2. Ympäristöjen ominaisuuksien vertailu.

Aiemmin esitetyn vertailun pohjalta voidaan sanoa, että Microsoftin ympäristö on tiiviisti integroitu paketti, jonka avulla ei tarvita useita eri ympäristöjä pyörittämään rinnakkain. Laitteistovaatimuksiltaan se on riittävän kevyt nykyaikaisille tietokoneille (Microsoft 2015). Siinä on yksinkertaisesti ja tehokkaasti toteutettu drill-down ominaisuus, jota oppii nopeasti käyttämään. Visuaalisuus on vertailun paras, koska perinteisten kaavioiden lisäksi tässä ympäristössä on monia erilaisia mittareita ja indikaattoreita jotka ovat hienoja ja todella hyvin kuvaavia. Microsoft SQL Serverin hankintakustannukset ovat noin 3000–4000 euroa, ja siinä on vertailun monipuolisimmat ominaisuudet ja paras visuaalisuuden hyödyntäminen, mutta tätä ympäristöä voi käyttää vain Windows -käyttöjärjestelmillä.

Jaspersoftin ympäristöt (Jaspersoft ETL sekä JasperReports) joutuu asentamaan erikseen. Laitteistovaatimuksiltaan ne ovat vertailun raskaimmat ympäristöt, mutta nykyaikaisilla tietokoneilla ei pitäisi tulla ongelmia. (Jaspersoft 2015c.) Jos käytössä on vanhempaa tekniikkaa, ei kannata asentaa Jaspersoftin ympäristöjä. Moniulotteisen datan käsittelyä varten joudutaan asentamaan Jaspersoftin OLAP -palvelu. Tiedonlouhintamahdollisuutta ei suoraan ole. Drill down -ominaisuutta ei ole tehty ympäristöön suoraan, vaan se joudutaan tekemään työläästi parametrien ja hyperlinkkien kautta. Visuaalisuutta on hyödynnetty hyvin, sillä perinteisten kaavioiden lisäksi tässä ympäristössä on erilaisia indikaattoreita kuvausta varten. Jaspersoftin suurin etu onkin se, että se tukee monia eri käyttöjärjestelmiä (Windows, Mac, Linux). Toinen Jaspersoftin vahvuus on ympäristöjen Community Edition, joka on täysin ilmainen. Jaspersoftilla on myös kaupallisia versioita tarjolla, mutta niiden hinta on selvitettävä myyjiltä.

Talendin dataintegraatio -ympäristö on hyvin samanlainen kuin Jaspersoftin vastaava ympäristö, joten siinä on myös hyvin samanlaiset laitteistovaatimukset. Ainoa ero on levytilan vaatimus. Koska Talend Data Integration keskittyy vain dataintegraatioon, se vie paljon vähemmän levytilaa. Ominaisuuksiltaan Talendin ympäristö on täysin identtinen Jaspersoftin ETL -ympäristön kanssa, ja myös tätä ympäristöä voidaan käyttää eri käyttöjärjestelmillä (Windows, Mac, Linux). Myös Talendin Open Studio for Data Integration on täysin ilmainen. Ta-

lendilla on myös kaupallisia versioita tarjolla. Näiden versioiden hintaa täytyy kysyä erikseen heidän myyjiltään. Koska Talend Open Studio for Data Integration -ympäristö keskittyy puhtaasti dataintegraatioon, niin siinä ei ole moniulotteisen datan käsittelyä varten työkaluja eikä tiedonlouhintatyökaluja. Jos tarvitaan vain pelkästään dataintegraatio -ympäristö, niin silloin Talend on varteenotettava vaihtoehto.

Pentahon ympäristöt (Kettle ja Report Designer) joutuu asentamaan erikseen kuten Jaspersoftin ympäristöt. Laitteistovaatimuksiltaan Pentahon ympäristöt ovat helposti vertailun kevyimmät, joten jos käytössä on vain vanhentunutta tekniikkaa, niin silloin Pentaho on paras vaihtoehto. (Pentaho 2015c.)

Moniulotteisen datan käsittelyä varten joudutaan asentamaan Schema Workbench -ympäristö. Myöskään tiedonlouhinta mahdollisuutta ei ole suoraan vertailussa käytetyissä ympäristöissä, vaan tätäkin varten on erillinen Wekaniminen ympäristö. (Pentaho 2015d.)

Kuten vastaavassa Jaspersoftin tuotteessa, drill down -ominaisuutta ei ole tehty ympäristöön suoraan, vaan se joudutaan tekemään työläästi parametrien ja hyperlinkkien avulla. Ominaisuuden toteutus on vain vielä työläämpää kuin Jaspersoftin tuotteella. Pentahon ympäristöissä visuaalisuuden hyödyntäminen on vertailun heikointa, sillä siinä on vain perinteiset kaaviot käytettävissä, eikä mitään hienompia indikaattoreita tai mittareita. Aivan kuten Jaspersoftin ja Talendin ympäristöt, myös Pentaho tukee useita eri käyttöjärjestelmiä (Windows, Mac ja Linux), ja kuten Jaspersoftin ja Talendin tuotteet, myös Pentahon ympäristöjen ”Standard Edition” on täysin ilmainen, ja muiden versioiden hintaa täytyy kysyä suoraan heidän myyjiltään. Näissä maksullisissa versioissa ei tule muuten mitään uutta, niissä on vain luvattu antaa tukea käyttöön tarvittaessa.

Vertailun toiset täysin kaupalliset ympäristöt Microsoftin lisäksi ovat Oraclen ympäristöt. Nämä ympäristöt eivät tule samassa paketissa kuin Microsoftilla, vaan ne täytyy ostaa ja asentaa erikseen. Laitteistovaatimuksiltaan ne ovat hieman kevyempiä kuin Microsoftin ja Jaspersoftin ympäristöt, mutta raskaampia kuin Pentahon. (Oracle 2015b).

Ominaisuuksiltaan Oraclen tuotteet ovat muuten erinomaiset, mutta jälleen moniulotteisen tiedon käsittelyä varten on oma ympäristönsä, joka on tässä tapauksessa Oracle OLAP -ympäristö. Myös tiedonlouhintaa varten on oma Oracle Data Mining (ODM) komponentti, eikä vertailussa olevissa ympäristöissä ollut näitä toimintoja. Drill down -ominaisuus on toteutettu erinomaisesti ja yksinkertaisesti, sillä tietoihin voidaan porautua syvemmälle yhdellä hiiren painalluksella ilman mitään työlästä prosessia, aivan kuten vastaavassa Microsoftin tuotteessa. Myös visuaalisuuden hyödyntäminen on hyvää tässä ympäristössä, sillä siinä on perinteisten kaavioiden lisäksi käytettävissä myös mittareita. On mainitsemisen arvoinen asia, että Oraclen ympäristöt ovat selvästi vaikeimmat asentaa, eikä niitä välttämättä Windowsin peruskäyttäjät osaa asentaa. Myös Oraclen ympäristöt tukevat eri käyttöjärjestelmiä kuten Windowsia ja Linuxia, mutta vertailun ilmaisyympäristöjen tukeman Mac -käyttöjärjestelmien tilalla on UNIX -tuki. Oraclen ympäristöjen suurin heikkous vertailussa onkin niiden hinta; näiden ympäristöjen hinnat voivat suorastaan kauhistuttaa pienempiä organisaatioita. Pelkkä Oraclen Data Integrator ympäristö maksaa noin 5000–26000 euroa riippuen siitä, kuinka monta vuotta ohjelmistoa haluaa käyttää. (Oracle 2015c.)

Mikäli haluaa käyttää Oraclen Business Intelligence -ympäristöä, niin sekin maksaa 1290 euroa per käyttäjä. (Moonsoft 2015). Oraclella onkin omat ympäristöt kaikkia asioita varten, ja ne kaikki maksavat. Joten jos haluaisi mahdollisimman monipuolisesti toimintoja, niin ympäristöjen hankintakustannukset nousevat hurjasti kilpailijoiden yläpuolelle.

4.2 Käyttöliittymien vertailu

Tässä luvussa käsitellään eri ympäristöjen käyttöliittymiä ja kuinka helppokäyttöisiä eri ympäristöt ovat, toisin sanoen kuinka helppoa peruskäytön oppiminen on. Peruskäytön oppimiseen vaikuttaa keskeisimmin ympäristön selkeys ja visuaalinen ilme sekä se, onko käyttäjällä jo kokemusta muista järjestelmistä, jotka ovat pääpiirteittäin samankaltaisia. Myös se vaikuttaa, kuinka hyvin löytyy

materiaalia oppimisen tueksi. Taulukossa 3 on käytettävyyden vertailu, eli kuinka helppokäyttöinen ympäristö on. Vertailussa käytettäviä kriteereitä arvioidaan samaan tapaan kuin aikaisemmassa luvussa, eli yhdestä (1) kolmeen (3) tähdellä. Pisteytykset perustuvat peruskäytön oppimiseen eivätkä kehittyneempien osaajien näkemyksiin. Pisteytykset pohjautuvat esimerkiksi siihen, kuinka selkeältä käyttöliittymä vaikuttaa aloittelijan silmin. Tai kuinka helposti oppimista tukevia materiaaleja on saatavilla tai minkä laatuista (helposti ymmärrettävää aloittelijalle) tämä materiaali on.

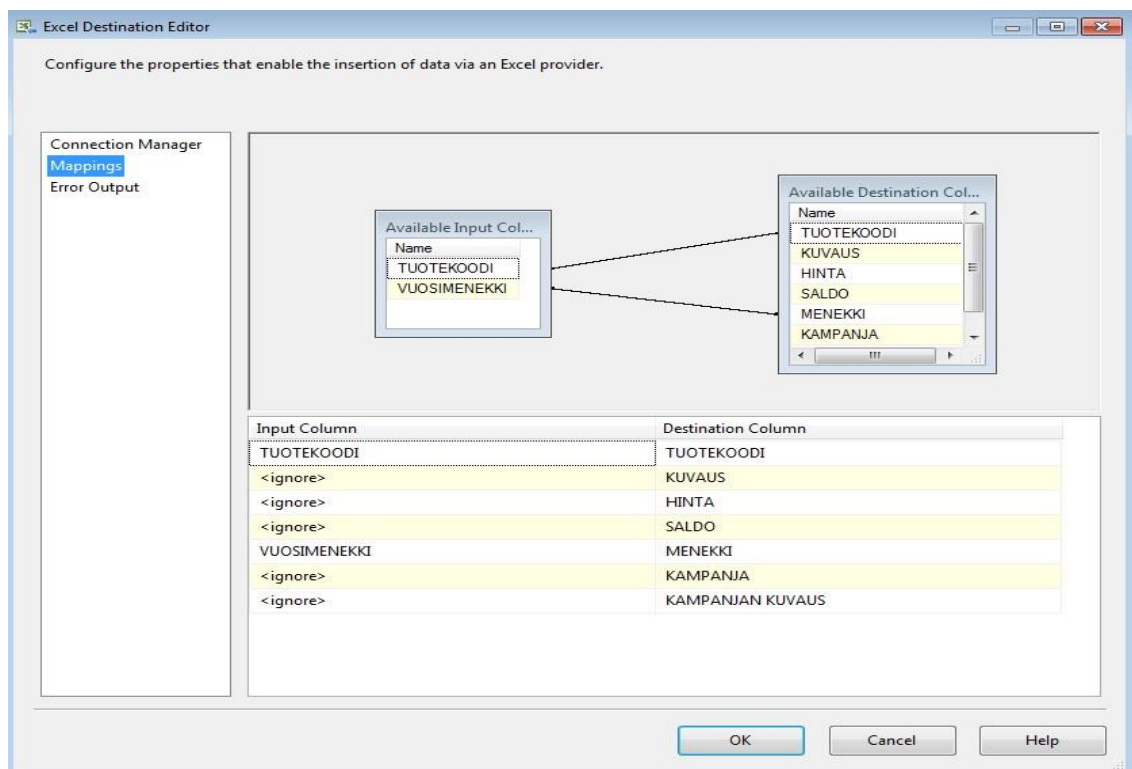
	Microsoft SQL Server	Jaspersoft ympäristöt	Talend Data Integration	Pentaho ympäristöt	Oracle ympäristöt
Ympäristön selkeys	***	**	**	***	*
Ympäristön visuaalisuus	***	**	**	*	**
Manuaalien saatavuus	**	**	**	**	***
Video Tutoriaalien saatavuus	***	**	***	**	***
Ympäristön peruskäytön oppiminen	***	**	**	***	*
Oppimiskäyrää tukevan/helpottavan materiaalin laatu	***	**	**	**	*

Taulukko 3. Ympäristöjen käytettävyyden vertailu.

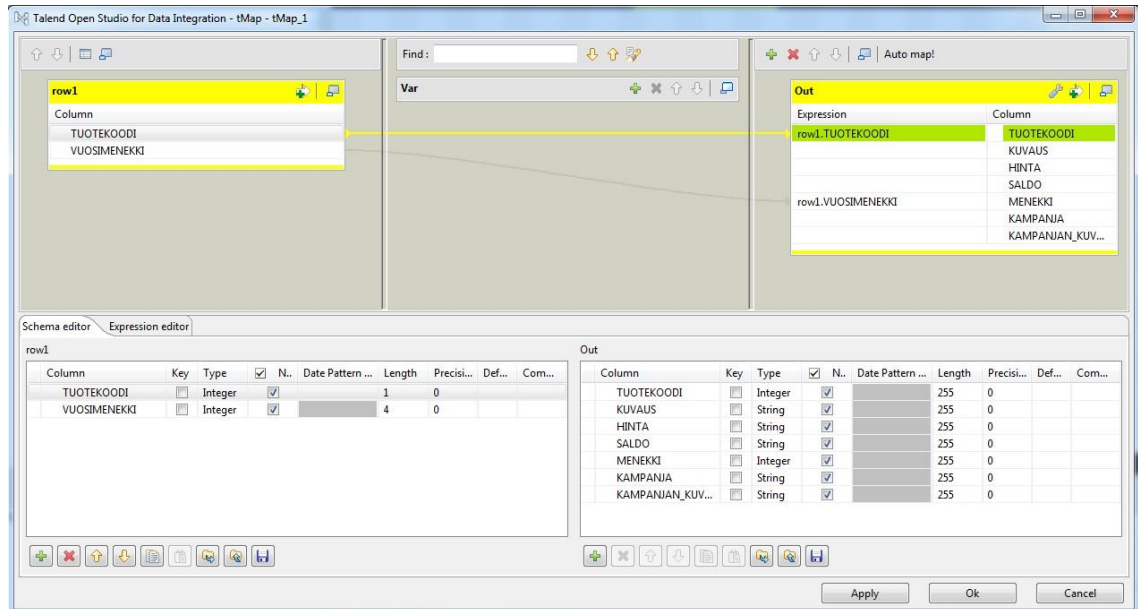
Nämä vertailun pisteytykset pohjautuvat käyttäjäkokemuksiin. Vertailussa yleisesti ottaen parhaat pisteet saanut Microsoftin ympäristö vaikuttaa olevan aloittelijaystävällisin sekä käyttöliittymän selkeyden että oppimista tukevien materiaalien laadun kannalta. Kaikissa ympäristöissä on saatavilla paljon sekä tekstimuotoisia manuaaleja, että videotutoriaaleja, joiden avulla voi opetella ympäristöjen peruskäyttöä. Materiaalien laatu vaihtelee melko paljon riippuen ympäristöstä.

Microsoftilla on oman käyttäjäkokemukseni perusteella selkeimmät videot, joiden avulla oppii peruskäytön melko vaivattomasti. Microsoftin videomateriaaleissa mennään selvästi askel kerrallaan eteenpäin. Oraclella taas on paljon materiaalia, mutta mielestäni peruskäytön oppiminen on vaikeaa materiaalien kanssa tai ilman. Oraclen materiaali on vaikeaselkoista, jos ei omaa paljoa kokemusta dataintegraatiosta tai business intelligencestä. Jaspersoftin, Talendin ja Pentahon ympäristöjen peruskäytön pystyy oppimaan ilman apumateriaaliakin, jos on edes hieman kokemusta Microsoftin ympäristöstä, mutta materiaalien laatu on kuitenkin riittävä, jos ongelmia ilmaantuu. Varsinkin Pentahon ympäristö vaikuttaa aloittelijalle sopivalta.

Käyttöliittymien eroja havainnollistetaan kuvien avulla. Kuvassa 14 on Microsoftin ympäristön, ja kuvassa 15 on Talendin ympäristön (vastaa myös Jaspersoftin) dataintegraatioiden ”mapping”. Molemmat toimivat hyvin samalla tavalla, kuvissa näkyvillä viivoilla linkitetään taulujen tiedot keskenään.

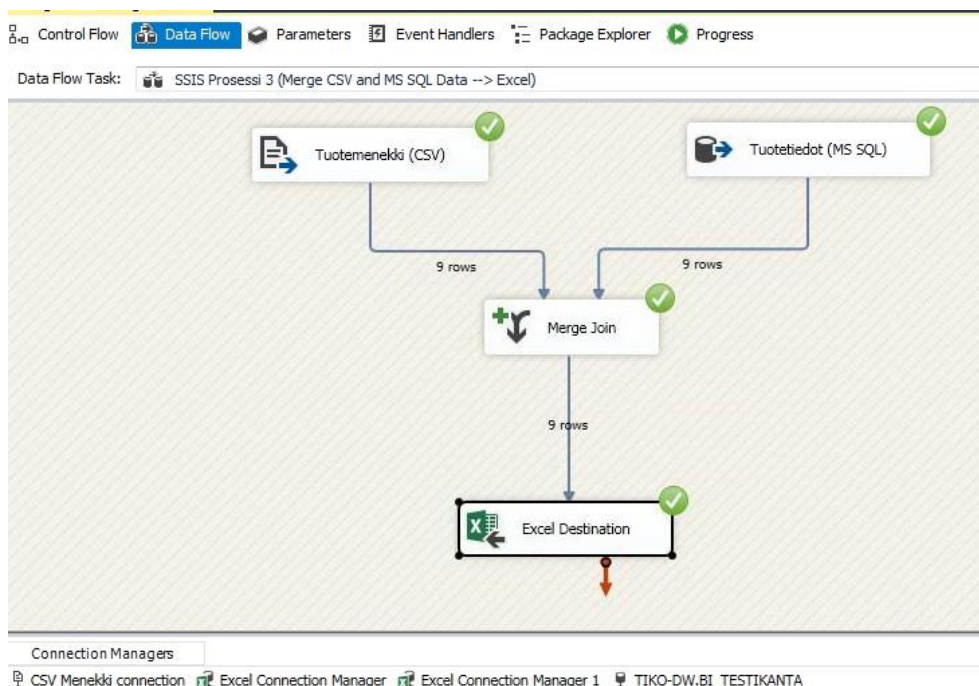


Kuva 14. Microsoftin dataintegraatio (Kuvakaappaus)



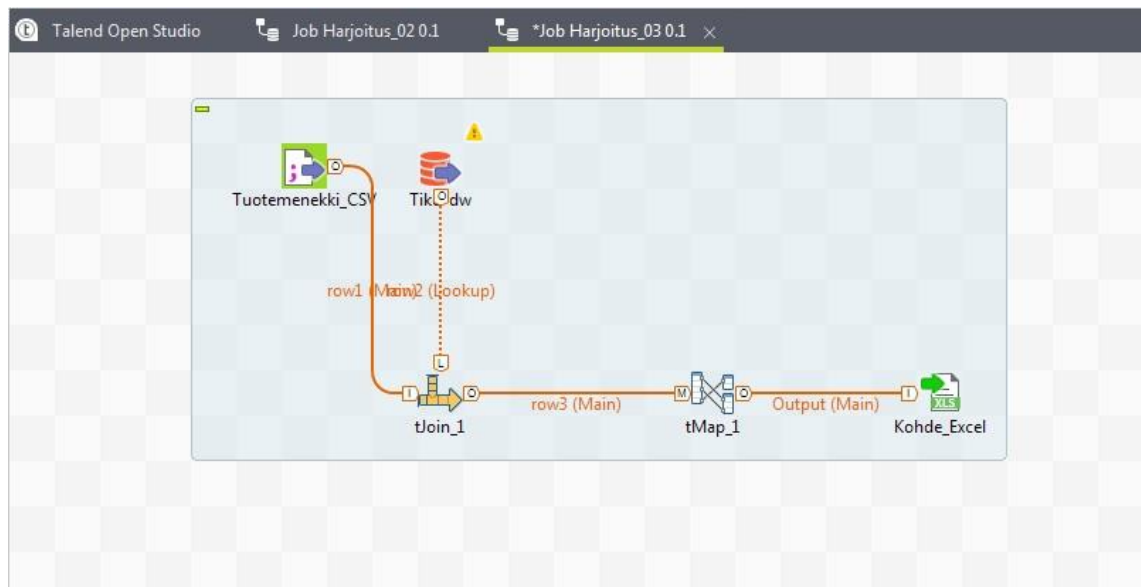
Kuva 15. Talendin (ja Jaspersoftin) dataintegraatio (Kuvakaappaus)

Yksi dataintegraation peruspilareista on eri lähteiden tietojen yhdistely, eli ”Merge Join”-operaatio. Kuvissa 16–19 on esitetty yksinkertainen Merge Join-operaatio tehtynä jokaisella vertailussa käytetyllä ympäristöllä, jotta käyttöliittymien eroavaisuudet tulisivat ilmi. Koska Jaspersoftin ympäristö on täysin samanlainen kuin Talendin, niin siitä ei ole kuvaa erikseen.



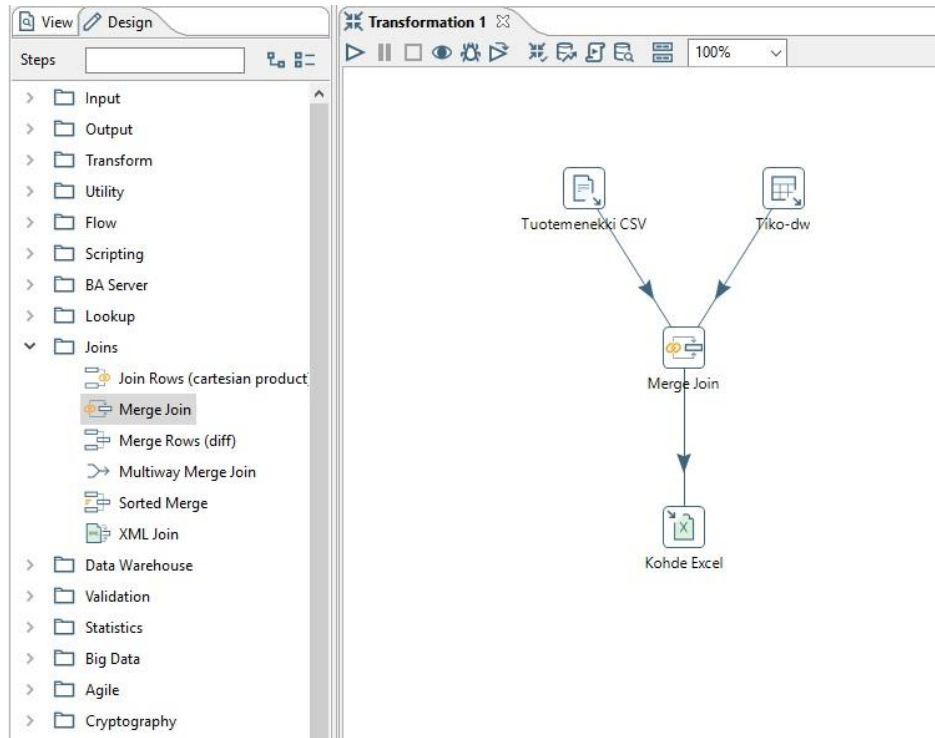
Kuva 16. Microsoft SQL Server -ympäristöllä tehty Merge Join -operaatio (Kuvakaappaus)

Kuvassa 16 on yksinkertainen kahden eri lähteen tietojen yhdistely, eli ”Merge Join” -operaatio Microsoft SQL Server -ympäristöllä toteutettuna. CSV-tiedostosta ja MS SQL -palvelimelta tulee tietoa, joka yhdistetään ennen kuin se siirretään kohteeseen, joka tässä tapauksessa on Excel-tiedosto. Käyttöliittymä on todella selkeä ja ymmärrettävä aloittelijallekin. Graafisia elementtejä on hyödynnetty hyvin, ja lähteiden linkitys onnistuu yksinkertaisesti vetämällä laatikosta nuolen toiseen.



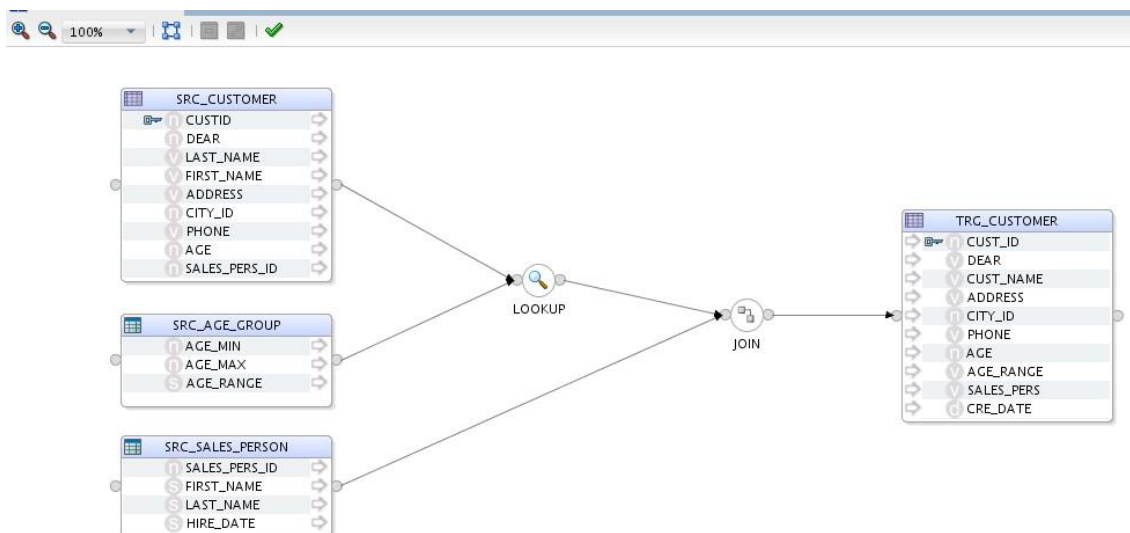
Kuva 17. Talend Open Studio for Data Integration (ja myös Jaspersoft ETL) ympäristöillä tehty Merge Join –operaatio (Kuvakaappaus)

Kuvassa 17 on toisena ”Merge Join” esimerkkinä vastaavan tyyppinen, yksinkertainen kahden eri lähteen tietojen yhdistely, tällä kertaa toteutettuna Talend Open Studio -ympäristöllä. Tässäkin tapauksessa on CSV-tiedosto ja palvelin, joiden tiedot yhdistetään ennen kuin ne siirretään kohteeseen. Toisin kuin Microsoftin ympäristössä, Talendilla täytyy kuitenkin tehdä ”mapping” ennen tiedon lopullista siirtoa. (kuvassa näkyvä tMap_1 on tätä varten) Jaspersoftin ympäristö toimii täysin samalla periaatteella, joten siitä ei ole erillistä kuvaa. Käyttöliittymä on hieman vaikeampi hahmottaa aloittelijan näkökulmasta verrattuna Microsoft SQL Serveriin, koska siinä on paljon välivaiheita jotka täytyy ottaa huomioon (esim. aiemmin mainittu mapping).



Kuva 18. Pentaho Data Integration (Kettle) -ympäristöllä tehty Merge Join – operaatio (Kuvakaappaus)

Kuvassa 18 on kolmantena ”Merge Join” esimerkkinä yksinkertainen kahden eri lähteen tietojen yhdistely, Pentaho Data Integration (Kettle) -ympäristöllä toteuttuna. CSV-tiedostosta ja SQL Server 2012 -palvelimelta tulee tietoa, joka yhdistetään ennen kuin se siirretään kohteeseen, joka on myös tässä tapauksessa Excel-tiedosto. Tämä on varsin samanlainen operaatio kuin Microsoft SQL Serverillä, ja käyttöliittymä on todella selkeä ja yksinkertainen. Jos osaa perusteet Microsoftin tuotteesta, oppii varmasti myös Pentahon -ympäristön todella nopeasti. Komponenteista vedetään vain yksinkertaisesti nuoli toiseen, ja nuolen suunta näyttää mistä mihin tieto kulkee.



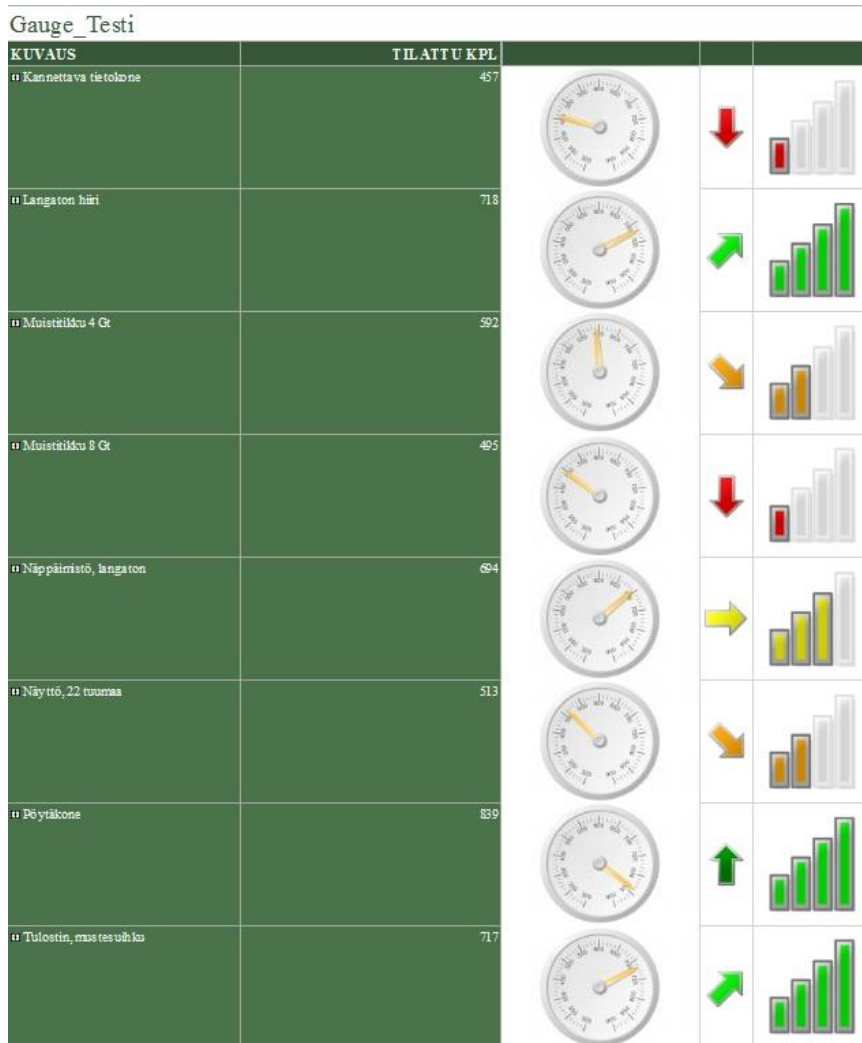
Kuva 19. Oracle Data Integrator -ympäristöllä tehty Merge Join -operaatio (Kuvakaappaus)

Kuvassa 19 on vastaava, yksinkertainen eri lähteiden tietojen yhdistely, tällä kertaa toteutettuna Oracle Data Integrator -ympäristöllä. Lähteenä toimivat tietokannan eri taulut, joista yhdistetään tietoja ja siirretään toiseen tietokantaan. Käyttöliittymä on huomattavasti erilaisempi kuin kilpailijoiden tuotteissa, ja se muistuttaa enemmän tietokannanhallintaa ja on vaikeaselkoinen aloittelijan näkökulmasta. Tämän vuoksi Oraclen ympäristön oppiminen on vaikeampaa aloittelijoille kuin muiden vastaavien ympäristöjen. Sillä ei pysty vain ottamaan lähettä, ja vetämään komponentteja yksinkertaisesti hiirellä haluamaansa paikkaan, vaan sen käytössä pitäisi olla jo kokemusta tietokannanhallinnasta.

Microsoftin, Talendin (ja Jaspersoftin) sekä Pentahon ympäristöt ovat hyvin samankaltaisia, kun taas Oracle on hyvin erilainen. Yksinkertaiset käyttöliittymät, joilla voidaan hiirellä vetää komponentteja paikasta toiseen, näyttävätkin olevan nykypäivänä suosiossa, mutta Oraclen käyttöliittymä muistuttaa tänäkin päivänä vielä enemmän tietokannanhallintaa (ja sen oppiminen on huomattavasti vaikeampaa aloittelijalle).

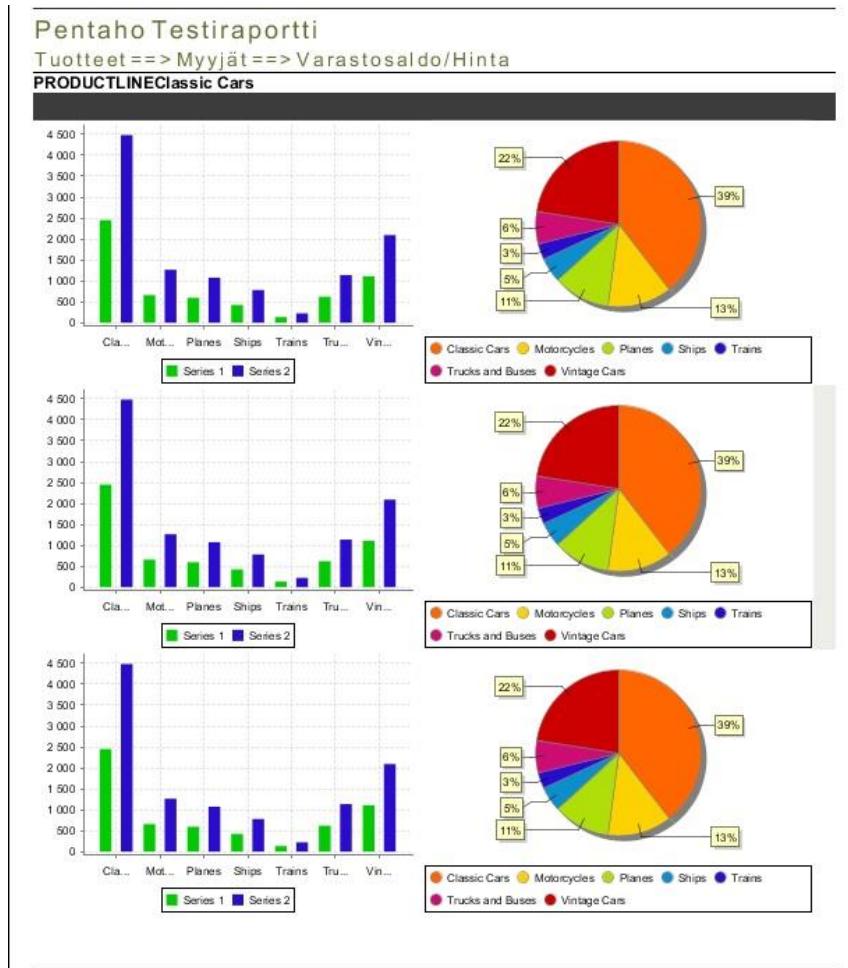
Toinen tärkeä asia on tiedon esittelyn yhteydessä käytettävät graafiset elementit. Varsinkin business intelligence -ympäristöjen kanssa erilaiset visuaaliset komponentit tehostavat (ja helpottavat) tietojen tarkastelua, kun ei tarvitse vain selata suuria numeromassoja. Kuvissa 20–22 on esitelty Microsoftin, Pentahon

ja Oraclen graafisia elementtejä. Näiden kuvien perusteella voidaan vertailla eri ympäristöjen visuaalista hyödyntämistä.



Kuva 20. Microsoft SQL Server Reporting Services (SSRS) -ympäristössä käytettäviä graafisia indikaattoreita ja mittareita (Kuvakaappaus)

Kuvassa 20 on esimerkkinä Microsoft SQL Serverillä toteutettuja mittareita ja indikaattoreita, joiden avulla on helpompi tarkastella tietoja. Microsoft SQL Server sai vertailun pisteilyssä parhaan tuloksen visuaalisuuden hyödyntämisessä, koska nämä graafiset komponentit ovat tyylikkäitä ja helposti ymmärrettäviä. Myös drill down -ominaisuuden toteutus on yksinkertaista ja onnistunutta. (tietoja voisi tarkastella tarkemmin painamalla kuvassa näkyviä painikkeita tuotenimien vieressä).



Kuva 21. Pentaho Report Designer -ympäristössä käytettäviä perinteisiä kaavioita (Kuvakaappaus)

Kuvassa 21 on esimerkkinä Pentaho Report Designer -ympäristöllä toteutettuja perinteisiä kaavioita. Tämä kyseinen ympäristö sai visuaalisuuden hyödyntämisessä heikot pisteet, koska nämä graafiset komponentit ovat hyvin yksinkertaisia, ja niitä pystyy tuottamaan mistä tahansa muustakin business intelligence ohjelmasta (esim. Excelistä). Drill down -ominaisuuden toteutus on kuitenkin kömpelöä, eikä se kuvassa olevassa testissä onnistunut.



Kuva 22. Oracle Business Intelligence -ympäristössä käytettäviä mittareita (Kuvakaappaus)

Kuvassa 22 on esimerkkinä Oracle Business Intelligence -ympäristöllä toteutettuja mittareita. Tämä kyseinen ympäristö sai visuaalisuuden hyödyntämisessä hyvät pisteet, koska toisin kuin aiemmin esitelty Pentaho, tämä ympäristö pystyy tuottamaan mittareita ja muita erikoistuneempia graafisia komponentteja. Mutta nämä graafiset komponentit eivät kuitenkaan ole niin tyylikkäitä kuin Microsoftin vastaavassa tuotteessa (ja indikaattorit puuttuvat), joten tämän vuoksi Oraclen pisteytys oli näiden kahden ympäristön väliltä. Drill down -ominaisuuden toteutus on yksinkertaista ja tehokasta, aivan kuten Microsoftin vastaavassa tuotteessa. Painamalla kuvassa näkyviä brandeja (esim. BizTech), voitaisiin tarkastella lukuja tarkemmin.

5 Vertailun tulokset ja johtopäätökset

5.1 Microsoft SQL Server

Vertailussa Microsoft SQL Server oli vahvoilla monissa eri asioissa. Vertailun perusteella voidaankin todeta, että MS SQL Server + Visual Studio ovat aloitteli-

jaystävällisimmät ympäristöt, mitä vertailussa oli. Selkeä käyttöliittymä ja monipuoliset ominaisuudet sekä integroidut ympäristöt tekevät Microsoftin ratkaisusta tehdyn vertailun perusteella parhaimman, jos käytössä on Windows – käyttöjärjestelmä. Microsoft SQL Serverin suurin heikkous on se, että se ei tue mitään muuta käyttöjärjestelmää kuin Windowsia. Jos organisaatiolla on käytössään Windows – käyttöjärjestelmällä varustettuja työasemia ja palvelimia, Microsoft SQL Server – ympäristöä voidaan suositella lämpimästi. Microsoft SQL Server soveltuu hienosti koulutukseen tai asioiden perusteiden opetteluun sekä pienehköille (mutta ei mikroyrityksille) että keskikokoisille yrityksille. Kymmenen tai useamman henkilön yrityksille 3000–4000 euron investointi ei ole enää este, mutta aivan pienille (mikro) yrityksille tämä voi olla ongelmallista. Tämä muutaman tuhannen investointi on kuitenkin melko pieni, kun ajatellaan, että saadaan kerralla kolme ympäristöä (integrointi-, analyysi- ja raportointipalvelut) käytettäväksi.

5.2 Jaspersoft ETL & JasperReports

Vertailuissa kävi ilmi, että Jaspersoftin ympäristöt ovat hieman raskaampia kuin muut, joten jos organisaatiolla on laitteistoresursseiltaan riittämätöntä tekniikkaa käytössään, kannattaa harkita tarkkaan, hankkiiko näitä ympäristöjä. Nämä Jaspersoftin tarjoamat ympäristöt täytyy asentaa erikseen, eivätkä ne ole valmiissa paketissa. Jaspersoftin vahvuutena on kuitenkin eri käyttöjärjestelmien tuki (varsinkin Mac, jota muut ympäristöt eivät kovin hyvin tue) sekä Community Editionin hinta, sillä ympäristö on ilmainen. Ominaisuuksiltaan se on riittävä, mutta kaikkea ei löydy mitä vastaavasta Microsoftin ympäristöstä löytyy. Varsinkin business intelligenen kannalta tärkeän drill down -ominaisuuden toteutus on kömpelöä. Jos organisaatiolla on käytössään muita käyttöjärjestelmiä kuin Windows ja jos ilmaisuus houkuttelee, Jaspersoftin ympäristö voi olla hyvä ratkaisu. Jaspersoftin ympäristö soveltuukin parhaiten pienille tai keskisuurille yrityksille, joilla on hyvin teknisiä työntekijöitä ja jotka eivät halua investoida da-taintegraatio- ja business intelligence -järjestelmiin.

5.3 Talend Open Studio for Data Integration

Vertailussa todettiin, että Talendin dataintegraatoratkaisu on identtinen Jaspersoftin ETL-ympäristön kanssa. Sen käyttö kuitenkin tuntuu hieman helpommalta, ja Talend tuntuu antavan vähemmän virheilmoituksia kuin Jaspersoftin ETL. Talendin ympäristö kuitenkin keskittyy vain dataintegraatioon, eikä sillä voi tehdä raportteja. Kuten Jaspersoftin ympäristö, myös Talend tukee monia eri käyttöjärjestelmiä, ja se on täysin ilmainen. Jos organisaatio tarvitsee vain dataintegraatioympäristön, voidaan suositella Talendin Open Studio for Data Integrationia, koska se vie vähän levytilaa ja sillä pystyy suorittamaan kaikki dataintegraatioon liittyvät tehtävät. Talendin ympäristö soveltuisi hyvin koulutukseen (jos on vain pelkkään dataintegraatioon liittyvä kurssi) sekä pienille yrityksille.

5.4 Pentaho Kettle & Report Designer

Pentahon ympäristöt ovat vertailun kevyimmät, ja tämän vuoksi niitä voidaan suositella myös sellaisille organisaatioille, joilla on toimintaympäristössään vaatimattomammat laitteistoresurssit käytettävissään. Kuten Jaspersoftin vastaavat tuotteet, nämä ympäristöt täytyy asentaa erikseen. Pentahon vahvuuksina ovat samat kuin Jaspersoftin vastaavassa tuotteessa, eli eri käyttöjärjestelmien tuki (Windows, Mac, Linux) sekä Standard -editionin ilmaisuus. Ominaisuuksiltaan se on myös melko samanlainen kuin Jaspersoftin tuote, ja drill down -ominaisuuden toteutus on kömpelöä. Pentaho on kuitenkin huomattavasti aloittelijaystävällisempi ja sillä on selkeä käyttöliittymä, minkä vuoksi Pentahon ratkaisuja voidaan suositella varsinkin koulutukseen tai asioiden perusteiden opetteluun sekä pienille yrityksille, ja kenties keskikokoisillekin yrityksille, jos on tiukka budjetti eikä organisaatio halua investoida liikaa dataintegraatio- ja business intelligence -ratkaisuihin.

5.5 Oracle Data Integrator & Business Intelligence

Oraclen on ns. ”raskaan sarjan” tekijä DW/BI-alalla. Oraclen ja sen ympäristöjen suurin vahvuus onkin niiden maine. Oracle itsessään on varsin tunnettu, todennäköisesti jokainen IT-alalla työskentelevä on jossain vaiheessa kuullut Oraclesta, vaikka ei sen ympäristöjä käyttäisikään. Tämä on samalla myös heikkous, sillä tämä maine on selvästi vaikuttanut hintoihin. Oraclen tuotteet ovatkin vertailun kalleimmat, ja hintaero kilpailevaan kaupalliseen tuotteeseen eli Microsoft SQL Serveriin on todella huomattava. Toisaalta Oraclen tuotteiden perusominaisuudet ovat myös hyvät, ja drill down -ominaisuus on toteutettu hyvin. Käyttöjärjestelmien tuki on hieman erilainen kuin muilla järjestelmillä, koska Mac-käyttöjärjestelmien tilalla on UNIX-käyttöjärjestelmät. Oraclen ympäristöt ovat myös vertailun vaikeimmat käyttää ja asentaa, joten asennusta varten tarvitaan kokeneita IT-alan ammattilaisia. Oraclen ympäristöä siis ei voida suositella muille kuin suurille organisaatioille, joille tuotteiden suuret hinnat eivät ole este. Suuret organisaatiot, joiden palveluksessa työskentelee kokeneita IT-ammattilaisia ja jotka käyttävät Linux- tai UNIX-käyttöjärjestelmiä (varsinkin palvelinpuolella), ovat Oraclen ympäristöille parhaiten soveltuva kohde.

6 Pohdinta

Opinnäytetyön toimeksiannossa tavoitteena oli tehdä vertailu erilaisten tietovarastointi- ja business intelligence -ympäristöjen välillä. Tätä varten piti tehdä kartoitus eri ympäristöjen ominaisuuksista ja käytännössä kokeilla, kuinka ympäristöt toimivat. Tämän lisäksi täytyi purkaa auki aiheen keskeisiä käsitteitä ja tekniikoita, ennen kuin varsinaista vertailua pystyi tekemään. Ennen opinnäytetyötä tietoperustassa käsittelemäni aiheet olivat minulle täysin tuntemattomia, ja lähteiden lukeminen ja tietoperustan kirjoittaminen auttoi minua ymmärtämään mistä on kyse. Varsinkin eri ympäristöjen kokeileminen käytännössä oli mielenkiintoinen ja ehkä hieman haastavakin prosessi, kun aiempaa kokemusta asiasta ei ollut. Opinnäytetyöprosessin loppupuolella huomasin, kuinka paljon ymmärrystä olin saanut aiheeseen. Osasin käyttää kaikkia perustoimintoja kaikilla

vertailussa olevilla ympäristöillä, joten tästä kokemuksesta on varmasti hyötyä jatkon kannalta.

Opinnäytetyöni aihe on mielestäni varsin ajankohtainen, sillä tietovarastointi- ja business intelligence -ratkaisut ovat olleet niin sanotussa toisessa nousussa. Pikemminkin voitaisiin sanoa, että nämä ympäristöt ovat jo 1980-luvulta asti tehneet tuloaan, mutta vasta nyt teknologia alkaa olla siinä pisteessä, että toteutuksessa ei tule pahoja pullonkauloja.

Opinnäytetyön käytännön osuuden (ympäristöjen kokeilun) oli tarkoitus olla hieman laajempi, mutta aikataulullisista syistä (ja opinnäytetyön rajauksen vuoksi) jouduin tekemään hieman suppeammat testaukset. Keskityin sitten eri ympäristöjen perusominaisuuksiin ja siihen, kuinka helppo aloittelijan on oppia käyttämään ympäristön näitä kyseisiä perusominaisuuksia.

Opinnäytetyössäni raapaistaan vain tietovarastointi- ja business intelligence -aiheiden pintaa. Tarkoituksena olikin vertailla eri ympäristöjä ja saada näkemystä eri ympäristöjen soveltuvuudesta erilaisille organisaatioille, eikä mennä kovin syvälle aiheessa.

Raportin alussa mainitsin, että opinnäytetyön tarkoituksena oli siis vastata seuraaviin kysymyksiin:

- Mitä Data Warehouse- ja business intelligence -ympäristöt ovat?
- Millaisia tekniikoita näihin ympäristöihin sisältyy?
- Millaisia ominaisuuksia näistä ympäristöistä löytyy?
- Millaisia eroja eri yritysten DW/BI -ohjelmistoissa on?

Nämä kysymykset ohjasivat lähteiden ja materiaalien etsimistä eteenpäin, kunnes sain itseäni tyydyttävät vastaukset näihin kysymyksiin. Kun tietoperusta oli valmis, alkoi ympäristöjen konkreettiset kokeilut. Tässä vaiheessa mukaan tuli myös Pentahon ja Oraclen ympäristöt, jotka laajensivat opinnäytetyön seuraavaa vaihetta, eli vertailuja. Asetin vertailujen kriteereiksi pääasiassa sen, kuinka helppoa ympäristöjen peruskäytön oppiminen on, ja kuinka selkeitä käyttöliittymät olivat aloittelijan näkökulmasta. Muita tärkeitä kriteereitä olivat erilaisten

ominaisuuksien tukeminen ja visuaalisuuden hyödyntäminen eri ympäristöissä. Toki vertailut ovat aina osittain mielipidekysymyksiä. Vaikka vertailulle on omat kriteerinsä, niin toisen henkilön mielipide voi poiketa minun mielipiteestäni. Mielestäni varsinkin toisille aloittelijoille nämä vertailut ovat luotettavia. Uskon vahvasti, että dataintegraatio- ja business intelligence -aloittelijat ovat samaa mieltä näiden vertailujen tuloksista tehtyjen johtopäätösten kanssa.

Tietovarastoinnin ja business intelligencen tulevaisuus tulee muuttumaan. Uskon, että nämä ympäristöt siirtyvät yhä enemmän pilvipalveluiden käyttöön. Toki jatkossakin käytetään näitä perinteisempiä tiedonsiirtorajapintoja, kuten Exceliä, mutta nykyinen trendi vaikuttaa siltä, että manuaalinen tietojenkäsittely (esim. kiintolevyillä) on siirtymässä syrjään pilvipalveluiden edessä. Ehkä varsinaiset tietovarastointi ja business Intelligence -ohjelmistotkin siirtyvät puhtaasti selaimessa toimiviksi kokonaisuuksiksi, osittain ne tänäpäivänäkin toimivat jo selaimessa. Varsinkin dataintegraatio tehdään vielä omassa ympäristössään eikä selainpohjaisella käyttöliittymällä. Automaatio näissäkin ympäristöissä varmasti lisääntyy 2020-luvun aikana huomattavasti.

Lähteet

Astera. 2015. Challenges of ETL.

<http://www.astera.com/solutions/technology-solutions/etl2> 14.9.2015

Dataintegration.info 2015. Data Integration. <http://www.dataintegration.info/data-integration>. 18.9.2015.

Data Warehouses. 2015a. Introduction to BI & DW. <http://data-warehouses.net/guide/introduction.html>. 16.9.2015.

Data Warehouses. 2015b. Architecture Overview. <http://data-warehouses.net/architecture/overview.html>. 16.9.2015.

Honkavaara, J.-P. 2013.

ETL-prosessin parhaita käytäntöjä tietovaraston rakentamisessa. Tampereen yliopisto.

<https://tampub.uta.fi/bitstream/handle/10024/84751/gradu06854.pdf?sequence=1>.

4.9.2015.

Jaspersoft. 2015a. Jaspersoft ETL.

<http://community.jaspersoft.com/project/jaspersoft-etl/resources%20>
30.11.2015

Jaspersoft. 2015b. JasperReports Server.

<http://community.jaspersoft.com/project/jasperreports-server>
29.12.2015

Jaspersoft. 2015c. System Requirements.

<http://community.jaspersoft.com/documentation/jasperreports-server-install-guide/v561/system-requirements>

28.12.2015

Kallio, T. 2012. SQL Server Integration Services –ETL prosessien kehittäminen. Savonia ammattikorkeakoulu.

http://www.theseus.fi/bitstream/handle/10024/41605/Kallio_Timo.pdf?sequence=1. 7.9.2015.

Microsoft. 2015. Hardware and Software Requirements for Installing SQL Server 2014.

<https://msdn.microsoft.com/en-us/library/ms143506%28v=sql.120%29.aspx>
28.12.2015

Moonsoft. 2015. Oracle Business Intelligence Standard Edition One.

<http://www.moonsoft.fi/products/000623.aspx>
30.12.2015

- Oracle. 2015a. Oracle Data Integrator.
<http://www.oracle.com/technetwork/middleware/data-integrator/overview/index.html>
29.12.2015
- Oracle. 2015b. Oracle Fusion Middleware System Requirements and Specifications.
https://docs.oracle.com/html/E38687_01/12c_fusion_requirements.htm
29.12.2015
- Oracle. 2015c. Oracle Store – Oracle Data Integrator.
https://shop.oracle.com/pls/ostore/product?p1=OracleDataIntegratorEnterpriseEdition&p2=&p3=&p4=&p5=&intcmp=ocom_Oracle_Data_Integrator_Enterprise_Edition
30.12.2015
- Pentaho. 2015a. Data Integration – Kettle.
<http://community.pentaho.com/projects/data-integration/>
22.12.2015
- Pentaho. 2015b. Installing Kettle.
<http://wiki.pentaho.com/display/EAI/01.+Installing+Kettle>
22.12.2015
- Pentaho. 2015c. Components Reference.
<https://help.pentaho.com/Documentation/5.2/0D0/160/000>
29.12.2015
- Pentaho. 2015d. Data Mining – Weka.
<http://community.pentaho.com/projects/data-mining/>
29.12.2015
- Rouse, M. 2015. Business Intelligence (BI) definition.
<http://searchdatamanagement.techtarget.com/definition/business-intelligence> 20.9.2015
- Talend. 2016. Talend Products - Data Integration.
<https://www.talend.com/products/data-integration> 19.1.2016
- Wikipedia. 2015a. Business Intelligence.
https://fi.wikipedia.org/wiki/Business_intelligence
20.9.2015
- Wikipedia. 2015b. Asiakkuudenhallinta.
<https://fi.wikipedia.org/wiki/Asiakkuudenhallinta>
20.9.2015
- Wikipedia. 2015c. Data warehouse.
https://en.wikipedia.org/wiki/Data_warehouse
20.9.2015

Wikipedia. 2015d. Toiminnanohjausjärjestelmä
<https://fi.wikipedia.org/wiki/Toiminnanohjausj%C3%A4rjestelm%C3%A4>
20.9.2015

Wikipedia. 2015e. Extract, transform, load
https://en.wikipedia.org/wiki/Extract,_transform,_load
20.9.2015

Wikipedia. 2015f. SQL
<https://fi.wikipedia.org/wiki/SQL>
20.9.2015

Wikipedia. 2015g. Data Mart. https://en.wikipedia.org/wiki/Data_mart 10.9.2015

Wikipedia. 2015h. Data Integration.
https://en.wikipedia.org/wiki/Data_integration
10.9.2015

Wikipedia. 2015i. SQL Server Integration Services.
https://en.wikipedia.org/wiki/SQL_Server_Integration_Services
20.10.2015

Wikipedia 2015j. Microsoft Analysis Services.
https://en.wikipedia.org/wiki/Microsoft_Analysis_Services
25.10.2015

Wikipedia. 2015k. SQL Server Reporting Services.
https://en.wikipedia.org/wiki/SQL_Server_Reporting_Services
26.10.2015