

Opinnäytetyö (AMK)
Tietotekniikka
Ohjelmistoliiketoiminta
2016

Kimmo Santamaa

DATAN VISUALISOINTI ORGANISAATIOSSA



Kimmo Santamaa

DATAN VISUALISOINTI ORGANISAATIOSSA

Tämän opinnäytetyön tavoite oli toteuttaa ns. POC eli Proof of Concept TeliaSoneralle QlikView-sovelluksen soveltuvuudesta työkaluna organisaation datan visualisointiin. QlikView on Qlikin kehittämä Business Intelligence -alusta, joka sopii sen in-memory-pohjaisen teknologian ansiosta suurille yrityksille. Suuryritysten tarpeina on löytää ratkaisu suurien datamäärien käsittelyyn. POC:n aikana pyrittiin arvioimaan QlikViewin soveltuvuutta mahdollisimman hyvin.

Valmis sovellus jaettiin kolmeen eri näkymään. Nämä näkymät ovat dashboard, tapahtumat ja henkilöstö. Eri näkymien on tarkoitus esittää määritelty data eri tavoilla. Dashboard on ensimmäinen näkymä, joka on suunnattu siten että käyttäjä näkee tärkeät KPI-arvot mahdollisimman selkeästi ja nopeasti. Kaikkien tapahtumien seuranta tapahtuu tapahtumat-näkymästä. Henkilöstö-näkymän mittareiden tarkoitus on seurata henkilöstön tehokkuutta, jolloin voidaan kohdistaa tulevat palvelupyynnöt ylikuormittamatta henkilöstöä. Tärkeänä ominaisuutena määriteltiin se, että käyttäjä voi porautua yksittäiseen tapahtumaan raakadataan asti.

Työ toteutettiin QlikView-ohjelmistolla pääosin ohjelmistosta valmiiksi löytyvistä komponenteista. Kaavojen ja laskujen teossa käytettiin Microsoft Expression Languagea. Datan rajausta, latausta ja muu määrittely tapahtui SQL:n avulla. Sovellus julkaistiin yrityksen sisäiseen verkkoon käyttäen QlikView Serveriä ja siihen liitettyä AccessPoint-portaalia, joka toimii Ajax-tekniikalla.

Työtä pidettiin erittäin onnistuneena, ja sen jatkokehityksestä keskusteltiin. Työ vaatii kuitenkin lisää kehitystä, mutta antoi hyvän pohjan jatkokehitykselle.

ASIASANAT:

Business Intelligence, datan visualisointi, SQL, tietomallit, tietovarastot, QlikView

Kimmo Santamaa

DATA VISUALIZATION IN AN ORGANIZATION

The aim of this thesis was to implement a Proof of Concept data visualization application for TeliaSonera, a telecommunications operator, to evaluate the suitability of QlikView as a tool for visualizing the organization's data. QlikView has been developed by Qlik as a Business Intelligence platform, which is suitable for large enterprises because of its in-memory-based technology capabilities. Large enterprises need to find a suitable solution to heavy data processing. The goal of the Proof of Concept was to validate the different capabilities that QlikView has to offer for data visualization.

The work was carried out with QlikView, mainly using Microsoft Expression Language and the ready-made objects within the software. SQL was used to define what data was to be loaded to the application. The application was released internally within the organization by using QlikView Server and the AccessPoint portal which works with Ajax technology.

The finished application is divided into three different sections. These sections are: Dashboard, Incidents and Personnel. The aim of the sections is to present the defined data in different ways. Dashboard is the first section which was made for the user to see the most important Key Performance Indicator values as clearly and quickly as possible. The monitoring of all incidents happens from incidents section. The purpose of the personnel section is to monitor the effectiveness of the personnel which can be used to allocate service requests without overloading the personnel. One of the important features is to be able to drill-down to a single incident in the raw data level.

The project was considered as a success. The application requires more development but it works well as a base for further development.

KEYWORDS:

Business Intelligence, data modeling, data visualization, data warehousing, SQL, QlikView

SISÄLTÖ

KÄYTETYT LYHENTEET TAI SANASTO	6
1 JOHDANTO	7
2 DATAN KERÄÄMINEN JA VARASTOINTI	9
2.1 Massadata	9
2.2 Tietovarastot	10
2.3 Lähdejärjestelmät	12
2.3.1 Data streaming-prosessi	12
2.3.2 Tietojärvi	12
2.4 ETL-prosessi	13
2.5 Data mart -kerros	14
3 DATA VISUALISOINNIN PERIAATTEET	15
3.1 Business Intelligence	15
3.2 BI- ja Datan visualisointityökalut	16
3.2.1 Tableau-työkalu	16
3.2.2 QlikView-työkalu	17
3.2.3 QlikSense-työkalu	18
3.3 Datan mallinnus	19
3.4 Käyttöliittymän suunnittelu	19
3.4.1 Suunnitteluperiaatteet	20
3.4.2 Kaaviot ja diagrammit	20
3.5 Sidosryhmät	22
3.5.1 Ohjattu analytiikka	22
3.5.2 Itsepalvelu visualisointi	22
3.5.3 Tiedonlouhinta	23
3.6 Raportointi	23
4 PALVELUPYYNTÖJEN VISUALISOINTI	24
4.1 Tietokantayhteyden luominen	24
4.2 Tietomallin rakentaminen	25
4.3 Sovelluksen rakentaminen	28
4.3.1 Dashboard-näkymä	28

4.3.2 Tapahtumat-näkymä	29
4.3.3 Henkilöstö-näkymä	29
4.4 Käyttöönotto ja testaus	30
5 YHTEENVETO	31
LÄHTEET	32

KUVAT

Kuva 1. Massadatan määritelmä (SQLAuthority 2013).	10
Kuva 2. Esimerkki tietovarastosta (Oracle 2016).	11
Kuva 3. Tähtimallin rakenne (IBM 2016).	11
Kuva 4. Gartnerin arvio BI- ja visualisointi työkaluista (Gartner 2016).	16
Kuva 5. Esimerkki Tableaulla tehdystä dashboardista (Technology Advice 2016).	17
Kuva 6. Esimerkki QlikView'illä tehdystä dashboardista (Business Intelligence Singapore 2015).	18
Kuva 7. Esimerkki QlikSensellä tehdystä dashboardista (UBM 2014).	19
Kuva 8. Oikeaoppinen piirasdiagrammi (DCIELTS 2010).	21
Kuva 9. Oikeaoppinen pylvädiagrammi (DCIELTS 2012).	21
Kuva 10. Väärin esitetty viivadiagrammi (ScribbleLive 2016).	22
Kuva 11. Tietokantayhteyden avain.	24
Kuva 12. Tietokantayhteyden luominen.	25
Kuva 13. SQL-kysely tietokannasta.	26
Kuva 14. Tietomallin muokkaaminen.	27
Kuva 15. Tietomallin lopputulos.	27
Kuva 16. Dashboard-näkymä	28
Kuva 17. Tapahtumat-näkymä.	29
Kuva 18. Henkilöstö-näkymä.	30

TAULUKOT

Taulukko 1. Tietovaraston ja tietojärven erot. (KDnuggets 2015.)	13
--	----

KÄYTETYT LYHENTEET TAI SANASTO

BI	Business Intelligence, yrityksen liiketoimintatiedon hallinta
CRM	Customer Relationship Management, asiakastietojen hallinta työkalu
Dashboard	Graafinen näkymä, mihin ollaan listattu kaikki tärkeimmät mitattavat arvot
ETL	Extract, Transform, Load. Prosessi jota käytetään laajasti tiedonhallinnassa
Expression Language	Ohjelmointikieli, jolla rakennetaan kaavoja ja lausekkeita
Jäsennetty data	Tieto joka muutetaan luettavampaan muotoon
Jäsentämätön data	Tieto joka tulee suoraan lähteestä ilman muutoksia
KPI	Key Performance Indicator, liiketoiminnan kannalta tärkeimpiä mittareita
Käyttökate	Liikevaihdosta jäljelle jäävä osuus kiinteiden ja muuttuvien kustannusten jälkeen
POC	Proof of Concept, yleinen termi kun tehdään demo-sovellus konseptin tai uuden teknologian esittelyyn
Relaatiotietokanta	Kokoelma tietokantatauluista jotka ovat yhdistetty toisiinsa avaimilla
Scorecard	Neliönmuotoinen kuvake mihin on asetettu tai laskettu määrättyjä arvoja
SLA	Service Level Agreement, palvelusopimus jonka mukaan palvellaan asiakkaita
SQL	Kyselykieli relaatiotietokannoille

1 JOHDANTO

Viime vuosina informaation kasvu on ollut räjähdysmäistä. Tämä on luonut uusia liiketoimintamahdollisuuksia niin pienille kuin suurillekin yrityksille. Ongelmana on kuitenkin informaation epämuodollisuus ja suuruus. Määrältään kasvavaa informaatiota kutsutaan massadataksi. Massadata on luonut uusia liiketoimintamahdollisuuksia, mutta suurin osa yrityksistä ei ole kuitenkaan vielä osannut hyödyntää kaikkea sitä määrää dataa, joka on saatavilla. (Sanastokeskus 2013.)

McKinsey Global Institute tutki, miten massadatan tulo ja informaation kasvu vaikuttavat viiteen eri sektoriin Yhdysvalloissa ja Euroopassa. Näihin sektoreihin kuuluvat mm. Yhdysvaltojen terveydenhuolto ja julkinen sektori Euroopassa. Voidaan siis helposti sanoa, että massadatan tulo vaikuttaa kaikkiin sektoreihin ympäri maailmaa. Esimerkiksi vähittäiskaupat voisivat hyötyä oikein käytetystä datasta ja pienentää käyttökattoutaan yli 60 prosenttia. McKinseyn mukaan myös Yhdysvaltojen terveydenhuolto voisi tuottaa yli 300 miljardin dollarin verran säästöä joka vuosi, mikäli dataa osataan käyttää tehokkaasti. (McKinsey & Company 2011.)

Tämä opinnäytetyö käy läpi datan keräämisen, varastoimisen ja sen visualisoinnin perusteet. Data on vain hyödyllistä niin pitkään kuin sitä osataan käyttää. Pelkkien numeroiden katsominen ei vain riitä, jonka vuoksi data pitää visualisoida eri menetelmillä. Tämä todettiin myös TeliaSoneran organisaatiossa.

Opinnäytetyön toimeksiantajana toimi TeliaSonera. Työ suoritettiin määrä-aikaisen sopimuksen aikana. TeliaSonera tarjoaa kuluttajille, yhteisöille ja yrityksille tietoliikenne-, viestintä ja televisiopalveluja. Eri palveluja löytyy laaja-alaisesti, ja niiden määrä on myös kasvamassa lähitulevaisuudessa. Vaikka TeliaSonera tunnetaan puhelinliittymistä ja verkkoyhteyksistä, TeliaSonera on laajentumassa ahkerasti uusille liiketoimintalueille. Yhtenä näistä palveluista on esimerkiksi data-palvelut. Suomessa organisaatiota pidetään yhtenä markkinajohtajista ja samoin voidaan puhua koko Pohjoismaiden osalta. Euroopan laajuudella TeliaSonera on kehittynyt viidenneksi suurimmaksi operaattoriksi. Suomessa TeliaSoneralla on noin 3 500 työntekijää. (Sonera 2016.)

Tarve TeliaSoneralle järjestelmien uudistamisen lisäksi oli saada hyödynnettyä olemassa olevaa dataa aikaisempaa paremmin. Suurin osa datasta on määrittelemätöntä tai on kiinni vanhoissa järjestelmissä. Tähän tarvittiin väliin työkalu, joka pystyy luke-

maan erityyppistä ja kokoista dataa mahdollisimman monesta eri lähdejärjestelmästä. QlikView valittiin datan visualisointityökaluksi sen monipuolisten ETL-ominaisuuksien ansioista, joista tullaan käymään läpi tarkemmin myöhemmissä osioissa.

Toteutus opinnäytetyössä tehtiin Proof of Conceptina, jonka tuloksena oli visualisointisovellus TeliaSoneralle. Tarkoituksena oli esitellä QlikView'n eri ominaisuuksia sekä sitä, mihin QlikView BI-alustana suoriutui. Sovellus tehtiin pääosin Microsoft Expression Languageella, sekä QlikView'n sisältä löytyvistä valmiista objekteista. Datan lataus määriteltiin SQL:n avulla. Sovelluksen luonteen takia myös datan määrä rajoitettiin, jonka vuoksi kehitystyö oli nopeampaa. Visualisointiin kiinnitettiin paljon huomiota käytettävyyden ja ulkoasun kannalta. Valmiiseen sovellukseen rakennettiin kolme eri näkymää.

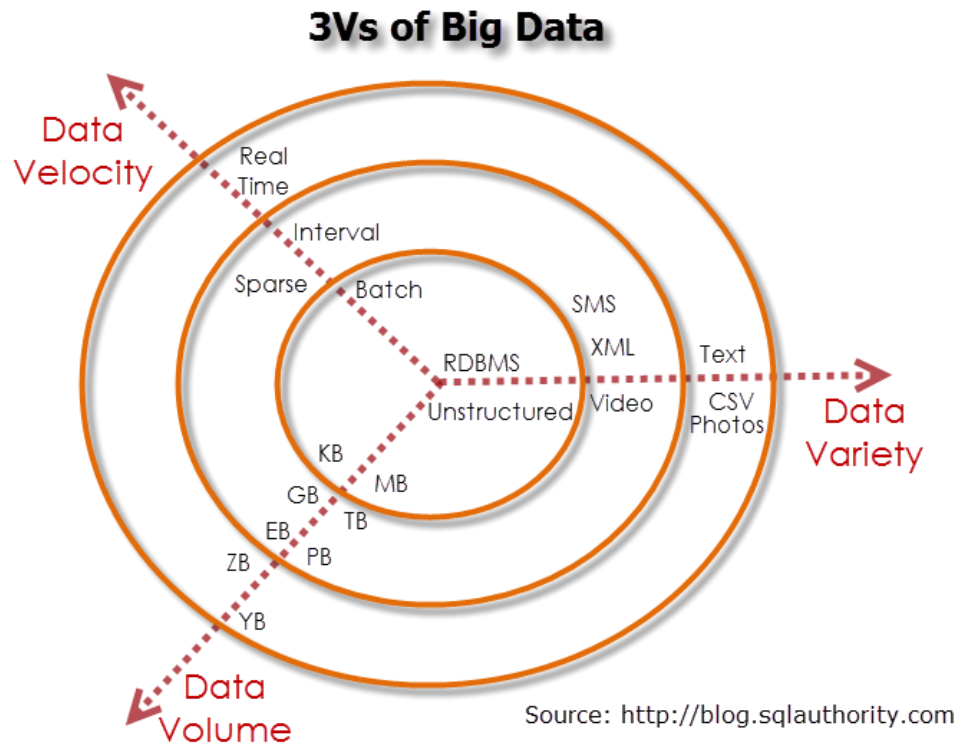
2 DATAN KERÄÄMINEN JA VARASTOINTI

Yrityksillä on jo pitkään ollut tärkeää kerätä omasta liiketoiminnasta muodostuvaa dataa, koska kehittääkseen kilpailukykyään on pystyttävä tekemään entistä tehokkaampia ratkaisuja. Datan kerääminen ja sen varastointi ei ole varsinaisesti uutta, mutta uusien teknologioiden johdosta sitä pystytään nyt hyödyntämään entistä paremmin. NykYTEknologian ansioista voidaan laskea miljoonia rivejä sekunnissa ja tietovarastojen merkittävyys on kasvanut yhtä nopeasti kuin dataliikenteen siirtonopeudet. Tutkimusten mukaan datan määrä tuplaantuu kahden vuoden välein. Tämän vuoksi yritykset ovat uudistamassa omia infrastruktuurejaan. (EMC 2014.)

2.1 Massadata

Massadataksi kuvataan dataa mikä on jäsenettyä tai jäsentämätöntä. Nykypäivänä tietokantaan tallennettava data on kooltaan suurta ja tyypeiltään erilaisia. Tämä luo uusia ongelmia tavallisille tietokantatyypeille. Esimerkiksi moni yritys on osoittanut hylkkyuteensa saada keräämänsä datan käyttöön reaaliajassa, johon heidän vanhat tietokantarakenteet eivät kykene.

Massadata koostuu ns. kolmesta V:stä eli Volume, Velocity ja Variety. Termiä käytetään yleisesti nykypäivänä. Hyödyntääkseen massadataa yritykset ovat siirtyneet tietovarastoihin, joiden on tarkoitus pystyä käsittelemään dataa yhä nopeammin ja tehokkaammin verrattuna aikaisempiin tietokantaratkaisuihin. Kuva 1 näyttää selkeästi, mitä tarkoitetaan massadatan kolmella V:llä. (SQLAuthority 2013.)

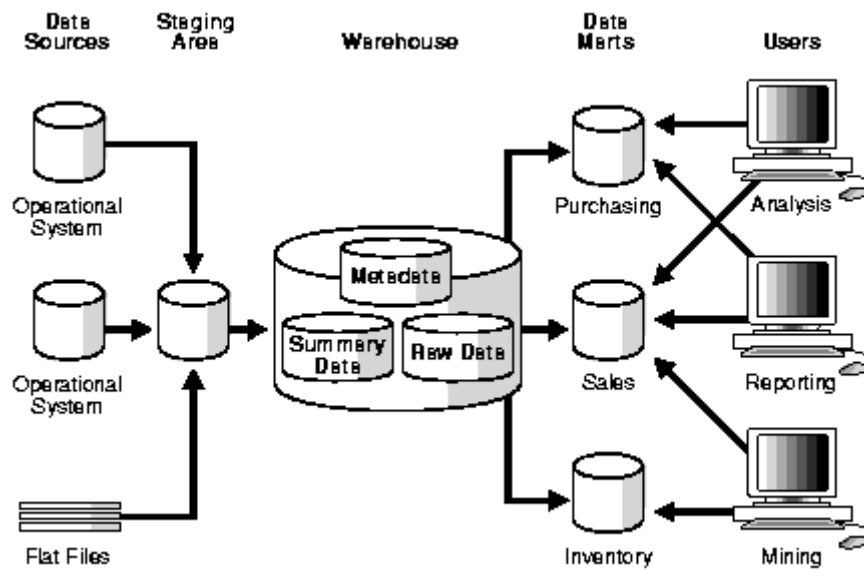


Kuva 1. Massadatan määritelmä (SQLAuthority 2013).

2.2 Tietovarastot

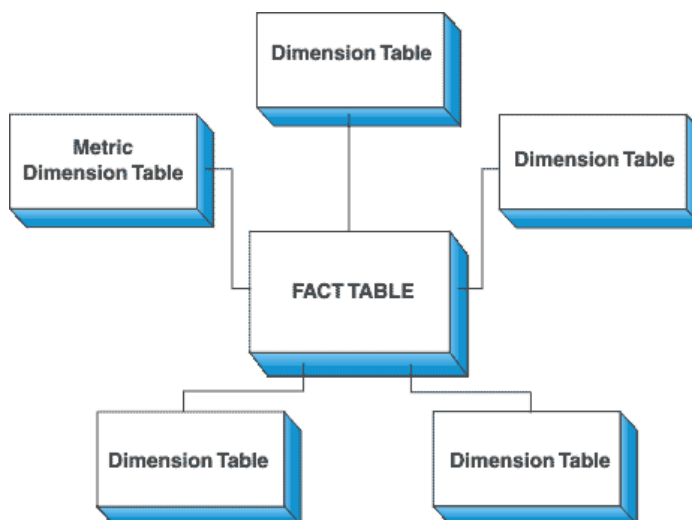
Tietovarasto on relaatiotietokanta, joka on tarkoitettu datan varastointiin. Tietovarastoja ei ole tarkoitettu suoranaisesti datan prosessointiin vaan data varastoidaan ns. ETL-prosessin mukaan ennen tietovarastoon vientiä. Ongelmana on prosessin pituus, joka saattaa viedä useita tunteja riippuen datan koosta ja laadusta. Datan varastointi tapahtuu yleensä yön aikana muun dataliikenteen takia. Prosessin pituuden vuoksi dataa on vaikeaa saada haettua tietovarastosta reaaliajassa.

Datan hyvä muoto tekee analytiikoille tietovarastosta tärkeän järjestelmän. Heillä on helppo työ analysoida valmiiksi hyvin paketoitua dataa. Tietovarastot optimoidaan tiedon hakua varten, jonka pohjana käytetään yleensä tähtimalliarkkitehtuuria. Tietovarastojen arkkitehtuureja on monenlaisia, joten tähtimalli ei ole ainoa ratkaisu. (Oracle 2016.)



Kuva 2. Esimerkki tietovarastosta (Oracle 2016).

Tähtimallia käytetään yleisesti tietovaraston rakenteen mallina, joka koostuu fakta ja dimensiotauluista. Faktataulu sisältää suurimman osan datasta ja siihen kirjataan muuttuvia arvoja esim. myyntitapahtumia. Dimensiotaulut ovat sisällöltään staattista ja tarkentavaa dataa esim. tuotekategoria. Dimensiotaulu liitetään faktatauluun avaimilla. Kuvasta 3 näkee tähtimallin rakenteen. (IBM 2016.)



Kuva 3. Tähtimallin rakenne (IBM 2016).

2.3 Lähdejärjestelmät

Lähdejärjestelmiksi kutsutaan järjestelmiä, jotka keräävät dataa. Dataa voidaan lähettää suoraan tietokantaan tai tietovarastoon. Lähdejärjestelmiä voi olla yrityksen koosta riippuen jopa satoja. Tämmöisissä tapauksissa yritysten käyttöön tarvitaan tietovarasto, joka pystyy käsittelemään suuren dataliikenteen. Esimerkiksi CRM-järjestelmä kerää asiakkaiden tiedot. Yleensä tiedot säilyvät lähdejärjestelmään tarkoitettussa tietokannassa. Tietokanta voi riittää yksittäisiin järjestelmiin, mutta suuremmilla yrityksillä on tarve keskittää omat datansa yhteen paikkaan, jolloin parhaana vaihtoehtona on tietovarasto. (Techopedia 2016.)

2.3.1 Data streaming-prosessi

Data streaming-prosessi käsitettä käytetään kun halutaan käsitellä tai siirtää dataa reaaliaikaisesti suoraan lähdejärjestelmistä. Tämä auttaa yrityksiä tekemään nopeampia päätöksiä. Dataa voidaan analysoida lennossa tai se voidaan tallentaa suoraan tietovarastoon tai tietokantaan. Data voidaan myös siirtää suoraan jäsentämättömänä, jolloin sen prosessointi tehdään myöhemmässä vaiheessa. (Infochimps 2016.)

2.3.2 Tietojärvi

Tietojärveä käytetään myös datan varastointiin. Erona on se, että tietovarastoihin kerätään vain jäsenettyä dataa kun taas tietojärveen tarkoitus on varastoida myös jäsentämätöntä dataa. Tietovarastoon verrattuna etuna voidaan pitää prosessin ketteryyttä. Tiedonlouhijoiden on helppo analysoida dataa ja muotoilla se haluamalla tavallaan kun taas tietovarastoihin data on jo valmiiksi jäsenellyn analyttikkoja varten. Arkkitehtuurissa tietojärveä käytetään yleensä jäsentämättömän datan varastointiin, josta se muotoillaan, siirretään tai analysoidaan suoraan tiedonlouhijoiden puolesta. Taulukossa 1 on kuvattuna tietovaraston ja tietojärven eroja.

Taulukko 1. Tietovaraston ja tietojärven erot. (KDnuggets 2015.)

DATA WAREHOUSE	vs.	DATA LAKE
structured, processed	DATA	structured / semi-structured / unstructured, raw
schema-on-write	PROCESSING	schema-on-read
expensive for large data volumes	STORAGE	designed for low-cost storage
less agile, fixed configuration	AGILITY	highly agile, configure and reconfigure as needed
mature	SECURITY	maturing
business professionals	USERS	data scientists et. al.

2.4 ETL-prosessi

ETL-prosessilla tarkoitetaan datan siirtoa toisesta tietokannasta toiseen. Tämän aikana data voidaan siirtää suoraan sellaisenaan kuin se on tai sen muotoa voidaan muuttaa toisenlaiseksi. ETL-työkaluja on monenlaisia ja suurimmalla osalla pystyy suorittamaan samat toiminnot. Tyypillisenä esimerkkinä voidaan pitää datan arvojen tai tyyppien muunnoksia. Eri lähdejärjestelmissä käytetään eri formaatteja, joten datan muuntaminen on isossa roolissa tietovaraston kannalta. ETL-prosessin vaiheet ovat seuraavat:

1. Extract – Data haetaan lähteestä ja siirretään esimerkiksi väliaikaiseen tietokantaan tai välimuistiin muunnosta varten.
2. Transform – Tässä vaiheessa data muotoillaan sellaiseen muotoon kuin se halutaan. Yhtenä esimerkkinä on desimaalien poistaminen. Mitä vähemmän numeroissa on desimaaleja, sen vähemmän tilaa se vie varastoinnin kannalta.
3. Load – Data ladataan tietovarastoon. Tietovarastoissa ETL on yleinen käytäntö, koska tietovarastoihin halutaan varastoida dataa mikä on muunneltu mahdollisimman hyvään muotoon jatkokäsittelyä varten. (Webopedia 2016).

2.5 Data mart -kerros

Data mart toimii seuraavana kerroksena tietovaraston arkkitehtuurissa. Kun tietovarastoja on vain yksi, data martteja voi olla kymmeniä tai satoja. Data martit ovat pienempiä kokonaisuuksia, joiden tarkoitus on jaotella data esimerkiksi osaston, kategorian tai liiketoiminnan osa-alueen mukaisesti. Data on selkeästi pakattu, rajattu ja data martista saadaan ladattua nopeasti data vähäisellä vaivalla analyysia varten. Tämän kerroksen jälkeen tulevat käyttäjät, raportit ja datan visualisointityökalut. (nModal Solutions Inc. 2008.)

3 DATA VISUALISOINNIN PERIAATTEET

Tietovarastoarkkitehtuurissa viimeisenä kerroksena ovat käyttäjät, raportointi ja datan visualisointi. Data visualisointi osuuden tarkoitus on selventää millaisia käyttäjiä, raportointi ja datan visualisointityökaluja on saatavilla, sekä data visualisoinnin yleisiä käytäntöjä. Datan visualisoinnissa on tärkeää että datan laatu on korkea ja sitä ei ole muunnettu lähteestä väärin. Jos datassa on virheitä, visualisoinnista on helppo tehdä vääriä johtopäätöksiä. Laadunvarmistus on erittäin tärkeä osa tätä prosessia. Datan visualisoinnin tarkoitus on auttaa yrityksen päätöksenteossa, kilpailukyvyn ylläpidossa ja johtamisessa.

3.1 Business Intelligence

BI eli Business Intelligencellä tarkoitetaan yrityksen liiketoimintatietojen hallintaa, jonka tarkoitus on auttaa yrityksen eri sidosryhmiä sekä käyttäjiä tekemään liiketoimintapäätöksiä. Tämän kautta saadaan myös tarkempi kuva mihin suuntaan yrityksen liiketoiminta on menossa tai voidaan löytää syitä miksi liiketoiminta ei ole mennyt suunnitellulla tavalla. Muita hyötyjä ovat esimerkiksi prosessien parantaminen, uusien liiketoimintamahdollisuuksien löytäminen, kilpailukyvyn ylläpitäminen tai parantaminen. Yrityksillä on monia syitä osallistua tiedonhallintaan liittyviin projekteihin, koska niistä saadaan liiketoiminnalle elintärkeitä tietoja. (TechTarget 2014.)

BI-ratkaisujen monipuolisuuden vuoksi jokaisen yrityksen liiketoimintatiedonkäsittely tehdään eri tavalla. Tietovarastot suunnitellaan aina yrityksen infrastruktuurin ja tarpeiden mukaisesti. Jokaisen yrityksen tarvitsee käydä läpi tietosuoja niin oman tarpeiden kuin lain mukaisesti. Esimerkiksi puheluista ja tekstiviesteistä saadun datan käytön suhteen Suomen valtiolla on erittäin tarkat säännöt. Tätä kutsutaan teletunnistetiedoksi. (FiCom 2016.)

Liiketoiminnasta kerättyä dataa voidaan analysoida eri tavoilla riippuen ketä tai kenelle data esitetään, sekä millä tavalla. BI-ratkaisuja löytyy nykyään jopa kymmeniä. Jokaisella työkalulla on omat vahvuutensa ja heikkoutensa. Johtoryhmän tulee tehdä päätös yrityksen datan käytön suhteen, jonka mukaan yritys valitsee sopivimman työkalun heidän käyttöönsä. Työkalun valintaa tehdessä on tärkeää että yritys vastaa seuraaviin kysymyksiin:

1. Millaista dataa keräämme?
2. Millainen infrastruktuuri meillä on?
3. Miten analysoimme dataa?
4. Mitkä ovat meidän omat vahvuutemme ja kompetenssit?
5. Vastaako tämä ratkaisu kaikkia meidän tarpeita?

3.2 BI- ja Datan visualisointityökalut

Datan visualisointiin on monia eri lähestymistapoja. Työkalun valitsemisessa täytyy selvittää millaisia käyttäjiä yrityksen sisältä löytyy, jonka jälkeen käyttäjät jaotellaan lähestymistavoiltaan eri sidosryhmiin. Kuvassa 4 Gartner on listannut yleisimmät BI- ja visualisointityökalut. Jokaisella työkalulla on omat vahvuutensa eri sidosryhmiä kohtaan.



Kuva 4. Gartnerin arvio BI- ja visualisointi työkaluista (Gartner 2016).

3.2.1 Tableau-työkalu

Tableaun suurin vahvuus on työkalun selkeys ja helppokäyttöisyys. Käyttöliittymään tottuu nopeasti ja se ei vaadi suurempaa kehittäjä taustaa. Tämän ansiosta työkalu

sopii täydellisesti analyytikoille. Yhteisö auttaa ongelmatilanteissa nopeasti, joka on kehittäjille tärkeää. Tableau keskittyy itsepalvelu tyyppiseen analyytiikkaan.

Valmiina pakettina Tableau sopeutuu monen yrityksen tarpeisiin ja se tarjoaa paljon valmiita ajureita tietokantayhteyksiä varten. Intuitiivisia ja monipuolisia kaavoja on tarjolla heti käyttöönoton jälkeen. Huonoina ominaisuuksina on varsinaisen ETL-toiminnon puuttuminen ja datan muotoilun vaikeus. Laskentateho on kuitenkin nopea ja tehokas. Tableaun heikkoutena on varsinaisen raportointiominaisuuden puuttuminen.



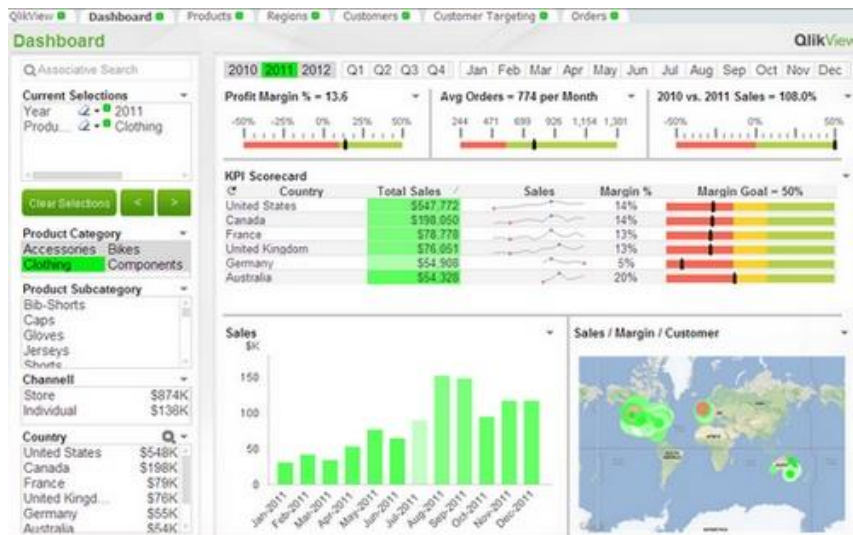
Kuva 5. Esimerkki Tableauilla tehdystä dashboardista (Technology Advice 2016).

3.2.2 QlikView-työkalu

Qlikview:ssä ei ole kovin paljon valmiiksi rakennettuja kaavoja, mutta lisä-osia löytyy moneen eri tarpeeseen. Tämän ansiosta QlikView toimii kokonaisuutena monipuolisempänä työkaluna kuin Tableau. Käyttöönotto yrityksen tasolla on helppoa ja tarjolla on monia demosovelluksia, koulutuksia ja tutoriaaleja.

Huonomman käyttöliittymän ja ylimääräisten ominaisuuksien takia QlikView saattaa antaa sekavan kuvan heti alussa verrattuna Tableauhun. Kuitenkin pienen käytön jälkeen QlikView'n hyvät ominaisuudet tulevat esille. Välimuistissa tapahtuvien prosessien käsittely auttaa sovelluksen sisällä tapahtuvien laskujen laskennassa, sekä suurien tiedostojen käsittelyssä. Toisena tärkeänä ominaisuutena on datan käsittely.

QlikView osaa lennosta määrittellä, mitkä datat ovat yhteydessä toisiinsa. Tämä luo uusia näkökulmia datan analysointiin. QlikView'stä löytyy myös ETL-ominaisuus, joka tuo lisää monia eri vaihtoehtoja analyysien tekemiseen. Vaikka QlikView on monipuolinen työkalu, QlikView-sovellukset vaativat paljon kehitystyötä. Suurin painopiste on siinä, että kehittäjät rakentavat sovelluksia loppukäyttäjille ja ne ovat räätälöityjä käyttäjien tarpeiden mukaan.

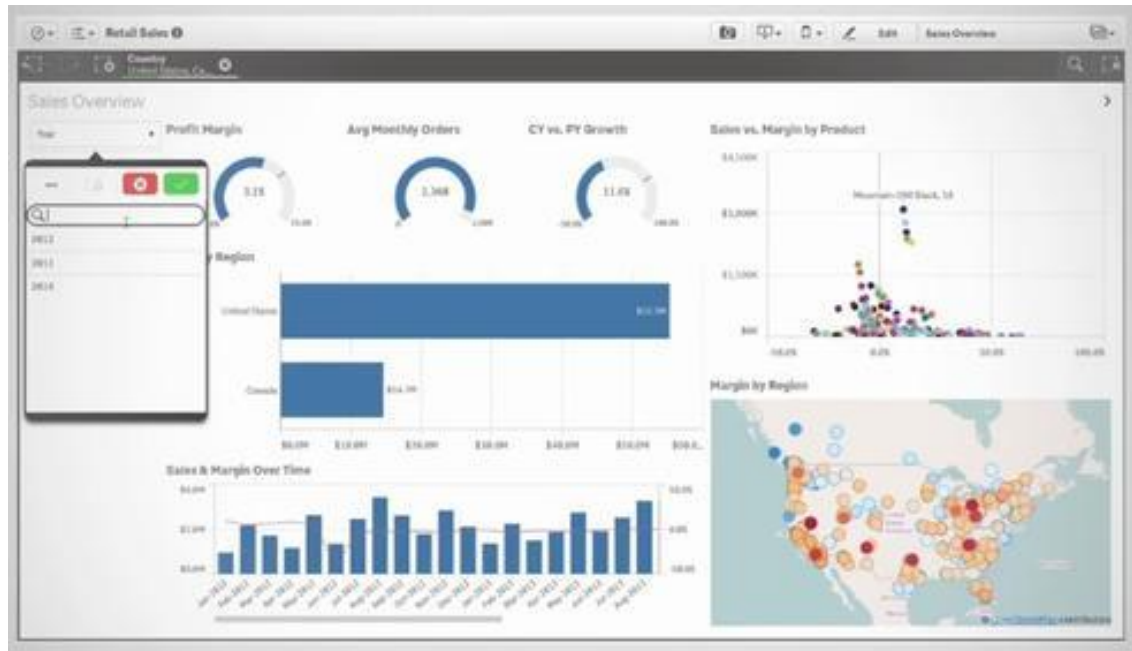


Kuva 6. Esimerkki QlikView'illä tehdystä dashboardista (Business Intelligence Singapore 2015).

3.2.3 QlikSense-työkalu

QlikSense yrittää seurata Tableaun jalanjälkiä. Se kehitettiin alun perin korjaamaan QlikView:ssä todettuja ongelmia, kuten responsiivisuutta ja sekavuutta. QlikView-sovelluksia on hankala kääntää eri laitteille, ja sovellukset täytyvät suunnitella aina erikseen eri resoluutioille. Sengen käyttöliittymä on paranneltu, ja se on helppo oppia. Laskennan tehokkuus ja muut ominaisuudet on pyritty pitämään entisellään, ja Sense toimiikin hyvin QlikView'n ohella.

Isoimpana ongelmana Sengen kannalta on se, että työkalun kehitys on vielä alkuvaiheessa. Tämän vuoksi monia ominaisuuksia ei ole vielä tehty valmiiksi. Tableau tarjoaa ainakin toistaiseksi paremman alustan itsepalveluanalytiikan puolella, kun taas QlikView on parempi ohjatun analytiikan ja raportoinnin puolella. Kokonaisuudessa Qlik tarjoaa paremman BI- ja datan visualisointi mahdollisuudet yrityksille.



Kuva 7. Esimerkki QlikSensellä tehdystä dashboardista (UBM 2014).

3.3 Datan mallinnus

Kuten tietokannan rakentamisessa, ensimmäinen vaihe datan visualisoinnissa on datan mallinnus. Yhdessä mallissa saattaa olla useita tauluja. Mallin avulla kehittäjät tietävät mikä data kuuluu mihinkin ja miten taulut ovat yhdistetty toisiinsa. Vaikka ETL-prosessi suoritetaan ennen datan siirtoa tietovarastoon, voidaan myös datan visualisoinnissa joutua muokkaamaan data parempaan muotoon. Sovellusta on mahdoton rakentaa ilman tietämättä miten data on yhdistetty toisiinsa, varsinkin kun kehittäjät ja analyytikot haluavat yhdistää useita eri data-lähteitä yhteen sovellukseen. (Webopedia 2016.)

3.4 Käyttöliittymän suunnittelu

Riippuen sidosryhmästä käyttöliittymällä saattaa olla suuri merkitys. Jos kehitetään sovellusta loppukäyttäjälle, tulee käydä läpi parhaat käytännöt käyttöliittymän suunnittelun osalta. Mikäli datan analysointia tehdään itsenäisesti, voidaan tehdä sovelluksesta oman tarpeen mukainen. On kuitenkin hyvä ottaa huomioon käytettävyyden periaatteet käyttöliittymää suunniteltaessa. (Usabilitygeek, 2013.)

3.4.1 Suunnitteluperiaatteet

BI- ja datan visualisointi-sovelluksia kehittäessä tärkeimpänä sääntönä on järjestelmällisyys ja selkeys. Ihmisen aivot voivat muistaa vain tietyn määrän asioita kerrallaan, joten on suositeltavaa näyttää maksimissaan 5-7 kaavaa tai diagrammia yhtäaikaisesti. Yhdellä välilehdellä ei saa missään tapauksessa olla liikaa tietoa. Yksinkertaisuus on hyvä muistaa sovelluskehityksessä, varsinkin jos kehitetään sovellusta loppukäyttäjälle. (Livescience, 2008.)

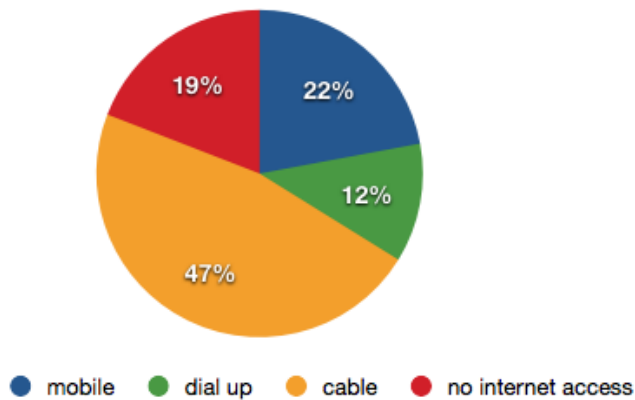
Yleisesti ihmiset lukevat vasemmalta oikealle. Tämän vuoksi suodattimet ja muut valinnat olisi hyvä jättää ruudun vasemmalle puolelle tai ylälaitaan. Tämä luo helppouden ymmärtää sovellusta kun käyttäjä katsoo sovellusta ensimmäistä kertaa. Samanlainen lajittelu ja sommittelu tulisi pitää kaikissa eri välilehdissä. Yhdessä sovelluksessa voi mahdollisesti olla kymmeniä eri välilehtiä, jotka kaikki analysoivat dataa eri näkökulmista. Yhtenäisyys on tärkeää jotta vältetään sekaannusta välilehdeltä toiselle siirtyessä. Tämä tekee sovelluksesta helposti seurattavan ulkopuolisen silmistä. (99designs, 2014.)

Värien osalta seurataan myös samoja periaatteita. Sovelluksessa ei tulisi käyttää liikaa eri värejä ja ne täytyvät pysyä samoina jokaisella välilehdellä. Värisokeus saattaa olla ongelmana joillain käyttäjille, joten punainen ja vihreä ei ole välttämättä aina paras väri kuvaamaan hyviä tai välttäviä arvoja. Tässä tapauksessa voidaan käyttää eri sävyjä kuten punaisen sijaan oranssia. Toinen tapa on vaihtaa tekstin kokoa. Tällöin arvot joita halutaan käyttäjien huomaavan paremmin, muutetaan isommaksi kuin muut arvot. Kolmas tapa on näyttää nousevia ja laskevia trendejä nuolilla. Näitä käytäntöjä tulee ottaa huomioon sovellusta suunnitellessa. (Usability 2016.)

3.4.2 Kaaviot ja diagrammit

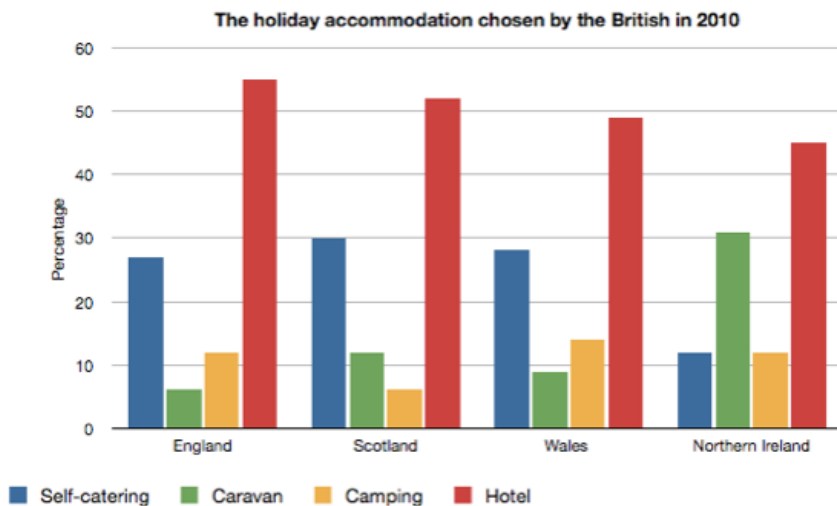
Käyttöliittymää suunniteltaessa tulee huomioida myös se, millaisilla kaavoilla ja diagrammeilla data esitetään, sekä se, että niillä on myös käytäntöjä, joita tulee ottaa huomioon. Prosenttiosuudet on helpoin näyttää piirasdiagrammin avulla. Tässä kannattaa pitää mielessä rajoituksena se, että enempää kuin 7 eri arvoa ei tule näyttää samanaikaisesti. Ihmisen on tämän jälkeen vaikea erottaa osuuksia toisistaan. (Scribble-Live 2016.)

How the British accessed the internet in 2010



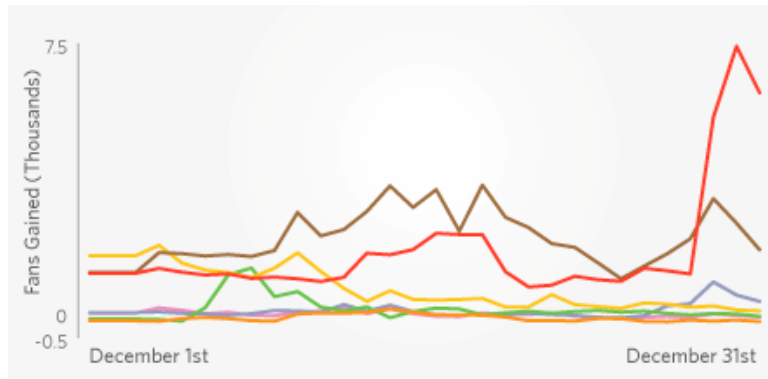
Kuva 8. Oikeaoppinen piirasdiagrammi (DCIELTS 2010).

Vertailuarvoihin käytetään yleensä pylväsdiagrammia. Se on helppo tapa erottaa mitkä ovat suurimmat ja pienimmät arvot. Oikeaoppisessa pylväsdiagrammissa dimensioiden arvoja kannattaa rajata riippuen diagrammin koosta. Mikäli arvoja on liikaa kokoon nähden, pylväät muuttuvat liian pieniksi ja silloin on vaikea erottaa arvoja toisistaan. Kuvassa 9 on esitetty pylväsdiagrammi oikeaoppisesti. (ScribbleLive 2016.)



Kuva 9. Oikeaoppinen pylväsdiagrammi (DCIELTS 2012).

Viivadiagrammia suunniteltaessa tulee ottaa huomioon samat käytännöt kuin pylväsdiagrammiin. Liiallinen määrä viivoja tekee diagrammista sekavan. Myös värit menevät sekaisin ja niitä on vaikea erottaa toisistaan, mikäli arvoja on liikaa. Kuvassa 10 on esitetty viivadiagrammi väärin.



Kuva 10. Väärin esitetty viivadiagrammi (ScribbleLive 2016).

3.5 Sidosryhmät

Sidosryhmät sisältävät erityyppisiä käyttäjiä ja näitä voivat olla niin tavalliset työntekijät, ulkoiset asiakkaat tai johtohenkilökunta. Sidosryhmät lajitellaan kolmeen eri kategoriiaan heidän analytiikka tarpeiden tai osaamisen mukaisesti. Nämä kolme kategoriata ovat ohjattu analytiikka, itsepalvelu visualisointi ja tiedonlouhinta. Sovellukset rakennetaan käyttäjien tarpeiden mukaan sopivalla tavalla.

3.5.1 Ohjattu analytiikka

Ohjatulla analytiikalla tarkoitetaan BI-sovelluksia, jotka kehittäjät tekevät loppukäyttäjille. Nämä sovellukset ovat suunnattu käyttäjille, jotka omaavat vähäistä kokemusta sovelluskehityksestä. Sovellus keskittyy tavallisesti yhteen liiketoiminta-alueeseen, jota käytetään koko osaston toimesta. On hyvin yleistä että osaston johtajat käyttävät sovellusta liiketoiminta-alueen kannattavuuden ja toiminnan seurantaan. (Qgate, 2015.)

3.5.2 Itsepalvelu visualisointi

Itsepalvelu visualisoinnilla tarkoitetaan kun käyttäjät ja analyttikot tekevät visualisointeja itsenäisesti. Tämä vaatii kuitenkin työkalun, joka on helppokäyttöinen ja sopii yrityksen analytiikka tarpeisiin. Tällöiset työkalut vaativat vähäistä kokemusta sovelluskehityksestä. Toiminnollisuudet rakennetaan yleensä työkalusta valmiiksi löytyvillä objekteilla. (Qgate, 2015.)

3.5.3 Tiedonlouhinta

Tiedonlouhijat suorittavat omia analyysejä suoraan tietojärvestä tai tietovarastosta ja käsittelevät suuria määriä dataa yhtäaikaisesti. Tämän vuoksi jokainen BI-työkalu ei välttämättä sovi heidän tarkoituksiinsa. Yleensä liiketoiminnan eri osastot antavat tiedonlouhijoille kysymyksiä, joihin he etsivät vastauksen datasta. Kysymykset ovat tärkeitä ja kiireellisiä liiketoiminnan sujuvuuden kannalta, ja ovat luonteeltaan sellaisia joihin pelkkien analyytikkojen on vaikea etsiä vastauksia. (Techopedia, 2016.)

3.6 Raportointi

Yrityksen johdon, ulkoisten asiakkaiden ja eri osastojen ajan tasalla pitämiseen käytetään raportointia. Liiketoiminnan eri osastot raportoivat ylemmälle johdolle tapahtumista ja mihin suuntaan liiketoiminta on menossa. Raportointia voidaan tehdä myös ulkoisille asiakkaille. Tämä on yleensä tehty Excelin avulla, mutta nykypäiväiset BI-ratkaisut kykenevät automatisoimaan raportoinnin. Data saadaan visualisoitua suoraan raporteille, ja raportit voidaan lähettää eteenpäin päivittäin, viikoittain tai kuukausittain.

Qlikillä on oma raportointityökalu nimeltään Qlik NPrinting, joka voidaan ajastaa hakemaan automaattisesti halutut objektit QlikView-sovelluksista. Jokaiselle objektille on annettu tunniste, joka voidaan määrittellä NPrinting-sovellukseen. Raportin luominen menee seuraavan prosessin mukaisesti:

1. Valitaan mistä QlikView-sovelluksesta tehdään raportti.
2. Määritellään tunnisteiden perusteella objektit raportille.
3. Sommitellaan raportin ulkoasu.
4. Valitaan kuinka usein raportti lähetetään.
5. Valitaan kenelle raportti lähetetään sähköpostiosoitteiden perusteella.
6. Valitaan tiedostotyyppi esim. pdf tai xls.

(Qlik, 2016.)

4 PALVELUPYYNTÖJEN VISUALISOINTI

Palvelupyyntöjen seurantaan käytetään OIVA:ksi nimettyä järjestelmää, johon asiakasyritykset pystyvät kirjaamaan erityyppisiä palvelupyyntöjä. Tässä työssä pyynnöt rajattiin ”incident” tapahtumiin. Datan visualisointiin valittiin QlikView. Rakennetun sovelluksen tarkoituksena oli seurata että TeliaSonera pystyy palvelemaan asiakasyrityksiä SLA eli Service Level Agreement arvojen puitteissa. Arvojen ylittyessä QlikView-sovelluksen kautta tulee pystyä porautumaan yksittäisistä tapahtumiin aina raakadata tasolle asti, jotta TeliaSonera pystyy selvittämään syyn miksi SLA arvot mahdollisesti ylittyivät. QlikView-sovellus yhdistettiin OIVA:n tietokantaa, josta rakennettiin analyyttinen sovellus palvelupyyntöjen seurantaan.

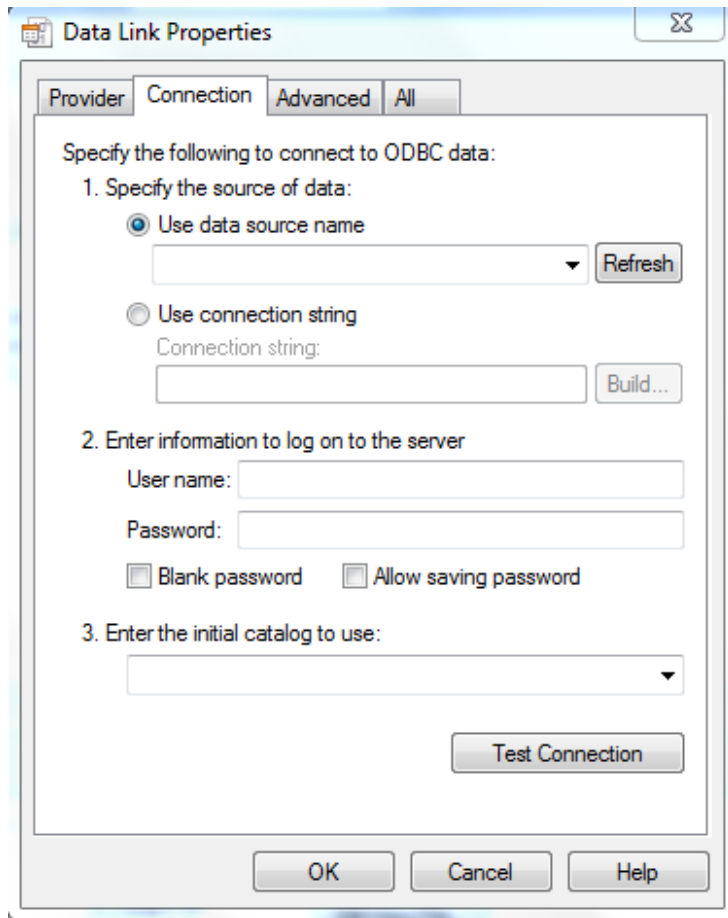
4.1 Tietokantayhteyden luominen

QlikView tukee ODBC eli Open DataBase Connectivity -ajuria, jonka avulla QlikView-sovelluksesta voidaan luoda suora yhteys kaikkiin yleisimpiin tietokantoihin. ODBC toimii avoimena rajapintana, joka mahdollistaa yhteyden luomisen eri tietolähteisiin. Kun tietokantayhteys on luoto niin QlikView luo sille oman avaimen. (Microsoft, 2016.)

```
12 | //Connection Strings  
13 | ODBC CONNECT TO [OIVA;DBQ=OIVA] (XUserId is RJHRYIRJJJaSMVEVIXSQB, XPassword is ONHIVVRNJLMKT...
```

Kuva 11. Tietokantayhteyden avain.

Kuvassa 12 näkyy ikkuna, minkä kautta luodaan yhteys tietokantaan. Yhteyden luomiseksi määritettiin tietokannan osoite ja se, millä tunnuksilla sitä luetaan. Tämän jälkeen voidaan hakea tarvittavat taulut tietokannasta SQL-kyselyjen avulla.



Kuva 12. Tietokantayhteyden luominen.

4.2 Tietomallin rakentaminen

Parhaimpina käytäntönä QlikView:ssä pidetään sitä, että dataa ei suoraan ladata sovellukseen tietokannasta vaan tallennetaan ensiksi erilliseen QVD-tiedostoon. QVD tarkoittaa QlikView Data -formaattia, joka on optimoitu QlikView'ille. Tämän ansiosta QlikView pystyy lukemaan suuria määriä dataa huomattavasti nopeammin ja kevyemmin. QVD-kerrosta käytetään tietomallin muokkaamiseen lopullista sovellusta varten.

Käyttäjille tarkoitettu sovellus hakee datan lopullisesta QVD-tiedostosta. Sovellusten ja datan päivittäminen tapahtuu ajastetusti QlikView Management Consolen kautta käyttäen QlikView Serveriä. Mikäli tietokanta on laajassa käytössä ja sisältää suuria määriä dataa, ajastus kannattaa tehdä yön yli inkrementaalisesti. Inkrementaalinen lataus tarkoittaa että ladataan vain uudet tai muuttuneet tiedot ja lisätään ne alkuperäiseen QVD-tiedostoon. Kun lataus tehdään yön yli, se ei vaikuta päivittäisten käyttäjien toi-

mintaan. OIVA:n kannalta ei ollut tärkeää saada dataa reaaliajassa, joten datan lataus tehtiin öisin. Kuvassa 13 on pieni osa SQL-kyselystä, jota käytettiin datan latauksessa.

```

21 //Load only modified and new records from Oiva
22 OIVA:
23 SQL SELECT "ACCOUNTING_CODE",
24           //"ARRIVAL_TIME",
25           TO_DATE('1970-01-01','YYYY-MM-DD') + arrival_time / 86400 + 2/24 As "ARRIVAL_TIME",
26           "ASSIGNED_TO_GROUP_",
27           "ASSIGNED_TO_INDIVIDUAL_",
28           "CASE_ID_",
29           "CASE_TYPE",
30           "CATEGORY",
31           "CONFIRM_RESOLVED_CASE",
32           "MODIFIED_TIME",
33           //"SERVICE_RESTORED",
34           TO_DATE('1970-01-01','YYYY-MM-DD') + service_restored / 86400 + 2/24 as "SERVICE_RESTORED",
35           SITE
36
37 FROM AR."HELPDESK" where modified_time > $(LastModified);
38 Store OIVA_NEW into 'C:\abc\OIVA_new.qvd';
39
40 drop table OIVA_NEW;
41 //Exit script;

```

Kuva 13. SQL-kysely tietokannasta.

SQL-kyselyn jälkeen haettu data tallennetaan seuraavasti:

1. SQL-kyselyn jälkeen data tallennetaan Store-komennolla QVD-tiedostoon.
2. Taulukko pudotetaan pois Drop Table -komennolla, koska QlikView suorittaa komennot muistissa. Tämä vaikuttaa latauksessa käytettävän tiedoston kokoon.

Seuraavassa tietomallin rakentamisvaiheessa muokataan QVD-tiedostoon tallennettua dataa sellaiseen muotoon, että siitä on helppo tehdä laskemia kaavoihin ja diagrammeihin. Pitää myös huomioida että laskut ja muunnokset menevät oikein. Datan laatua on tärkeää mitata jokaisen latauksen jälkeen, jotta säästytään virheilä. Kuvassa 14 on yksinkertainen esimerkki tietomallin muokkaamisesta.

```

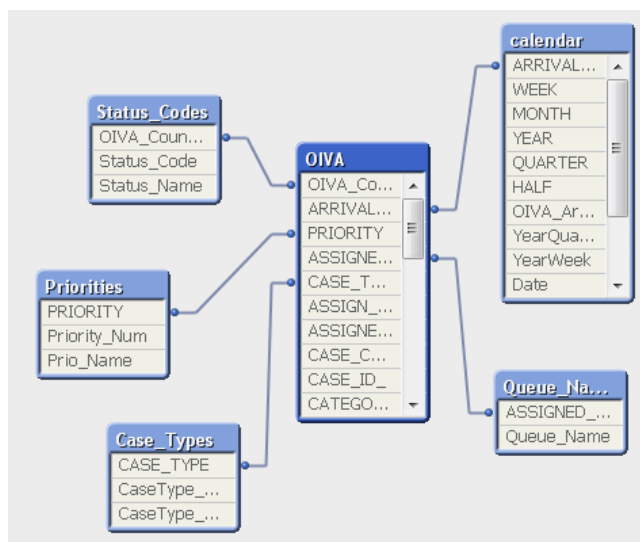
1 OIVA_NEW:
2 LOAD
3 // ACCOUNTING_CODE,
4 ARRIVAL_TIME,
5 ASSIGNED_TO_GROUP_,
6 ASSIGNED_TO_INDIVIDUAL_,
7 CASE_ID_,
8 CASE_TYPE,
9 CATEGORY,
10 CONFIRM_RESOLVED_CASE,
11 If(IsNull(SERVICE_RESTORED), 1, 0)As OIVA_Count_Open,
12 If(IsNull(SERVICE_RESTORED), 0, 1)As OIVA_Count_Resolved,
13 If(IsNull(SERVICE_RESTORED), Interval(Today() - ARRIVAL_TIME, 'D'), 0) As OIVA_Open_Time,
14 SITE
15
16 FROM
17 C:\abc\OIVA_new.qvd(qvd)
18 Where
19 CASE_TYPE = 'Incident';
20
21 STORE OIVA_NEW into 'C:\abc\OIVA_all.qvd';
22
23 Exit script;

```

Kuva 14. Tietomallin muokkaaminen.

1. Ladataan valitut kentät.
2. Suoritetaan laskuja tai muutoksia kenttiin.
3. Palvelutapahtumat rajattiin pelkkiin incidentteihin käyttämällä Where -komentoa.
4. Tallennetaan uuteen QVD-tiedostoon.

Kaikkien muutosten ja laskujen jälkeen tietomalli muokattiin tähtimallin mukaiseksi. Kuvassa 15 on lopullinen tulos tietomallista.



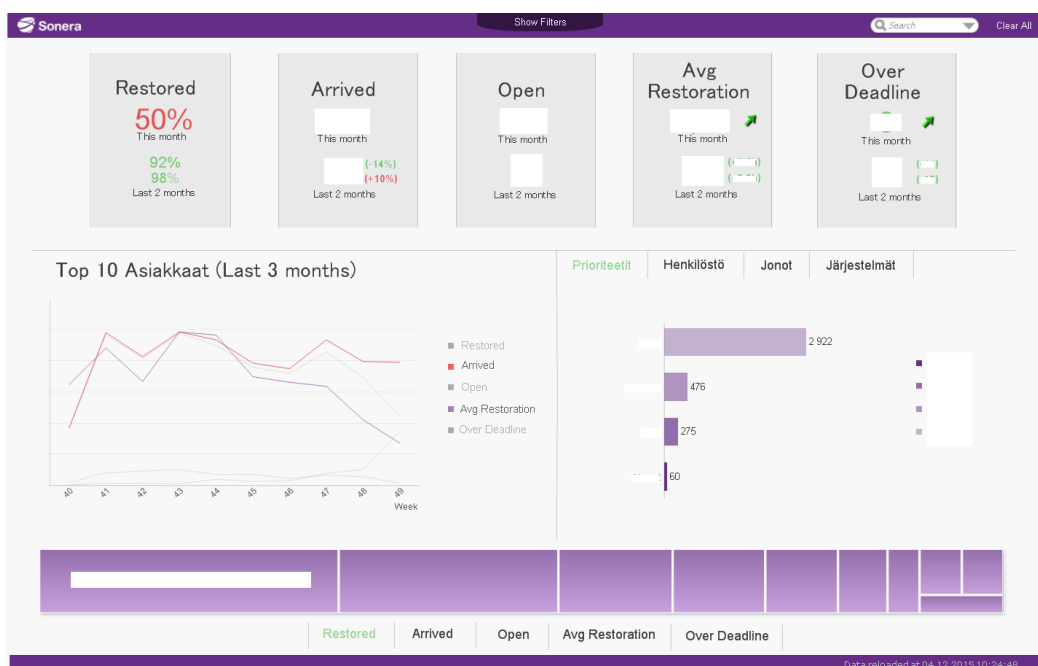
Kuva 15. Tietomallin lopputulos.

4.3 Sovelluksen rakentaminen

Sovelluksen rakentaminen aloitettiin määrittelemällä loppukäyttäjien tarpeet. Projektin edetessä sovellusta esitettiin eri tahoille, ja kehitystä jatkettiin iteratiivisesti palautteen perusteella, sekä lisäämällä uusia ominaisuuksia. Lopulliseen sovellukseen saatiin valmiiksi kolme eri näkymää, jotka ovat dashboard, tapahtumat ja henkilöstö. Jokaisella näkymällä on eri tarkoitus vaikka niissä on käytetty paljon samoja menetelmiä. Sovellus tehtiin Proof of Conceptina, jonka tarkoitus oli tuoda esille QlikView'n eri ominaisuuksia.

4.3.1 Dashboard-näkymä

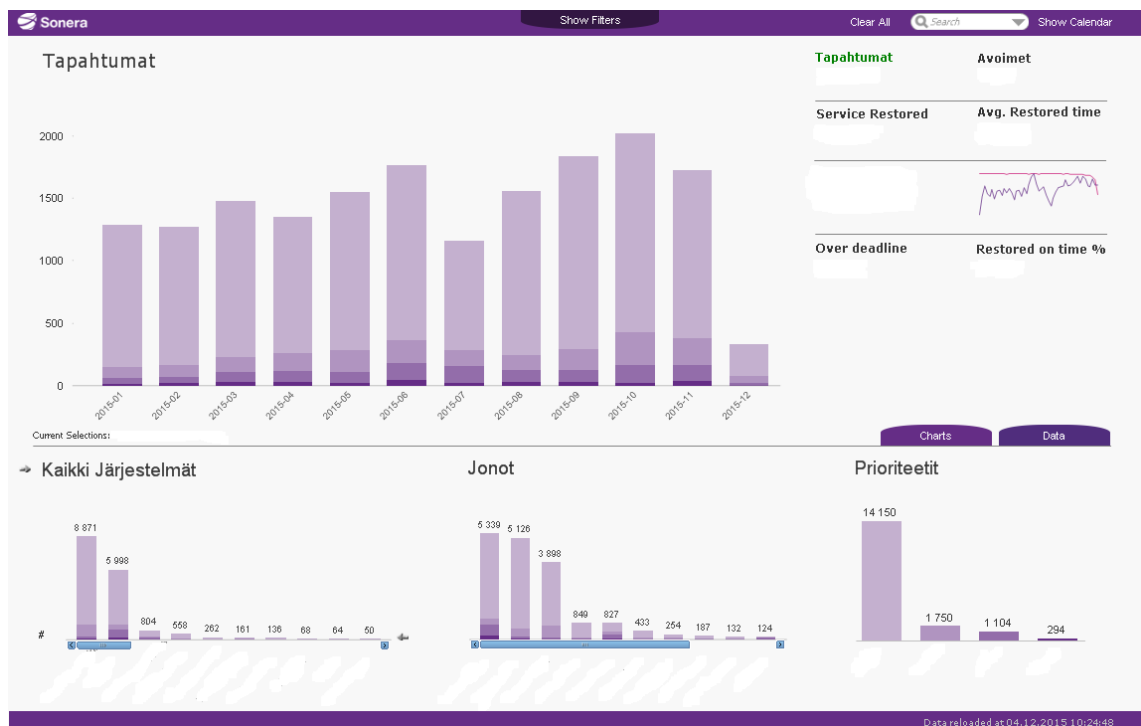
Dashboard-näkymä antaa nopean katsauksen edellisen kahden kuukauden tapahtumiin. Scorecardit tehtiin pääkäyttäjien määritelmien mukaisesti esittämään tärkeimpiä KPI eli Key Performance Indikaattoreita. Näistä pääkäyttäjät saavat nopeasti tiedon, mihin suuntaan kirjatut palvelutapahtumat ovat menossa. Näkymässä keskityttiin viiteen tärkeimpään arvoon, ja ne rajattiin edelliseen kuukauteen.



Kuva 16. Dashboard-näkymä

4.3.2 Tapahtumat-näkymä

Tapahtumat-näkymästä sovelluksen pääkäyttäjät pystyvät seuraamaan kaikkia saapuneita palvelupyynnöitä, ja suorittamaan analyysejä nopeasti. Kaavoihin porautuminen tapahtuu painamalla pylväsiagrammien pylväitä tai valitsemalla suodattimia sovelluksen ylälaudasta. Sovelluksen alalaidassa on myös Data-painike, joka avaa taulukon valittuihin arvoihin raakadataan asti. Näin pääkäyttäjät pystyvät näkemään palvelupyynnöt tunnisteiden avulla suoraan valittujen arvojen perusteella. Näkymän yläkulmasta saadaan avattua kalenteri, jonka avulla käyttäjät voivat valita halutun aikajakson tapahtumien seurannalle.

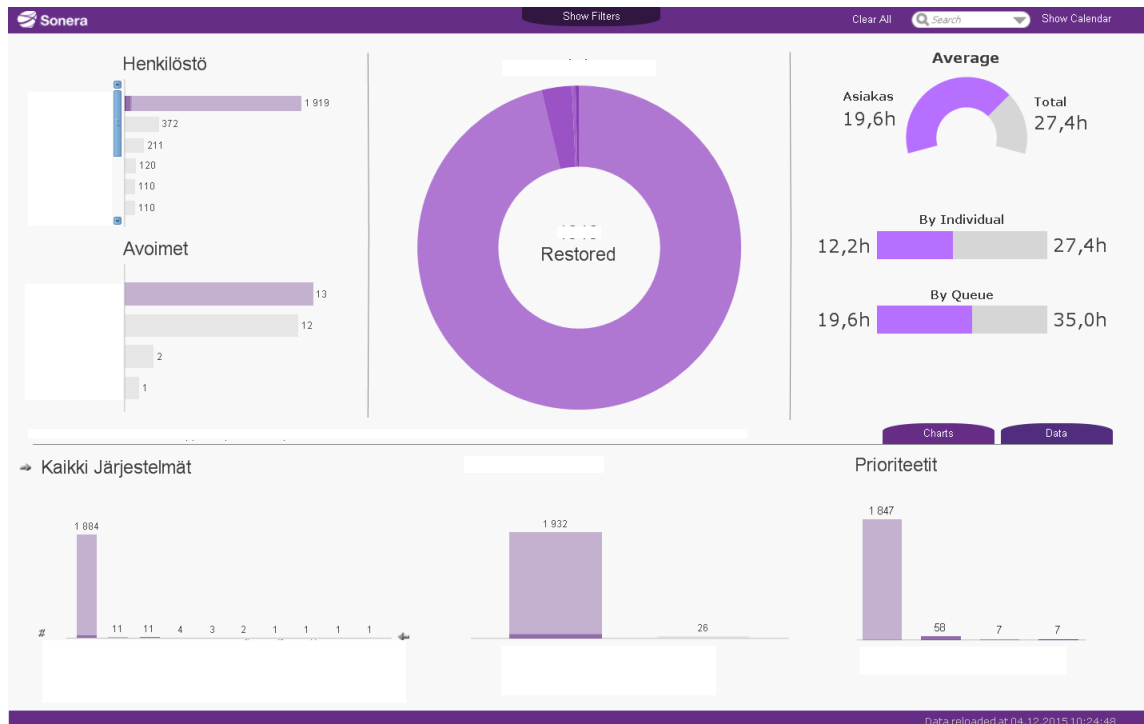


Kuva 17. Tapahtumat-näkymä.

4.3.3 Henkilöstö-näkymä

Henkilöstö-näkymän valinnat toimivat samalla periaatteella kuin tapahtumien näkymässä. Tässä näkymässä tapahtumien sijaan keskitytään henkilöstön tehokkuuden parantamiseen. Pääkäyttäjät näkevät nopeasti palvelupyynnöiden ratkaisuaikat. Näitä voidaan sen jälkeen verrata henkilöstöön tai jonoon. Tämän perusteella voidaan tehdä nopea analyysi millaisia palvelupyynnöitä, mihin järjestelmään, sekä mille asiakkaalle

pyynnöt kuuluvat. Huomioitavaa on se, että mittareiden vertailuarvoksi asetettiin kaikkien asiakkaiden, jonojen ja henkilöstön keskiarvo. Lopullisessa sovelluksessa näitä arvoja tullaan aina vertaamaan SLA-arvoihin. POC:n kannalta ei ollut kuitenkaan tärkeää käyttää SLA-arvoja.



Kuva 18. Henkilöstö-näkymä.

4.4 Käyttöönotto ja testaus

Julkaisu tapahtui QlikView Management Consolen kautta. Sovellus esiteltiin Telia-Soneran sisällä muutamalle eri taholle. Sovellus siirrettiin testi-palvelimelle testattavaksi, johon annettiin pääsy muutamalle pääkäyttäjälle. Julkaistu sovellus toimii websivuston sisällä, joka on QlikView Serverillä. Erillistä asennusta ei käyttäjien tarvitse tehdä. Pääsy tapahtuu portaalin kautta, joka toimii TeliaSoneran sisäverkossa. Datan päivitys asetettiin tapahtuvaksi öisin.

5 YHTEENVETO

Opinnäytetyön tavoitteena oli tehdä datan visualisointiin liittyvä POC-sovellus käyttäen QlikView-työkalua. Data haettiin OIVA-järjestelmään kirjatusta palvelupyynnöistä. QlikView-sovelluksia oli jo aikaisemmin tehty yrityksen sisällä, mutta yrityksen laajuuden vuoksi kaikki osastot eivät olleet vielä ehtineet ottamaan QlikView'tä käyttöön.

Tavoitteena oli luoda sovellus, joka tyydyttää loppukäyttäjien määrittelemiä tarpeita. Sovelluksessa oli tärkeää pystyä porautumaan aina raakadataan asti ja sovelluksessa haluttiin näytettävän määrättyjä KPI-mittareita. Sovellus julkaistiin sisäverkkoon testattavaksi, sekä sitä pidettiin onnistuneena. Toteutetun sovelluksen jatkokehityksestä puhuttiin esittelyn jälkeen.

Opinnäytetyön teoriaosuudessa käytiin läpi, mikä on massadataa, miten dataa kerätään ja miten sitä varastoidaan. Datan ymmärtäminen on lähtökohtana sen visualisointiin. Jos ei tiedetä, miten dataa haetaan tietovarastoista, on mahdoton tehdä analyyttisiä datan visualisointisovelluksia. Opinnäytetyön aikana käytiin läpi tietovaraston rakennetta, koska datan visualisoinnin taustalla on aina tietovarasto tai tietokanta. Datan ymmärtämisen lisäksi on tärkeää tietää datan visualisoinnin periaatteet ja hyvät käytännöt, jota käytiin myös opinnäytetyössä läpi.

Opinnäytetyön aikana tuli muutamia haasteita. Tietovarastosta haettu data ei täsmännyt täysin lähdejärjestelmän tietokannassa olevaan dataan. Tämän selvittely ei kuitenkaan haitannut opinnäytetyön tekoa, koska datan visualisointi ja sovelluksen rakentaminen pystyttiin aloittamaan vaikka data ei täsmännyt täysin. Toisena haasteena oli toisen data visualisointityökalun testaaminen, jota pohdittiin opinnäytetyön alkuhetkellä. Päätöksenä pysyi kuitenkin se, että keskitytään vain yhteen datan visualisointityökaluun.

Opinnäytetyön määrittely muuttui tämän vuoksi kahden eri datan visualisointityökalun vertailusta pelkkään QlikView'iin, sekä QlikView-sovellusten kehittämiseen. Vaikka opinnäytetyön aikana määrittely muuttui, se ei vaikuttanut lopputulokseen. Yhtenä haasteena oli myös se, että opinnäytetyön kirjoittajalla ei ollut aikaisempaa kokemusta datan visualisoinnista. Työkalut eivät olleet myöskään entuudestaan tuttuja. Ongelmista ja muutoksista huolimatta opinnäytetyön tavoitteet saavutettiin, sekä työ saatiin valmiiksi aikataulun mukaisesti.

LÄHTEET

99designs 2014. 6 principles of visual hierarchy for designers. Viitattu 24.5.2016 <https://99designs.com/blog/tips/6-principles-of-visual-hierarchy>.

Business Intelligence Singapore 2015. QlikView. Viitattu 24.5.2016 <http://singaporebusinessintelligence.blogspot.fi/2015/04/free-hands-on-qlikview-workshop-in.html>.

DCIELTS 2012. Bar charts. Viitattu 24.4.2016 <http://www.dcielts.com/task-1-2/ielts-barcharts-patterns>.

DCIELTS 2010. Pie charts. Viitattu 24.4.2016 <http://www.dcielts.com/ielts-writing/academic-task-1-pie-charts-2>.

EMC 2014. The digital universe of opportunities: Rich data and the increasing value of the internet of things. Viitattu 10.4.2016. <http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>.

FiCom 2016. Teletunnistetieto. Viitattu 22.4.2016 http://www.ficom.fi/tietoa/tietoa_4_1.html?Id=1130934583.html.

Gartner 2016. Magic Quadrant for Business Intelligence and Analytics platform. Viitattu 23.4.2016 <https://www.gartner.com/doc/reprints?id=1-2XXET8P&ct=160204>.

IBM 2016. Components of a star schema. Viitattu 12.4.2016 http://publib.boulder.ibm.com/tividd/td/TEDW/SC32-1497-00/en_US/HTML/srfmst158.htm.

Infochimps 2016. Streaming data and real-time analytics. Viitattu 12.4.2016 <http://www.infochimps.com/infochimps-cloud/cloud-services/cloud-streams>.

KDnuggets 2015. Data Lake vs Data Warehouse: Key Differences. Viitattu 12.4.2016 <http://www.kdnuggets.com/2015/09/data-lake-vs-data-warehouse-key-differences.html>.

Livescience 2008. Mind's Limit Found: 4 Things at Once. Viitattu 1.6.2016 <http://www.livescience.com/2493-mind-limit-4.html>.

McKinsey & Company 2011. Big Data: The next frontier for innovation, competition, and productivity. Viitattu 09.4.2016 <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>.

Microsoft 2016. Microsoft Open Database Connectivity (ODBC). Viitattu 24.5.2016 <https://msdn.microsoft.com/en-us/library/ms710252%28v=vs.85%29.aspx>.

nModal Solutions Inc. 2008. Datawarehouse vs Data mart. Viitattu 17.4.2016 <http://www.datamartist.com/data-warehouse-vs-data-mart>.

Oracle 2016. Datawarehousing concepts. Viitattu 12.4.2016 https://docs.oracle.com/cd/B10500_01/server.920/a96520/concept.htm.

Qgate, 2015. Business Intelligence: Guided Analytics v. Self-Service Analytics. Viitattu 1.6.2016 <http://www.qgate.co.uk/blog/business-intelligence-guided-analytics-v-self-service-analytics/>.

Qlik 2016. QlikView NPrinting. Viitattu 24.5.2016 <http://www.qlik.com/products/qlikview/nprinting>.

- Sanastokeskus 2013. Massadata. Viitattu 1.6.2016 http://www.tsk.fi/tsk/termitalkoot/hakemistot-267.html?page=get_id&id=ID352&vocabulary_code=TSKTT.
- ScribbleLive 2016. Best practices: Maximum elements for different visualization types. Viitattu 24.4.2016 <http://www.scribblelive.com/blog/2012/03/29/maximum-elements-for-visualization-types>.
- Sonera 2016. Sonera – TeliaSonera tytäryhtiö Suomessa. Viitattu 9.4.2016 <https://www.sonera.fi/teliasonera>.
- SQLAuthority 2013. Big Data. Viitattu 11.04.2016 <http://blog.sqlauthority.com/2013/10/02/big-data-what-is-big-data-3-vs-of-big-data-volume-velocity-and-variety-day-2-of-21>.
- TechnologyAdvice 2016. Tableau review. Viitattu 22.4.2016 <http://technologyadvice.com/products/tableau-reviews/>.
- Techopedia 2016. Data Science. Viitattu 1.6.2016 <https://www.techopedia.com/definition/30202/data-science>.
- Techopedia 2016. Data Source. Viitattu 1.6.2016 <https://www.techopedia.com/definition/30323/data-source>.
- TechTarget 2014. Business intelligence (BI). Viitattu 22.4.2016 <http://searchdatamanagement.techtarget.com/definition/business-intelligence>.
- UBM 2014. Qlik takes on Tableau in data visualization. Viitattu 22.4.2016 <http://www.informationweek.com/big-data/software-platforms/qlik-takes-on-tableau-in-data-visualization/a/d-id/1317946>.
- Usability 2016. User interface design basics. Viitattu 23.04.2016 <http://www.usability.gov/what-and-why/user-interface-design.html>.
- Usabilitygeek 2013. User Experience: What Is It And Why Should I Care? Viitattu 24.3.2016 <http://usabilitygeek.com/user-experience>.
- Webopedia 2016. Data Modeling. Viitattu 22.4.2016 http://www.webopedia.com/TERM/D/data_modeling.html.
- Webopedia 2016. ETL – Extract, Transform, Load. Viitattu 15.4.2016 <http://www.webopedia.com/TERM/E/ETL.html>.