

Juuso Paloniemi

**TIETOJÄRJESTELMIEN JA BIG DATA-ANALYTIIKAN  
HYÖDYT METSÄTEOLLISUUDEN LIKETOIMINNAN  
HALLINNASSA**



JYVÄSKYLÄN YLIOPISTO  
TIETOJENKÄSITTELYTIETEIDEN LAITOS  
2018

# TIIVISTELMÄ

Paloniemi, Juuso

Tietojärjestelmien ja Big Data-analytiikan hyödyt metsäteollisuuden liiketoiminnan hallinnassa

Jyväskylä: Jyväskylän yliopisto, 2018, 86 s.

Tietojärjestelmätiede, Pro Gradu - tutkielma

Ohjaaja: Pekka Makkonen

Big Data ja sen analysointi on tälläkin hetkellä yksi tietojärjestelmätieteen ajankohtaisimpia tutkimusaiheita. Se tarjoaa ennennäkemättömiä mahdollisuuksia tallentaa, käsitellä, varastoida ja analysoida tietoa niin yksityishenkilöille, kuin organisaatioille. Moderneista tietojärjestelmistä juuri Big Data-sovelluksia pidetäänkin yhtenä kulmakivenä seuraavassa digitaalisessa vallankumouksessa. Metsäteollisuudessa tieto- ja toiminnanohjausjärjestelmiä on uudistettu lähinnä suurempien kone- ja laitteistoinvestointien yhteydessä, mutta nyt on mahdollista integroida tietojärjestelmiin ja Big Data-analytiikkaa uusien liiketoimintamuotojen, tehokkuuden ja kustannussäästöjen mahdollistamiseksi. Ongelmana on kuitenkin se, mitä tietoa kannattaa ottaa prosessista talteen ja kuka määrittelee mitä tietoa yksittäisestä prosessista tarvitaan, jotta liiketoimintaa voidaan hallita ja kehittää mahdollisimman tehokkaasti.

Tässä työssä tutkitaan, kuinka nykyaikaisia tietojärjestelmiä ja Big Data - analytiikkaa voisi parhaiten käyttää hyväksi metsäteollisuuden konelinjastoilla parhaan mahdollisen arvonlisän saamiseksi. Työssä keskityttiin havaitsemaan tutkimuksen keinoin niitä mahdollisuuksia, jotka tuotannon asiantuntijat ja päälliköt kokevat mahdollisiksi ja tavoittelemisen arvoisiksi. Lisäksi selvitettiin tietojärjestelmien ja liiketoiminnan hallinnan yhteenlinjauksen haasteita asiantuntija- ja esimiesorganisaatioissa. Näistä teemoista muodostui liiketoiminnan hallinnan ja Big Data-tietojärjestelmäanalytiikan kokonaisuus, joka luo uutta ymmärrystä siitä, kuinka valtavasti lisääntyntä tietomäärää voidaan käyttää hyväksi uutta liiketoimintaa kehitettäessä sekä nykyistä toimintaa tehostettaessa. Tätä ymmärrystä johdettiin myös tutustumalla niihin teknologioihin (mm. Hadoop, HDFS, MapReduce) jotka mahdollistavat tämän uuden, entistä kattavamman ja tarkemman tavan kerätä, tallentaa ja analysoida prosessista saatavaa tietoa. Tämä oli erittäin tärkeää, jotta tulosten pohjalta pystytään analysoimaan niitä teknologioita, joiden jatkokehitys on organisaation kehityksen kannalta kriittisintä.

Työssä käsiteltiin ensin erilaisten massadataan perustuvien tietojärjestelmäteknologioiden käsitteistöä ja toiminnallisuuksia, jonka jälkeen siirryttiin liiketoiminnan hallintaan ja näiden kahden yhteenlinjauksen aiheuttamiin haasteisiin. Kyseistä aiheista rakennetun kirjallisuuskatsauksen avulla haettiin ymmärrystä tutkimusongelmaan. Tämän jälkeen suoritettiin laadullinen tapaustutkimus valikoidulle asiantuntijajoukolle, jossa vastauksia haettiin poh-

jautuen asiantuntijoiden haastatteluihin. Tämän jälkeen pohdittiin empiirisen tutkimuksen havaintoja ja vertailtiin niitä kirjallisuuden havaintoihin, jonka avulla tutkimusongelmaan pystyttiin rakentamaan lisää ymmärrystä.

Asiasanat: Tietojärjestelmä, Big Data, Big Data-analytiikka, Toiminnanohjaus, Toiminnanohjausjärjestelmä, Metsäteollisuus, Kunnossapito, ERP-järjestelmä, Liiketoiminnan hallinta, Liiketoiminnan ja IT:n yhteenlinjaus

## **ABSTRACT**

Paloniemi, Juuso

The benefits of ICT and Big Data-analytics in forest based industries business management

Jyväskylä: University of Jyväskylä, 2018, 86 p.

Information Systems, Master's Thesis

Supervisor: Pekka Makkonen

Big Data and analysing it as part of business intelligence (BI) systems is at this very moment one of the most popular research topics in the field of Information Systems Science. It offers unprecedented opportunities to store, handle, store and analyse information for both private individuals as for organizations. Big Data -systems that are implemented to business intelligence systems are considered to be one of the cornerstones in the next digital revolution. In the forest based industry Enterprise Resource Planning systems have been revised mainly with larger investments in machinery and equipment, but it is now possible to integrate the ERP system and Big Data analytics to build new kind of business as well as improve efficiency and bring costs to a lower level. However the problem is what information should be used and how the process of extracting the information generated by individual processes can be processed so it can be analysed as a whole. Also an important part of the process is to research if the users are collecting usable information and whether or not they know what they are supposed to do with that information.

This study examines how Big Data - analytics could best be utilized in business intelligence - systems to obtain the forest based industries operational sites to bring the best possible added value. The study focused find and locate the opportunities that engineers and managers found possible and worth pursuing. The research also studied the potential benefits of investments for customers as well as suppliers of raw materials in the form of a purely economic added value, but also in the form of increased knowledge and harmonization of operating cultures. These issues formed the Big Data analytics package, which creates a new understanding of how the data explosion can be used for the benefit of new business development as well as making current operations more efficient.

The study explored not only the Big Data systems theoretical and technological background (e.g. Hadoop, HDFS & MapReduce) by means of a literature review, but also the challenges in implementing and gathering usable information from forest based industries BI-systems. That gave more understanding to the research problem. This was followed by a qualitative case study relevant group of experts in which the answers were sought based on the interviews of those experts and professionals. After this the empirical research was discussed and compared to the literature findings. With that method it was possible to build more understanding to the research problem.

Keywords: Big Data, Big Data Analytics, ERP, Enterprise Resource Planning System, Forest Based Industry, Paper Machine, Board Machine, Pulp Mill, Maintenance

## KUVIOT

KUVIO 1 Tutkimuksen rakenne.....	16
KUVIO 2 Tiivistelmä tutkimusmetodologiasta.....	19
KUVIO 3 Datasta viisauteen.....	24
KUVIO 4 Tiedon eri tasojen arvo yritykselle.....	25
KUVIO 5 Määrä, nopeus, monimuotoisuus.....	27
KUVIO 6 Hadoopin arkkitehtuuri.....	33
KUVIO 7 Flume-agentin arkkitehtuuri.....	36
KUVIO 8 Informaation merkitys eri päätöksenteon tasoilla.....	44
KUVIO 9 Liiketoimintatiedon hallintaprosessin vaiheet.....	46
KUVIO 10 Liiketoimintatiedon hallinnan arkkitehtuuri.....	48

## TAULUKOT

TAULUKKO 1 Liiketoimintatiedon hallinnasta aiemmin esitettyjä määritelmiä.....	42
TAULUKKO 2 Haastateltavien profiilit.....	62

## SISÄLLYS

	TIIVISTELMÄ.....	2
	ABSTRACT.....	3
	KUVIOT.....	4
	TAULUKOT.....	4
	SISÄLLYS.....	5
1	JOHDANTO.....	8
	1.1 Tutkimuksen tausta.....	10
	1.2 Tutkimuksen motivointi.....	12
	1.3 Tutkimusongelma ja tutkimuskysymykset.....	13
	1.4 Rajoitukset ja soveltaminen.....	15
	1.5 Tutkimuksen rakenne.....	16
2	TUTKIMUKSELLISET MENETELMÄT.....	18
	2.1 Menetelmät.....	18
	2.2 Kirjallisuuskatsaus.....	19
	2.3 Empiirinen metodi.....	20
3	BIG DATAN TEOREETTINEN PERUSTA.....	23
	3.1 Tietopääoma.....	25
	3.2 Datan monimuotoisuus.....	26
	3.3 Big Datan määritelmä.....	26
	3.4 Big Data 3V - käsitelmä.....	27

3.4.1	Määrä .....	28
3.4.2	Nopeus .....	28
3.4.3	Monimuotoisuus .....	29
3.5	Big Datan hyödyntäminen .....	29
4	BIG DATA TEKNOLOGIAT JA NIIDEN MAHDOLLISUUDET METSÄTEOLLISUUDESSA .....	31
4.1	Kunnossapito .....	31
4.2	Tekniikat ja teknologiat Big Datan hyödyntämiseen .....	32
4.3	Hadoop .....	32
4.4	MapReduce .....	33
4.5	HDFS .....	34
4.6	Apache Spark .....	35
4.7	Massadatan kerääminen .....	35
4.7.1	Apache Flume .....	35
4.7.2	Muut massadatan keräämiseen tarkoitetut ohjelmistot .....	36
4.8	Massadatan analysointi reaaliajassa .....	36
4.8.1	Apache Storm .....	37
4.9	Muut massadatan onnistuneen käyttöönoton kannalta oleelliset ohjelmistot .....	37
4.9.1	Apache HBase .....	37
4.9.2	Apache Hive .....	37
4.9.3	Apache Pig .....	38
4.10	Savcor Wedge .....	39
4.11	Teknologioiden hyödyt organisaatiolle .....	39
5	LIIKETOIMINTATIEDON HALLINTA .....	41
5.1	Liiketoimintatiedon hallinta käsitteenä .....	42
5.2	Sisäinen ja ulkoinen liiketoimintatieto .....	43
5.3	Liiketoimintatiedon hallintaprosessi ja sen vaiheet .....	45
5.4	Prosessin menetelmät/järjestelmät .....	46
5.5	Prosessin hyödyt ja strategiaan yhdistäminen .....	46
5.6	Liiketoimintatiedon hallinnan järjestelmät .....	47
5.6.1	Tiedon keräys .....	48
5.6.2	Tiedon varastointi .....	49
5.6.3	Tiedon analysointi ja raportointi .....	50
5.7	Toiminnanohjausjärjestelmät metsäteollisuudessa .....	50
5.8	Käyttöönottoprosessi .....	51
6	KIRJALLISUUSKATSAUKSEN YHTEENVETO .....	52
7	TUTKIMUKSEN TARKOITUS, TAVOITTEET JA TUTKIMUSKYSYMYKSET .....	54
8	AINEISTO JA MENETELMÄT .....	57
8.1	Kohderyhmä ja aineistonkeruu .....	57

8.2	Teemahaastattelu aineiston hankinnassa .....	58
8.3	Aineiston analysointi.....	59
9	TUTKIMUKSEN TULOKSET .....	61
9.1	Haastateltavien taustatiedot.....	61
9.2	Tietojärjestelmien ja Big Data-sovellusten käytettävyys liiketoimintaympäristössä .....	63
9.3	Tiedon keräämisen ja käyttämisen selkeys organisaatiossa.....	65
9.4	Liiketoiminnan ja tietojärjestelmien yhteiset rajapinnat organisaatiossa .....	67
10	POHDINTA JA YHTEENVETO .....	70
10.1	Tulosten analysointi ja johtopäätökset .....	70
10.2	Tutkimuksen onnistuminen.....	73
10.3	Tulosten hyödyntäminen tulevaisuudessa sekä jatkotutkimuksen aiheet.....	74
	LÄHTEET .....	75
11	LIITE 1 HAASTATTELUTEEMAT.....	85

# 1 JOHDANTO

Metsäteollisuus on konservatiivisena alana saanut historiansa aikana odotella melkein pä kaikkia teknologisia läpimurtoja pidempään kuin muut raskaan prosessiteollisuuden alat. Valtavasti lisääntyneet datamäärät kuitenkin ovat pakottaneet myös metsäteollisuusklusterin tutkimaan tuon datan käsittelyä ja siitä mahdollisesti saatavia hyötyjä. Tällä hetkellä ei ole vielä yksimielisyyttä siitä kuinka Big Data määritellään, mutta muutamia yhteisiä teemoja on. (Franks, 2012) Termin alle voidaan liiketoiminnasta riippuen tuoda monia erilaisia tietojärjestelmiä sekä niiden osia ja jotkut yritysten nykyisistä järjestelmistäkin voidaan laskea Big Dataan kuuluvaksi, vaikka niiden tuottamaa tietoa ei vielä järjestelmällisesti analysoidaisikaan. Vaikka alan termistöstä ei vielä olekaan päästy täyteen yksimielisyyteen, voidaan Big Dataa kuitenkin kuvata tietokokonaisuudeksi, joita ei voida perinteisillä tietokantatyökaluilla kaapata, säilöä, käsitellä tai analysoida niiden suuresta koosta johtuen. (Yin & Kaynak, 2015)

Big Datan hyödyistä puhutaan valtavasti, eikä vielä kukaan tiedetä kaikkia mahdollisia hyötyjä joita se käyttäjilleen voi tuoda. Tämä tiedonkäsitteelyn mullistuminen tulee saattamaan monia perinteisiä prosesseja yrityksissä tarpeettomaksi ja pakottamaan yritykset muokkaamaan niin omia tietojärjestelmiään, kuin myös tarkastelemaan asiakkaiden tarpeita uudessa valossa. Erittäin raskaassa teollisuudessa erilaiset toiminnanohjausjärjestelmät ovat yrityksen päivittäisen toiminnan keskeisin osa. Ne ovat läsnä kaikessa toiminnassa ja niiden avulla ohjataan paitsi tuotantoa, niin myös tarkastellaan sitä mihin suuntaan toimintaa tulisi kehittää jotta tuottavuus parantuisi. Strategiset IT-investoinnit ovatkin lähes poikkeuksetta koko organisaation kattavia hankkeita ja ne vaativat yritykseltä valtavia panostuksia niin henkilöstöressurssien, kuin myös taloudellisten resurssien muodossa. Niiden toteutuneet hyödyt voivat kuitenkin olla vaikeasti mitattavissa ja tämä ongelma luokin mahdollisuuksia niin uusien organisaatio-, kuin myös liiketoimintamallien lanseerauksella. (Hallikainen, Kivijärvi & Nurminen, 2002)



Kuinka sitten toiminnanohjausjärjestelmiä pystytään päivittämään vastaamaan Big Datan nostamiin haasteisiin metsäteollisuuden eri toimialoilla? Metsäteollisuuden asiakkaat, esimerkiksi pakkaus- ja elintarviketeollisuus haluavat saada tarkempaa tietoa asiakkaistaan ja samalla myös tavarantoimittajistaan. Metsäteollisuuden onkin vastattava tähän haasteeseen ja implementoitava Big Data osaksi järjestelmiään asiakaslähtöisesti siten, että jokainen kumppani saisi analysoidusta tiedosta itselleen lisäarvoa. Tämä työ käsitteleeekin Big Datan hyödyntämistä raskaan prosessiteollisuuden parissa sekä tutkia liiketoiminnan ja tietotekniikan yhteenlinjausta globaalissa teollisuusorganisaatiossa.

Big Data-analytiikalla tavallisesti tarkoitetaan ennustavaa ja ohjailevaa analytiikkaa. Ennustava analytiikka pyrkii ennustamaan todennäköisyyksiä tuleville tapahtumille ja ohjailevalla analytiikalla pyritään seulomaan ne jatkotoimenpiteet, joilla organisaatio pääsee parhaisiin mahdollisiin tuloksiin liiketoiminnassaan. (Minelli, 2013). Ennustavasta analytiikasta on hyötyä erityisesti prosessiteollisuuden kunnossapidon puolella, jossa voitaisiin ennustaa mahdollisia laiterikkoja entistä tarkemmin jolloin niitä voitaisiin ennaltaehkäistä suunniteltujen kunnossapitoseisakkien yhteydessä. Metsäteollisuuden erityispiirteenä on valtavan kokoiset koneet ja laitteet, joiden käynnistysajat laiterikosta toipuessa ovat minimissään tunteja, mahdollisesti jopa päiviä. Näitä laiterikkoja vähentämällä pystymme nostamaan merkittävästi konelinjan ajettavuutta ja siten tuottavuutta. Tähän pystyäksemme meidän on kuitenkin pystyttävä analysoimaan tämä valtava määrä dataa, sillä ainoastaan silloin tuo data on arvokasta. On myös tärkeää, että prosessia pystytään analysoimaan sen ollessa käynnissä, toimivatko venttiilit ja anturit oikein ja voidaanko ratakatkon todennäköisyys johtaa johonkin tiettyyn sakeuteen tai venttiilikuormaan prosessissa? Vuonna 2013 tehdyn yritysjohtajille suunnatun tutkimuksen mukaan 58% heistä totesi data-analytiikan elintärkeäksi yrityksensä liiketoiminnalle. Tutkimuksissa on silti havaittu, että organisaatiot kokevat usein datan määrän musertavana, eivätkä osaa muuttaa sitä tuloksiksi. (Evans, 2015)

Metsäteollisuus on Suomen kansantalouden kannalta erittäin merkittävä ala. Se työllistää suoraan 82 000 suomalaista ja tuo viidenneksen maamme vientituloista. Suunniteltuja metsäteollisuuden investointeja on vuosille 2015-2017 kaavailtu 2,0 mrd eurolla (Lehtoviita, Mäki & Tenhola, 2016) Näiden jo sinällään merkittävien lukujen ulkopuolelle jää vielä usein se seikka, että metsäteollisuus käyttää 63,7 miljoonaa kuutiometriä kotimaista puuraakaainetta joka vuosi ja kantorahatuloja metsänomistajille kertyi 1,6 miljardia euroa. (Lehtoviita ym, 2016) Näin ollen voidaan todeta, että metsäteollisuus paitsi työllistää, niin myös jakaa hyvinvointia tasaisesti metsänomistajille ympäri Suomea, jolloin myös arvoketjun alapäähän käytetty raha jää Suomeen ja luo työtä ja toimeliaisuutta välillisesti. Tuo puun jalostuksesta saatu lisäarvo on tärkeää jättää tulevaisuudessakin Suomeen, sen sijaan että toimisimme raaka-aineittana ulkomaisille yrityksille, haluamme säilyä kilpailukykyisenä puunjalostusmaana ja erilaisten analyysityökalujen haltuunotolla on suuri rooli tuon kilpailukykyyn ylläpitämisessä.

Metsäteollisuusorganisaatiot ovat olleet jatkuvassa rakennemuutokset kuluvat 15 vuotta, niissä kaikissa on tapahtunut valtavia muutoksia niin strategian, organisaatioiden kuin IT:n suhteen erittäin lyhyessä ajassa. Idänkaupan tyrehtyminen ja kilpailijoiden nousu uusilla liiketoiminta-alueilla niin Aasiassa kuin Etelä-Amerikassakin on johtanut teknologiseen kilpajuoksuun, joka yhdessä turvatus raaka-aineen saannin kanssa muodostaa kilpailuvaltit nykypäivän metsäteollisuusyritykselle. Entistä tarkempi erikoistuminen ja tuotepalettien sirpaloituminen yhdessä alati tiukkenevien ympäristö- ja laatuvaatimusten kanssa nostavat organisaatioille asetettuja vaatimuksia. Metsäteollisuusorganisaatiot ovat myös käyneet läpi merkittäviä tuotannollisia uudistuksia ja siirtyneet kohti vaativampia, korkeamman katteen tuotteita ja jopa hankkiutuneet eroon kannattamattomasta liiketoiminnasta joko yrityskaupoin tai konelinjojen alasajoilla vain vastatakseen paremmin muuttuneeseen liiketoimintaympäristöön ja tulevaisuuden tarpeisiin.

Tässä työssä pyritään tutkimuksen keinoin selvittämään, kuinka konelinjojen insinöörit ja päälliköt näkevät liiketoiminnan ja tietojärjestelmien rajapintojen löytämisen vaikuttavan heidän kykyynsä tehdä sellaisia päätöksiä, joista on lisäarvoa paitsi yritykselle, niin myös sen sidosryhmille. Kuinka raaka-aineen toimittajat hyötyvät lisääntyvästä datan määrästä ja kuinka asiakkaat voivat käyttää paremmat tiedot lopputuotteen laadusta hyväkseen. Kaikesta maailman datasta noin 90% on syntynyt viimeisen kahden vuoden aikana ja dataa kuvataankin liiketoiminnan ”uudeksi öljyksi.” (Krushe-Lehtonen, 2014) Johdantoluku antaakin yleiskuvan opinnäytetyöhön ja selittää paitsi, taustaa niin myös tutkimuksen rakennetta. Liiketoiminnan ja IT:n yhteenlinjauksen rakennuspalikoita käsitellään työn jokaisessa vaiheessa, termin määritelmän ollessa; ”se piste jossa IT-sovelluksia, infrastruktuuria ja organisaatiota voidaan käyttää liiketoimintastrategian ja prosessien tukena, sekä se prosessi jolla tämä pystytään ymmärtämään.” (Silvius, 2009) Kyseessä on laaja määritelmä, mutta se onnistuneesti kiteyttää IT:n ja liiketoiminnan yhteenlinjauksen ja sen ydinvälineet.

## 1.1 Tutkimuksen tausta

Liiketoiminnan ja IT:n yhteenlinjauksella on vakiintuneet perinteet tutkimuksen saralla niin yritysmaailmassa, kuin akateemisen tutkimuksen saralla. Tämän väitteen toteamiseksi riittää, että katsoo vaikkapa suhteellisen tuoreita tutkimuksia, jotka vahvasti viittaavat siihen että on olemassa havaittava positiivinen suhde IT:n yhteenlinjauksen ja liiketoiminnan tehokkuuden välillä. Tämä selittää myös yritysmaailman kiinnostusta asiaa kohtaan. (esim. Oh & Pinnoseault, 2007 sekä Gerow, Thatcher & Grover, 2014)

Organisaatiot, joiden strategiset suunnitelmat ja liiketoiminnan prosessit ovat linjassa IT-operaatioiden ja -käytäntöjen kanssa pystyvät ottamaan käyttöön teknologioita innovatiivisemmin ja heidän IT-kulunsa on opti-

moitu paremmin, näillä on mahdollista saavuttaa selkeää kilpailuetua. (Peppard & Ward, 1999) Yrityksille siis yhteenlinjaus IT:n ja liiketoimintojen saralla johtaa suurempaan arvонуonttiin ja on siten tavoittelemisen arvoista suurienkin resursointipanostusten voimin.

Liiketoiminnan ja IT:n yhteenlinjaus on myös suhteellisen käsitelty aihe, kuten Coltman, Tallon, Sharma & Quieroz (2015) toteavat tutkimuksessaan; Strateginen IT:n yhteenlinjaus: 25 vuotta käynnissä, että sen syntyhistoria ulottuu 1980-luvulle. Vuonna 1984 aloitettiin hanke "MIT90s" ja sen tarkoituksena oli rakentaa runko ja puitteet sellaiselle toiminnalle, jolla pystyttäisiin tarkastelemaan kriittisimpien osien välisiä suhteita liittyen liiketoimintastrategiaan, rakenteeseen, teknologiaan, ihmisiin ja hallintoon. Tämä projekti laajeni koskemaan sen ajan kymmentä suurinta IT-organisaatiota ja se kesti yli kahdeksan vuotta. (Coltman ym., 2015)

Yksi tunnetuimmista IT:n ja liiketoiminnan yhteenlinjauksen malleista, Strategic Alignment Model (SAM) syntyi tämän projektin yhteydessä. (Henderson & Venkatraman, 1990 sekä Henderson & Venkatraman, 1993) SAM -mallia käytetään myös tässä työssä, sillä se on edelleen voimassa monin tavoin. Ennen tätä projektia ei IT:n ja liiketoiminnan yhteenlinjauksesta oltu tehty tutkimusta tässä mittakaavassa. (Coltman ym., 2015) Voidaankin sanoa, että MIT90s hanke oli alku liiketoiminnan ja IT:n yhteenlinjauksen tutkimuksella, joka on jatkunut aktiivisena tähän päivään asti. Toki tutkimus on kehittynyt ja kehittyä edelleen, strategioita ja lähestymistapoja liiketoimintasegmenteissä on tutkittu erittäin laajasti ja nuo tutkimukset ovatkin opettaneet paljon IT:n yhteenlinjauksesta. Nämä lähestymistavat saattavatkin olla liian kapeita moderneihin tarpeisiin, joten seuraava askel tutkimuksessa voisi olla monitasoisten liiketoiminnan ja IT:n yhteenlinjaamisen ristiin tutkiminen eri lähestymistapoja käyttäen. (Coltman ym., 2015)

Tutkimuksen kannalta on tärkeää muistaa myös liiketoiminnallinen puoli, jossa informaatioteknologialla on mahdollisuuksia saavuttaa tavoitteita myös strategisella puolella auttaen yritystä parantamaan kilpailukykyä, tuottavuutta sekä muokata organisaatiota ja luoda uusia tapoja johtaa sitä ja sen synnyttämiä uusia liiketoimintamahdollisuuksia. (Hallikainen ym., 2002)

Tässä työssä käytetään myös uudempia lähestymistapoja, kuten Reynolds & Yetton (2013), jotka teorisoivat että mikäli liiketoiminnan ja IT:n yhteenlinjausta lähestytään useammasta näkökulmasta, se johtaa kattavimpiin havaintoihin. Tätä Reynoldsin & Yettonin monimallilähestymistapaa testattiin Pekkolan ja Niemisen (2015) tapaustutkimuksessa ja he totesivat sen hyödylliseksi lähestymistavaksi. Monimallilähestymistapa käyttää kolmea erilaista mallia kolmesta liiketoiminnan ja IT:n yhteenlinjauksen ulottuvuudesta, joten sen tulisi antaa kokonaisvaltainen näkemys koko konseptiin. (Reynolds & Yetton, 2013) Monimallilähestymistapaa käsitellään tarkemmin myöhemmissä luvuissa.

Koska liiketoiminnan ja IT:n yhteenlinjaus on jo kohtalaisen kypsä aihe akateemisessa hallinnon tutkimuksessa, nousee esiin väitteitä että se on turhaa tai hyödytöntä. Onkin perusteltua väittää, että liiketoiminnan ja IT:n yhteenlinjauksen tulisi olla jo olennainen osa johtamista ja että siihen keskittymi-

sen ei tulisi enää antaa merkittävää kilpailullista etua. Chan ja Reich (2007) ovat listanneet muutamia näistä väitteistä;

1. Liiketoiminnan ja IT:n yhteenlinjauksen tutkimus ei pysty liian mekaanisena vastaamaan tosielämän vaatimuksiin
2. Liiketoiminnan ja IT:n yhteenlinjaus ei ole mahdollista, mikäli strategia on tuntematon tai jo käynnissä.
3. Liiketoiminnan ja IT:n yhteenlinjaus ei ole toivottava päämäärä, sillä liiketoiminnan tulee olla jatkuvassa muutoksessa.
4. IT:n tulisi jatkuvasti haastaa liiketoimintaa, ei seurata sitä.

Chan ja Reich (2007) huomauttavat myös, että yllä mainittuja vastaargumentteja tulisi pitää haasteina IT:n ja liiketoiminnan yhteenlinjauksen tutkimisessa enemmän kuin syinä lopettaa tutkimus tai jopa kyseenalaistaa koko aihepiirin mielekkyys. Tottakai liiketoiminnan ja IT:n yhteenlinjauksen voi toteuttaa huonosti ja jotkin mahdollisista virheistä on lueteltu yllä. Tämä opinnäyte jakaa Chanin ja Reichin (2007) näkemyksen siitä, että yritysten IT:n ja liiketoiminnan linjaus ei ole yksiulotteinen tekijä tai prosessi, vaan sen sijaan siinä on useita perspektiivejä ja monia tapoja tuoda lisäarvoa organisaatiolle.

## 1.2 Tutkimuksen motivointi

Metsäteollisuusorganisaatiot ovat olleet jatkuvassa rakennemuutoksessa kuluvat 15 vuotta, niissä kaikissa on tapahtunut valtavia muutoksia niin strategian, organisaatioiden kuin IT:n suhteen erittäin lyhyessä ajassa. Entistä tarkempi erikoistuminen ja tuotepalettien sirpaloituminen yhdessä alati tiukkenevien ympäristö- ja laatuvaatimusten kanssa nostavat organisaatioille asetettuja vaatimuksia. Metsäteollisuusorganisaatiot ovat myös käyneet läpi merkittäviä tuotannollisia uudistuksia ja siirtyneet kohti vaativampia, korkeamman katteen tuotteita ja jopa hankkiutuneet eroon kannattamattomasta liiketoiminnasta joko yrityskaupoin tai konelinjojen alasajoilla vain vastatakseen paremmin muuttuneeseen liiketoimintaympäristöön ja tulevaisuuden tarpeisiin.

Käsiteltäessä strategisia IT-investointeja ja -linjauksia on kuitenkin huomattava, että IT:n ja liiketoiminnan linjauksen taso on suoraan pääteltävissä kunhan vain IT-investointien arviointiin käytettävät menetelmät, niiden rajoitteet ja rajoitteiden aiheuttamat haasteet ovat tiedossa. Alan kirjallisuudessa onkin todettu, että mikäli organisaatio eteen tulevista haasteista ja vaikeuksista huolimatta pystyy toteuttamaan kattavan ja huolellisen arvioinnin IT-investoinneistaan, se kykenee samalla parantamaan yrityksensä IT:n ja liiketoiminnan yhteenlinjausta, jonka on tutkimuksissa todettu parantavan yrityksen liiketoiminnan arvoa. (Tallon, Kraemer & Gurbaxani, 2000)

Metsäteollisuudessa ja siinä tapahtuvassa IT:n ja liiketoiminnan yhteenlinjauksessa muutos itsessään ei kuitenkaan välttämättä ole se varsinainen haaste, vaan muutoksen ja toteutumisen välinen viive on se joka mahdollisesti asettaa haasteita organisaatioille. (Chan & Reich, 2007) Onkin kohtuullista olettaa, että muutokset aiheuttavat viiveitä organisaatioiden sisällä ja näiden viivekohtien paikallistaminen helpottaa varmasti organisaatioita kohdentamaan toimintaansa. Metsäteollisuuden organisaatioissa on tärkeää keskittyä yhteisiin tavoitteisiin ja liiketoiminnan ja IT:n yhteenlinjaus vastaakin kysymyksiin strategisesta painopisteestä, vaikeasti mitattavista tuotoista, investointien suhteesta strategiaan ja sekä pitkän toteuttamisaikavälin haasteisiin muuttuvassa liiketoiminta- ja tuotantoympäristössä. Nämä olivat pääkohdat sille, miksi tämä opinnäytetyö tehtiin.

### 1.3 Tutkimusongelma ja tutkimuskysymykset

Tämän työn tavoitteena on tutustua metsäteollisuusorganisaatioiden nykytilaan liiketoiminnan ja IT:n yhteenlinjauksen saralla. Tämä voidaan saavuttaa tunnistamalla liiketoiminnan ja IT:n yhteenlinjaukseen valittujen menetelmien haasteita ja sitä kautta auttaa organisaatioita määrittämään mitkä tavat ja menetelmät kannattaa ottaa jatkokehitykseen alan sisällä. Johtopäätökset tuovat perusteltuja löydöksiä tai faktoja yhteenlinjauksen esteenä oleville haasteille. Alan organisaatiot voivat tämän jälkeen tulkita tuloksia kuinka parhaaksi näkevät, tutkimuskysymyksen ollessa seuraava:

- Mitä tietojärjestelmien ja liiketoiminnan hallinnan yhteenlinjauksen haasteita voidaan tunnistaa metsäteollisuusorganisaatioissa?

Lisäksi työssä tutustutaan Big Data - menetelmien tarjoamiin mahdollisuuksiin, sekä kartoitetaan niiden taustalta löytyvää teoriaa. Kuinka on mahdollista analysoida dataa paremmin ja sitä kautta käyttää arvokkaita raaka-aineita, kuten puuta ja kemikaaleja tehokkaammin hyväksi paremman kustannuskilpailukyvyn saavuttamiseksi? Tämä tehtiin etsimällä relevantteja artikkeleita ensin alan tärkeimmistä julkaisuista, kuten MIS Quarterlystä, Information Systems Researchista sekä Journal of AIS:stä. Tämän jälkeen oli helppo laajentaa hakuja Google Scholariin sekä JYKDOK:n arkistoon, josta eri tekijöitä hakemalla saattoi löytää sekä mielenkiintoisia artikkeleita, että kyseisten artikkeleiden lähteistä etsimällä vielä mielenkiintoisempia artikkeleita. Lisäksi etsin hakusanoilla "metsäteollisuus", "Big Data", "massadata", "analytiikka", "analytics", "tietojärjestelmä" sekä "toiminnanohjaus." Ajallisesti en materiaalia rajannut, sillä Big Data - tutkimus on niin tuoretta, että kaikki materiaali on viimeisen kymmenen vuoden ajalta. Käytännössä siis etsitään teknologioita, joiden avulla voidaan paremmin ymmärtää prosessin syy-seuraus - suhteita sekä siten saavuttaa paitsi parempi-

laatuinen lopputuote, niin myös jatkuvatoimisempi prosessi vähempine häiriöineen ja parempine hyötysuhteineen. Voidaanko tiettyjä raaka-aineita käyttää vähemmän, vaikka lopputuote pysyisi ennallaan?

Monimallilähestymistapa liiketoiminnan ja IT:n yhteenlinjaukseen on suhteellisen uusi tapa, vaikkakin lähestymistavan käyttäminen itsessään onkin tärkeä osa tätä opinnäytetyötä. Monimallilähestymistavan teoreettinen oikeutus liiketoiminnan ja IT:n yhteenlinjaukseen tulee Reynoldsin ja Yeattonin (2013) konseptista, sekä Pekkolan ja Niemisen (2015) tapaustutkimuksesta. Tämä työ tulee olemaan hyvä lisä akateemiseen tutkimukseen aiheesta, vaikka alkaakin olla jo selvää että tällä lähestymistavalla on mahdollista päästä käytännön hyötyihin.

Tätä ennen on kuitenkin selitettävä liiketoiminnan ja IT:n yhteenlinjauksen konsepti, jotta on helpompi ymmärtää sekä monimalli-lähestymistapa, että siihen liittyvät löydökset.

Tämä johtaa neljään tutkimusongelmaan:

1. Mitä on tietojärjestelmien ja liiketoiminnan yhteenlinjaus?
2. Millaisen organisaation se vaatii toimiakseen?
3. Mitä Big Data merkitsee tietojärjestelmien ja liiketoiminnan yhteenlinjaukselle?
4. Minkälaisia ongelmia metsäteollisuusympäristössä voi tällä saralla kohdata?

Ensimmäiset kaksi tutkimusongelmaa tulee ymmärtää, jotta voidaan ymmärtää liiketoiminnan ja IT:n yhteenlinjauksen merkityksen organisaatioille. Kolmas tutkimusongelma kohdataan opinnäytetyöhön valitun teoreettisen kehyksen ympärillä ja neljännen tutkimusongelman tarkoitus on tunnistaa nykyiset haasteet alalla. Yhdistävänä tekijänä näille tutkimusongelmille on se, että yrityksillä on lukuisia tapoja kerätä tietoa eri prosessien avulla, mutta tuon datan valjastaminen liiketoiminnan tueksi on osa-alue, jota ei ole aiemmin hyödynnetty riittävästi. Metsäteollisuuden prosessien mittaaminen ja niiden mittausten hyödyntämisen potentiaalia ei ole käytetty itse liiketoiminnan johtamiseen. Lisäksi on ollut epäselvyyttä, kuka omistaa käytetyt prosessit ja kenelle prosessien tuloksista ja muutoksista raportoidaan. Vastaukset näihin kysymyksiin muodostavat vastauksen tutkimuskysymykseen.

## 1.4 Rajoitukset ja soveltaminen

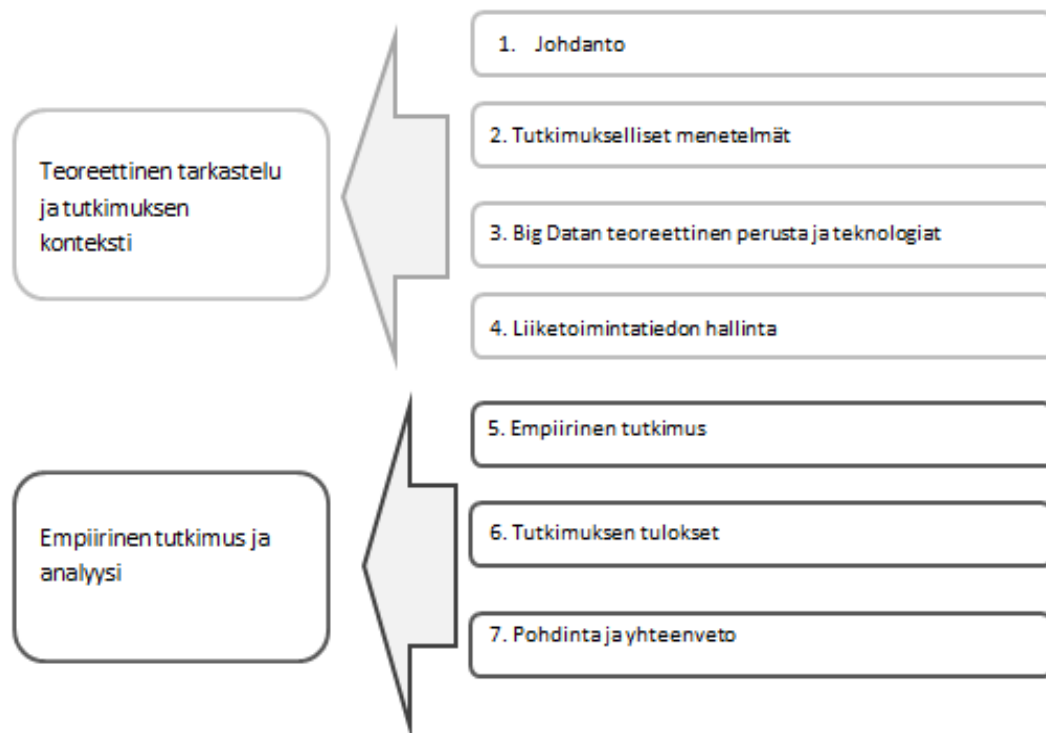
Tutkittaessa metsäteollisuuden suhdetta IT:n ja liiketoiminnan yhteenlinjaukseen, tässä opinnäytteessä päädyttiin tutkittavan ilmiön kvalitatiiviseen tutkimusmenetelmään tutkittavan ilmiön kompleksisuuden, moninaisuuden ja asiayhteyksien suuren määrän vuoksi. Tutkimusaiheen kokonaisuuden hahmottaminen, puhumattakaan tarkoista tuloksista olisi ollut käytännössä mahdotonta kvantitatiivisella tutkimusmenetelmällä aiheen monitahoisuuden vuoksi. Tämän opinnäytteen menetelmät sisältävät kirjallisuuskatsauksen niin Big Datasta, strategisista IT-investoinnista, kuin liiketoiminnan ja IT:n yhteenlinjauksesta sekä haastatteluita metsäteollisuusorganisaatioiden avainhenkilöiden kanssa.

Pidän tärkeänä, että opinnäyte on helposti yleistettävissä metsäteollisuudessa, myös globaalisti. Tämän vuoksi empiirinen metodi on osuva tähän tutkimukseen. Työn käytännön panoksena metsäteollisuudelle on auttaa rakentamaan tietämystä IT:n ja liiketoiminnan yhteenlinjauksesta. Tämä auttaa metsäteollisuusorganisaatioita tekemään parempia päätöksiä tulevaisuudessa. Myös tietämyksen lisääminen niin liiketoiminnan saralla IT:stä, kuin IT:n ja tietojärjestelmien saralla liiketoiminnasta voidaan nähdä eduksi organisaation päätöksenteolle tiedon ja ymmärryksen lisääntyessä organisaation laajuisesti. Teemahaastatteluiden avulla pystyttiin toteuttamaan laadullisia ja määrällisiä havaintoja liiketoiminnan ja IT:n yhteenlinjauksesta organisaatioissa ja tiedonkeruussa käytettiin asiantuntijahaastatteluita, joissa haastateltavat on valittu eri organisaatioista mahdollisimman hyvän tutkimuspeiton ja kattavan kokonais käsityksen saamiseksi.

Kirjallisuuskatsaus toimi empiirisen tutkimuksen pohjana ja auttaa löydösten oikeuttamisessa. Siinä pyritään vastaamaan tutkimuskysymykseen määrittämällä Big Data-palveluiden teoreettista perustaa ja mahdollisia käyttökohteita, niin myös liiketoiminnan hallinnan tärkeyttä metsäteollisuusyrityksen kannalta. Liiketoiminnan ja IT:n yhteenlinjauksen näkökulmasta tässä työssä hallitsevampana näkökulmana on IT. Tämä lähestymistapa aiheuttaa myös rajoituksia opinnäytetyöhön. Tämän kaltainen rajoitus jossa haasteet ja mahdollisuudet esitellään IT:n näkökulmasta, tekee monilta osin mahdottomaksi esittää perusteellisia ja syvällisiä organisaation laajuisia ratkaisuita tunnistettuihin haasteisiin. Oli kuitenkin mahdollista analysoida ja prosessoida löydöksiä pohjautuen kirjallisuuteen liiketoiminnan ja IT:n yhteenlinjauksesta siitä rakennettua viitekehystä apuna käyttäen. Tältä pohjalta oli mahdollista tehdä muutamia ehdotuksia. Tarkoituksena ei siis ole ehdottaa perusteellisia korjausehdotuksia organisaatioiden sisällä, vaan antaa lähtökohta arvioinnille ja nostaa esiin muutamia mahdollisia painopistealueita.

## 1.5 Tutkimuksen rakenne

Opinnäytetyön rakenne esitetään kuviossa 1 ja opinnäytetyön rakenne voidaan jakaa kahteen osaan. Ensimmäinen osa tarjoaa taustatiedot työhön sekä teoreettisen viitekehyksen, toisen osan keskittyessä empiiriseen tutkimukseen, löydöksiin ja johtopäätösten muodostamiseen.



KUVIO 1. Tutkimuksen rakenne

Ensimmäisessä luvussa, selitetään tutkimuksen tausta, motivaatio ja tavoitteet antaen samalla selkeän kuvan siitä, mitä opinnäytetyön on tarkoitus selittää sekä halutut tulokset. Toisessa luvussa käsitellään menetelmiä ja niiden taustalla olevia filosofioita ja ohjeita jotka jatkuvat yhdenmukaisena koko työn läpi. Menetelmäluvussa pyritään selventämään ajatusprosessit ja johdonmukaisuudet, jotka vaikuttavat opinnäytteen analyysiin ja löydöksiin. Tätä seuraavat menetelmiä käsittelevät kappaleet kolme, neljä ja viisi kytkevät kirjallisuuden työhön. Luvun kolme muodostaa Big Datan teoreettisen perustan, luvun neljä Big Datan teknologiasovellukset sekä niiden metsäteollisuuden käyttömahdollisuudet. Luvun viisi keskittyy liiketoiminnan hallintaan ja siinä tarkastellaan



sekä yritysten mahdollisuuksia, että akateemista konseptia ja malleja joita käytetään analyysin perusteena itse työssä.

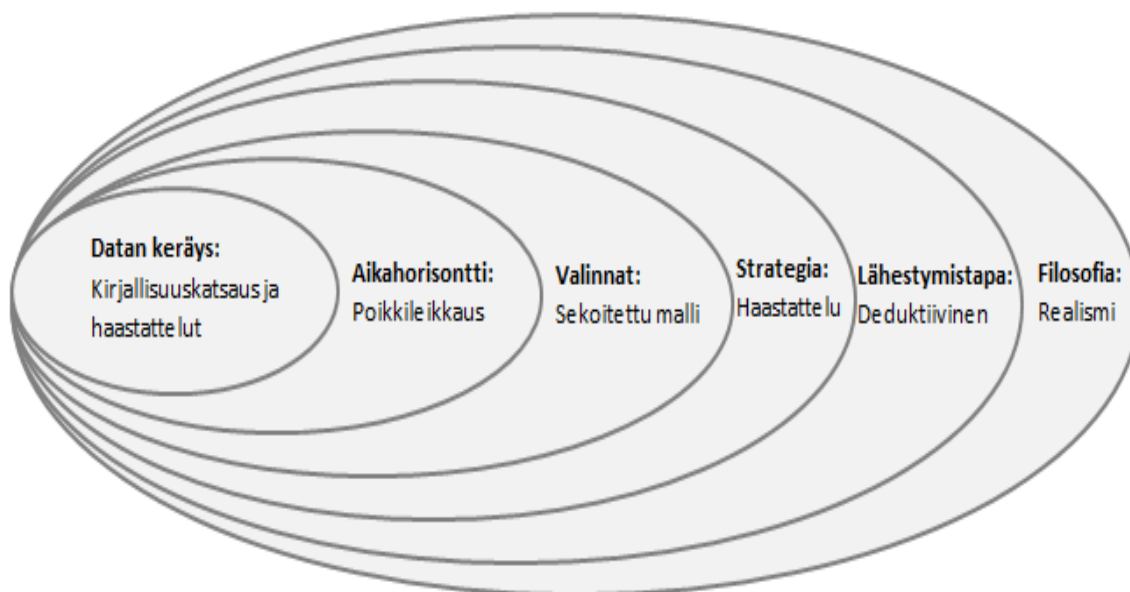
Työn toinen osa selventää empiirisen tutkimuksen osaa itse työssä. Luvussa kerrotaan raskaan teollisuuden vaatimuksista yksityiskohtaisemmin, jotta löydökset pystytään ymmärtämään oikeassa kontekstissa. Kuudennen kappaleen pääsisältö on kuitenkin haastattelumuodon ja haastateltavien esittely. Seitsemännessä luvussa käydään läpi ja luodaan yhteenveto tutkielman havainnoista ja yhdistetään ne työn ensimmäisen puoliskon kirjallisen osuuden teoreettiseen viitekehykseen. Luvussa kahdeksan otetaan kokonaisvaltaisempi ote havaintoihin ja tutkitaan haasteita hieman syvemmin sekä tarjotaan niihin mahdollisia ratkaisuja perustuen kirjallisuuskatsaukseen. Viimeisessä luvussa yhdeksän tehdään päätelmät, yhteenveto sekä vastataan tutkimuskysymykseen ja tutkimusongelmiin.

## 2 Tutkimukselliset menetelmät

Tässä luvussa kerrotaan opinnäytetyön tutkimusmenetelmien takana olevat filosofiat ja itse tutkimusmenetelmät esitellään yleisellä tasolla. Tämän yleisen teorian lisäksi empiiriset metodit on kuvattu kappaleen jälkipuoliskolla. Luku selvittää lukijalle työn tekijän ajatusprosessia valittaessa menetelmiä ja lähestymistapoja parhaan mahdollisen lopputuloksen aikaansaamiseksi.

### 2.1 Menetelmät

Opinnäytteen tutkimusmenetelmä on esitetty yhteenvetona kuviossa 2. (Saunders, Lewis & Thornhill, 2009). Saundersin tutkimussipuli tarjoaa riittävän mallin kaikista oleellisista kysymyksistä liittyen tutkimusfilosofiaan ja tutkimusmenetelmiin. (Hirsjärvi, Remes & Sajavaara, 2007) kuitenkin korostavat, että mikäli ymmärtää tutkimukseen liittyvät filosofiset kysymykset on helpompaa ymmärtää loogiset argumentit tutkimuksen löydösten takana. Tämä helpottaa valitsemaan oikeat tutkimusmenetelmät. Seuraavassa on kaikki tutkimussipulin eri kerrokset selitettynä.



KUVIO 2. Tiivistelmä tutkimusmetodologiasta (Saunders ym., 2009, s. 108 mukaan)

## 2.2 Kirjallisuuskatsaus

Vaikka onkin äärimmäisen tärkeää ymmärtää tämän työn punainen lanka, on myös oleellista ymmärtää kuinka tutkimus suoritettiin. Tutkimuksessa on paitsi teoreettista tutkimusta kirjallisuuskatsauksen muodossa, niin myös empiirinen haastattelututkimus. Alla molemmat lähestymistavat on selitetty tämän opin- näytetyön kontekstissa.

Saunders (2009), listasi kaksi merkittävää syytä sille, miksi kirjallisuuskatsaus on oleellinen osa mitä tahansa opinnäytettä tai jopa projektia yli- päättään. Ensinnäkin, kirjallisuuskatsaus on työkalu esitutkimukselle, joka vaa- ditaan jotta voidaan muodostaa hyvä tutkimusaihe ja ongelmat. Toinen syy on luoda pohja löydöksille ja analyysille. Tätä kutsutaan kriittiseksi kirjallisuuskat- saukseksi ja se on yleensä osana lopullista työtä tai projektia, kuten myös tässä työssä. Kappaleet kolme, neljä ja viisi keskittyvät Big Dataan, sekä Metsäteolli- suuden teknologian ja liiketoiminnan yhteenlinjaukseen. Näiden kappaleiden jälkeen kirjallisia lähteitä on käytetty todentamaan ja vahvistamaan löydettyjä tuloksia.

Lähteiden etsimiseen käytettyihin metodeihin lukeutuvat käytän- nössä ne sähköiset tietokannat, jotka Jyväskylän Yliopiston opiskelijoille ovat avoimna. Näitä ovat käytännössä JYKDOK/FINNA (<https://jyu.finna.fi/jykdok/>) ja Google Scholar (<https://scholar.google.fi/>), niiden kautta löytyi suurin osa lähteistä niiden käyttäessä tunnettuja kansainvä- lisiä tietokantoja materiaalin hakemiseen, kuten Springer Link (<http://link.springer.com/>), Science Direct (<http://www.sciencedirect.com/>), IEEE (<http://ieeexplore.ieee.org/>) tai ACM DL (<http://dl.acm.org/>). Tietoa

haettaessa käytettiin hakusanoina termejä ”Big Data”, ”Forest Based Industries”, ”Business and IT alignment”, ”Paperboard”, ”Digitalization” sekä ”IT alignment.” Näiden määritelmien lisäksi käytettiin kyseisten määritelmien johdannaisia. Metsäteollisuudesta oli hieman vähemmän tutkimusta, mutta liiketoiminnan ja IT:n yhteenlinjauksesta oli sitäkin enemmän laadukasta tutkimusta. Lopulta tulokset karsittiin julkaisuvuoden ja tekijöiden mukaan, joka antoi hyvän yleiskatsauksen prosessiteollisuuden, sekä Big Datan ja siihen liittyvien liiketoiminnan ja IT:n yhteenlinjauksen mahdollisuuksien kirjallisuuskatsauksen muodostamiseen.

Vaikka sähköiset artikkelit olivatkin käytetyin lähdemuoto, myös muita lähteitä ja kanavia käytettiin mikäli ne koettiin merkityksellisiksi. Esimerkiksi erilaisia ammattilehtiä oli mahdollista käyttää hyväksi paitsi aiheeseen tutustuessa, niin myös standardeja luodessa. Kuitenkin täytyy muistaa, että mitä tahansa lähdettä kriittisesti arvioitaessa tulee pohtia lähteen kykyä tuottaa luotettavia tuloksia ja sitä kautta löydöksiä. Hirsjärvi (2007), alleviivaa neljä osa-aluetta, joihin tulee keskittyä kun arvioidaan potentiaalista lähdetä.

Kirjoittajan maine ja hänen yleinen arvostuksensa  
Lähteen ikä ja alkuperä  
Lähteen uskottavuus ja julkaisijat  
Objektiivisuus ja taloudellinen riippumattomuus

Näitä kohtia tarkastellessa voidaan todeta, että toinen ja kolmas kohta ovat korkeimmin arvostettuja. Käytännössä tämä tarkoittaa sitä, että yritetään etsiä ajankohtaisimmat lähteet jokaiseen aiheeseen kunnioitetuilta julkaisijoilta. Lisäksi on todettava, että käytettäessä ainoastaan kunnioitettuja julkaisijoita, kirjoittajan arviointi ei ole niin tärkeää. Kuitenkin kunnioitetun kirjoittajan tekstiä saatetaan käyttää, mikäli se koetaan iästään huolimatta silti arvokkaaksi. Listan viimeisenä kohtana tarkoitetaan sitä, että vältetään lähteitä joista voi selkeästi nähdä, että niissä on yhtymäkohtia potentiaaliseen taloudelliseen hyötyyn. Esimerkiksi yritysten julkaisemat tai sponsoroimat tutkimukset ovat tiettyyn pisteeseen asti aina puolueellisia ja niitä voidaankin käyttää ainoastaan esimerkin muodossa. Nämä esimerkit on kuitenkin tuotu niin selkeästi esiin, että kriittisen ja läpinäkyvän kannan muodostaminen ei ole lukijalle vaikea tehtävä.

## 2.3 Empiirinen metodi

Koska tavoitteena on kerätä riittävä ymmärrys nykyisestä tilanteesta metsäteollisuuden eri organisaatioissa, empiirisen metodin tulee heijastella tuota tilannetta. Tutkijalla on vahva kokemus metsäteollisuudesta ja työskenteli IT-roolissa roolissa metsäteollisuudessa tutkimuksen valmistumisen aikana. Tämä helpotti

empiirisen lähestymistavan omaksumista. Pääasiallinen empiirinen metodi tutkimuksessa olivat haastattelut joihin pyydettiin osallistumaan teollisuudessa toimivia, insinööri - ja päällikkötasolla operoivia toimihenkilöitä. Tutkimuskysymyksen tarkoituksena on kartoittaa niitä haasteita, joita metsäteollisuuden avainhenkilöillä on Big Datan kanssa liiketoiminnan ja IT:n yhteensovittamisen kanssa. Avainhenkilöiden mukaan tällaista tutkimusta tarvitaan ja nämä haastattelut ovatkin erinomainen tapa kerätä tietoa asian tiimoilta.

Haastattelut ovat erittäin joustava tapa kerätä dataa ja tämä tapa antaa myös mahdollisuuden mukautua haastateltavan tietopääomaan. (Hirsjärvi, 2007) Haastattelut esimerkiksi antavat tutkijalle mahdollisuuden esittää samankaltaisen agendan kysymyksiä lukuisille eri ihmisille eri osista eri organisaatioita. Haastattelut antavat myös mahdollisuuden keskittyä tiettyihin kohdealueisiin, jotka kiinnostavat tutkijaa erityisesti. (Hirsjärvi, 2007) Vaikka opinnäytteen teemat ja kysymykset ovat samat kaikille, on hyödyllistä että on mahdollisuus keskittyä tiettyjen haastateltavien kanssa tiettyihin alueisiin, joista heillä on huomattavasti enemmän tietoa.

Joka tapauksessa haastatteluissa on muutamia luonteenomaisia ongelmia, jotka tulee tunnistaa. "Saunders (2009) nimeää kolme suurinta haastetta datan laadussa haastatteluiden osalta:

Luotettavuus

Puolueellisuus

Oikeellisuus ja yleistettävyys

Luotettavuus ja haastateltavan mahdollinen puolueellisuus ovat jossain määrin toisiinsa liittyviä ongelmia. Käytännössä se tarkoittaa sitä kysymystä, että voiko haastattelija luottaa että haastateltavat kertovat totuuden, eivätkä jätä kiusallisia tai henkilökohtaisia asioita mainitsematta. (Saunders, 2009) Tämän välttääkseen haastattelut on luvattu pitää täysin luottamuksellisina ja tulevat analyysit tullaan myös tekemään anonymisti. Datan mahdolliset luotettavuusongelmat minimoidaan siten, että haastattelut keskittyvät aitoihin ympäristöihin ja aitoihin ongelmiin. Kysymykset voivat liittyä oikeisiin ongelmiin joissa päätöksentekoon olisi tarvittu lisää tukea ja siten myös helpottaa varsinaisten ongelmien havaitsemista. Oikeellisuuden ja yleistettävyuden haasteisiin vastataan ottamalla useampia haastateltavia samankaltaisista työtehtävistä. Kysymällä samoja kysymyksiä eri henkilöiltä samasta työtehtävästä mahdollistetaan tulosten yleistäminen ja siten vahvistetaan niiden oikeellisuus.

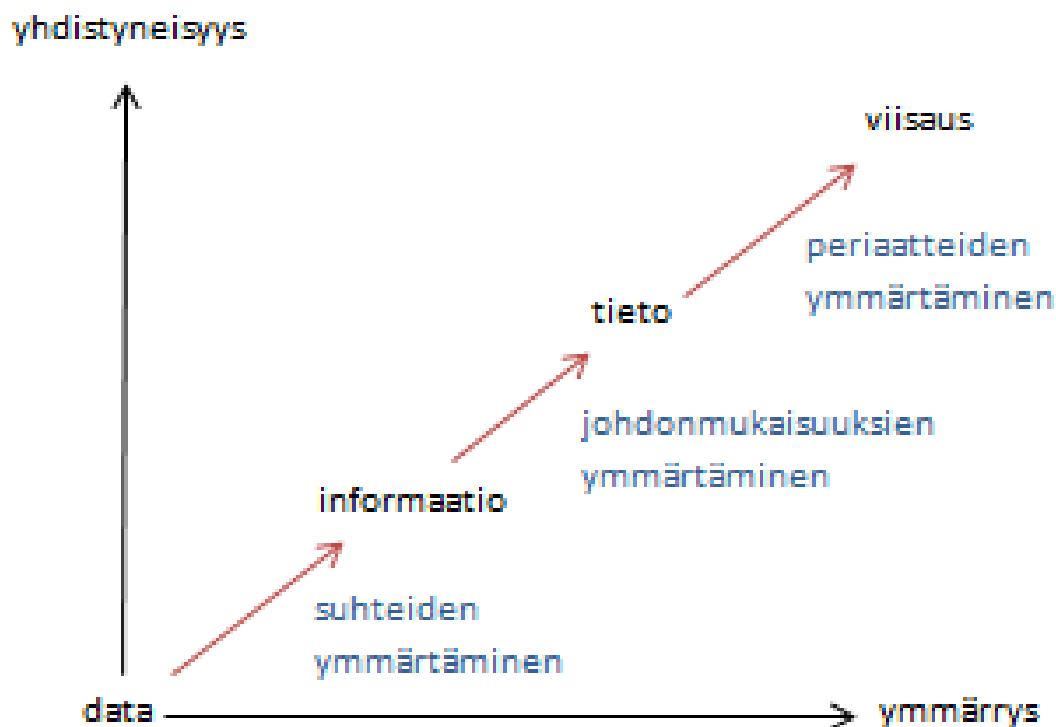
Kysymykset löytyvät työn viimeisiltä sivuilta liitteenä. Liiketoiminnan ja IT:n yhteenlinjaukseen Big Datan muodossa prosessiteollisuudessa keskitytään tässä opinnäytteessä suhteellisen yleisellä tasolla ja kuitenkin haastateltavien henkilöiden haasteet työelämässä voivat olla hyvinkin monitahoisia. Keskustelu haastatteluiden muodossa mahdollistaa tutkijalle oivan etenemispolun niille alueille, joissa haasteita eniten on. Haastattelu mahdollistaa myös laadullisen ja määrällisen analyysin haastatteludatasta. (Hirsjärvi, 2007) Haastatte-

luiden kerryttämän datan kategorisoinnilla on mahdollista muodostaa numeerista vertailua empiirisen datan perusteella.

### 3 BIG DATAN TEOREETTINEN PERUSTA

Tässä luvussa käydään läpi Big Datan perusominaisuuksia ja sitä, millaisia sovelluksia ja teknologioita tarvitaan, jotta raskaan teollisuuden organisaatioiden on mahdollista ottaa käyttöön Big Data-analytiikkaa. Lisäksi määritellään, mitä Big Data-analytiikka lopulta oikein on. Yrityksille ja yhteiskunnalle on muodostumassa haasteeksi se, kuinka valtavalla vauhdilla kasvavaa ja sen lisäksi vielä monipuolistuvaa dataa pystytään nyt ja tulevaisuudessa tallentamaan, hallitsemaan ja analysoimaan. Verkkoon, oli se sitten internetiin tai organisaation sisäiseen verkkoon kytketään nykyään laitteita jatkuvasti kiihtyvällä tahdilla ja ne myös tuottavat tietoa entisestään kiihtyvällä tahdilla. Teknologian kehittyminen mahdollistaa entistä tehokkaamman datan luomisen, tallentumisen ja siten datamäärät jatkavat kasvuaan. (Salo, 2013)

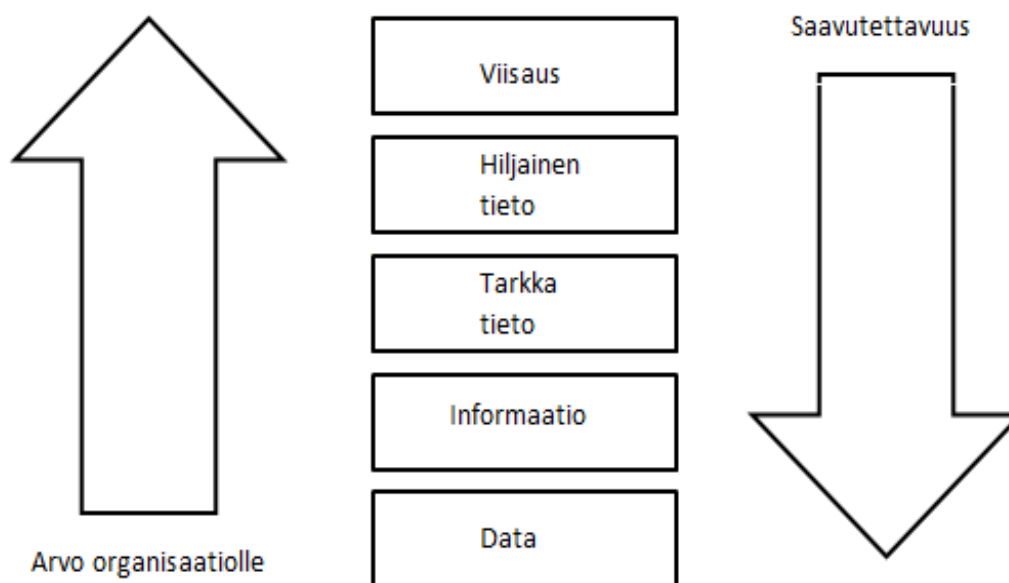
Datan määrä onkin tärkein tunnusmerkki sille, että "Big Data" -termiä voidaan käyttää. Normaalisti teratavua on pidetty Big Datan tilavuuden alarajana, mutta myös tiedostojen, asiakirjojen, taulukoiden tai transaktioiden lukumäärää voidaan käyttää datamäärän kuvaamiseen. (Russom, 2011) Datamäärienkään kuvaaminen ei vielä yksin riitä, vaan siitä tulisi pystyä määrittämään tietoa, jolla on useita tasoja. Brelade & Harman (2003) toteavatkin, että "eräs mahdollisuus viisauden, tiedon, informaation ja datan keskinäisten suhteiden määrittämisen selittämiseen on todeta, että datalla ei ole kontekstia. Vasta kun data sijoitetaan ennalta määritellyyn kontekstiin siitä syntyy informaatiota. Kun yhteydet informaation osien välillä on järjestetty ja kaavat ymmärretty, informaatio muuttuu tietämykseksi. Kaavojen periaatteiden ymmärryksen selittäminen taas muuttaa tietämyksen viisaudeksi." (Kuvio 3)



KUVIO 3. Datasta viisauteen (Brelaide & Herman, 2003, s. 113 mukaan)

Tällä hetkellä organisaatiot ympäri maailman miettivät, kuinka hallita heille kertyviä suuria datamääriä. Metsäteollisuudessa on vasta viime aikoina herätty tietojen analysoinnin hyötyihin, joka väistämättä johtaa uusiin haasteisiin kuten tehokkaaseen tallentamiseen ja käsittelyyn sekä tietojen tehokkaaseen analysointiin. (Lublinsky, Smith & Yakubovich, 2013) Tähän liittyy läheisesti käsite siitä, ketkä organisaatiossa pystyvät laittamaan saamansa datan oikeaan kontekstiin muodostaakseen siitä näin informaatiota. Tiedon arvo myös kasvaa, mikäli se on saatavilla mahdollisimman nopeasti ja helposti silloin kun sen tarve huomataan. Se että tieto saadaan mahdollisimman tehokkaaseen käyttöön vaatii ymmärrystä tietopääomasta ja sen johtamisesta. Lopulta voidaankin todeta, että data on havaintoja, informaatio on kontekstiin laitettua dataa ja tieto itsessään kumpuaa kokemuksista ja heijastuksista. Näin päästään lopputulokseen, jossa data itsessään ei tuo arvonlisää vaan se syntyy työntekijän kautta hänen tietojen ja taitojen lopputuloksena. (Zack, 1999) (kuvio 4, Wilson & O'Conor 2000).





KUVIO 4. Tiedon eri tasojen arvo yritykselle (Wilson & O’Conor, 2000, s.121)

### 3.1 Tietopääoma

Prosessia jossa datasta syntyy informaation kautta tietoa ja viisautta kutsutaan tietopääomaksi. Tämä tietämys mahdollistaa aineellisen ja aineettoman pääoman, eli substanssin muodostumisen. Tietopääoma voidaan siis Stähle & Grönroos (1999) sanoin määritellä ”yrityksen aineettomat omaisuuserät sekä kyky käyttää informaatiota ja osaamista uusien ideoiden ja innovaatioiden jatkuvaan tuottamiseen.” Aineettoman pääoman hankkiminen lisää kilpailukykyä innovoinnin ja strategisen syvyyden kautta, mutta sen syntymiseen vaaditaan inhimillistä pääomaa työntekijöiden muodossa. Metsäteollisuuteen sovellettuna voidaan käyttää esimerkkiä, jossa tietopääoma on osa yrityksen markkina-arvoa yhdessä taloudellisen pääoman kanssa. Tällöin on mahdollista laskea tietopääoman arvoksi kirjanpito- ja markkina-arvon erotus vaikka toki täytyy muistaa, että markkina-arvoon vaikuttavat tietopääoman lisäksi myös spekulatiot ja oletukset tulevaisuudesta. (Edvinson & Malone, 1997)

Vaikka tietopääomalle on mahdollista laskea arvo ja myös mahdollinen arvonlisä, myös massadataa hyödynnettäessä tulee pitää mielessä ettei tieto itsessään tuota arvoa, vaan arvo muodostuu lopulta siinä vaiheessa kun tavoitetta saavutettaessa tietoa tuota tietoa pystytään hyödyntämään. Tällöin voidaan puhua tiedon käytöstä osana päätöksentekoa, jolloin tiedon arvo saavutetaan vasta tiedonhallinnan prosessin myötä. Tieto, joka on oleellista ja oikealla instanssilla oikeaan aikaan tuottaa suurimman mahdollisen hyödyn. (Laihonon ym. 2013) Käsillä oleva esiin puristettu tietomäärä on usein ainoastaan pieni osa siitä tiedosta, joka organisaatiossa todella on. Tämä vaikeuttaa yrityksen todellisen arvon laskemista, sillä kuten edellä todettiin yrityksen arvo koostuu sekä

aineellisista, että aineettomista omaisuususeristä ja aineeton omaisuus on suurimmaksi osaksi näkymätöntä arvoa. (Hannula ym. 2013)

### 3.2 Datan monimuotoisuus

Datan olemus on monimuotoinen, massadatassa data on sekä strukturoitua, osittain strukturoitunutta sekä strukturoimatonta. (Russom, 2011) Strukturoidulla datalla viitataan sellaisiin tietoihin, joille on määritelty tietynlainen pituus ja koko, eli käytännössä selkeä rakenne. Strukturoitu eli jäsennelty data voi sisältää mitä tahansa numeroita ja kirjaimia sekä niiden muodostamia ryhmiä, näitä ryhmiä kutsutaan stringeiksi. Niiden rakenne pysyy samana ja sen vuoksi strukturoitu data on helposti käsiteltävää. Esimerkiksi tietokannat ovat strukturoidun datan kokoelmia. Datan rakenne on siis strukturoitua, sillä sen tallentaminen rakenteeseen on selkeää ja ennalta määriteltyä. (Rouse 2012) Strukturoitua dataa on mahdollista käsitellä perinteisin tiedonhallintamenetelmin ja sitä voidaan tallentaa perinteisiin relaatiotietokantoihin perustuviin järjestelmiin. (Siiramaa, 2012)

Jäsentelemätön, eli strukturoimaton data on rakenteetonta. Se voi olla vaikkapa tekstiä, videota tai ääntä, mitä tahansa koneen tai ihmisen luomaa tietoa. Vaikkapa organisaatiossa syntyvät raportit tai facebook-päivitys. Osittain strukturoitu data taas on kahden edellisen välimuoto, sisältäen molempia datatyyppisiä eli jonkin tasoista rakennetta, jota ei kuitenkaan voida suoraan analyttisesti hyödyntää. Lokitiedostot toimivat tyyppillisenä esimerkkinä osittain strukturoidusta datasta. (Zikopoulos, Eaton, Deroos, Deutsch, & Lapis, 2012)

### 3.3 Big Datan määritelmä

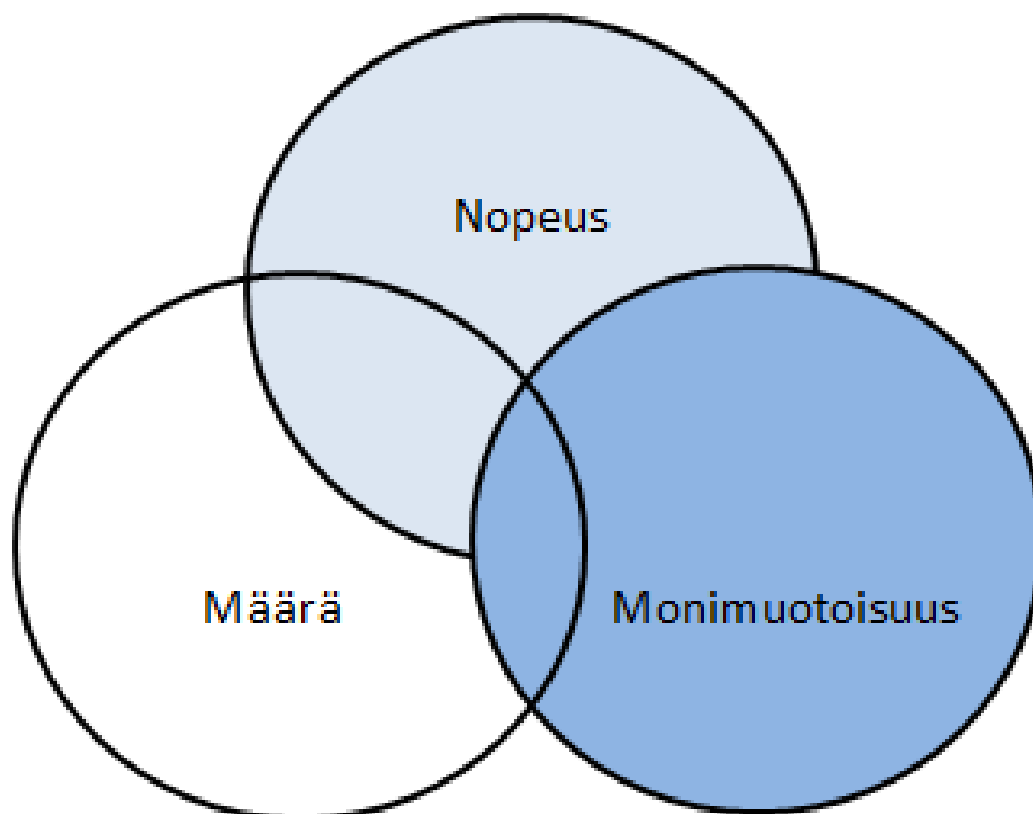
Termiä ei vielä kukaan määritellyt yleisesti ja tarkasti, useat eri tahot ovat kuitenkin pyrkineet määrittelemään sen ja vaikkapa Mikkelän (2012) mukaan käsitteellä tarkoitetaan valtavia tietomassoja, jotka sisältävät erittäin hajanaista, vaihtelevaa ja nopeaa dataa. Ne sisältävät monimuotoisia ja -sisältöisiä elementtejä, kuten ääntä, kuvaa, tekstiä, numeroita ja tilastoita. Toisaalta taas Jacobs (2009) toteaa big datan määrittelyn sidosteiseksi aikaan, 1990-luvulla big data tarkoitti eri asiaa kuin miten se nykyään määritellään.

Big Dataa on terminä valtavan kritisoitu, käsitteenä sille on ennustettu hyvinkin lyhyttä elinkaarta sen hankalan määriteltävyyden sekä yleisen sekavuuden johdosta. (Davenport, 2014) Siiramaa (2012) taas toteaa, että "Big Datalla tarkoitetaan missä tahansa muodossa olevaa, nopeasti muodostuvaa dataa, jota ei pystytä käsittelemään tehokkaasti perinteisillä järjestelmillä." On kuitenkin merkillepantavaa, että liiketoiminnassaan Big Dataa käyttävät yritykset eivät liiketoimintaansa kuvatessaan käytä Big Data - käsitteitä, vaan he pyr-

kivät tuomaan asian esille käytännönläheisemmin, kuten “Analytiikkamme poimii ylikuumenevat pisteet prosessista, jotta voisimme paremmin ymmärtää missä prosessin vaiheissa riskikohteet sijaitsevat.” Se on huomattavasti helpommin ilmaistu tavalliselle asiakkaalle kuin sanoa, että “Teemme Big Data -analyysiä.”

### 3.4 Big Data 3V - käsitelmä

Toinen tapa määritellä Big Data -terminä on siten, että jokaisen tietolähteen tulee täyttää 3V-käsitelmän tunnusomaiset piirteet, jotta se voidaan tulkita Big Dataksi. (Patil & Thia, 2013) Perinteisesti Big Dataa kuvataan 3V-käsitelmällä, jossa Volume on määrä, Variety on monimuotoisuus ja Velocity on nopeus. Big Data -käsitteeseen on mahdollista liittää myös neljäs V (Value) eli arvo (kuvio 5). Arvo kuvaa hyvin sitä liiketoimintahyötyä, jota Big Data tuo. (Remes, 2013) Hurwitz, Nugent, Halper & Kaufman (2013) tuovat luokitteluun vielä kaksi luokkaa lisää, Validityn eli datan oikeellisuuden ja tarkkuuden sen tarkoitukseen ja Volatility joka kertoo sen, kuinka kauan kyseistä dataa tulee varastoida.



KUVIO 5. Määrä, Nopeus, Monimuotoisuus (Salo, 2013, s. 23)

### 3.4.1 Määrä

Datan määrä on Big Datan ominaisimpia tuntomerkkejä ja kuten aiemmin tekstissäni totesin, Big Datan alarajana olevaa tietomäärää pidetään perinteisesti yhtenä teratavuna. Esimerkiksi Twitterin käyttäjät tuottavat noin 12 teratavua erilaisia twiittejä joka päivä. Kyseistä dataa analysoimalla on mahdollista selvittää ihmisten suhtautumista tutkimuksen alla oleviin palveluihin tai tuotteisiin. (Morabito, 2015) Edellisen esimerkin 12 teratavua päivässä saattaa tuntua valtavalta määrältä, mutta fakta on kuitenkin se että dataa tallentuu joka puolella ja keräämämme datan määrä kasvaa jatkuvasti. Elektroniikalla on mahdollista seurata mitä tahansa lähtien taloudesta ja lääketieteestä päättyen valvontaan ja ympäristöön. Lokitietoihin tallentuu jokainen klikkaus koneella ja organisaatioilla on lopulta tietopankeissaan valtavia määriä tietoa datan muodossa. Tätä dataa on osattava hallita tai dataa hallinnoiva organisaatio hukkuu siihen. Mikäli organisaatio osaa käyttää sopivia teknologioita ja työkaluja, suurestakin datamäärästä on mahdollista analysoida käytännössä kaikki hyödyllinen data. Tämän tuloksena on mahdollista syntyä parempi ymmärrys niin omasta liiketoiminnasta, kuin myös asiakkaista ja markkina-asemasta. (Zikopoulos, 2012)

### 3.4.2 Nopeus

Pelkistettynä nopeudella tarkoitetaan sitä vauhtia, jolla uutta dataa muodostuu järjestelmässä. Erikseen kannattaa huomioida, että Big Datassa nopeudella on muutakin merkitystä kuin datan saapumis-, tallennus- ja hakuaika. Perinteinen esimerkki nopeudelle on sosiaalisen median tietovirran ja tietolähteiden analysointi sekä yhdistely. (Prajapati, 2013) Datavirtojen vauhti onkin ensisijainen tarkastelun kohde nopeutta tarkasteltaessa ja nykypäivän organisaatioissa datavirrat ovatkin jatkuvia sensorimäärien (mm. RFID) kasvun ja muiden tietovirtojen johdosta. (Zikopoulos, 2012)

Nopeudessa etu kilpailijoihin tulee siitä, että organisaatio pystyy tunnistamaan kehityssuuntauksia, olivat ne sitten ongelmia tai mahdollisuuksia ennen kilpailijoitaan. Useimmissa liiketoiminnoissa tulee pystyä lähes reaaliaikaiseen data-analyysiin, jotta analyysistä on mahdollista puristaa kilpailuetua muihin verrattuna. Reaaliaikainen data-analyysi on tärkeää myös datan säilyvyyden vuoksi, sillä nykyään tuotetun datan käytettävyyss-/säilyvyysaika on lyhyt. (Zikopoulos, 2012) Uuden tiedon syntynopeudesta ja siihen liittyvistä haasteista hyvinä esimerkkeinä toimivat verkkomainokset, jotka tarkkailevat käyttäjien toimintaa tuhansia, jopa miljoonia kertoja sekunnissa. Toinen erinomainen esimerkki ovat äärimmäisen nopeat osakekauppa-algoritmit, jotka raportoivat markkinoiden muutoksia mikrosekunneissa. (MongoDB, 2016)

### 3.4.3 Monimuotoisuus

Monimuotoisuudella pyritään kuvaamaan datan olemusta ja siinä tapahtuvaa datan muuttumista yhä epäyhteneväisemmäksi. Big Datassa yhdistyy siis strukturoitu, strukturoimaton ja osittain strukturoitu data. Big Datan avulla voidaan tuoda kaiken tyyppinen data analysoitavaksi ja organisaatioiden hyödynnettäväksi. (Zikopoulos, 2012) Monimuotoisuutta ei kuitenkaan tule sekoittaa SAS:n (2013) Big Data-määritelmiin lisäämiin vaihtelevuuteen (Variability) ja monimutkaisuuteen (Complexity). Näistä vaihtelevuudella tarkoitetaan datavirtoja ja niiden jaksoittaisia piikkejä, joista syntyy epäjohdonmukaisuuksia. Monimutkaisuudella taas pystytään kuvaamaan eri data-kokonaisuuksien yhteenmukauttamisen vaikeutta. Jotta yritys pystyy yhdistelemään dataa eri liiketoimintayksiköiden ja järjestelmien välillä, tulee ymmärtää kuin eri lähteistä tuleva data on mahdollista liittää yhteen.

## 3.5 Big Datan hyödyntäminen

Big Datan hyödyntäminen vaatii paitsi teknistä taitoa ja mahdollisuutta käsitellä ja muokata nopeaa, monimuotoista ja määrältään valtavaa dataa. Näihin ominaisuuksiin kun yhdistetään fyysisen arkkitehtuurin vaatimukset alkaen käyttövarmuudesta ja skaalautuvuudesta päättyen tehokkuuteen ja joustavuuteen ollaan useiden erilaisten valintojen edessä. Mikä on kyselyn valmistumisen maksimiviive? Mikä on järjestelmän ylläälläoloprosentti? Entä laskentavoima ja kerättävän/tallennettavan datan määrä? Päivitetäänkö vanhat järjestelmät osaksi uusia järjestelmiä, vai aloitetaanko rakennetaanko kaikki alusta? Vanhoja järjestelmiä ei yleensä voi suoraan integroida tai analysoida jolloin mahdolliset yrityksen päivittää niitä voi olla erittäin haasteellista. Tällöin voi olla kustannustehokkaampaa jättää vanha tieto hyödyntämättä ja aloittaa modernin Big Data - järjestelmän rakentaminen puhtaalta pöydältä. (Manyika ym, 2009)

Millaisilla ohjelmistoilla Big Dataa sitten on mahdollista nykypäivänä hyödyntää? Kuinka reaaliaikaisen tiedon analysointi, piilossa olevan tiedon esiin kaivaminen ja prosessien automatisointi sekä ennakointi tuottavat organisaatiolle lisää sen kaipaamaa kilpailukykyä? Yleisesti arvioidaan päätöksenteon tuottavuuden kasvavan kuusi prosenttia kun prosessiin liitetään Big Data-analysointi. McKinsey arvioi Big Datan vaikuttavan tuottojen kasvuun jopa 60%. (Autio, 2014) Tuottojen kasvua ja liiketoimintaongelmien ratkaisemista data-analytiikan keinoin kuitenkin rajoittaa se, ettei siitä ole kirjoitettu vielä kovinkaan paljon tieteellisiä artikkeleita. (Martinez-Lopez & Casillas, 2013) Alan potentiaalia kuvaa hyvin se, että tällä hetkellä arviolta noin viisi prosenttia joka päivä tuotetusta 2,8 zettatavusta analysoidaan edes jollain lailla. Vaatiikin useiden uusien työkalujen kehittämistä, että mielekäs osa tästä datasta saadaan

rakeneellistettua ja analysoitua, jonka jälkeen voidaan alkaa miettimään mahdollista arvonlisää. Tämänkin jälkeen tulee ottaa huomioon, että arviolta 25% tuotetusta datamassasta on millään tapaa hyödyllistä. (Davenport, 2014)

## 4 BIG DATA TEKNOLOGIAT JA NIIDEN MAHDOLLISUUDET METSÄTEOLLISUUDESSA

### 4.1 Kunnossapito

Kunnossapidon liiketoiminnallinen merkitys kasvaa vuosi vuodelta kilpailun kiristyessä ja teollisuuden kunnossapito on kovien muutospainoiden kohteena. Kunnossapitotoimintoja ulkoistetaan ja erillisillä kunnossapitoyrityksillä pyritään hillitsemään kunnossapitokustannusten nousua. Selkeänä trendinä on kuitenkin ollut keskittyminen ennakoivaan ja ehkäisevään kunnossapitoon perinteisen korjaavan kunnossapidon sijaan. Lisäksi tietotekniikan merkitys on kasvanut vuosi vuodelta myös kunnossapidon saralla, tietojärjestelmät auttavat kustannusten hallinnassa ja toiminnan tehostamisessa. Käyttäjälähtöinen tietojärjestelmä on suuri kustannussäästö pääomavaltaisessa metsäteollisuudessa.

Kunnossapidon määrittelee parhaiten standardi SFS-EN 13306: ”Kunnossapito koostuu kaikista kohteen elinajan aikaisista teknisistä, hallinnollisista ja liikkeenjohdollisista toimenpiteistä, joiden tarkoituksena on ylläpitää tai palauttaa kohteen toimintakyky sellaiseksi, että kohde pystyy suorittamaan vaaditun toiminnon”. Kunnossapito ei ole mitenkään sidottu tiettyihin toimialoihin, vaan sitä esiintyy ainakin kaikilla toimialoilla, joissa on jonkinlaisia laitteita käytössä. (Järviö, Piispa, Parantainen & Åström, 2007)

Kunnossapito on olennaisesti sidottu tuotannon kokonaistehokkuuteen, jossa kunnossapidon tehtävänä on pitää koneiden ja laitteiden käyttövarmuus ja tehokkuus mahdollisimman korkeana. Kokonaistehokkuutta pystytään tarkastelemaan lähemmin analysoimalla KNL-menetelmän avulla prosessilaitteiden käytettävyyttä, laatua ja nopeutta. Käytettävyydellä tarkoitetaan järjestelmän mahdollisuuksia suoriutua käyttäjän määrittelemistä operaatioista määrättyinä ajankohtana. Käytettävyyys koostuukin kunnossapitovarmuuden ja kunnossapidettävyyden lisäksi myös toimintavarmuudesta (Leiviskä, 2009)

Huollon ja ehkäisevän kunnossapidon raja ei ole selkeä, mutta huolto voidaan mieltää laitetoimittajan määrittelemänä ennakkoon suunniteltuna toimenpiteenä, joka on välttämätön laitteiston toiminnan kannalta. Huollon tehtävänä

onkin pitää laite toimintakuntoisena tai palauttaa alentunut suorituskyky ennen mahdollista vikaa tai laiterikkoa. (Järviö & Lehtiö, 2012)

## 4.2 Tekniikat ja teknologiat Big Datan hyödyntämiseen

Tunnetuimmat tekniikat Big Datan hyödyntämiseen ovat Hadoop, NoSQL, pilvipalvelut ja muistinvarainen analytiikka. Näistä metsäteollisuuden kannalta mielenkiintoisin työväline on Hadoop, joka avoimena ohjelmistokehityksenä tarjoaa sekä kustannus-, että skaalaetuja jopa suurille metsäteollisuuden yrityksille. Lisäksi sen päälle voidaan rakentaa mitä erilaisimpia sovelluksia, joilla kerätty data pystytään muuttamaan muotoon, jota myös perinteiset analyysityökalut ymmärtävät. (Hammarsten, 2013) Massadatan hallintajärjestelmät ja sovellukset tarvitsevatkin uusia kehittyneitä tapoja hallita, analysoida ja visualisoida tallennettua dataa. (Chen, Chiang & Storey, 2012)

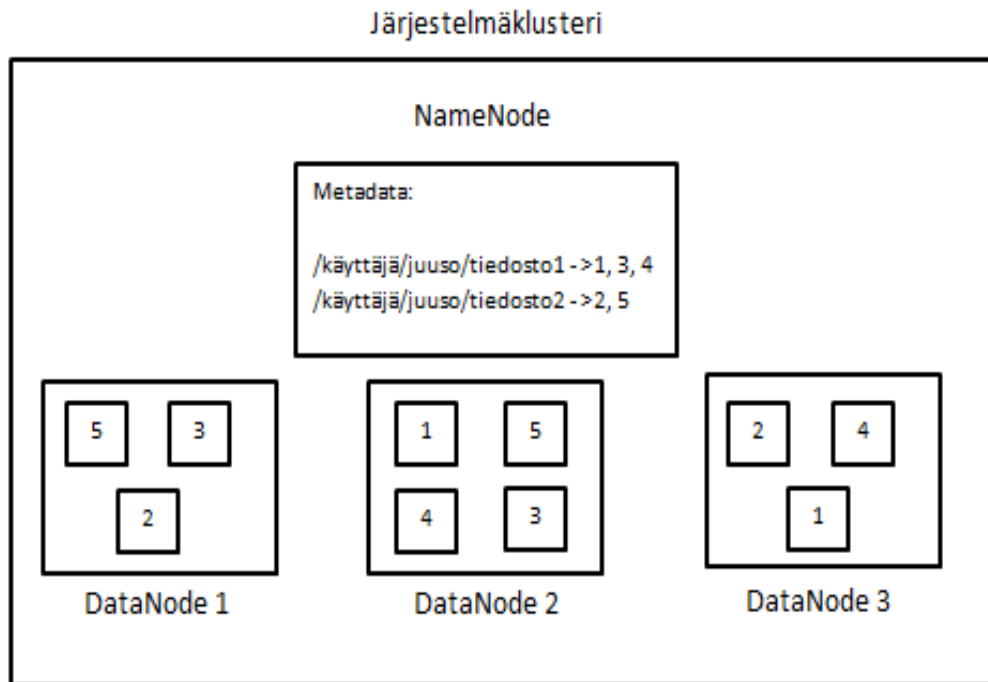
## 4.3 Hadoop

Hadoop on avoimen lähdekoodin ohjelmistoprojekti Big Datan kustannustehokkaaseen käsittelyyn. (Kuvio 6). Sitä käyttäen on mahdollista luoda palvelin-klusteri, jota voidaan käyttää suuren ja useita eri tiedostomuotoja sisältävän datamäärän tallentamiseen ja analysointiin alustasta riippumatta ja joustavasti. Hadoopissa tiedostot jaetaan lohkoiksi, jotka kopioidaan ja levitetään jokaiselle verkkoon kytkettyyn palvelimeen, jossa Hadoopia käytetään. Tämän jälkeen jokainen pyyntö jaetaan pienemmiksi tehtäviksi, jolloin pyyntöä suoritettaessa kaikki tehtävät ajetaan erillisillä palvelimilla ja tulos palautetaan käyttäjälle vasta, kun kaikki yksittäiset tehtävät on suoritettu. (Fischer, Su & Yin, 2011) Data myös tallennetaan oletuksena kolmena kopiona toimintavarmuuden lisäämiseksi. (Siiramaa, 2012) Hadoop-projekti sisältää neljä moduulia joista Hadoop Common sisältää muita moduuleja tukevat yleiset palvelut, HDFS eli hajautettu tietojärjestelmä, YARN-ohjelmistokirjasto jota tarvitaan tehtävien aikatauluttamiseen ja klustereiden resursointiin sekä MapReduce moduuli datan käsitteelyyn rinnakkaisesti.

Hadoopissa käytetty hajautettu tiedostojärjestelmä (HDFS) on suunniteltu tallentamaan ja edelleenvälittämään erittäin suuria datajoukko, sen arkkitehtuuri koostuu yhdestä nimisolmusta ja lukuisista datamolmuista. Tiedostojärjestelmässä sovelluksen data tallennetaan eri datamolmuille, metadatan tallentuessa nimisolmuun. Järjestelmän palvelimet keskustelevat toistensa kanssa TCP-pohjaisen protokollan avulla. (Shvachko, Kuang, Radia & Chansler, 2010). Järjestelmät vaativat korkeita tiedonsiirtonopeuksia peräkkäisiä luku- ja kirjoitusprosesseja varten. Järjestelmän piirteinä tehokkaan toiminnan saavut-



tamiseksi onkin tallennettavien tiedostojen suuri koko. Tiedostojen oletuskoko jaetuille lohkoille onkin 64 mt. (Botharkur, 2007)



KUVIO 6. Hadoopin arkkitehtuuri tiedostolohkojen kopiointikertoimen ollessa 2

Koska Hadoopin ydinteknologiat HFDS (Hadoop Distributed File System) sekä MapReduce ovat vapaan lähdekoodin ohjelmistoja, niitä on helppo kehittää eteenpäin myös pienemmissä yksiköissä. Täten on mahdollista, että jokainen metsäteollisuuden toimipiste pystyy kehittämään omia, hieman muokattuja versioita näistä järjestelmistä. Tämä helppo muokattavuus lisää entisestään järjestelmän houkuttelevuutta, sillä järjestelmän pohjalta muodostuu jatkuvasti innovaatioita, joita muut voivat ottaa käyttöön yksiköstä riippumatta.

#### 4.4 MapReduce

MapReduce on Hadoopin ydintoimintoja, se on laskentamalli jonka avulla pystytään ohjelmoiden käsittelemään ja analysoimaan suuria datamääriä tavallisista palvelimista muodostuvista palvelinklustereista sekä purkamaan niihin kertynyt datamassa pienemmiksi palasiksi, jotka jaetaan klusterin muodostavaksi solmuiksi. (Schmidt & Phillips, 2013) MapReducen rajapinta koostuu kahdesta funktiosta, kartoita (map) ja supista (reduce), jotka mahdollistavat mallin käyt-

tämisen lukuisissa sovelluksissa, esim. koneoppimisessa, hakukoneissa ja datan prosessoinnissa. (Yang, Dasdan, Hsiao & Parker, 2007)

Hyvänä esimerkkinä MapReducen käytöstä toimii säätietojen analysointi. Sääanturit keräävät jatkuvasti tietoa ympäri maailmaa kooten samalla valtavan määrän lokitietoja, jotka ovat semistrukturoitua ja tallennepainotteista dataa, eli hyvä vaihtoehto MapReduce-analyysin kohteeksi. (White, 2012) MapReduce-kehyyksen tarkoituksena on mahdollistaa ohjelmoijan ilmaista pelkät suoritettavat laskutoimitukset piilottamalla kehyyksen taakse kätkeytyvät monimutkaisemmat prosessit, kuten kuormantasaus, vikasieto, rinnakkaisajo ja tiedonjakelu. Inspiraatio tämän kehyyksen luomiselle tuli funktionaalisista ohjelmointikielistä, jotka sisältävät kartoita ja supista funktiot. (Dean & Ghemawat, 2008)

MapReducen tarkoituksena on siis tuoda Big Data-analytiikan hyödyt myös ei data-analytiikan ammattilaisten ulottuville. Tästä olisi merkittävä hyöty metsäteollisuudessa, sillä valtava osa työntekijöistä ei osaa käyttää dataa analyytikon tavoin. Mitä helpommassa muodossa data on saatavilla, sitä paremmin ja innokkaammin metsäteollisuuden operaattorit sitä hakevat ja soveltavat. Lisäksi maallikolle on käytön kannalta helpompi ymmärtää, että HDFS-klusterin sisällä ajetaan ohjelmakoodia jolloin data ei liiku, vaan koodi liikkuu datan luokse.

## 4.5 HDFS

Hadoop Distributed File System on Hadoopin ydinkomponentteja ja suunniteltu suurten datamassojen hajauttamiseen klusterin palvelimille. Perinteisessä käytössä data kirjoitetaan ainoastaan kerran, mutta luetaan useasti vaikkapa analysointia varten. Tallennettava data jaetaan lohkoihin klusteriin, jolloin niitä on mahdollista käsitellä tehokkaammin sekä rinnakkaisprosessoida mahdollistaen nopeammat analyysit. (Lublinsky ym., 2013.) Analyysien nopeus on kriittistä metsäteollisuudessa, sillä vaikka vasteajat ovat pitkät niin prosessissa tapahtuu jatkuvasti kymmeniä, jopa satoja muutoksia venttiilien asentojen ja sitä myötä eri sakeuksien johdosta. Ketjureaktiot joissa asia vaikuttaa toiseen voi tehdä suuria muutoksia pidemmällä aikavälillä, tämän vuoksi dataa tulisi pystyä analysoimaan rinnakkain jotta pystytään varautumaan mahdollisimman moneen eri skenaarioon kun prosessissa muutoksia tehdään. HDFS:n vahvuutena on myös investointikustannusten maltillisuus, sillä järjestelmän ajamiseen ei tarvitse räätälöityjä erikoisjärjestelmiä. Tämä toki näkyy hieman korkeampana mahdollisuutena virheilmoituksiin, sekä korkeampana viiveenä tietojä käsitellessä.

## 4.6 Apache Spark

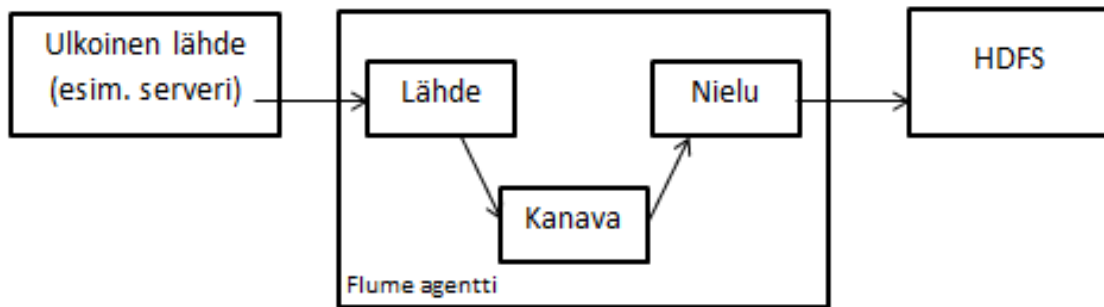
Spark on Hadoopin tapaan myös avoimen lähdekoodin ohjelmistokehitys hajautettuun laskentaan tietokoneklustereissa. Sen pääabstraktioksi kutsutaan sinnäkstä hajautettua datajoukkoa, joka kuvaa kokoelman usealle tietokonesolmulle ositettuja vain luku -objekteja. Kyseiset datajoukot on mahdollista varastoida välimuistiin ja niitä voidaan käyttää uudelleen MapReducen kaltaisia rinnakkaisia tehtäviä varten. Spark on parhaimmillaan Hadoopia jopa 20 kertaa tehokkaampi ja yhdelle teratavulle dataa sillä voi suorittaa kyselyitä 5-7 sekunnin viiveellä. (Zaharia, Chowdhury, Das, Dave, Ma, McCauley, Franklin, Shenker & Stoica, 2012) Nopeat viiveet tarkoittavat nopeaa reagoitua, päällystyskoneella vaikkapa yhden gramman täyteaineen lisäys neliometrille vaikuttaa jo merkittävästi sekä painettavuuteen, että opasiteettiin. Näiden lähes reaaliaikainen seuranta ja rinnakkaissimulointi on ensiarvoisen tärkeää korkeamman jalostusasteen tuotteissa.

## 4.7 Massadatan kerääminen

### 4.7.1 Apache Flume

Apache Flume on luotettava ja suosittu hajautetun datan syöttötyökalu, jolla pystytään syöttämään ja keräämään suuria määriä dataa eri lähteistä yhteen, keskitettyyn datavarastoon. (Apache, 2012) Flumea voi käyttää useiden, hyvin monimuotoisten datalähteiden kuten sosiaalisen median sisältöjen, palvelinlokien tai tietoverkkojen tuottaman tiedon käsittelyyn. Perinteisellä ryhmään kytkemättömälläkin tietokonelaitteistolla Flume pystyy käsittelemään datahävikkeettä kymmeniä tuhansia tapahtumia sekunnissa. (Wang, Rayan & Schwan, 2012)

Flumen pääkomponentit ovat tapahtumat, lähteet, kanavat, nielut ja agentit. Agentti siis koostuu lähteestä, kanavasta ja nielusta, joka joko vastaanottaa tai noutaa dataa lähdekomponenttiinsa välittäen sitä kanavakomponenttiensa kautta nielukomponenteille. Nielukomponentti taas välittää dataa eteenpäin tallennettavaksi tiedostojärjestelmään tai seuraavalle Flume-agentille.



KUVIO 7. Flume-agentin arkkitehtuuri (The Apache Software Foundation, 2012, Flume User Guide)

#### 4.7.2 Muut massadatan keräämiseen tarkoitettut ohjelmistot

Massadataa voidaan kerätä myös monilla muilla ohjelmistoilla, joista tunnetuimpina voidaan pitää Apachen Sqoopia, Apache Kafkaa ja Apache Chukwaa. Sqoop on relaatiotietokantoja Hadoopiin yhdistävä komentorivityökalu ja sitä käytetään, kun toimitaan Hadoop-ympäristön ulkopuolella esimerkiksi asiakaspalvelin yhteyksissä. (Zburivsky, 2013) Kafkaa taasen käytetään pääosin sellaisissa ympäristöissä, joissa eri järjestelmien tehokas keskinäinen kommunikointi on tärkeää. Chukwa taasen muistuttaa toimintaperiaatteeltaan Flumea, sillä sen arkkitehtuuri koostuu myös agenttiprosesseista ja kerääjäprosesseista. Datalähteistä tieto siirtyy agenttiprosessin hallinnoiman sovitinmoduulijoukon kautta kerääjäprosesseille ja sitä kautta HDFS-nielutiedostoihin. (Boulon, Konwinski, Qi, Rabkin, Yang E. & Yang M., 2008)

### 4.8 Massadatan analysointi reaaliajassa

Nykyiset analysointijärjestelmät ovat erittäin käyttäjälähtöisiä ja vaativat siten lyhyttä vasteaikaa. Usein onkin ongelmana se, kuinka jatkuvaa datavirtaa pystytään käsittelemään, kun eräajoanalyysimenetelmät eivät tarjoa ratkaisua informaation ajantasaiseen käsittelyyn. Dataa tuleekin siis analysoida välittömästi sitä mukaa, kun sitä saadaan kerättyä talteen järjestelmään. Apache Storm on tästä hyvä esimerkki ja yhdessä eräajotyökalujen kanssa se auttaa käsittelemään dataa huomattavan joustavasti vaatimuksista riippuen.

Metsäteollisuudessa eräajoanalyysit nousevat entistä suurempaan arvoon, sillä osaprosesseja on myös pystyttävä analysoimaan erilaisten simulaatioiden avulla. Voi olla että joku prosessin osa saa hyvinkin pienillä muutoksilla hyvinkin erilaisia vaikutuksia aikaan kokonaisprosessia ajatellen, jolloin olisi

tärkeää pystyä simuloimaan nopeasti pienten kokonaisuuksien vaikutuksia paitsi rinnakkain, niin myös syy-seuraussuhteen avulla koko järjestelmään.

#### **4.8.1 Apache Storm**

Apache Stormin on reaaliaikainen laskentajärjestelmä suurten datalähteiden käsittelyyn. Se mahdollistaa hyvin skaalautuvan ja vikaa sietävän työkalun, jolla jatkuvalla datavirralla voidaan suorittaa hajautettua laskentaa ja analyysiä samaan tapaan, kuin MapReduce-tehtävillä voidaan tehdä suurille dataerille tiedostojärjestelmien sisällä. Stormin keskeiset käsitteet ovat rännit ja pultit jotka yhdessä muodostavat dataa käsittelevän laskentaverkoston eli topologian. Rännit lukevat datalähteitä ja ohjaa ne topologiaan, jossa pultit ottavat datavirran vastaan ja aloittavat laskennan, jonka seurauksena syntyy ulos tuleva datavirta. (Apache, 2014) Storm olisi pääasiallinen työkalu konelinjakäytössä, etsitään ne asiat prosessista jotka eivät toimi halutulla tavalla ja korjataan ne, mikäli siten pystytään vaikuttamaan positiivisesti prosessiin.

### **4.9 Muut massadatan onnistuneen käyttöönoton kannalta oleelliset ohjelmistot**

#### **4.9.1 Apache HBase**

HBase on Hadoopin päälle rakennettu sarake-orientoutunut NoSQL-tietokantakomponentti massadatan varastointiin ja reaaliaikaiseen tai satunnaiseen lukemiseen. HBase ei sisällä toissijaisia indeksejä tai kehittyneiden sarakkeiden tietotyypppejä, mutta soveltuu hyvin miljardien rivien tallentamiseen. HBase pohjautuu Googlen BigTable-arkkitehtuuriin. (Chang, Dean, Ghemawat, Hsieh, Wallach, Burrows, Chandra, Fikes & Gruber, 2006) Järjestelmän taulut tallentuvat hajanaisena, moniulotteisena karttana joka sisältää rivejä ja sarakkeita. Jokaisessa solussa on myös oma aikaleimansa. Hadoopin hajautettuun tietojärjestelmään HBase siis tuo skaalautuvan, jaetun ja vikoja hyvin sietävän tietokannan sallien samalla satunnaisen pääsyn tallennettuun massadataan. (Taylor, 2010)

#### **4.9.2 Apache Hive**

Apache Hive antaa mahdollisuuden käsitellä Big Dataa SQL:n menetelmin. Sen avulla voidaan luoda massadatasta tietokantatauluja sarakkeineen ja kirjoittaa HDFS:ään natiivisti peräkkäistiedostoina. Hive-kysely luovat MapReduce-prosesseja joita ajetaan massadataa vasten rinnakkaisesti. Näin ollen Hive on paitsi datan varastointiteknologia, niin myös infrastruktuuri sille. (Hotti, 2012)

Tietokantamalli Hivessa voidaan jakaa kolmeen tasoon, jotka ovat taulukot, ositukset ja lokerot. Taulukoista jokainen osoittaa yhteen HDFS-hakemistoon ja jokaisella taulukolla voi olla yksi tai useampi ositus valittujen sarakkeiden mukaan, jolloin jokainen sarakkeen yksittäinen arvo muodostaa oman alihakemistonsa. Hakua voidaan siis nopeuttaa rajoittamalla haku tiettyyn alihakemistoon, joka vastaa sarakkeen arvoa. Ositukset on myös mahdollista jakaa vielä lisäksi lokeroihin, jolloin taulukosta poimitaan lokeroitavat sarakkeet ja kyseisten sarakkeiden arvoille lasketaan lokeroon osoittava indeksi. Alihakemistossa yksi tiedosto vastaa yhtä lokeroa. (Thusoo, Sarma, Jain, Shao, Chakka, Antohony, Liu, Wyckoff & Murthy, 2009)

Hive-kyselyt muutetaan Map Reducen Map- ja Reduce-tehtäviksi suoritusympäristössä. Tietokantamallinsa kolmen eri tason tiedot Hive varastoi erilliseen metadata-varastoon, joka toteutetaan tiedostojärjestelmänä tai perinteisenä relaatiotietokantana. Perinteisenä relaatiotietokantana dataa voidaan käyttää myös perinteisemmällä datankäsittelytyökaluilla, kuten Wedgellä. Tällöin esimerkiksi mobiililaitteilla pystytään analysoimaan tietoa tarkastelun kohteena olevan toimilaitteen vieressä.

### 4.9.3 Apache Pig

Apache Pig on massadata-analyysin tarpeisiin kehitetty tietovirtakieli, jonka avulla pystytään käsittelemään massadataa ilman Javaa. Se hyödyntää paitsi HDFS:ää, niin myös MapReduce-käsittelyjärjestelmää. Sen avulla voidaan tuottaa valtavista massadatamääristä järjestäytyneitä ja analysoitua tietoa. Pig Latin - ohjelmointikieli sisältää perinteisiä toimintoja, kuten liitä, lajittele ja suodata. Kielen käyttäjät voivat myös kehittää omia lukemistoimintojaan sekä käsitellä ja kirjoittaa tietoa. (Gates, 2011)

Apache Pig - ohjelmistoalustaan kuuluu myös Pig-kääntäjä, joka luo MapReduce-tehtäviä Hadooppille Pig Latin - ohjelmointikielillä kirjoitetusta koodista. Pig-kääntäjä myös optimoi samalla skriptejä nopeuttaakseen Hadoopin suorituskykyä. Pig Latin kielenä kuitenkin yksinkertaistaa MapReduce-ohjelmien kirjoittamista, sillä se yleistää erilaisia rinnakkaistehtäviä korkeamman tason käskyjen taakse. (Olston, Reed, Srivastava, Kumar & Tomkins, 2008)

## 4.10 Savcor Wedge

Jäsentelemättömän datan hyödyntämiseen tarvitaan uudenlaisia työkaluja, joista Stora Ensolla käytetään Savcor Wedge-sovellusta prosessin analysointiin. Data voidaan muuttaa Wedgen sovellusten avulla muotoon, jossa sitä voidaan käsitellä perinteisillä tietojärjestelmätieteen työkaluilla. Wedge sisältää valtavan määrän työkaluja, jotka toimivat yhdessä ja siten sitä voi käyttää paitsi lukuisiin eri prosessiteollisuuden tehtäviin, niin myös hyvin yksilöllisesti ja käyttäjälähtöisesti.

Ohjelmisto tarjoaa työkalut ja keinot prosessihäiriöiden ratkaisemiseksi järjestelmästä kerättyjen tietojen avulla, se pystyy paitsi keräämään tietoa, niin myös yhdistelemään niitä. Wedge on erityisesti kartonkikoneen analysointia varten luotu järjestelmä, jolla tavoitellaan paitsi tuotannon tehokkuuden parantamiseen, niin myös lopputuotteen laadunhallinnan parantamiseen. Ohjelman avulla pystytään osoittamaan mahdolliset prosessihäiriöt ja näin ollen löytämään optimaalisin ajotapa koko prosessille. Prosessia analysoidaan erilaisten matemaattisten työkalujen avulla, joita käyttämällä on mahdollista järjestelmällisesti analysoida prosessia. (Kaukonen, 2010)

## 4.11 Teknologioiden hyödyt organisaatiolle

Teknologioita oikein käyttämällä on mahdollista saavuttaa kilpailullista hyötyä markkinoilla, jotka ovat maailmanlaajuiset. Vaikka suomalaisella metsäteollisuudella on kilpailuvaltteina raaka-aineen suhteellisen helppo saatavuus, alhainen energian hinta sekä korkeasti koulutettu työvoima, tarvitaan niiden tueksi nykyisessä markkinaympäristössä myös kykyä tehdä organisaation jokaisella tasolla tehokkaita ja tarkkoja päätöksiä ripeästi ja oikea-aikaisesti. Kilpailukyvyn kannalta on kriittistä, että massadata-analytiikkaa pystytään hyödyntämään erilaisia datamassoja käsitellessä ja sitä kautta löytämään korrelaatioita jo tapahtuneiden asioiden välillä, sekä mahdollisesti jopa ennustamaan tulevaisuuden odotusarvoja. Mikäli analytiikka pystytään implementoimaan kokonaisvaltaisesti päivittäiseen toimintaan, on mahdollista että organisaatiot voivat jatkuvasti tehdä analysoituun tietoon perustuvia päätöksiä liiketoimintansa tueksi. (Apte, Hong, Natarajan, Pednault, Tipu & Weiss, 2003)

On myös tärkeää muistaa, että luvussa mainitut teknologiat ovat hyvinkin suoraviivaisia työkaluja niin datan analysointiin, sen keräämiseen ja tallentamiseen sekä tulosten ennustamiseen. Mitä enemmän, nopeammin ja tehokkaammin dataa pystytään prosessista keräämään, sitä enemmän voidaan havaita erilaisia kaavoja ja korrelaatioita massadata-ainestojen välillä. (Bose, 2009) Käsitellyt teknologiat auttavat organisaatiota lukemattomilla tavoilla, pystymme paitsi optimoimaan ja mallintamaan paremmin prosessia, niin myös löytämään uusia tärkeitä datalähteitä joita analysoida. Tämä johtaa kerrannais-

vaikutuksiin, joissa tietomäärä kasvaa ja analysoidun tiedon laatu paranee. Tämä taas johtaa parempaan liiketoimintatietoon ja sitä kautta oman yrityksen markkinoiden kasvamiseen sekä tulojen nousuun.

Apache-projektin sekä sen pääosin Hadoopin ympärille rakennettujen työkalujen suurimmat liiketoiminnalliset hyödyt tulevatkin esiin silloin, kun halutaan löytää kilpailuetua louhimalla dataa omista tai ulkoisista prosesseista, jotta voimme löytää mahdollisia trendejä tai tilastoja. Lisäksi metsäteollisuuden konservatiivisessa liiketoimintaympäristössä ja keski-ikältään korkeassa organisaatorakenteessa on tärkeää, että tuotetut data-analyysit pystytään visualisoimaan helposti ymmärrettävään muotoon, sekä myös tuottamaan näistä visualisoinneista tarvittaessa vielä laadultaan tarkempia analyyssejä organisaatioiden yksilöllisten tarpeiden mukaan. Krishnan (2013) toteaa, että Big Data-analytiikan tavoitteena on luoda paitsi uusia innovaatioita, niin myös viedä muutosta avoimempaan ja määritellympää suuntaan, jossa tuloksia pystytään ennustamaan sekä niihin voidaan reagoida niiden vaatimalla tavalla. Nämä kaikki edistysaskeleet niin teknologian, kuin yleisen ajattelun tasolla ovat erittäin tervetulleita tuulettamaan osin hyvinkin vanhakantaisia ajatusmalleja metsäteollisuuden piirissä.



## 5 LIKETOIMINTATIEDON HALLINTA

Yritysten prosesseista syntyy päivittäin valtava määrä dataa, jonka hyödyntämisestä osana liiketoimintaa tulee koko ajan tärkeämpää. Tietoa tulisi siis käyttää osana päätöksentekoa, mutta siten että päästäisiin käsiksi paitsi oleelliseen informaatioon, niin sen tulisi olla myös helposti saavutettavissa. Ghoshal ja Kim (1986) painottivat keskittymään liiketoimintaympäristön tarkkailuun globaaleissa, epävakaisissa ja niin teknologisesti, kuin johtamiskyvykkyyksiltään eriytyvissä liiketoimintaympäristöissä. He näkivät edellä mainittujen seikkojen johtavan arvaamattomiin muutoksiin paitsi liiketoimintaympäristössä, niin myös lainsäädäntö-, kilpailu- ja sosiaalisessa ympäristössä.

Suuret yritykset pystyvätkin nykyään varastoimaan suuria datamääriä, mutta organisaatioille on muodostumassa haasteeksi tuon samaisen datamäärän analysointi ja soveltaminen käytäntöön. Tällöin ongelmana on tiedon ja kyvykkyyden puute siitä, mitä kaikella tuolla datalla tulee tehdä. (Davenport & Harris, 2007) Tehokkaan tiedonhallinnan kannalta onkin olennaista, että organisaatio tunnistaa oman tietotarpeensa mahdollisimman tarkasti ja välttää siten keräämästä liikaa tietoa. Päätävälle elimille ohjatun tiedon tulisi sisältää ainoastaan yhdessä sovittujen tietotarpeiden mukaista tietoa liian yksityiskohtaisen datamassan sijaan. (Pirttimäki, 2007)

Tietotarpeet voidaankin helposti tulkita tarkoittamaan kilpailukyvyyn ja menestymisen lähtökohtaa. Kuinka hyvin tietoa pystytään käyttämään hyväksi liittyen niin yrityksen resursseihin, kilpailijoihin, markkinoihin tai omaan jokapäiväiseen tuotantoon. Lisäksi on tärkeää pystyä yhdistelemään näitä tietoja mahdollisimman useasta eri lähteestä, jotta tiedon analyysi on mahdollisimman tarkkaa ja kattavaa. Liiketoimintamahdollisuuksia on paljon helpompaa ja turvallisempaa analysoida & priorisoida, kun sitä koskeva informaatio on tarkkaa ja kattavaa. Tällaisen tiedon avulla pystytään myös vähentämään liiketoiminnan kustannuksia ja siihen kohdistuvia riskejä. (Vaarnas & Virtanen, 2001) Tällaista, helposti analysoitavaa dataa ovat esimerkiksi erilaisten internet-portaalien sisällöt, markkinatrendit ja alaa käsittelevät julkaisut. Kun niiden tieto tuodaan päätöksentekijöille, niistä on mahdollista puristaa liiketoiminnan

kannalta tärkeitä tietoja vaikkapa uusien tuotteiden tai jopa kokonaan uusien kilpailijoiden muodossa.

## 5.1 Liiketoimintatiedon hallinta käsitteenä

Liiketoimintatiedon hallintaa käsitteellisellä tasolla on tulkittu ja tarkasteltu useista eri näkökulmista useilla eri tasoilla (taulukko 1). Suppeimmat määritelmät toteavat liiketoimintatiedon hallinnan olevan lähinnä perinteistä raportointia laajempien näkökulmien linkittäessä mukaan myös muita johtamis- ja analyysityökaluja kuten tietojohdamisen tai suorituskyvyn mittaamisen. (Herschel & Jones, 2005) Yleisesti määritelmässä huomioidaan kuitenkin laajemmin eri ominaisuuksia, joissa liiketoimintatiedon hallinta käsitetään koko organisaation kattavaksi strategiatyökaluksi, jolla pystytään tukemaan päätöksentekoa ja raportointia organisaation eri tasoilla. (Abukari & Jog, 2003) Useat kaupallisten sovellusten tarjoajat taasen korostavat analyysityökaluja liiketoimintatiedon hallinnan määritelmässään. (Azvine, Cui & Nauck, 2005)

TAULUKKO 1 Liiketoimintatiedon hallinnasta aiemmin esitettyjä määritelmiä

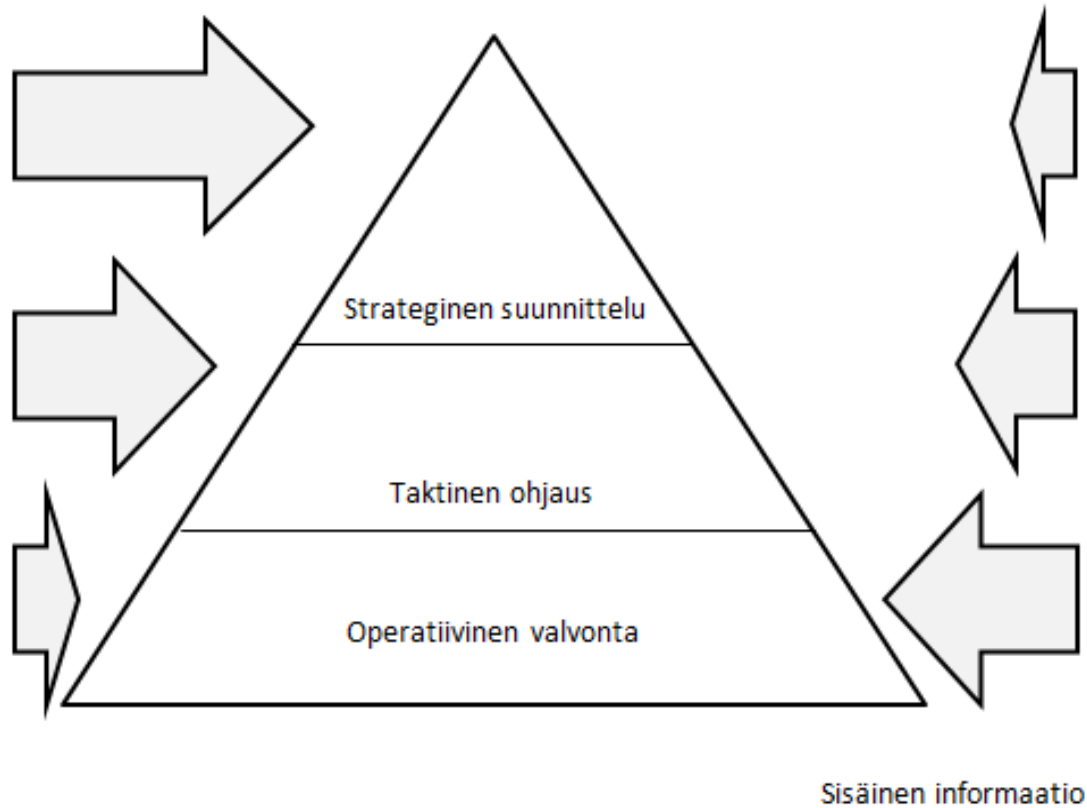
Lähde	Käytetty määritelmä
Abukari ja Jog, 2003, s. 15	Liiketoimintatiedon hallinta kuuluu päätöksenteon tukijärjestelmiin. Se on faktapohjainen lähestymistapa päätöksentekoon.
Chung, Chen ja Nunamaker 2003, s. 1	Liiketoimintatiedon hallinta auttaa sisäisen ja ulkoisen toimintaympäristön ymmärtämisessä
Collins 1997, s. 14	Liiketoimintatiedon hallinta on prosessi, jolla tuetaan päätöksentekoa liiketoiminnassa.
Halliman 2000, s. 3-7	Liiketoimintatiedon hallinnalla pyritään hallitsemaan liiketoiminnan tulevaisuutta
Negash ja Gray, 2003, s. 3190 ja 3197	Liiketoimintatiedon hallinnan tarkoituksena on auttaa yritystä ymmärtämään paremmin omaa asemaansa, asiakkaitaan ja kilpailijoitaan
Marshall, McDonald, Chen ja Chung 2004, s. 873	Liiketoimintatiedon hallinnan tarkoituksena on tukea päätöksentekoprosessia.
Salonen ja Pirttimäki 2005, s. 662	Liiketoimintatiedon hallinta on systemaattinen prosessi, jolla kerätään, analysoidaan ja levitetään organisaatiota koskevaa strategista ja operatiivista informaatiota.
Thomas Jr. & James H 2001, s. 48	Liiketoimintatiedon hallinnan tarkoituksena on kehittää systemaattinen prosessi, jolla kerätään tietoa ja analysoidaan kilpailuympäristöä

## 5.2 Sisäinen ja ulkoinen liiketoimintatieto

Menestyvän liiketoiminnan rakentaakseen on perinteisesti tullut analysoida dataa ja sen pohjalta ennakoida suuntaa, johon markkinat ovat liikkumassa. Moderni trendi tässä kehityksessä on välittömästi saatavilla oleva reaaliaikainen data, jota rahoitus- ja markkinointiosastot ovat tulkinneet. Tähänkään yritykset eivät voi loputtomasti tukeutua, sillä markkinoiden muutostahti kasvaa jatkuvasti ja menneen datan perusteella tehty analyysi markkinoista ei ole enää täysin luotettavaa. (Azvine, Cui, Nauck & Majeed, 2006)

Kuinka sitten erotella erilaisia päätöksentekotilanteita sisäisessä ja ulkoisessa liiketoimintatiedossa. Pirttimäki (2007) toteaa liiketoimintatiedon hallinnan eroavan niin strategisella, taktisella kuin operatiivisellakin tasolla. Hän jakaa organisaation informaatiotarpeet ulkoiseen ja sisäiseen liittäen samalla mukaan niiden tarpeet päätöksentekotasosta riippuen (kuvio 8). Kuvasta voidaan huomata, että liiketoimintatiedon tyypin pinta-alan suuruus on suoraan verrannollinen sen vaikuttavuuteen päätöksentekotilanteessa.

Ulkoinen informaatio



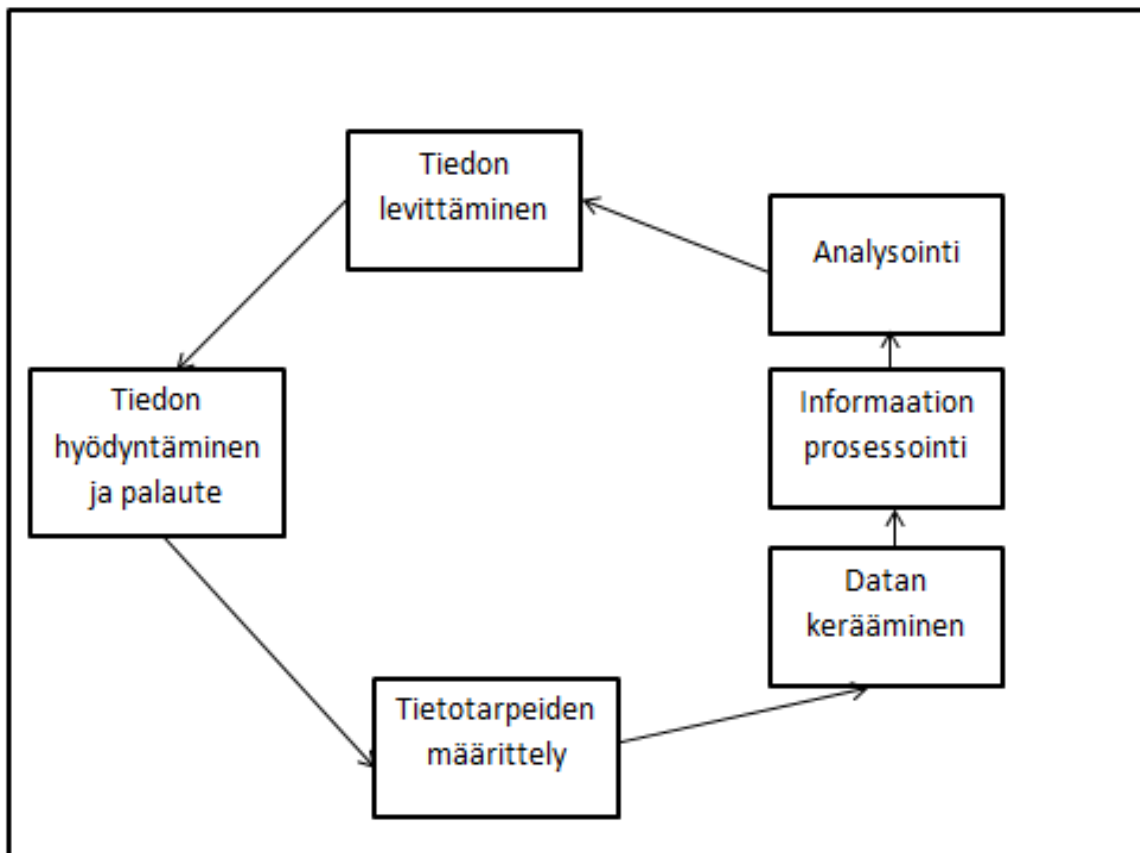
KUVIO 8. Informaation merkitys eri päätöksenteon tasoilla (Uusi-Rauva, 1994, s. 6)

Sisäinen liiketoimintainformaatio kerätään sisäisistä tuotantoprosesseista ja niiden tuloksista sekä analyysistä. Organisaation henkilöstön tietotaito, hiljainen tieto sekä muu arvokas tietopääoma ovat myös tärkeä osa sisäistä liiketoimintatietoa. Perinteisellä tietovarastoinnilla kerätään kuitenkin ainoastaan eksplisiittistä tietoa, joten muut liiketoimintatiedon muodot tulee kerätä sosiaalisin keinoin. (Alavi & Leidner, 2001)

Pirttimäen ja Hannulan (2002) määritelmä ulkoisen liiketoimintatiedon hallinnasta tutkimuksellisessa tarkoituksessa sisältää ajatuksen jatkuvasta ja systemaattisesta prosessista, jonka tavoitteena on tuottaa ja kommunikoida organisaation ulkopuolelta ennakoivaa ja hyödyllistä tietoa tukemaan päätöksentekoa. Suomalaisissa suuryrityksissä painotetaan enenevässä määrin kilpailijatiedon merkitystä yhdessä oman toimialan ja sen asiakkaita koskevien tietojen kanssa. (Halonen & Hannula, 2007) Ulkoisen tiedon analysointi mahdollistaa tarkemman ennakkoinnin liittyen markkinoiden kehityssuuntiin ja tämän lisäksi tarjoaa myös mahdollisuuden peilata oman organisaation toimintaa muiden yritysten tekemiä päätöksiä vasten. (Salonen & Pirttimäki, 2006)

### 5.3 Liiketoimintatiedon hallintaprosessi ja sen vaiheet

Liiketoimintatiedon hallintaprosessi on systemaattinen ja jatkuva prosessi, jossa data muokataan hyödynnettävissä olevaan tietomuotoon. Toimeenpanemalla näitä prosesseja organisaatio pystyy analysoimaan tietotarpeitaan järjestelmällisemmin, sekä kykenee nopeampiin päätöksentekoprosesseihin. Seuraamalla jatkuvasti omaa liiketoimintaympäristöään on mahdollista havaita heikkojakin signaaleita yrityksen liiketoiminta-alueella. Näiden suorien liiketoimintahyötyjen lisäksi myös muutosten ennakointi ja yllätysten välttäminen helpottuvat jatkuvan seurannan ansioista kuvion 9 mukaisesti. (Hannula & Pirttimäki, 2003) Prosessi voidaan jakaa vaiheisiin ja tuleekin muistaa, että liiketoimintatiedon hallinnan prosessi on täysin hyödytön, mikäli prosesseja ei viedä osaksi toimintaa ja käytäntöjä. (Thierauf, 2001)



KUVIO 9. Liiketoimintatiedon hallintaprosessin vaiheet (Hannula & Pirttimäki, 2003, s. 90 mukaan)

Liiketoimintatiedon hallinnan prosessi voidaan jakaa kuuteen eri vaiheeseen, joista ensimmäinen on selvitys siitä, mitä tietoa tarvitaan ja missä muodossa sen tulee olla. Tulee siis määrittää tietotarpeet. (Laihonen ym. 2013). Tarpeen määrittämisestä ovat yleensä vastuussa paitsi tiedon käyttäjät, niin myös päätöksentekijät. Tarvittavat tietotarpeet voidaan määrittellä vaikkapa sen mukaisesti, millaisia tiedonhankintakohteita on käytettävissä. (Thomas Jr., 2001) Kun tiedonhankintatarpeet on tunnistettu, toisessa vaiheessa keskitytään keräämään kyseistä tietoa sopivaksi katsotuista lähteistä. Tämän jälkeen seuraa kolmas vaihe, jossa kerätty tieto varastoidaan ja prosessoidaan, jonka jälkeen dataa voidaan alkaa neljännessä vaiheessa analysoidaan paitsi käytettävyyden, niin myös luotettavuuden ja käytettävyyden perusteella. Näissä vaiheissa on tärkeää ymmärtää, että analysoitu tieto tulee pystyä pakkaamaan ja muotoilemaan niin, että prosessissa muodostuneilla eri tietotuotteilla ja palveluilla organisaatio pystyy luomaan lisäarvoa myös niiden käyttäjille ja käyttävien tahojen välille (Choo, 2002). Viimeisessä vaiheessa prosessista muodostunut tietopääoma muutetaan organisaatiossa päätöksiksi ja annetaan samalla palaute kerätyn aineiston hyödyllisyydestä sekä siitä, kuinka hyvin tietotarpeet ovat täyttyneet kerätyllä tiedolla. (Laihonen ym., 2013).

## 5.4 Prosessin menetelmät/järjestelmät

Liiketoimintatiedon hallinnan prosessia tukevat järjestelmät sijaitsevat usein prosessin alussa erillään, prosessin edetessä niitä on kuitenkin kannattavaa alkaa yhdistelemään järjestelmiä jotka eivät normaalisti ole yhteydessä toisiinsa. Tämä on kriittistä sen varmistamiseksi, että eri järjestelmissä sijaitseva raakadata saadaan muutettua helposti käytettäväksi ja jaettavaksi informaatioksi organisaation käyttöön. (Abukari & Jog, 2003)

Työkalut, joilla liiketoimintatietoa voidaan analysoida ja kerätä voidaan jakaa pääosin kahteen eri kategoriaan, sovelluksiin ja palveluihin. Sovellukset sisältävät tiedon tallentamisen, operatiiviset raportointiin tarkoitettujen ohjelmien, erilaiset portaalit sekä analyysityökalut. Palveluita taas ovat datan louhinta ja analysointi, datan tuottaminen ja prosessien rakentaminen. Näihin kaikkiin on markkinoilla tarjolla yksittäisiä kaupallisia sovelluksia niin strategia-, kuin raportointinäkökulmastakin. (Negash & Gray, 2003)

## 5.5 Prosessin hyödyt ja strategiaan yhdistäminen

Jotta yritys saisi kaiken irti tehokkaista prosesseista liittyen liiketoimintatiedon hallintaan, tulee yrityksessä vallita avoin kulttuuri ja mutkaton sekä läheinen vuorovaikutus eri yksilöiden ja osastoiden välillä. Avoin kulttuuri johtaa jaetun informaation määrän kasvamiseen ja samalla myös hiljaisen tiedon lisääntymiseen.

seen. Samalla on mahdollista lisätä paitsi organisaation tiedon määrää, niin myös kustannustehokkuutta sillä omia organisaation sisäisiä asiantuntijoita pystytään hyödyntämään enemmän, eikä tietoa tarvitse kerätä ulkopuolelta. (Hannula & Pirttimäki, 2003.)

Prosessien päämääränä ei ole ainoastaan tuottaa hyödyllistä tietoa liittyen organisaation päätöksentekoon, mutta myös luoda älykkäitä ratkaisuita organisaatiokohtaisesti mahdollistaen mahdollisimman tehokkaan käytön kerätylle aineistolle. Tärkeimmät päämäärät prosesseille ovat Thomaksen (2001) mukaan uhkien ja mahdollisuuksien parempi tunnistaminen, reaktioajan pidentyminen, yllätysten välttäminen ja kilpailulliset tekijät kuten tietopääoman suojaaminen. Niiden avulla voidaan myös vertaisarvioida, muodostaa graafeja ja taulukoita sekä optimoida prosessia myös yleisellä tasolla. Useat suuryritykset toimivat näillä erittäin kilpailuilla markkinoilla mukaanlukien mm. SAP, IBM, Oracle, SAS ja viimeisimpänä tulokkaana Microsoft. (Henschen, 2011)

Varsinainen strategiatyöskentely sisältää strategian muodostamisen, implementoinnin ja kontrolloinnin. Strategiaa muodostettaessa tulee määritellä missio ja sen mukaiset päätavoitteet sekä analysoida organisaation sisäinen ja ulkoinen ympäristö joiden perusteella voidaan valita sopivimmat strategiavaihtoehdot. Strategiaprosessin implementoinnilla ja läpiviennillä on vaikutus yrityksen suorituskykyyn, suorituskyky on suoraan verrannollinen strategiaprosessin formaaliuteen. (Falshaw, Glaister & Tatoglu, 2006)

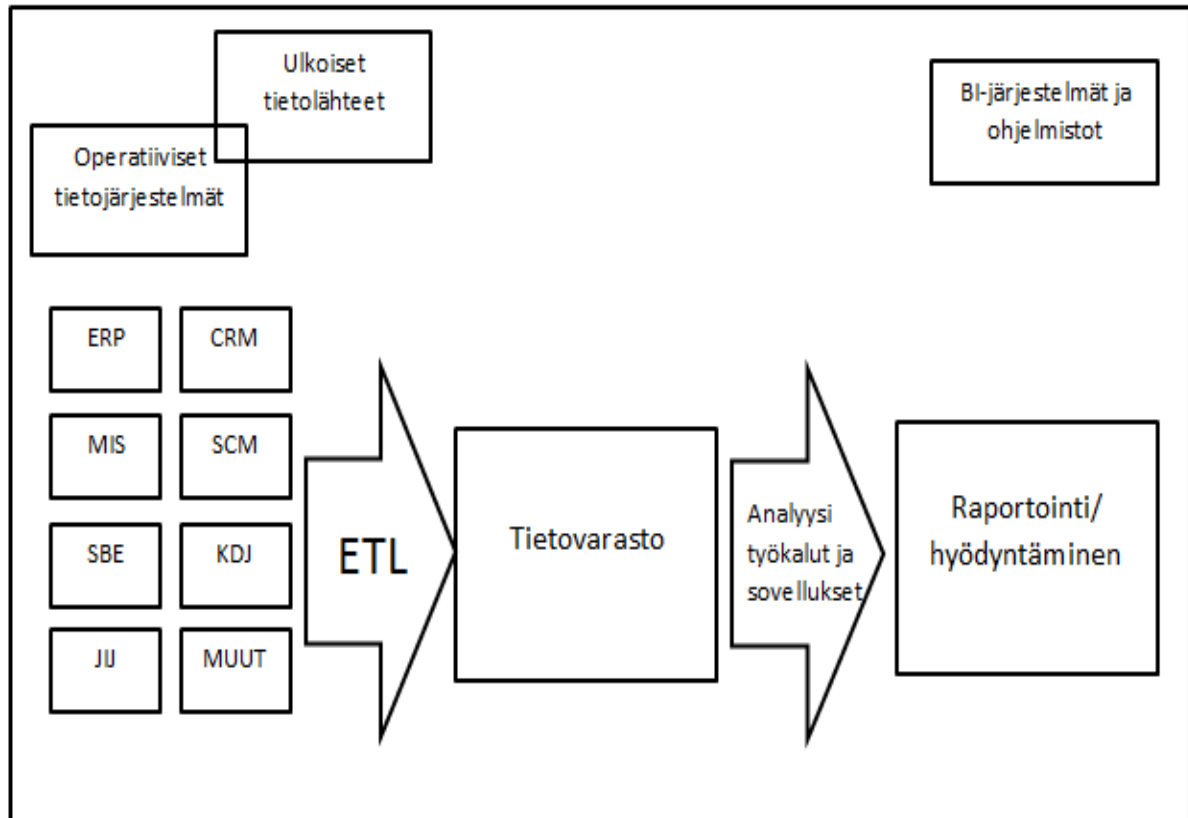
Erityisesti strategian muodostamisvaiheessa liiketoimintatiedon hallinta auttaa työskentelyssä. Systemaattinen tapa toteuttaa ympäristön analyysi ja prosessin eri vaiheet datan keräämisestä prosessointiin ja analysointiin tukevat tätä vaihetta strategiaprosessissa. Liiketoimintatiedon hallinnalla yrittään vastata kysymyksiin strategiaprosessissa kysymyksiin mahdollisista muutoksista, optimaalisesta toteutuksesta ja mahdollisesta tulevaisuuden strategias- ta. Saavuttaakseen mahdollisimman suuren tuen strategiaprosessille, liiketoimintatiedon hallintaa tulee toteuttaa yhdessä strategiatyöskentelyn kanssa eikä erillisinä toimintoina. (Fahey, 2007)

Strategiatyöskentely ja liiketoimintatiedon hallinta ovatkin vuorovaikutteisia prosesseja ja suurin haaste on juuri niiden toteuttaminen rinta rinnan. Ei siis kannata perustaa omaa liiketoimintatiedon hallinnasta vastaavaa yksikköä, vaan implementoida se osaksi strategiasta vastaavaa yksikköä. Näin voidaan käyttää käytännöllisempiä analysointitekniikoita ja voidaan ottaa kokonaisvaltaisesti kantaa projektiin sen eri vaiheissa. Lisäksi on mahdollista, että mikäli informaatiolla pystyy saavuttamaan merkittävää yrityksen sisäistä hyötyä, sen kulkua saatetaan rajoittaa oman edun nimissä. (Ghoshal ym., 1986)

## 5.6 Liiketoimintatiedon hallinnan järjestelmät

Liiketoimintatiedon hallinnan organisaatiosovellusten perusarkkitehtuuriin kuuluu operatiiviset tietojärjestelmät, tiedon integrointi eli ETL-prosessit, tietovarastot ja erilaiset business intelligence raportointi- ja ohjelmistojärjestelmät.

(Hovi, Hervonen & Koistinen, 2009) Kuviossa 10 voidaan nähdä, kuinka operatiivisten tietojärjestelmien ja ulkoisten lähteiden tuottama tieto ohjataan ETL-prosessien kautta tietovarastoon, jossa niitä voidaan analysoida ja lopulta raportointisovellusten avulla käyttää päätöksenteon tukena.



KUVIO 10. Liiketoimintatiedon hallinnan arkkitehtuuri (Hovi ym., 2009; Davenport & Harris, 2009, s. 102 mukaan)

### 5.6.1 Tiedon keräys

Kuvio lähtee datan keräämisestä ja organisaatiot keräävät operatiivisten järjestelmien avulla tietoa perusdatasta. Aho (2011) toteaa operatiivisiksi tietojärjestelmiksi toiminnanohjausjärjestelmät (ERP), jotka sisältävät organisaation kaikki tietojenkäsittelytarpeet. Asiakkuudenhallintajärjestelmät (CRM), jotka tukevat asiakashallintaa, myyntiä ja markkinointia. Tuotannonohjausjärjestelmät (MES) jotka tukevat tuotantoketjua sekä siihen läheisesti liittyvät toimitusketjun hallintajärjestelmät (SCM) jotka optimoivat logistiikkaketjuja. Muita talous- ja johtamisjärjestelmiä ovat mm. Suunnittelu-, budjetointi-, ja ennustamisjärjestelmät (SBE), konsolidointijärjestelmät (KDJ) jotka helpottavat talousraportointia sekä johdon informaatiojärjestelmät (JI) jotka tukevat organisaation päätöksentekoa tarjoamalla historiadataa yrityksen menneistä operaatioista.



Ulkoisina datalähteinä käytetään Hovin ym. (2009) mukaan ulkoisia tietopankkeja ja muita julkisia tietolähteitä, joista on saatavilla strukturoimattomia laadullista dataa. Oli kyse sitten sisäisistä tai ulkoisista tietolähteistä, hyödyntääkseen dataa tulee se joka tapauksessa riittää ETL-prosessin lävitse. Prosessissa tieto integroidaan tietojärjestelmistä tietovarastoon muokkaamalla haettavien tietolähteiden tiedostomuoto sopivaksi ja lopulta ohjataan sekä tallennetaan itse tietovarastoon. (Kimball & Caserta, 2004)

### 5.6.2 Tiedon varastointi

Tietovarasto määritellään organisaation jäsenten yhteiskäyttöiseksi tietokannaksi, jonka avulla eri tietojärjestelmistä kerätyjä tietoja voidaan kerätä tukemaan liiketoimintatiedon käyttöä. Tietovaraston tärkein etu on potentiaalinen sijoitetun pääoman tuotto, kilpailullinen etu sekä kasvava tuottavuus. (Connolly & Begg, 2005) Onnistuneen tietovaraston päämääriä ovat Kimballin ja Rossin (2013) mukaan tietovaraston saatavuus, jossa lähtökohtana on mahdollisimman helposti saatavilla oleva organisaation tieto. Lisäksi organisaation jäsenten tulee osata käyttää tuota tietoa eli sen tulee olla ymmärrettävää ja tarpeeksi selkeää, jotta sitä voidaan käyttää myös liiketoiminnallisiin tarkoituksiin. Datan tulee olla luettavaa ja sitä tulee voida käyttää mahdollisimman monimuotoisesti yhdistellen ja erotella eri osia helppojen työkalujen avulla.

Tietovaraston tulee olla myös sekä johdonmukainen, mukautuvainen että joustava. Data on laadukasta ja puhdistettua, eri prosessien tiedot vastaavat samoin merkittyä toisen prosessin tietoa eli kaikki tieto on määritelty yhden mukaisella tavalla järjestelmän sisällä. Yhdenmukaisuuden lisäksi käyttäjien erilaisiin tarpeisiin on pyrittävä vastaamaan, tiedon tulee olla mukautuvaista erilaiselle käytölle ja sitä tulee pystyä muokkaamaan tiedostomuodosta toiseen. Lisäksi tulee huolehtia luotettavuudesta, organisaation tieto on kilpailullisesti äärimmäisen arvokasta, joten sitä pitää pystyä käyttämään turvallisesti apuna päätöksenteossa ilman pelkoa siitä että ne valuvat ulkopuolisille. Luotettavuuteen liittyy myös se, että ihmisten on koettava tieto hyödylliseksi ja tarpeelliseksi. Tietojärjestelmä toimii päätöksenteon perustana ja organisaation kannalta on äärimmäisen tärkeää, että tietovarasto otetaan käyttöön myös päätöksentekijöiden tasolla. Tällöin tiedon keräämisestä on eniten hyötyä organisaatiolle, tätä hyväksymisprosessia voidaan jouduttaa kunnollisella testaamisella.

### 5.6.3 Tiedon analysointi ja raportointi

Mietittäessä potentiaalisia analysointityökaluja tulee määritellä kuinka syvälle itse liiketoimintaprosessiin päätöksenteon voi viedä. Tämän jälkeen tulee tutkia, halutaanko päätösten tulla järjestelmästä automaattisesti vai avustetusti siten, että päättävässä asemassa oleva analysoi itse dataa ennen päätöksen tekemistä. Moderneja tiedon analysointityökaluja ovat moniulotteinen analysointi, tilastolliset tai kvantitatiiviset algoritmit, tiedonlouhintaohjelmistot, tekstinlouhintaohjelmistot sekä simulaatiotyökalut. (Hovi ym., 2009) Kun sopiva työkalu on valittu, voidaan siirtyä prosessin tärkeimpään vaiheeseen, eli raportointiin jossa loppukäyttäjät sekä koostavat että lukevat raportteja ja käyttävät niitä päätöksentekonsa tukena. Työkalujen kirjo on laaja, osa käyttää eriteltyä dataa ja toiset taasen strukturoimatonta dataa. (Davenport & Harris, 2007)

## 5.7 Toiminnanohjausjärjestelmät metsäteollisuudessa

Yrityksen ohjaustoimia, joita käytetään liiketoiminnan ohjaamiseen, tukemiseen ja ylläpitämiseen kutsutaan yrityksen toiminnanohjaukseksi. Toiminnanohjauksessa käytetään tiedonhallintajärjestelmiä, eli toiminnanohjausjärjestelmiä. Toiminnanohjausjärjestelmiä, eli ERP-järjestelmiä voidaan määritellä monin eri tavoin, mutta Beheshtin (2006) mukaan ne ovat useamman eri liiketoimintasovelluksen kokonaisvaltainen järjestelmä, joka yhdistää kaikkien eri liiketoimintayksiköiden tiedot yhteen kokonaisuuteen organisaation sisällä. Toiminnanohjausjärjestelmät sisältävät tavallisesti mm. Henkilöstöhallinnon, tuotannon, kirjanpidon sekä logistiikan sekä raaka-ainevirtojen hallinnan.

Tarve toiminnanohjausjärjestelmille tuli ensimmäisenä esiin 1960-luvulla, jolloin varastoiden hallintaan alettiin suunnitella järjestelmiä, joihin ennen pitkää alettiin kehittämään lisäominaisuuksia tilausten hallintaa silmälläpitäen. Yritykset kehittivät näitä ohjelmistoja ensin yksittäisten yritysten tarpeita varten ja kaupalliset, tarpeeksi monipuoliset toiminnanohjausjärjestelmät joita voitiin myydä yleisesti eri yritysten tarpeisiin yleistyivät vasta 1970-luvulla. Nykyaikaiset ERP-järjestelmät syntyivät vasta kuitenkin 1980-1990 - lukujen aikana, jolloin myös tuotannonohjaus- ja taloushallinnon ominaisuudet tulivat lisäyksi materiaalinhallinnan ohjelmistoihin. (Kettunen & Simons, 2001)

ERP- eli toiminnanohjausjärjestelmillä on jatkuvasti kasvava vaikutus organisaation kannattavuuteen ja kykyyn kilpailla markkinoilla. Oikein käytettynä toiminnanohjausjärjestelmät paitsi parantavat yrityksen resurssienhallintaa, niin myös laskevat kustannuksia ja parantavat kykyä palvella asiakkaitaan. (Vilpola & Kouri, 2006)

## 5.8 Käyttöönottoprosessi

Käytännössä kaikki yritykset tarvitsevat ERP-järjestelmiä toimintansa tueksi, sillä kilpailu on radikaalisti muuttunut liiketoimintaympäristöjen ja markkinoiden laajennuttua. (Umble, Haft & Umble 2003) He jatkavat, että tämän johdosta asiakkaat painostavat yrityksiä pienentämään kustannusrakennettaan, kehittämään asiakaspalvelua sekä parantamaan laatua ja toimitusvarmuuttaan lisäten vielä, että kilpailijoiden kehittyessä organisaatioiden on pakko kehittää omia käytäntöjään ja prosessejaan pysyäkseen kilpailukykyisinä. Tieto on pystyttävä jakamaan sidosryhmien kanssa, olivat nämä sitten asiakkaita tai tavarantoimittajia.

## 6 KIRJALLISUUSKATSAUKSEN YHTEENVETO

Kirjallisuuskatsauksessa tutkittiin tietojärjestelmiä, sekä erityisesti Big Data - teknologiasovellusten mahdollisuuksia modernissa metsäteollisuusympäristössä. MapReduce - tekniikan ja erityisesti Apache Storm - teknologian avulla on mahdollista ottaa kaikki mahdolliset datapisteet kartonkikoneesta ja mallintaa koneen prosesseja niiden perusteella. Tätä dataa olisi mahdollista myös tallentaa, jolloin talteenotetun datan perusteella voitaisiin ajaa simulaatiota jota verrattaisiin koneella paraikaa tapahtuvaan ajotapahtumaan, tällöin koneelle reaaliajassa tehtäville muutoksille olisi jatkuva vertailukohta, jota vasten muutoksen toiminnallisuutta voitaisiin jatkuvasti verrata. Laskentatehon ollessa riittävä olisi myös mahdollista muuttaa useita eri parametrejä ja ajaa näitä simulaatioita rinnakkain MapReduce-toiminnallisuuksien avulla.

HDFS-hajauttamisen hyödyt taas tulevat siinä, että kerran hajautettua dataa on nopeampi käsitellä useammasta eri lokaatiosta. Esimerkiksi suurempi yritys, kuten Stora Enso voisi jakaa kaikkien tehtaidensa datan keskenään, jolloin olisi mahdollista löytää parhaat mahdolliset ajotavat kaikille koneille vertailemalla eri paikkakuntien toimintamalleja. Tähän kun lisätään Sparkin tehonlisä, niin eri paikkakunnat voivat ajaa hyvinkin nopealla aikataululla omia simulaatioitaan toisistaan riippumatta.

Apache Flumen avulla taasen voisi kerätä myös muuta kuin prosessidataa, sää-/lämpötila, työntekijöiden tyytyväisyyskyselyt ja mahdolliset logistiikkatiedot voitaisiin niputtaa yhteen kasaan HDFS:n kautta, olisi mahdollista saada pelkkää prosessin tilaa kattavampi kuvaus koko integraatin tapahtumista. Nämä kaikki tiedot tallennettaisiin HBase tietokantakomponentilla ja niitä pystyttäisiin Hiven kautta käsittelemään ja jopa ryhmittelemään pienempiin palasiin SQL-tauluiksi, joita pystytään analysoimaan perinteisemmillä työkaluilla tehdaspaikkakunnalla esimerkiksi mobiilityökaluilla optimointia vaativan moottorin tai venttiilin vierellä.

Näihin paikalliseen käyttöön erittäin hyvin soveltuviin työkaluihin kuuluu myös Savcor Wedge, jonka avulla on mahdollista etsiä koko prosessin mittauspisteet graafisessa, helposti esitettävässä muodossa. Tämä mahdollistaa Big Data - työkalujen avulla rajattujen, muutaman kymmenen kriittisimmän

mittauspisteet suhteiden analysoinnin, joka saattaa osoittautua elintärkeäksi prosessin pitkien vasteaikojen syy-seuraus - suhteita etsittäessä.

Kirjallisuuskatsauksessa käsiteltävänä olleet teknologiat auttavat suomalaista metsäteollisuutta pysymään maailman kärjessä kilpailukyvyssä niin innovaatioiden, kuin kohdennetun datankeräyksen mahdollistaman parantuvan kustannuskilpailukyvyn johdosta. Oikein käytettynä nämä teknologiat mahdollistavat datan yhä tehokkaamman analysoinnin yhä kasvavasta datapistemäärästä, jonka avulla pystytään saavuttamaan positiivinen kierre tiedon hyödyntämisessä. Tätä tietomassaa on helppo hallita paitsi tehdastasolla, niin myös suuremmissa organisaatioissa maatasolla sekä edelleen konsernitasolla. Tiedon analysoinnista päästään nykyaikaisilla työkaluilla helposti korrelaatioiden etsimiseen ja erilaisten skenaarioiden odotusarvojen muodostamiseen.

Näihin edellä mainittuihin teknologioihin yhdisteltiin myös liiketoiminnan hallinnan teoriaa ja tutkimuksessa oli tarkoituksena selvittää liiketoiminnan hallinnan ja tietojärjestelmien yhteenlinjausten implementoinnin haasteita teollisuuden liiketoimintaympäristössä. Tutkimuksella pyritään rakentamaan kuvaa siitä, kuinka tietojärjestelmillä ja siihen yhdistetyllä Big Data -analytiikalla pystytään luomaan lisäarvoa sekä organisaation sisälle, että myös asiakkaalle asti. Lisäksi pyrittiin selvittämään niitä asioita, jotka järjestelmän implementoinnissa tulee ottaa huomioon aina suunnitteluvaiheesta alkaen. Kirjallisuuskatsauksessa tutustuttiin myös erilaisiin Big Data -analytiikan tekniikoihin, että myös niiden teoreettiseen taustaan. Loppupuolella lähdettiin rakentamaan teoriaa liiketoimintatiedon hallinnalle ja käytiin läpi erilaisia käsitteitä ja järjestelmiä aiheeseen liittyen. Aivan viimeisenä pureuduttiin tiedon kulkuun ja varastointiin organisaation tietojärjestelmässä ja pohdittiin tietojärjestelmien käyttöönottoon liittyviä riskejä ja mahdollisuuksia. Kirjallisuuskatsauksen vahvana pohjaviestinä oli, että kerätyn tiedon tulee olla johdonmukaista ja liiketoimintaa tukevaa. Sen käyttäjien tulee tietää miksi sitä kerätään ja mihin sitä käytetään, jotta siitä voidaan saada organisaatioille liiketoiminnallista hyötyä.

## 7 TUTKIMUKSEN TARKOITUS, TAVOITTEET JA TUTKIMUSKYSYMYKSET

Metsäteollisuutta koskeva voimakas kilpailuasetelma niin globalisaation, kuin tiukkenevien ympäristönormien johdosta vaatii metsäteollisuuden toimijoilta paitsi houkuttelevampia tuotteita ja innovaatioita, niin myös suuria panostuksia itse tuotannon tehokkuuden parantamiseen. Tutkimuksen kohteena ovat suomalaiset metsäteollisuuden ammattilaiset, joiden vastuulla paine liiketoiminnan tehostamisesta pääosin lankeaa. Tämän tutkimuksen tarkoituksena on kuvata tekijöitä, jotka vaikuttavat konelinjojen kykyyn tehdä tulosta modernissa bioteollisuusympäristössä. Tutkimuksen tavoitteena on tuottaa konkreettista tietoa arkielämän tiedon keräämisestä ja käytöstä. Tätä tietoa voidaan myöhemmin hyödyntää paitsi päätöksentekoprosessien kehittämisessä, niin myös sekä kokonaan uusien, että nykyistä laitekantaa ylläpitävien investointien suunnittelussa.

Saundersin tutkimussipuli tarjoaa riittävän mallin siihen, että kaikkiin olennaisiin tutkimuskysymyksiin saadaan vastattua metodologisesta ja tutkimusfilosofisesta näkökulmasta. (Saunders, 2009) Sipulin ulommilla kehillä kuvaillaan filosofista lähestymistä aiheeseen. Kaksi suosittua tapaa lähteä lähestymään tutkimusfilosofiaa ovat tietoteoria ja ontologia. (Saunders, 2009). Tietoteoria keskittyy siihen mikä on hyväksyttyä tietoa ja mitkä ovat hyväksyttävät keinot kerätä sitä. Ontologia taasen keskittyy siihen mikä on totta tai todellisuutta. (Hirsjärvi, 2007) Tämän johdosta voitaneen todeta, että mikäli tutkimuksessa asettaisiin kyseenalaiseksi joko Big Data - analytiikka tai metsäteollisuus, tulokset eivät olisi kovinkaan merkityksellisiä. Sen sijaan tutkimus keskittyy löytämään parhaat keinot informaation keräämiseen, yhdistämiseen ja arviointiin jotta Big Data - analytiikan ja metsäteollisuuden yhteensulautuminen olisi mahdollisimman tavoitteenmukaista ja sen vaikutus tuotantoon olisi niin positiivista kuin mahdollista. Näin ollen tietoteoria on riittävä lähestymistapa tutkimusfilosofiaan.

Mitä tulee tietoteorian puitteisiin, niin kriittinen realismi on tähän tarkoitukseen parhaiten sopiva. Kriittisessä realismissa esimerkiksi metsäteollisuutta, Big Dataa sekä niiden yhteensulautumista pidetään totena ja sellaisina asioina, jotka ovat olemassa myös käsitteiden ulkopuolella. Tämän jälkeen ihmismieli muodostaa oman kuvansa asiasta ja tätä asiaa voi muokata mikäli asian rakenne on ymmärretty. (Bryman, 2012) Tämän tutkimuksen lähestymistapa on deduktiivinen, eli testataan olemassa olevaa teoriaa ja tutkimusstrategiana on kyselytutkimus. Kyselytutkimuksessa tavoitteena on vastata kysymyksiin, kuten kuka, mitä, missä, kuinka paljon ja kuinka moni. (Saunders, 2009) Tätä osaa voidaan käyttää suoraan työn empiirisessä osuudessa ja tavoitteena on paikallistaa ne haasteet joita paperikonelinjastolla Big Data -analytiikan omaksumisen suhteen on. Tässä tutkimuksessa on myös paljon tapaustutkimuksen elementtejä, mutta kyselytutkimuksen filosofia on lähempänä työn tavoitteita. Yleensä tämän tyyppisissä tutkimuksissa on käytetty strukturoitua datankeruutapaa, mutta tässä tapauksessa erittäin strukturoitu lähestymistapa ei ehkä ole paras keino hankkia tuloksia. Big Datan ja liiketoiminnan yhdistäminen on melkoisen abstrakti aihe ja vaikka on malleja mitata näin abstraktin asian onnistumista (Henderson & Venkatraman, 1993), kyseisiä kriteereitä voi olla vaikea löytää yksiselitteisesti. Täten puolistrukturoitu metodi on tässä tapauksessa parempi vaihtoehto. Tutkimus käyttää tiedon keräämiseen kirjallisuuskatsausta ja haastatteluita.

Tutkimusmenetelmänä käytin laadullista, eli kvalitatiivista menetelmää. Pysin tutkimuksessani etsimään ja ymmärtämään ilmiöitä, joiden avulla tuotannon tehokkuutta pystyttäisiin lisäämään ilman, että muutosvastarintaa organisaatiossa esiintyisi tai että avainhenkilöt kuormittuisivat liikaa.

Tutkimuksessa oletettiin, että kaikki haastateltavat ovat työnantajan edustajia ja työsuhteessa metsäteollisuuden toimijan kanssa. Perusoletuksena oli se, että asiantuntijalle ja esimiehelle oli jonkinasteinen peruskäyttökoulutus annettu käytettyihin järjestelmiin, jolloin erot muodostuisivat alan kulttuurin ja henkilöiden ominaisuuksien kautta. Tutkimuksessa käsiteltiin ainoastaan metsäteollisuuden prosessien ohjaukseen käytettäviä ohjelmistoja sekä niiden yhteenlinjausta liiketoiminnan hallinnan kanssa.

Laadullista menetelmää käyttämällä saan erinomaisen kuvan erilaisista haasteista ja näkemyksistä Big datan käytöstä tällaisessa ympäristössä. Uskon löytäväni tällä tutkimusmenetelmällä ne yhdistävät tekijät eri yritysten ja kone-linjojen välillä, joiden avulla Big datan käyttöä voitaisiin lähteä hyödyntämään.

Tuloksien odotan keskittyvän parempaan tiedonhallintaan, järjestelmä joka näyttäisi suoraan mitkä parametrit ovat ns. "punaisella" eli poikkeavat normaaliarvoistaan milläkin hetkellä, jotta tuotantoon vaikuttavien ongelmien jäljille päästäisiin nopeammin. Kun puhutaan tuhansista muuttujista jotka vaikuttavat sellumassan tai paperi-/kartonkikoneen laatuun ongelman korjaaminen voi lopulta olla yllättävänkin helppoa, vian paikallistamiseen sen sijaan voi mennä tunteja ja sen jälkeen vasta päästään miettimään mahdollisia korjaustoimia ja niiden ajoittamisia.

Jos edellä kuvattuja tuloksia löytyy, niitä voidaan hyödyntää ennen kaikkea jo olemassa olevien toimintatapojen yhdenmukaistamisessa sekä niiden hyödyntämisessä tiedonkeruun tehostamiseksi ja sen saattamiseksi useamman henkilön saataville entistä helpommin. Tutkimustuloksia voidaan hyödyntää ennen kaikkea jo olemassa olevien toimintatapojen yhdenmukaistamisessa sekä niiden hyödyntämisessä tiedonkeruun tehostamiseksi ja sen saattamiseksi useamman henkilön saataville entistä helpommin.



## 8 AINEISTO JA MENETELMÄT

Tutkimuksen tavoitteena oli kartoittaa metsäteollisuuden henkilöstön kokemuksia IT:n ja liiketoiminnan yhteenlinjauksesta ja siitä, kuinka selkeitä tiedon keräys ja sen käyttäminen eri organisaatioissa ovat. Tutkimuksen menetelmänä toimii puolistrukturoitu teemahaastattelu, kvalitatiiviseen menetelmään päädyttiin, sillä tutkimuksen tavoitteisiin pääsy loi paineen löytää kokemuksia, näkökulmia ja muut kokemusmateriaalia koko metsäteollisuuden asiantuntijakentältä. Haastatteluilla saatiin kerättyä tuota tärkeää empiiristä aineistoa metsäalan esimiehiltä ja asiantuntijoilta. kvalitatiivisen tutkimuksen aineistosta ei voi tehdä laajasti yleistettäviä päätelmiä, yksittäisistä tapauksista on tutkimuksen aikana kuitenkin mahdollista havaita kokonaisilmiössä toistuvat, merkitsevät asiat. (Hirsjärvi ym. 2007)

### 8.1 Kohderyhmä ja aineistonkeruu

Tutkimuksen kohderyhmänä olivat metsäteollisuuden asiantuntijat edustaen useita yrityksiä, paikkakuntia ja ikäryhmiä. Kaikki ovat alan koulutuksen saaneita ja toimivat joko insinööri-, asiantuntija- tai päällikkötason tehtävissä metsäteollisuuden tuotanto-organisaatiossa. Tutkimuslupa kysyttiin jokaiselta henkilöltä erikseen ja haastattelut suoritettiin joko henkilökohtaisesti tai puhelimitse. Tutkija sopi joko puhelimitse tai sähköpostilla haastatteluajankohdan ja -paikan sekä haastattelun keston. Tutkimukseen osallistuneet henkilöt haastateltiin välillä 02.04.2017-23.06.2017. Kysymykset laadittiin soveltaen teoriaosuuden tärkeimpiä menestystekijöitä sekä liiketoiminnanhallinnan kannalta tärkeitä teemoja. Yhden henkilön haastattelun kesto vaihteli vajaasta puolesta tunnista aina yhteen tuntiin asti. Osa haastateltavista vastaili kysymyksiin erittäinkin suorasti ja tarkasti, mutta isomman osan kanssa syntyi enemmän polveilevaa keskustelua. Teemat käytiin haastatteluissa poikkeuksetta läpi, eikä lisäkysymyksiä tarvinnut esittää.

## 8.2 Teemahaastattelu aineiston hankinnassa

Teemahaastattelurunko perustuu kirjallisuuskatsaukseen, eli siihen mitä asiasta valmiiksi tiedetään. Haastattelun teemat liittyvät Big Datan ja liiketoiminnan hallintaan tarkoitettujen tietojärjestelmien käyttöön ja mahdollisuuksien kuvaamiseen, sekä tuotanto-organisaatioissa että sidosryhmien kanssa. Mahdollisuuksien kuvaamisen tarkoituksena olisi selvittää, mistä asioista voitaisiin saada sellaista hyötyä, että niiden analysointi ja niihin investointi olisi kannattavaa. Tämän kokonaisuuden selvittäminen tuo ymmärrystä data-analytiikan takana olevien teknologioiden priorisoinnista mietittäessä potentiaalisia vaikutuksia tiedon analysointiin ja teollisuuden toiminnanohjaukseen useiden eri toimijoiden ja sidosryhmien näkökulmasta. Se kuinka henkilöt kokevat prosessien tehokkuuden ja käytettävyyden on mahdollista johtaa takaisin käytettyihin teknologioihin, jolloin saadaan tärkeää tietoa siitä mitä osia tietojärjestelmästä tulee kehittää paremman käyttäjäkokemuksen turvaamiseksi. On myös tärkeää todeta, että tietoa on vaikea kerätä mikäli tiedon kerääjät kokevat keräämisen vaikeana ja jopa vastenmielisenä. Tämä mielekkyys voidaan johtaa suoraan myös analysoinnin mielekkyyteen, jolloin mahdollinen analysoinnista saatava hyöty saattaa valua täysin hukkaan organisaation kokiessa uudet tehokkaat työkalut hyödyttömiksi. Tieto-opillisesti haastattelutyypilliset tutkimukset kuuluvat empirismiin, jolloin haastatteluista saatava tieto pohjautuu haastateltavien kokemuksiin. Haastattelututkimus ei siis varsinaisesti tutki todellisuutta, vaan haastateltavien kokemusta itse todellisuudesta. (Kurkela, 2014)

Viikko ennen haastatteluiden alkua esihaastateltiin metsäteollisuudessa työskennellyt diplomi-insinööri, joka toimii nykyään opetustehtävissä oman toimensa ohella. Esihaastattelun tarkoituksena oli selvittää haastattelutilanteen kesto sekä tuoda esiin mahdollisia puutteita teemahaastattelurungossa, joka toimii haastattelutilanteessa muistilistana ja samalla ohjaa keskustelua. Esihaastattelulla testattiin myös aihepiiriä ja kysymysten muotoa. (Hirsjärvi & Hurme, 2000)

Esihaastattelun perusteella selvisi myös, että teemahaastattelurungolla saatiin vastauksia tutkimuskysymyksiin vaikka teemoista kerrottiin haastateltaville jo ennen haastattelun alkua. Lisäkysymyksiinkin oli varauduttu, mikäli haastateltava olisi kokenut oman näkemyksensä liian suppeaksi, jolloin kerättävän tiedon määrä olisi jäänyt rajalliseksi, jopa pinnalliseksi.

Tutkimusaineisto kerättiin yksilöhaastatteluilla haastatteleamalla 10-20 paperiteollisuuden ammattilaista, joilla on taustallaan korkeakouluopintoja. Haastattelun alkuvaiheessa tutkija kertasi vielä haastattelun teemat. Lisäksi kerrotaan haastattelun kulku, haastattelun mahdollinen nauhoittaminen mikäli haastattelu tapahtuu kasvokkain sekä kerrataan tutkimuksen tarkoitus ja tutkimuksen luottamuksellisuus. Tutkija kertoo myös alustavan aikataulun raportin valmistumiselle sekä ajankohdan, jolloin raporttiin pääsee tutustumaan.

Ennen haastattelun alkua haastateltavan taustatiedot kerätään strukturoidulla kyselylomakkeella, jota ei esitellään kysymysten yksinkertaisen luonteen ja rajallisten vastausvaihtoehtojen vuoksi. Haastattelut tulevat etenevästi teeman mukaisesti, mutta teemat eivät ole tiukasti rajattuja ja vastauksetkin voivat olla hieman kertomuksen luonteisia. Tulee kuitenkin huolehtia, että kaikki teemat käsitellään ja kysymyksiin saadaan mahdollisimman täsmälliset vastaukset. Eniten huolta tutkijalle aiheuttaa se, että (Hurme & Hirsjärvi, 2009) korostavat haastateltavan motivaation keskeisyyttä haastattelun onnistumiselle. Mikäli en onnistu motivoimaan haastateltavaa, tai hän ei vain yksinkertaisesti ole motivoitunut vastaamaan kysymyksiin vaikkapa kiireen tai muun syyn vuoksi, on vaikea saada mielekkäitä ja syviä vastauksia tutkimuksen kannalta. Hajamielisyys tai vastausten pintapuolisuus tulevat myös mieleen mahdollisina ongelmina tutkimuksen luotettavuuden ja vastausten mielekkyyden kannalta.

Teemoja on kolme, sillä liian monta teemaa voi olla haitallista haastateltavalle sillä se voi uuvuttaa haastateltavan vaikuttaen vastausten sisältöön (Auerbach & Silverstein, 2003) Teemoina ovat siis organisaation ja sidosryhmien hyödyt Big Datasta ja tietojärjestelmistä, sekä niiden kehittäminen. Tutkimus kuitenkin pitäytyy kolmessa teemassa, sillä niistä saa riittävän määrän kysymyksiä aikaiseksi ja teemoittamisen tarkoituksena on kuitenkin jaotella tiedonkeruussa ilmi tulleet tiedot valittujen teemojen mukaan. Esihaastattelun pohjalta ja sen perusteella muokattu haastattelu esitettiin kaikille haastateltaville. Aiemmin mietittyjä lisäkysymyksiä ei tarvinnut ottaa esille, sillä nykyisten kysymysten parista saatiin enemmän kuin riittävästi analysoitavaa tietoa paitsi tutkimuskysymyksistä, niin myös niiden välittömästä läheisyydestä.

### 8.3 Aineiston analysointi

Aineisto analysoidaan laadullisesti induktiivisella eli aineistolähtöisellä sisälönanalyysillä. Laadullisen analyysin tarkoituksena on rakentaa kokonaisvaltainen kuva tutkittavan asian laadusta, ominaisuuksista ja erilaisista merkityksistä. (Lähdesmäki, Hurme, Koskimaa, Mikkola & Himberg, 2012) Analyysia ohjaavat tutkimuksen kysymykset. Ennen analyysiä tulee päättää tarkastellaanko ilmisältö vai myös piilossa olevat viestit. (Tuomi & Sarajärvi, 2004). Tässä nimenomaisessa tutkimuksessa tarkastellaan ainoastaan ilmisältöä, sillä tutkijan tulkinta vaikuttaa piilosisällön analyysissä merkittävästi tutkimuksen todennäköisyyteen. (Kyngäs & Vanhanen, 1999) Aineistoa pyritään analysoidaan mahdollisimman objektiivisesti ilman ennako-oletuksia.

Henkilökohtaisen haastattelutilanteen haluavien nauhat kuunnellaan ja litteroidaan kokonaisuudessaan heti haastattelupäivänä. Näin kaikki puheen vivahteet taukoineen ja äänenpaineineen ovat tutkijalla tuoreessa muistissa.

Litteroinnin jälkeen aineisto käydään useaan kertaan läpi ja sieltä etsitään ne asiat, jotka vastaavat tutkimuskysymyksiin ja tutkimuksen tarkoitukseen.

## 9 TUTKIMUKSEN TULOKSET

### 9.1 Haastateltavien taustatiedot

Haastatteluun osallistui 10 metsäteollisuuden ammattilaista, jotka toimivat sekä insinööri-, että päällikkötason tehtävissä haastattelun aikana. Haastateltavista kaksi ovat naisia ja loput miehiä iän vaihdellessa 28-44 ikävuoden välillä, keski-ikä ollessa 35 vuotta. Haastatteluun osallistuvat työskentelevät tuotantoorganisaatioissa eri metsäalan yrityksissä ja he valmistavat kaikkea kuitupohjaista aina sellusta pakkauskartonkiin asti. He toimivat poikkeuksetta esimiehinä tai asiantuntijatehtävissä ja niiden lisäksi heillä on myös tulosvastuuta. Kaikki ovat kokoaikatyössä. Haastateltavilla on metsäteollisuuden työkokemusta 6 vuodesta 20 vuoteen keskiarvon ollessa 9 vuotta. Kaikki haastateltavat olivat vakituisessa työsuhteessa. Heistä 4 on työskennellyt nykyisessä yrityksessä yhtäjaksoisesti yli 10 vuotta ja loput alle 6 vuotta. Haastateltavista kolme tekee myös vuorotyötä, loput ovat päivätöissä. Jokainen haastateltava on työskennellyt kuluneen vuoden aikana tuotannon/kunnossapidon kehittämissuhteissa. Haastateltavien profiilit esitellään taulukossa 2.

Ensimmäinen innokas tutkimuskohde oli 36-vuotias asiantuntija (Haastateltava H36), joka ohjaa matriisiorganisaatiossa laatuvaatimusten toimintaa toimien liiketoimintatiedon hyödyntäjänä liittyen asiakkaiden laatuvaatimuksiin ja vie niitä tehtaan raportointijärjestelmien kautta kentälle konemiestien tietoon. Toinen haastateltava oli 31-vuotias esimies (H31), joka hyödyntää tietovarastojen historiatietoja työssään ja laatii niiden pohjalta myös raportteja. Hän on myös ollut mukana rakentamassa oman alueensa raportointimalleja. Kolmas haastateltava on 50-vuotias logistiikan asiantuntija (H50), jonka tehtävänä on kerätä paitsi tietoa järjestelmään, myös koostaa sitä raporteiksi johtajatasoa varten. Neljäntenä ja viidentenä suoritettiin yhteishaastattelu, johon osallistui kaksi tuotannon esimiestä joiden toimenkuvaan kuuluu liiketoiminnalle arvokkaan informaation kaivaminen päivittäisestä tuotantodatasta. He ovat molemmat naisia ja iältään 29- & 28-vuotiaita. (H29 & H28) Kuudes haastateltava on 40-

vuotias käyttöpäällikkö (H40), jonka tehtävä on linjaorganisaation johtamisen lisäksi kehittää konelinjansa digitalisointia ja tiedonkeruuta. Hän kuitenkin kokee, ettei tiedä tarpeeksi digitalisaation ja analysoinnin mahdollisuuksista käyttäköseen niitä tehokkaasti päätöksenteon tukena. Seitsemäntenä haastateltavana toimii kartonkilinjan entinen käyttöinsinööri (H33), joka oli vastuussa erilaisen kemikaalien käytöstä kartonkilinjalla. Työkseen hän analysoi kartongin laatua ja ajettavuutta kartonkikoneella ja sitä, kuinka kemikaalien sekoitussuhteet siihen vaikuttavat. Nykyään hän työskentelee suuren kemikaaliyrityksen palveluksessa toimittaen kemikaaleja eri metsäteollisuusyrityksille. Kahdeksas haastateltava (H38) on toiminut eri tehtailla päällikkönä jatkuvan kehittämisen projekteissa ja siten ollut mukana useissa eri tuotannon tehostamisprojekteissa, joihin kaikkiin on liittynyt automaation lisäämistä, kuin myös digitaalisia hyppäyksiä eteenpäin. Yhdeksäs haastateltava (H42) on toiminut pitkään paperikoneen linjaorganisaatiossa sekä vuoro-, että päiväorganisaatiossa. Hänellä on esimieskokemusta ja hän on ollut avainhenkilönä esimerkiksi kokonaisten pakkaamoiden ja logistiikkajärjestelmien uusinnassa. Hänellä on taustaa automaatiikasta ja robotiikasta, sekä näiden kahden asian tuomisesta tuotantoympäristöön. Kymmenes haastateltava (H44) on toiminut käyttöpäällikkönä eri yrityksissä niin kartonki- kuin paperikoneilla. Hänellä on kokemusta myös kehityspuolelta ja hän haluaa aktiivisesti tuoda uusia tietojärjestelmiä metsäteollisuuden pariin. Hän ymmärtää data-analytiikan tuomat hyödyt metsäteollisuudelle ja käyttää mielellään analytiikkaa päätöksentekonsa tukena.

#### TAULUKKO 2 Haastateltavien profiilit

Sukupuoli	Ikä	Asema	Liiketoimintaympäristö
Mies	36	Asiantuntija	Laatupalvelut
Mies	31	Esimies	Tuotanto
Mies	50	Asiantuntija	Logistiikka
Nainen	29	Esimies	Tuotanto
Nainen	28	Esimies	Tuotanto
Mies	40	Päällikkö	Tuotanto
Mies	33	Asiantuntija	Tuotanto
Mies	38	Päällikkö	Kehitys
Mies	42	Asiantuntija	Kehitys
Mies	44	Päällikkö	Tuotanto

## 9.2 Tietojärjestelmien ja Big Data-sovellusten käytettävyys liiketoimintaympäristössä

Ensimmäinen teema sisältää itse järjestelmien käytettävyyden käsittelyä perinteisessä liiketoimintaympäristössä. Onko uusista Big Data- järjestelmistä todellista hyötyä hetkessä liiketoimintaympäristössä ja antavatko ne apua päätöksentekoon erityisesti tarkasteltaessa liiketoiminnan kannalta merkittäviä päätöksiä. Tuleeko ensin pitää huolta siitä, että järjestelmät ovat kokemattomamman käyttäjän kannalta mielekkäitä käyttää, vai onko tärkeämpää että ne jotka osaavat käyttää järjestelmää pystyvät hyödyntämään sitä mahdollisimman tehokkaasti? Teemassa pyrittiin selvittämään, käyttävätkö päättävissä asemassa olevat henkilöt tietojärjestelmiä muuhun kuin tilannetarkistuksiin, varsinkaan uudemman sukupolven massadatajärjestelmiä, joissa tilannekuva muodostuu useammasta eri lähteestä. Haastattelujen perusteella voidaan todeta, että käyttäjät ovat jakautuneita käytön suhteen.

Lisäksi muutamat kokivat käytön kankeaksi varsinkin, kun pitäisi tehdä nopeasti kestäviä päätöksiä liittyen järjestelmästä saatuun raakadataan.

H33: Jos vaikka sileys heikkenee yhtäkkiä koneella ja pitää löytää syy joka sen aiheuttaa, kysyn edelleen mieluummin koneenhoitajalta joka pystyy saman tien sanomaan jotain josta ongelmaan voidaan lähteä ratkaisemaan. Tietokoneelta saman asian katsominen vie aikaa, jota vikatilanteissa ei aina ole.

Käytännössä siis järjestelmien käytettävyyttä tulisi parantaa, jotta käyttäjä pystyisi mahdollisimman helposti paikallistamaan vian syyn. Toisaalta taas ongelmien jälkitarkasteluun tietojärjestelmät koettiin erittäin hyödyllisiksi. Niistä pystyi löytämään paitsi katkojen ja muiden vikojen syitä, niin myös käymään vikatilanteita läpi miehistön kanssa ja oppimaan yhdessä niistä.

H40: Voi olla, että miehistöä ei näe pariin viikkoon, kun itse on päivähommissa ja vuoro kiertää omalla painollaan. Vikatilanteet kuitenkin näkee aina koneen ruudulta ja niitä voi käydä jälkikäteenkin läpi vuoron kanssa ja katsoa, mikä silloin meni pieleen. Vuoromestarilla menee niin paljon aikaa henkilöstön pyörittämiseen, että on hyvä antaa lisätukea kentälle vikatilanteissa. Siihen tarkoitukseen meidän järjestelmät ovat hyviä.

Myös ohjeistuksen merkitys erityisesti uran alkuvaiheessa tuli selkeästi esille, jos ei osannut käyttää järjestelmiä oli pakko luottaa työkavereihin jotka osasivat tai sitten kysellä kentältä ja luottaa kenttämiehistön kokemukseen. Kynnys aloittaa järjestelmien käyttö on korkea, mutta ilmeisten etujen tultua esille useimmat aloittivat ainakin perusominaisuuksien käytön eri tietojärjestelmissä.

H38: En mä voinut olla ukkojen vietävänä koko aikaa, oli pakko opetella ite käyttään järjestelmää jotta tietää missä mennään vikaan ja onko ne viat toistuvia. Itse opettelin käyttämään ja kyllähän siitä apua on ollut, oppimiskynnys oli korkea mutta sitten kun perusasiat hahmotti, niin alkoi helpottamaan. Kaikki tieto siellä on mitä tarvitsee, ja paljon enemmänkin, se pitää vain osata etsiä.

Ristiriitaisia näkemyksiä oli sen suhteen, kuinka sokeasti järjestelmään tulee luottaa ja voiko kaikkia päätöksiä tehdään sen perusteella. Varsinkin iäkkäämällä vastaajilla oli kokemuksia sokean tietojärjestelmäuskon vallassa tehdyistä karkeista virhearvioinneista prosessin ajovaiheessa, mutta konsensusta löydettiin siitä että järjestelmät ovat viime vuosina menneet huomattavasti parempaan suuntaan. Liiketoimintatieto oli monilta osin rajoittunut lähinnä siihen, millaisia täyteaineita asiakkaat halusivat päällysteisiinsä tai siihen, kuinka paljon enemmän uudesta tuotelaadusta voidaan asiakkaalta laskuttaa, mikäli koeajossa päästään esimerkiksi asetettuihin opasiteetti- tai sileystavoitteisiin.

H42: Vaikeahan se on välillä kommunikoida sitä mihin suuntaan myynti haluaisi että tuotteet kehittyvät. Paperimiehille kentällä kuitenkin on tärkeintä että on selkeät tavoitteet joihin koneella on mahdollista ammattitaidolla päästä. Sitä kautta niitä tavoitteita tulee asettaa, että katsotaan järjestelmästä millä parametreilla viime ajoa on ajettu ja millaista laatua silloin saatiin aikaiseksi. Sitten mietitään yhdessä että mitä voidaan parantaa ja mistä kohdin konetta.

Tuotannon tukitoiminnoissa taasen tietojärjestelmät ja varsinkin uudemmat massadataa hyväkseen käyttävät järjestelmät koettiin erittäin hyödyllisiksi. Varsinkin logistiikan saralla eri historiatiedot yhdistävät järjestelmät olivat kullanarvoisia, sillä muutoksia oli helpompi hallinnoida vaikka useampi eri osapuoli operoi järjestelmiä.

H50: Meillä kuljetusliikkeet pääsevät myös osaan tehtaan logistiikkatietokannasta varaamaan itselleen kuljetuksia sen mukaan, kuinka paljon tuotantoa koneelta tulee. On äärimmäisen helppo pysyä tilanteen tasalla, kun samalta näytöltä näkee paitsi historian siitä kuinka kauan keskimäärin aiemmin johonkin asiakastoimitukseen on käytetty aikaa laivalla, junalla tai autolla. Sen jälkeen näkee että saman kuljetusliike on tulossa hakemaan samantyyppistä kuormaa, joten voi hyvillä mielin olla asiakkaaseen yhteyksissä ja luvata heille, että 99% todennäköisyydellä heidän tilaamansa tuote on aikataulun mukaan perillä.

Haastateltavien kanssa keskusteltiin myös yleisesti siitä, kuinka uusia järjestelmiä on yleisesti helpompi käyttää kuin vanhoja järjestelmiä, vaikka saatavissa olevan tiedon määrä on aivan toisella tasolla. Haastateltavat pitivät äärimmäisen tärkeänä, että käytettävissä olevat järjestelmät ovat mahdollisimman joustavia ja monikäyttöisiä. Käytännössä niin, että oli mikä tahansa ongelma liittyen heidän työhönsä, niin sen voisi yhden käyttöliittymän takaa ratkaista. Tämän lisäksi myös vasteajat koettiin tärkeiksi, pilvipalveluiden käytöissä laskentatehoa ei koettu ongelmaksi, vaan sitä jos suuri datamäärä tuli siirtää pilveen laskentaa varten. Monella tehdaspaikkakunnalla nettiyhteydet eivät ole edes



100mb/s - luokassa, joten laskentatehoa haluttiin mahdollisimman ”lyhyen piuhan” päähän.

### 9.3 Tiedon keräämisen ja käyttämisen selkeys organisaatiossa

Toisessa teemassa keskityttiin itse tietoon ja sen käyttöön organisaatiossa. Kuinka hyvin henkilöt tiesivät mitä tietoa heidän tulisi käyttää ja toisaalta minä tiedon varassa heidän esimiehensä tekivät päätöksensä. Vastaako informaatio heidän tarpeitaan ja pystyvätkö he lisäksi kehittämään palautteellaan järjestelmiä haluamaansa suuntaan. Onko prosessi ja prosessin mukaisesti kerätty tieto tärkeämpää, vai annetaanko enemmän painoarvoa sille että organisaation toimijat pääsevät vaikuttamaan keräämänsä tiedon laatuun ja sen pohjalta tehtyihin päätöksiin? Lisäksi haluttiin tietää, ymmärtävätkö haastateltavat tiedon liikkumista omassa organisaatiossaan, sitä kuinka se lopulta vaikuttaa päätöksentekoon vai vaikuttaako se lopulta lainkaan. H44 totesi, että hänelle on ongelmallista motivoida vuorohenkilöstöä käyttämään raportointijärjestelmiä, sillä he eivät näe sitä prosessia kuinka heidän raportoimaansa tietoa käytetään ylemmillä portilla erittäin tärkeänä päätöksenteon välineenä. Vuorohenkilöstö käyttää siis tietojärjestelmistä saamaansa tietoa kyllä prosessin välittömään ohjaukseen ja muutoksiin reagoimiseen, josta he myös kokevat saavansa hyötyä oikeiden päätösten myötä. Se, kuinka nuo päätökset tulisi raportoinnissa selittää auki ja varsinkin se, kuinka paljon lopulta tietojärjestelmien tuottaman informaatio lopulta vaikutti päätöksentekoon jää pimeäksi, sillä vuorohenkilöstöllä ei ole halua tai kykyä kertoa ylemmille tahoille kokonaisvaltaisesti siitä, kuinka nuo järjestelmät prosessinohjauksessa heitä lopulta auttavat. Tähän mahdollista vaikuttaa ottamalla vuorohenkilöstö mukaan niihin palavereihin, joissa heidän keräämänsä tiedon pohjalta tehdään päätöksiä.

Saman suuntaisia ajatuksia oli myös tuplahaastateltavilla H28 & H29:llä. He kokivat että heidän esimiehensä painostavat heitä keräämään tietoa kentältä järjestelmiin esimerkiksi synkronoimalla katkokameroiden aikaleimoja järjestelmän virtausmittausten aikaleimoihin, mutta he eivät tieneet oliko tämä heidän oman työnsä edistämiseksi vai käyttikö tätä informaatiota myös muut tahot organisaatiossa.

H28: Kyllä järjestelmiä pitäisi kehittää siihen suuntaan, että jokainen saisi saman tiedon ja osaisi sitä tulkita. Nyt ollaan kuitenkin vielä hyvin pitkälle jokainen omissa poteroissaan ja jos joku ehtii, niin hän pystyy sitten yhdistelemään tiedonmurusia. Onhan meillä valmiita ohjelmamalleja, joilla saa kaikki käyrät yhdelle kuvaajalle ja pystyy sieltä etsimään samankaltaisuuksia, mutta kun miehistöresurssi on lopussa niin ei ole vain mahdollisuutta alkaa tutkimaan käyriä. Vuoron esimiehenä lähinnä sammuttelen tulipaloja ympäri tehdasta, aina on vain tärkeintä että saadaan kone takaisin ajolle katkon jälkeen.

Heitä häiritsi myös se, että usein uusia järjestelmiä käyttöön otettaessa jokin vanhasta järjestelmästä tuttu, hyväksi havaittu ominaisuus poistuu. Lisäksi pitäisi tietää jo etukäteen kuinka järjestelmät kehittyvät ja mitä tietoa jatkossa prosessista halutaan onkia irti.

H29: Onhan se turhauttavaa, että opettelee käyttämään järjestelmää ja sitten se järjestelmä korvataan. Yleensä vieläpä niin, että järjestetään kyselyitä joissa kysytään tärkeimpiä ominaisuuksia työn kannalta ja ne luvataan säilyttää. Sitten kun tulee käyttöönoton aika niin niitä joko ei ole lainkaan, tai ne on piilotettu niin hankalan käyttöliittymän taakse, että ne ovat käytännössä käyttökelvottomia. Ei se tiedon puute ole ongelma, vaan se että voidaan ajaa suoraan sellaisia raportteja kuin halutaan mutta aina ei tiedetä, että mitkä ne tärkeimmät parametrit siinä raportissa ovat.

Tietojärjestelmien kehittäminen on siis vaikeaa myös organisaatiolle, joka on muutoksen kohteena. Heillä kuluu jo entisestään ohuen organisaation resursseja ylimääräiseen koulutautumiseen ja uuden opetteluun. Lisäksi välikädessä olevat henkilöt, jotka paitsi vastaavat uusien järjestelmien käyttöönotosta, niin myös haluavat muokata niitä järjestelmiä vastaamaan käyttäjien tarpeita kokevat, että heille säilytetään liian iso kuorma käyttöönottoprojektien onnistumisesta. Prosessit ja vaatimukset kehittyvät jatkuvasti, tietojärjestelmien tulisi pysyä niiden mukana tai mielellään jopa edellä.

H36: Jatkuvasti halutaan enemmän tietoa puristettavaksi irti järjestelmästä ja oikeastaan nyt laitevalmistajat tarjoavat jo asioita, joihin meidän järjestelmät eivät ole valmiita vielä ainakaan viiteen vuoteen. On vaikea hallita prosessia ja liiketoimintaa, kun olisi mahdollista tehdä vaikka mitä, mutta nykyiset järjestelmät eivät siihen pysty ja jos järjestelmiä lähdetään päivittämään, niin siihen liittyy riski paitsi osaamisen katoamisesta niin myös tuotannollisista katkoksista. Metsäteollisuuden seisakit ovat pisimmillään noin viikon mittaisia, mikä on tietojärjestelmien testausykyllissä juuri ja juuri asennukseen riittävä aika. Testaus on pakko tehdä muualla ja järjestelmää asennettaessa laitettava polvia myöten kaikki ristiin, että järjestelmä toimii ja sieltä saadaan kaikki se tieto ulos, mitä halutaankin saada.

Lisäksi käyttäjiä on useita erilaisia, varsinkin isoissa organisaatioissa käyttäjiä ja tarpeita voi olla satoja. Erilaisia konelinjastoja on isoissa yrityksissä jopa satoja ja ne ovat kaikki eri vuosikymmeniltä, puhumattakaan myyntiorganisaatioista joissa vastuut kulkevat ristiin rastiin. Mikä toimii yhdelle ei siis välttämättä toimi toiselle ja siinä missä joku haluaa analysoida tietoa porukalla, toinen haluaa keskittyä omassa rauhassaan syvällisesti. Tähän päälle toki vielä käyttöoikeuksien hallinta, sillä vaikka joku tieto olisi prosessin kannalta äärimmäisen tärkeä, ei siihen ehkä ole liiketoiminnallisista syistä ole oikeutta kuin pienellä joukolla ihmisiä.

H33: Meillä pääsi kerran päällystettyä kanaaliin, eikä ollut täysin varmaa tietoa siitä, mitä kaikkea mahdollisesti ympäristölle haitallista kyseinen päällyste sisälsi. Siinä kun lähdettiin sitten tutkimaan, että kuinka vakava ympäristöilmoitus tästä tulee tehdä, niin huomattiin että omat laajat käyttöoikeuteni yrityksen intranettiin eivät olleetkaan tarpeeksi. Kemikaalin sisältämät aineet oli julistettu liiketoimintasalaisuuk-

siksi, sillä tietojärjestelmien uusimisprojektissa tietokantoja oli yhdistelty siten, että samoilla tunnuksilla pääsi entistä useampaan tietokantaan. Oikeuksia oli kuitenkin samalla rajattu sillä vaikka tietokantojen väliset esteet poistuivatkin, ei kaikille silti voitu jakaa kaikkea tietoa. Lopulta esimieheni pääsi omilla tunnuksillaan järjestelmään ja selvisi, että aine hajoaa luonnossa aivan normaalisti sekunneissa eikä vaaraa ympäristölle ole.

Kehitysideoita tuli myös runsain mitoin eri yrityksissä työskenteleviltä ja eri organisaatioitasoilta. Ne keskittyivät lähinnä raportoinnin yksinkertaistamiseen ja siihen, että olisi joku täsmähenkilö, joka ohjaisi raportoidut tiedot oikealle paikalleen ja oikeille henkilöille. Käytännössä siis itse raportointiin haluttaisiin ”tyhjä kangas”, johon jokainen asia kirjoitettaisiin eikä siihen kaikessa kiireessä tarvitsisi edes laittaa mitään hakusanoja tunnisteeksi. Kankaasta näkisi kuka on kirjoittanut ja mitä on kirjoitettu, jonka jälkeen yksi henkilö joka tehtaalla pystyisi allokoimaan tätä tietoa sinne, mistä siitä eniten olisi hyötyä.

H38: Itsellä ainakin on jatkuvasti kiire ja kun on päällekkäisiä järjestelmiä, niin ei sitä oikein tiedä mihin sitä kiireessä kirjoittaa jos kirjoittaa. Varsinkin kun raportointi on joissain ohjelmissa tehty niin hankalaksi. Joskus tuntuu että olisi vaikka yhteinen Paint-kangas johon vain kirjoittaisi asiansa ja joku vähän katsoisi perään että tietääkö ne kenen vastuulla asia on sen, mitä sinne on kirjoitettu. Poistuisi tämä järjestelmien välinen sekamelska joskus olisi mahdollisuus kirjoittaakin jotain muistiin ylös, niin olisi apua seuraavassa vikatilanteessa.

Haastatteluiden perusteella voidaankin siis todeta, että edistyksellisten tietojärjestelmien esiinmarssi on tuonut ongelmia siihen, että käyttäjät eivät mikään kaikki tieto heidän tulisi raportoida eteenpäin. Näin ollen voidaan todeta, että tietojärjestelmäinvestointien lisäksi myös käyttäjien koulutukseen ja organisaation kehittämiseen tulisi panostaa resursseja.

## 9.4 Liiketoiminnan ja tietojärjestelmien yhteiset rajapinnat organisaatiossa

Kolmannessa teemassa haluttiin löytää niitä asioita, joissa tietojärjestelmistä oli mahdollista saada runsasta kilpailuetua omaan liiketoimintaan. Kuinka liiketoimintaa johdettiin tietojärjestelmien näkökulmasta ja kuinka tuo johtaminen näkyi arkipäiväisissä valinnoissa. Onko edes mahdollista saavuttaa henkilöstön nykyisellä osaamistasolla merkittäviä kilpailukykyetuja tietojärjestelmien päivittämisellä? Metsäteollisuudessa ei tuotannon puolella pysty tekemään kovinkaan nopeita peliliikkeitä ajettavien laatuojen tai käytettävien materiaalien suhteen. Tilauskirjat ovat pitkiä ja toimitusajat sitäkin pidempiä. Jos haluaa saada kartonkirullan Kiinaan, tulee varautua vähintään kuukauden merimatkaan. Laivausajat satamissa nostavat kuljetusajat tehtaalta asiakkaalle helposti yli kahdenkin kuukauden. Liiketoimintaan voidaan kuitenkin saada hyötyjä välillisesti vaikkapa hakemalla säästöjä raaka-aineista ja energiasta, sekä optimoi-

malla kuljetusreitit ja seisokkiaikatauluja. Tuoko henkilöstön koulutus suurempia etuja kuin päivittäminen uusimpaan järjestelmään? Voivatko organisaation työntekijät johtaa itse itseään tehokkaasti jopa liiketoimintarajapintojen yli ja saavutetaanko sen saavuttamisella edes teoriassa merkittävää hyötyä suhteessa panostukseen?

H31: Liiketoimintarajapintoja on suhteellisen vaikea löytää tuotannon puolelta, luonnollisesti on mahdollista saavuttaa synergiaetuja vaikkapa logistiikan osalta, mutta kovat bisnespäätökset, jotka olisivat suoraan yhteydessä tuotantoon tehdään oikeastaan ainoastaan kunnossapidon ja investointien puolella. Investointilaskelmissa tietojärjestelmiä totta kai käytetäänkin runsaasti, on helppo laskea että jos puusta kuoritaan puita liikaa, eli myös kuitupuuta menee muun kuoren mukana polttoon, niin kuinka paljon kuorimisaikaa pienentämällä voidaan saada lisäsäästöjä. Toki aina on syy miksi nuo kuorimisajat on tehty niin pitkiksi, mutta jos voidaan laskea että sulamisveden lämpötilaa nostamalla jäisen tukin kuorinta-aikaa voidaan lyhentää, niin kyllä se kannattaa laskea ja tehdä investointiehdotus takaisinmaksuaikoihin.

Käytännössä metsäteollisuudessa raaka-aineiden hinnat tarkastetaan noin kerran vuodessa joko ylös- tai alaspäin. Raaka-aineet ovat erittäin merkittävä kuluerä, mutta niiden hintoihin ei tietojärjestelmillä juuri voi vaikuttaa. Lisäksi energia on toinen suuri kuluerä, mutta myös sen hinta on melkoisen riippumaton yksittäisistä tietojärjestelmä uudistuksista. Se mihin voidaan vaikuttaa on tuotannon laatu ja toiminta poikkeustilanteissa, reagoinnin nopeus ja jopa ennakointi. Kun toimitaan valtaviin tuotantoihin ja prosessien kanssa, ensisijaisesti pyritään välttämään mahdolliset ongelmat sillä virheiden korjaus ja tuotannon alasajo on äärimmäisen kallista.

H50: Liiketoimintarajapintaan oma kosketukseni on kohtalaisen suuri, näen logistiikka-asiantuntijana tavaravirtamme ympäri maailmaa ja mikäli joku kuljetus on myöhästymässä, pystyn arvioimaan kuinka paljon se yrityksellemme tulee maksamaan. Meillä on toki sopimukset mahdollisten myöhästymisten varalle, mutta välillä joudumme laskennallisesti ottamaan erittäin kalliita rekkakyytejä vain, jotta pitäisimme asiakkaamme tyytyväisinä. Käytännössä pystyn siis painostamaan tuotantoa tekemään jotain toista laatua, mikäli tiedän että toisen laadun asiakkaan laiva on myöhässä ja se tilaus joudutaan joka tapauksessa kuljettamaan rekoilla perille. On paljon järkevämpää pitää huoli, että muut tilaukset ovat ajoissa perillä, kuin että tilaus jolla menee kaksi viikkoa laivalla meneekin yhtäkkiä viidessä päivässä rekalla triplahinnalla. Sen rekalla menevän tilauksen ehtii ajaa myöhemminkin, silti kuitenkin ajoissa jotta se lopulta rekalla ehtii perille.

Jatkossa näitä rajapintoja on mahdollista kuitenkin löytää lisää, tällä kertaa kuitenkin kenenkään haastateltavan tehtaalla ei ollut käytössä koko tehtaan kattavia ERP-järjestelmiä tai muita koko tehtaan kattavia massadataa hyödyntäviä ohjelmistoja. Raportoinnin helpottamisen lisäksi myös yhteistoiminnan helpottaminen nousi monessa keskustelussa tärkeään rooliin. Se kuinka käyttäjät saataisiin omista siiloistaan näkemään myös se, mitä ympärillä tapahtuu auttaisi huomattavasti jokaista prosessin osa-aluetta. Ympärillä olevat toimijat osaisivat reagoida ja jopa ennakoita mahdollisia vaikeuksia ja näin ollen pystyisivät no-

peammin toimimaan poikkeustilanteiden sattuessa. H28 ja H29 kertovat, että suurimpia riskejä liiketoiminnan kannalta on se, että loppukäyttäjän tarpeita ei kuunnella ja että käyttöönotto on aivan muualla kuin varsinainen käyttö.

H28: Tuotannossa tuntuu välillä muutenkin siltä, että monissa organisaation sisäisissä asioissa taistellaan tuulimyllyjä vastaan. Asioita on vaikea saada vietyä eteenpäin jos ei tiedä juuri sitä oikeaa polkua kyseisen aihepiirin yksittäisen asian eteenpäinviemiseksi. Lisäksi totta kai on hyvä, että käyttäjiä kuunnellaan kun muutoksia aletaan valmistelemaan, mutta huomattavasti parempi olisi, jos muutokset olisivat käyttäjälähtöisiä. Meidänkin miehistö on tehnyt vuosien varrella satoja kehitysehdotuksia nykyiseen järjestelmään, pyrkien siihen että nykyistä, kaikkien edes jollain tasolla osaamaa järjestelmää voitaisiin kehittää pienin askelin eteenpäin vastaamaan yhä paremmin meidän tarpeitamme. Sen sijaan kerran viidessä vuodessa ylhäältä tulee käsky, että nyt täytyy järjestelmä vaihtaa ja tällaisen saatte tilalle, kertokaa mitä osia vanhasta järjestelmästä haluatte siihen sisällyttää niin yritetään saada ne sopimaan. Tämä ei tee hyvää moraalille, eikä sille että työntekijät kokisivat olevansa vaikuttavassa asemassa oman työnsä suhteen. Tämän vuoksi tuntuu vähän hassulta puhua liiketoimintaympäristöön vaikuttamisesta tietojärjestelmien kanssa, kun emme aina saa vaikuttaa edes omaan tietojärjestelmäämme puhumattakaan, että voisimme sillä hirveästi lisäarvoa tuottaa liiketoiminnalle olemattoman koulutusresurssimme puitteissa.

Tutkimuksen tuloksissa muina tietojärjestelmien ja liiketoiminnan yhteenlinjauksen avaintekijöinä löydettiin viestinnän ja selkeyden lisäksi loppukäyttäjien kuunteleminen prosessin alusta asti, sekä tarpeen selkeän määrittämisen jälkeen johdon tuki, vahva ohjaus ja sitoutuminen toteutuvaan muutokseen.

## 10 POHDINTA JA YHTEENVETO

Tutkimuksen tarkoituksena oli selvittää tietojärjestelmien, erityisesti suurten datamassojen käsittelyn mahdollistavien teknologioiden ja liiketoiminnan hallinnan yhteenlinjaukseen, sekä tietojärjestelmien yleiseen käytettävyyteen ja käytännöllisyyteen liittyviä kehitystarpeita eri yrityksistä koostuvien metsäteollisuuden ammattilaisten joukossa. Tutkimusta varten luotiin viitekehys, jonka pohjalle empiirinen tapaustutkimus rakentui. Tiedonkeruuvälineinä toimivat yksilöhaastattelut. Tässä luvussa sekä analysoidaan tutkimuksen tutkimuksesta saatuja tuloksia, että tehdään niiden perusteella johtopäätöksiä. Lisäksi empiirisen osion tuloksia vertaillaan kirjallisuuskatsauksen havaintoihin ja esitetään päätelmiä sekä vastaus tutkimuskysymykseen. Luvun lopussa käsitellään vielä tutkimuksen onnistumista arvioimalla saatujen tulosten validiteettia ja reliabiliteettia sekä pohditaan mahdollisia jatkotutkimusaiheita ja tapoja, joilla jo saatuja tuloksia voidaan hyödyntää.

### 10.1 Tulosten analysointi ja johtopäätökset

Tässä pro gradu - tutkielmassa ratkaistavaksi esitettiin seuraava tutkimuskysymys :

- Mitä tietojärjestelmien ja liiketoiminnan hallinnan yhteenlinjauksen haasteita voidaan tunnistaa metsäteollisuusorganisaatiossa?

Tutkimuskysymykseen haettiin ensin vastausta kirjallisuuskatsauksen kautta tutustumalla nykyaikaiset massadatajärjestelmät mahdollistaviin teknologioihin, sekä niiden potentiaalsiin hyötyihin metsäteollisuuden parissa. Teknologioihin tutustumalla pystyttiin johtamaan potentiaalisia hyötyjä data-analytiikan paris-

ta sekä löytämään niitä ominaisuuksia, jotka ovat kriittisiä massa-dataa hyödyntävien tietojärjestelmien onnistumiselle niin organisaation kilpailukyvyn, kuin pidemmällä aikavälillä koko teollisuudenalan kilpailukyvyn kannalta. Tämän jälkeen oli empiirisen tapaustutkimusosuuden vuoro. Tapaustutkimus toteutettiin yksilöhaastatteluina, joissa haastateltiin yhteensä kymmentä alan ammattilaista jotka ammatillisessa mielessä kuuluvat asiantuntija- sekä päällikkötasolle omissa organisaatioissaan. Kirjallisuuskatsauksen ja haastatteluiden perusteella saatiin tehtyä havaintoja, joista pystyttiin rakentamaan vastaus itse tutkimuskysymykseen.

Kirjallisuuskatsauksen tärkein havainto oli, että suurimmat haasteet liittyen tietojärjestelmien tehokkaaseen ja oikeaoppiseen käyttöön liiketoiminnan hallinnassa liittyvät tiedon puutteeseen siitä, mitä kaikkea kaikella sillä kerätyllä datalla tulisi tehdä (Davenport & Harris, 2007). Toiseksi tärkeäksi näkökulmaksi voitiin nostaa myös se seikka, että päättävälle elimille ohjattu tieto tulisi olla käytäntöjen mukaista, sellaista mistä on yhdessä sovittu erikseen organisaation tietotarpeita määriteltäessä. Käytännössä siis käytännöllistä, analysoitua tietoa liian yksityiskohtaisen datamassan sijaan. (Pirttimäki, 2007)

Nämä kirjallisuuskatsauksessa esiintyneet haasteet näkyivät voimakkaasti myös tutkimuksen kohdehenkilöiden vastauksissa. Epäselvyys siitä kuka tietoa käyttää ja mitä tarkoitusta varten tietoa ylipäätään kerätään, olivat suurimpia kysymysmerkkejä aihepiirin sisällä. Jopa tutkijan itsensä kuitenkin yllätti se, kuinka paljon itse tiedon keruuprosesseista löydettiin parannettavaa ja kuinka paljon epäselvät käytännöt saivat palautetta osakseen organisaatiotasolla.

Muiden käytäntöjen lisäksi kritiikkiä sai liiketoimintapuolelta se, että jo kerättyyn tietoon on vaikeaa päästä käsiksi. Tämä korostuu erityisesti silloin, kun henkilöt vaihtavat työpaikkaa yrityksen sisällä ja vielä enemmän henkilön hypätessä kokonaan toisen yrityksen palvelukseen. Tämä aiheuttaa haasteita, sillä silloin tietojärjestelmien ja liiketoiminnan yhteenlinjautumisen sijaan ”pyörä joudutaan keksimään uudelleen”, eli organisaation hallussa jo oleva tietopääoma joudutaan hakemaan ja rakentamaan järjestelmästä uudestaan.

Tietokantojen puutteellinen käyttökokemus nähtiin näissä tapauksissa vaikeimmaksi seikaksi, eli tiedettiin että tieto on järjestelmiin tallentunut ja se olisi tutkittavassa muodossa, mutta siihen ei vain päästä riittävän helposti käsiksi. Lisäksi osa haastateltavista nosti ongelmaksi sen, että heillä oli tiedossa liiketoimintaprosessi, mutta ei tuotantoprosessia. Eli käytännössä he kyllä tiesivät kuinka tuotteella tuotetaan asiakkaille lisäarvoa, mutta eivät sitä kuinka itse tuote valmistetaan. Tällöin on hankala innovoida uutta, kun ei tiedä mitä kaikkea oma tuote mahdollistaa.

Uutena asiana nousi myös tarve paikallisen tukihenkilön tarpeelle liiketoiminnan datan hallitsemisen opettamiseen. Vanhoja excel-taulukoita asiakastapaamisten menneistä ajankohdista ja käsitellyistä aiheista olisi kyllä pystynyt viemään uuteen järjestelmään, mutta se olisi tullut tehdä manuaalisesti tiedostoja kopioimalla eikä ohueksi venytetyllä organisaatiolla ollut siihen toimintaan osoittaa resursseja. Tämän tyyppisen liiketoimintatiedon siirtäminen osaksi uutta järjestelmää mahdollisimman kivuttomasti olisi kuitenkin äärimmäisen

tärkeää. Kirjallisuuskatsauksessakin todetaan, että kun informaatio on tarkkaa ja kattavaa uudessa järjestelmässä, pystytään vähentämään liiketoiminnan kustannuksia ja siihen kohdistuvia riskejäkin. (Vaarnas & Virtanen, 2001)

Useat haastateltavat toivat myös esiin sen seikan, että vaikkakin tietojärjestelmiä on yhdenmukaistettu vuosien varrella, niin silti tuntuu että tiedolle on liian monta kanavaa, joita on vaikea priorisoida ja yhdistää. Kokonaisvaltaisuus puuttuu ja monet kokivatkin, että liiketoimintatiedon hallinta oli Herschel & Jonesin (2005) toteamaa suppeasti määriteltyä hallintaa joka tukeutuu hyvin pitkälti perinteiseen raportointiin. Lisäksi resurssien puute haittaa myös koulutuspuolella kunnollisen tilannekuvan muodostamista, sillä liiketoiminnan ja tietojärjestelmien yhteenlinjauksessa tiedon analysointi on keskiössä. Mikäli käyttäjillä ei ole mahdollisuutta oppia edes käytön perusteita Hovin ym. (2009) määrittelemistä moderneista tiedon analysointityökaluista, (moniulotteinen analysointi, algoritmit, tiedon louhinta & simulaatiot) niin kuinka heidän on mahdollista käyttää näihin samoihin työkaluihin perustuvia sovelluksia menestyksekkäästi työssään?

Ihmetystä aiheutti myös se, että nämä ongelmat ovat olleet toistuvia jo haastateltavien henkilöiden työurien aikana. Keskusteluiden lomassa tuli kuitenkin ilmi, että kokeneemmat työntekijät ovat raportoineet samoista ongelmista tiedon siirtymisen ja kommunikaation puutteen osalta aina heidän työuriensa alusta asti, mitä tukee Kettusen & Simmons (2001) fakta siitä, että nykyaikaiset ERP-järjestelmät yleistyivät vasta 1980-90 - lukujen vaihteessa. Kukaan haastateltavistakaan ei silti kiistä sitä tosiasiaa, että oikein käytettynä erilaiset tietojärjestelmät ja varsinkin ERP-järjestelmät ovat erittäin tehokas työkalu ja paitsi parantavat yrityksen resurssienhallintaa, niin myös laskevat kustannuksia ja parantavat asiakaspalvelukykyä. (Vilpola & Kouri, 2006)

Edelliseen liittyen esitettiin myös kiitosta siitä, että on ymmärretty luopua vanhoista järjestelmistä tietojen katoamisen uhallakin, sillä usein kokonaan alusta aloitettu ratkaisun hakeminen oli tuottanut paremman lopputuloksen kuin vanhojen järjestelmien päivittäminen. Tätä tuki kirjallisuudessa myös Manyika ym. (2009) jossa todettiin että vanhojen järjestelmien suora integrointi ja analysointi tai päivitys on usein erittäin haasteellista. Jolloin on kustannustehokkaampaa aloittaa koko tietojärjestelmän rakentaminen puhtaalta pöydältä.

Kokonaisuutena voidaankin havaita, että haastateltavien esiin tuomat tapaukset ja tilannekuvaukset liittyivät järjestelmien vajavaisuuksien lisäksi lähes poikkeuksetta puutteelliseen ohjeistukseen ja johtamiseen. Tutkijana olikin vaikea vetää rajaa niiden puutteiden välille, jotka johtuivat puutteellisesta tietotaidosta ja toisaalta taas huonosta järjestelmäsuunnittelusta. Jos henkilö totesi järjestelmän olleen vaikeasti käytettävä, tai että "tällä järjestelmällä ei voi saada mitään aikaan." Niin on hyvin haasteellista lähteä selvittämään ennakoasenteiden roolia mahdolliseen oppimiseen tai ylipäänsä sitä onko kunnollista oppimismahdollisuutta järjestelmän käyttöön edes tarjottu. Voi hyvin olla, että iso osa järjestelmistä on hyvinkin järkevästi rakennettu ja muutaman tunnin koulutuspanoksella käyttäjät saivat huomattavasti enemmän irti järjestelmistä, kun oppisivat niiden logiikan ja sen, kuinka tieto kyseisillä järjestelmillä haetaan,



analysoidaan ja prosessoidaan. Liiketoimintarajapintojen löytäminen tuotannosta oli myös haasteellista, sillä tutkimushenkilöt olivat pitkälti omissa ”siloissaan” ja hyvin fokuoituneita siihen, kuinka hyvin heidän oma organisaationsa suoriutuu päivittäisessä toiminnassa.

Yhteenvedona voidaankin todeta, että ennen kuin metsäteollisuudessa voidaan ottaa askeleita kohti liiketoiminnan ja tietojärjestelmien kokonaisvaltaiseen implementointiin, tulee hoitaa perusasiat kuntoon johtamisesta, resursoinnista ja kommunikoinnista lähtien. Tietoa on ajanmukaisten järjestelmien johdosta saatavilla ylen määrin ja sitä pystytään asiantuntijatasolla käyttämään myös liiketoiminnan parantamiseen. Tämä vaatii kuitenkin paljon aikaresurssia ja varsinkin tuotannon ja logistiikan hektisessä maailmassa tuon ajan löytäminen on hankalaa, varsinkin ilman kunnollista tukihenkilöä tai koulutusmahdollisuutta. Sitoutuminen vanhoihin ”hyväksi havaittuihin” toimintamalleihin oli suurta, mutta useimmat uskoivat tämän muuttuvan mikäli vain oppisi käyttämään uusia järjestelmiä kunnolla. Suurin osa vastaajista koki, että järjestelmien käyttö helpottaisi varmasti heidän työkuormaansa pitkällä tähtäimellä, mutta harva oli valmis oman työnsä lisäksi alkaa opettelemaan täysin uusia järjestelmiä. Tähän siis kaivataan koulutusresurssia organisaatioiden tueksi.

Työntekijöillä oli kuitenkin yleisellä tasolla positiivinen asenne tietotekniikkaa kohtaan, etukäteisviestintään tulisi kuitenkin panostaa ja kertoa, minkä takia jotain tiettyä asiaa tehdään ja kuinka se auttaa yritystämme pidemmällä tähtäimellä. Koulutukseen tulee myös panostaa ja sen merkitys korostuukin, mitä monimutkaisempia järjestelmiä henkilöt joutuvat käyttämään.

## 10.2 Tutkimuksen onnistuminen

Hirsjärven ym. (2007) mukaan tutkimuksen onnistumista tulee arvioida reliabiliteetin ja validiteetin näkökulmasta. Reliabiliteetti arvioi tutkimustulosten toistettavuutta ja validiteetti sitä, mittaavatko tutkimuksessa käytetyt mittarit sitä mitä niiden tässä tutkimuksessa halutaan mittaavan. Koen tutkimuksen reliabiliteetin olevan kelvollinen, mutta näen myös että mikäli mukaan olisi otettu alle 10 000 henkeä työllistävissä yrityksissä työskenteleviä henkilöitä, niin sillä olisi todennäköisesti ollut jonkinlaista vaikutusta tuloksiin. Nythän kaikki haastattelut olivat töissä joko Stora Ensolla, UPM:llä tai Metsä-Groupilla. Lisäksi tutkijan tausta on enemmän tuotannon ja logistiikan puolella, joten taloushallinnon asioista ei pystynyt keskustelemaan haastateltavien kanssa yhtä syvästi. Tämä saattoi jättää peittoon joitain onnistumisia ja epäonnistumisia liittyen liiketoiminnan ja tietojärjestelmien yhteenlinjautumiseen.

Tutkimuksen validiteetti on myös kelvollinen, sillä haastattelun kysymykset laadittiin teoriapohjaan perustuen. Tutkija myös selkeästi määritteli käsitteet haastattelutilanteessa, jolloin väärinymmärryksen riski väheni. Yhteenvedona tutkimus voidaan nähdä onnistuneena, sillä tutkimuskysymykseen saatiin paitsi vastaus tutkimuskysymykseen, niin paljon myös muuta informaatiota jota

voidaan käyttää jatkotutkimuksien rakentamisessa. Tutkimus myös vastasi sille annettuihin tavoitteisiin ja tuloksista saatiin poimittua useita liiketoiminnanhallintaan sekä IT:n ja tietojärjestelmien yhteenlinjaukseen negatiivisesti vaikuttavia asioita. Tämän tutkimuksen perusteella onkin mahdollista laatia ohjeistus metsäteollisuuden yrityksille seuraavien tietojärjestelmäprojektien implementoinnista.

### **10.3 Tulosten hyödyntäminen tulevaisuudessa sekä jatkotutkimuksen aiheet**

Metsäteollisuuden tietojärjestelmätarpeista ei ole juuri lainkaan saatavilla ajantasaista tutkimustietoa. Liiketoiminnan hallinnan yhdistäminen tietojärjestelmiin on lähes ennenkuulumatonta alan tutkimuksessa. Se kuinka kaukana kuitenkin Suomen kansantalouden kannalta tärkeällä alalla oltiin tietojärjestelmien käytön optimoinnista ja kuinka arkipäiväisiä ja itse tietojärjestelmistä riippumattomia nuo syyt lopulta olivat, yllätti tutkijan. Tapaustutkimuksella saavutettiin tavoitteet ja luotiin uutta tietoa aiemmin tuntemattomasta aihepiiristä.

Jatkotutkimusaiheista esiin nousee perustutkimus tietojärjestelmien käytön omaksumisesta tehdasympäristössä. Kuinka henkilöt käyttävät tietojärjestelmiä ja ennen kaikkea se, kuinka hyvin he kokevat osaavansa käyttää järjestelmää, oikeasti osaavat käyttää järjestelmää ja kuinka paljon he ovat siihen saaneet opastusta. Lisäksi olisi mielenkiintoista tutkia kirjallisuuskatsauksesta esiin nousseita Big Datan vaikutuksia käyttäjien työkuormaan, mutta koen että metsäteollisuudessa tulee tehdä vielä paljon pohjatyötä käyttäjien tietojärjestelmäosaamisen kanssa, ennen kuin ala pystyy aloittamaan uusimpien tietojärjestelmien hyödyn täyden ulosmittauksen.

Tulevaisuudessa olisi myös äärimmäisen mielenkiintoista toteuttaa pitkästä tutkimus yrityksessä, jossa voitaisiin panostaa henkilöstön kunnolliseen kouluttamiseen ja sen jälkeen vertailla tehokkuus- ja työhyvinvointituloksia prosessin edetessä.

## LÄHTEET

Alavi, M. & Leidner, D.E. (2001). Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MISQ Review* 25(1), 107-136.

Abukari, K. & Jog, V. (2003). Business intelligence in action: Three examples of how it really works. *CMA Management*, March 2003, 15-18.

Apte, C.V., Hong, S.J., Natarajan, R., Pednault, E.P.D., Tipu, F.A. & Weiss, S.M. (2003). *Data-intensive analytics for predictive modeling*, *IBM Journal of Research & Development*. Vol. 47(1), 17-23.

Auerbach, C.F. & Silverstein, C.B. (2003). *An Introduction to Coding and Analysis Qualitative Data*. New York University, New York.

Autio, T. (2014). *Big Data hypestä todellisuuteen*. Haettu osoitteesta <http://www.bilot.fi/fi/big-data-hypesta-todellisuuteen/>

Azvine, B., Cui, Z., Nauck, D. & Majeed, B. (2006). Real Time Business Intelligence for the Adaptive Enterprise. *Proceedings of the 8th IEEE International Conference on E-Commerce Technology and the 3rd IEEE International Conference on Enterprise Computing, E-Commerce, and E-services (CEC/EEE'06)*. Washington, DC: IEEE Computer Society.

Azvine, B., Cui, Z. & Nauck, D.D. (2005). Towards real-time business intelligence. *BT Technology Journal*, Vol. 23(3), 214- 225.

Borthakur, D. (2007). *The Hadoop Distributed File System: Architecture and Design*. Haettu osoitteesta [https://svn.apache.org/repos/asf/hadoop/common/tags/release-0.16.0/docs/hdfs\\_design.pdf](https://svn.apache.org/repos/asf/hadoop/common/tags/release-0.16.0/docs/hdfs_design.pdf)

- Bose, R. (2009). *Advanced analytics: opportunities and challenges, Industrial Management & Data Systems*. Vol. 109(2), 155 – 172.
- Boulon, J., Konwinski, A., Qi, R., Rabkin, A., Yang, E. & Yang, M. (2008). *Chukwa: A Large-Scale Monitoring System*. Proc. Cloud Computing and its Applications (CCA'08). Chicago, IL, USA, 1-5.
- Brelade, S & Harman, C. (2003). *Practical guide to knowledge management*. Thorogood Publishing. Haettu osoitteesta <https://ebookcentral.proquest.com/lib/jyvaskyla-ebooks/reader.action?docID=308926>
- Bryman, A. (2012). *Social research methods*. Oxford University Press
- Chan, Y. & Reich, H. (2007). IT Alignment: What have we learned? *Journal of Information Technology*. Vol. 22, 297-315
- Chang, F., Dean, J., Ghemawat, S., Hsieh, W., Wallach, D., Burrows, M., Chandr, T., Fikes, A. & Gruber, R. (2006). *Bigtable: A Distributed Storage System for Structured Data*. Proc. OSDI'06: 7th Symposium on Operating System Design and Implementation. Seattle, USA, 205 – 218.
- Chen, H., Chiang, R.H.L. & Storey V.C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165-1188.
- Choo, C. (2002). Information management for the intelligent organization : The Art of Scanning the Environment. *ASIST Monograph Series (3rd edition)*. New Jersey : Information Today.
- Chung, W., Chen, H. & Nunamaker Jr. (2002). Business Intelligence Explorer: A Knowledge Map Framework for Discovering Business Intelligence on the Web. *Proceedings of the 36th Hawaii International Conference on System Sciences*, 1-10.
- Collins, B. (1997). *Better Business Intelligence – How to Learn More about Your Company*. Astron On-Line, Letchworth.
- Coltman, T., Tallon, P., Sharma, R. & Queiroz, M. (2015). Strategic IT alignment: twenty-five years on. *Journal of Information Technology*, 30(2), 91-100.
- Connolly, T. M & Begg, C. E. (2005). *Database Systems – A Practical Approach to Design, Implementation, and Management*. Harlow, England : Addison Wesley.

Davenport, T.H. & Harris, J.G. (2007). *Competing on Analytics: The New Science of Winning*, Harvard Business School Publishing Corporation. USA.

Davenport, T. H. (2014). *Big Data at work: dispelling the myths, uncovering the opportunities*. Harvard Business School Publishing Corporation. USA.

Davenport, T.H, (2015). *Three big benefits of big data analytics*. Haettu osoitteesta [http://www.sas.com/tr\\_tr/news/sascom/2014q3/Big-datadavenport.html](http://www.sas.com/tr_tr/news/sascom/2014q3/Big-datadavenport.html)

Dean, J. & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107-113.

Edvinsson, L. & Malone. S. (1997). *Intellectual Capital – Realizing your company's true value by finding its hidden Brainpower*. HarperBusiness, New York.

Evans, J. (2015). Modern Analytics and the Future of Quality and Performance Excellence. *The Quality Management Journal*. Vol. 22(4), 6.

Fahey, L. (2007). Connecting Strategy and competitive intelligence: refocusing intelligence to produce critical strategy outputs. *Strategy & Leadership*. Vol. 35(1), 4-12.

Falshaw, D., Glaister, A. & Tatoglu, Z. (2006). Evidence on Formal Strategic Planning and Company Performance. *Management Decision*, Vol. 44(1), 9-15.

Fischer, M. J., Su, X., & Yin, Y. (2010). Assigning tasks for efficiency in Hadoop. *In Proceedings of the 22nd ACM symposium on Parallelism in algorithms and architectures*, ACM, 30-39.

Franks, B. (2012). *Taming the Big Data Tidal Wave : Finding Opportunities in Huge Data Streams with Advanced Analytics*. Hoboken.

Gates, A. (2011). *Programming Pig*. O'Reilly. Sebastopol. CA. Haettu osoitteesta, <http://www.gocit.vn/files/Oreilly.Programming.Pig-www.gocit.vn.pdf>

Gerow, E., Thatcher, B. & Grover, V. (2014). Six Types of IT-Business Strategic Alignment: An investigation of the constructs and their measurement. *European Journal of Information Systems*. Vol. 24(3), 1-27.

Ghoshal, S. & Kim, S. K. (1986). Building Effective Intelligence Systems for Competitive Advantage. *Sloan Management Review*, Vol. 28(1), 49-58.

Halliman, C. (2001). *Business Intelligence Using Smart Techniques*. Information Uncover, Houston.

Hammarsten, H. (2013, 15. marraskuuta). Big data jyllää teollisuudessa. Haettu 4.7.2017 osoitteesta <http://www.talouselama.fi/kumppaniblogit/tieto/big-data-jyllaa-teollisuudessa-1241614>

Hallikainen, P., Kivijärvi, H. & Nurmimäki, K. (2002). Evaluating Strategic IT investments: An Assessment of Investment Alternatives for a Web Content Management System, *Proceedings of the 35th Hawaii International Conference on System Sciences, Hawaii University*

Halonen, P. & Pirttimäki, V. (2002). *Liiketoimintatiedon hallinta suomalaisissa suuryrityksissä 2002*. eBRC research reports 4.

Halonen, P. & Hannula, M. (2007). *Liiketoimintatiedon hallinta suomalaisissa suuryrityksissä 2007*. eBRC research reports 37.

Hannula, M. & Pirttimäki, V. (2003). Business Intelligence. Empirical Study on the top 50 Finnish Companies. *Journal of American Academy of Business*, 2(2), 593-599.

Henderson, J. & Venkatraman, N. (1990). Strategic Alignment: A model for organizational transformation via information technology. *Cambridge, MA: Sloan School of Management, Massachusetts Institute of Technology*.

Henschen, D. (2011). *BI and Information Management Trends*. InformationWeek Reports, November 2011.

Herschel, R. T. & Jones, N.E. (2005). Knowledge management and business intelligence: the importance of integration. *Journal of Knowledge Management*, Vol. 9(4), 45-55.

Hirsjärvi, S. & Hurme, H. (2000). *Tutkimushaastattelu, teemahaastattelun teoria ja käytäntö*. University Press, Helsinki.

Hirsjärvi, S. & Hurme, H. (2009). *Tutkimushaastattelu, teemahaastattelun teoria ja käytäntö*. Gaudeamus

Hirsjärvi, S., Remes, P. & Sajavaara, P. (2007). *Tutki ja kirjoita*, Kustannusosakeyhtiö Tammi. Helsinki.

Hotti, M. (28. marraskuuta 2012). *Pikaperehdytys Big Dataan: Mikä on Apache Hadoop, entä Hive?* Tivi. Haettu 20.6.2017 osoitteesta <http://www.tivi.fi/Arkisto/2012-11-28/Pikaperehdytys-Big-Dataan-Mik%C3%A4-on-Apache-Hadoop-ent%C3%A4-Hive-3196648.html>

- Hovi, A., Hervonen, H. & Koistinen, H. (2009). *Tietovarastot ja Business Intelligence*. Porvoo : WS Bookwell.
- Hurwitz, J., Nugent, A., Halper, F. & Kaufman, M. (2013). *Big Data for dummies*. John Wiley & Sons, Inc.
- Jacobs, A. (2009). The pathologies of big data. *Communications of the ACM*, 52(8), 36-44.
- Järviö J., Piispa T., Parantainen T. & Åström T., (2007). *Kunnossapito*. Helsinki: KPMedia
- Järviö, J & Lehtiö, T. (2012). *Kunnossapito - tuotanto-omaisuuden hoitaminen*, Kunnossapitoyhdistys Promaint. KP-Media Oy, Helsinki
- Kaukonen, M. (2010). *Tietämuspohjaisen järjestelmän käyttöönotto Inkeröisen tehtaalla*. (Opinnäytetyö). Saimaan Ammattikorkeakoulu.
- Kettunen, J. & Simons, M., (2001). *Toiminnanohjausjärjestelmän käyttöönotto pk-yrityksessä - Teknologia- ja tietotekniikan ajattelusta kohti tiedon ja osaamisen hallintaa*. Tummavuoren Kirjapaino Oy, Vantaa.
- Kimball, R. & Caserta, J. (2004). *The Data Warehouse ETL Toolkit - Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. Indianapolis, Indiana : Wiley Publishing, Inc. Haettu osoitteesta <http://users.itk.ppke.hu/~szoer/DW/Kimball%20&%20Caserta%20-The%20Data%20Warehouse%20ETL%20Toolkit%20%5BWiley%202004%5D.pdf>
- Kimball, R. & Ross, M. (2013). *The Data Warehouse Kit - The Definitive Guide to Dimensional Modeling*. Indianapolis, Indiana : John Wiley & Sons, Inc. Haettu osoitteesta <http://www.essai.rnu.tn/Ebook/Informatique/The%20Data%20Warehouse%20Toolkit,%203rd%20Edition.pdf>
- Krishnan, K. (2013). *Data Warehousing in the Age of Big Data*, Morgan Kaufmann.
- Krushe-Lehtonen, U. (2014). *Mistä sopivia tietoasiantuntijoita?* Tietoasiantuntija 1/2014
- Kurkela, R. (2014). *Tilastollinen tiedonkeruu*. Tilastokeskus ja Helsingin ammattikorkeakoulu Stadia. Haettu 02.01.2017 osoitteesta: <http://tilastokeskus.fi/virsta/tkeruu/>
- Kyngäs, H. & Vanhanen, L. (1999). Sisällön analyysi. *Hoitotiede* 11(1), 2-12

Laihonen, H. Hannula, M. Helander, N. Ilvonen, I. Jussila, J. Kukko, M. Kärkkäinen, H. Lönnqvist, A. Myllärniemi, J. Pekkola, S. Virtanen, P. Vuori, V. & Yliniemi, T. (2013). *Tietojohdaminen*. Tampereen teknillinen yliopisto. Juvenes Print. Tampere Tiedonhallinnan ja logistiikan laitos.

Leiviskä, K., (2009). *Papermaking Science and Technology. Book 14: Process and Maintenance Management*, Gummerus Oy, Jyväskylä

Lehtoviita, J., Mäki, P. & Tenhola, T. (2016). *Metsäbiotalouden arvoketjut - loppuraportti* (Julkaisusarjan osa 9). Tapio Oy

Lublinsky, B, Smith, K, Yakubovich, A. 2013. *Professional Hadoop Solutions*. Haettu osoitteesta

[http://proquest.safaribooksonline.com/book/databases/hadoop/9781118824184/chapter-1-big-data-and-the-hadoop-ecosystem/chapter01\\_html?](http://proquest.safaribooksonline.com/book/databases/hadoop/9781118824184/chapter-1-big-data-and-the-hadoop-ecosystem/chapter01_html?)

Lähdesmäki, T., Hurme, P., Koskimaa, R., Mikkola, L. & Himberg, T. *Metelmäpolkuja humanisteille*. Jyväskylän yliopisto, humanistinen tiedekunta. .

Manyika, J., Chui, M., Brown, B, Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A.H. (2009). *Big Data: The Next frontier for innovation, competition and productivity*. McKinsey Global Institute. Haettu ta <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>

Marshall, B., McDonald, D., Chen, H. & Chung, W. (2004). EBizPort: Collecting and Analyzing Business Intelligence Information. *Journal of the American Society for Information Science and Technology*, Vol. 55(10), 873-891.

Martinez-Lopez, F.J & Casillas, J. (2013). Artificial intelligence-based systems applied in industrial marketing: An historical overview, current and future insights. *Industrial Marketing Management* Vol. 42(4), 489-495.

Mikkelä, H. (2012). Big Data mullistaa tietomaailmaa? *Tietoasiantuntija* 4, 11-12

Minelli, M., Chambers, M. & Dhiraj, A. (2013). *Big Data Big Analytics: emerging business intelligence and analytic trends for today's businesses*. New Jersey: John Wiley & Sons.

MongoDB Inc. (2016). *What Is Big Data?* Haettu osoitteesta <https://www.mongodb.com/big-data-explained>

Morabito, V. (2015). *Big Data and Analytics. Strategic and Organizational Impact*. Haettu osoitteesta. <https://link.springer.com/book/10.1007%2F978-3-319-10665-6>



Oh, W. & Pinsonneault, A. (2007). On the Assessment of the Strategic Value of Information Technologies: Conceptual and analytical approaches. *MIS Quarterly*, Vol. 31(2), 239-265.

Negash, S. & Gray, P. (2003). Business Intelligence. *Ninth Americas Conference on Information Systems*, 3190-3199.

Olston, C., Reed, B., Srivastava, U., Kumar, R. & Tomkins, A. (2008) Pig Latin: A Not-So-Foreign Language for Data Processing. *Proc. SIGMOD '08: 2008 ACM SIGMOD International Conference on Management of Data*. Vancouver, BC, Canada, 1099-1110. Haettu osoitteesta <http://infolab.stanford.edu/~olston/publications/sigmod08.pdf>

Patil, M & Thia, F. (2013). *Pentaho for Big Data Analytics*. Haettu osoitteesta [Http://proquest.safaribooksonline.com/book/programming/java/9781783282159/3dot-churning-bigdata-with-pentaho/ch03s02\\_html?](http://proquest.safaribooksonline.com/book/programming/java/9781783282159/3dot-churning-bigdata-with-pentaho/ch03s02_html?)

Pekkola, S. & Nieminen, N. (2015). Comparison of Functional, Structural, and Dynamic Business-IT Alignment Models: A Case Study. *Proceedings of the 21st Americas Conference on Information Systems*.

Peppard, J., Ward, J. (1999). Mind the Gap: Diagnosing the Relationship Between IT Organization and the Rest of the Business", *Journal of Strategic Information Systems*, Vol. 8, 29-60

Pirttimäki, V. (2007). *Business Intelligence as a Managerial Tool in Large Finnish Companies*. (Väitöskirja). Julkaisu 646. Tampereen teknillinen yliopisto, Haettu osoitteesta <https://tutcris.tut.fi/portal/files/1661109/pirttimaki.pdf>

Prajapati, V. (2013). *Big Data Analytics with R and Hadoop*. Birghingam. Haettu osoitteesta [http://proquest.safaribooksonline.com/book/databases/hadoop/9781782163282/preface/pr07s04\\_html?](http://proquest.safaribooksonline.com/book/databases/hadoop/9781782163282/preface/pr07s04_html?)

Remes, M. (2013). *Big datasta kaikki hyöty irti*. KPMG-Finland. Haettu osoitteesta [Http://issuu.com/kpmgfinland/docs/asiakaslehti-view-1-2013/8?e=6625711/1463933](http://issuu.com/kpmgfinland/docs/asiakaslehti-view-1-2013/8?e=6625711/1463933)

Reynolds, P. & Yetton, P. (2013). Aligning Business and IT Strategies in Multi-Business Organizations. *International Conference for Information Systems*

Rouse, R. 2012. *Definition Database*. SQLServer. Haettu osoitteesta <http://searchsqlserver.techtarget.com/definition/database>

Russom, P. (2011). Big Data Analytics. *TDWI Best Practices Report, 4th quarter 2011*.

Salo, I. (2013). *Big Data, tiedon vallankumous*. Jyväskylä: Docento.

Salonen, J. & Pirttimäki, V. (2006). Outsourcing a business intelligence function. *Frontiers of e-business research 2005*, 661-675.

SAS. (2013). *Big Data Meets Big Data Analytics*, SAS. Haettu osoitteesta [http://www.sas.com/resources/whitepaper/wp\\_46345.pdf](http://www.sas.com/resources/whitepaper/wp_46345.pdf)

Saunders, M., Lewis, P. & Thornhill, A. (2009). *Research methods for business students*. Pearson Education. United Kingdom.

Schmidt, K & Phillips, C. (2013). *Programming Elastic MapReduce*. Canada: O'Reilly Media, Inc.

Shvachko, K., Kuang, H., Radia, S. & Chansler, R. (2010). The hadoop distributed file system. *Mass Storage Systems and Technologies (MSST), 2010 IEEE 26th Symposium*, 1-10.

Siiramaa, P. (2012). *HADOOP Ohjelmistokehitys Big Datan käsittelyyn*. (Pro gradu -tutkielma). Itä-Suomen Yliopisto.

Silvius, G. (2009). Business and IT Alignment: What We Know and What We Don't Know. *International Conference on Information Management and Engineering*, 448-563.

Stähle, P. & Grönroos, M. (1999). *Knowledge management – tietopääoma yrityksen kilpailutekijänä*. Werner Söderström Oyj, Helsinki, Porvoo, Juva

Tallon, P.P., Kraemer, K.L. & Gurbaxani, V. (2000). Executives' Perceptions of the Business Value of Information Technology: A Process-Oriented Approach, *Journal of Management Information Systems*, Vol. 16 (4), 145-173.

Taylor, R. C. (2010). An overview of the Hadoop/MapReduce/Hbase framework and its current applications in bioinformatics. *BMC bioinformatics*, 11(12), 1.

The Apache Software Foundation. (2012). *Flume User Guide*. Apache Foundation. Haettu osoitteesta <https://flume.apache.org/FlumeUserGuide.html>

The Apache Software Foundation. (2014). *Apache Storm – Tutorial*. Apache Foundation. Haettu osoitteesta <http://storm.apache.org/releases/current/Tutorial.html>

Thierauf, R.J. (2001). *Effective Business Intelligence Systems*. Quorum books. Westport. USA

Thomas, Jr. & James H. (2001). Business Intelligence - Why? *eAI Journal*, July 2001, 47-49.

Thusoo, A., Sarma, J., Jain, N., Shao, Z., Chakka, P., Anthony, S., Liu, H., Wyckoff, P. & Murthy, R. (2009) Hive - A Warehousing Solution Over a Map-Reduce Framework. *Proceedings of the VLDB Endowment*, 1626-1629.

Tuomi, J. & Sarajärvi, A. (2004). *Laadullinen tutkimus ja sisällönanalyysi*. Jyväskylä: Gummerus Kirjapaino Oy

Umble, E., Haft, R. & Umble, M. (2003). Enterprise resource planning: Implementation procedures and critical success factors. *Research and Practical Issues of Enterprise Information Systems. IFIP International Federation for Information Processing, vol 205. Springer, Boston, MA, 319-324*

Uusi-Rauva, E. (1994). *Ohjauksen tunnusluvut ja suoritusten mittaus*. Tampere: Tampereen teknillinen korkeakoulu, tuotantotalouden osasto, teollisuustalouden laitos.

Vaarnas, M. & Virtanen, J. (2001) *Markkinatieto yrityksen kansainvälistymisessä, toinen painos*. Tammer-Paino Oy, Tampere.

Vilpola, I. & Kouri, I., (2006). Toiminnanohjausjärjestelmän hankinta C-CEI-menetelmän avulla. *Teknologiategollisuuden julkaisu nro 2/2006*. Teknologiainfo Teknova Oy, Helsinki

Wang, C., Rayan, I. & Schwan, K. (2012). Faster, Larger, Easier: Reining Real-Time Big Data Processing in Cloud. *Proc. Middleware'12: Posters and Demo Track*. Montreal, QC, Canada,

White, T. (2012). *Hadoop: The Definitive Guide, 3rd Edition*. CA: O' Reilly Media, Inc.

Wilson, P & O'Conor, D. (2000). *Managing for knowledge*. Scitech Educational Haettu osoitteesta  
[https://books.google.fi/books/about/Managing\\_for\\_Knowledge.html?id=SU3-SgAACAAJ&redir\\_esc=y](https://books.google.fi/books/about/Managing_for_Knowledge.html?id=SU3-SgAACAAJ&redir_esc=y)

Yang, H. C., Dasdan, A., Hsiao, R. L. & Parker, D. S. (2007). Map-Reduce-Merge: Simplified Relational Data Processing on Large Clusters, *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*. ACM, 1029-1040.

Yin, S. & Kaynak, O. (2015). Big Data for Modern Industry: Challenges and Trends. *Proceedings of the IEEE*, 103(2), 143 - 146.

Zack, M.H. (1999). Managing codified knowledge. *Sloan Management Review*, Vol. 40(4), 45-58.

Zaharia M, Chowdhury M, Das T, Dave A, Ma J, McCauley M, Franklin M, Shenker, S. & Stoica, I. (2012). Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing. *Proc. NSDI'12: 9th USENIX conference on Networked Systems Design and Implementation*, 15-29. San Jose, CA, USA.

Zburivsky, D. (2013). *Hadoop Cluster Deployment*. Birmingham UK: Pack Publishing. Haettu osoitteesta [https://s3-eu-west-1.amazonaws.com/ebooks.sequenceiq.com/Hadoop%20Cluster%20Deployment%20\[eBook\].pdf](https://s3-eu-west-1.amazonaws.com/ebooks.sequenceiq.com/Hadoop%20Cluster%20Deployment%20[eBook].pdf)

Zikopoulos, P., Eaton, C., Deroos, D., Deutsch, T. & Lapis, G. (2012). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. USA: McGraw-Hill.

## 11 LIITE 1 HAASTATTELUTEEMAT

### **Käytettävyys**

Pyritään selvittämään, kuinka mielellään käyttäjät tietojärjestelmiä käyttävät ja kokevatko he niitä ylipäätään hyödyllisiksi

- Onko tietojärjestelmiä mielestäsi vaikea käyttää?
- Tarjoavatko järjestelmät mielestäsi relevanttia tietoa liittyen liiketoimintaan?

### **Prosessit ja niiden kehitys**

Pyritään selvittämään, ovatko prosessit selkeitä ja millaisia kehityskohteita niissä on

- Onko organisaatiossanne selkeät roolit liittyen tietojärjestelmien käyttöön ja raportointiin?
- Ymmärrätkö selkeästi miksi tietoa kustakin asiasta kerätään ja raportoidaan?
- Miten kehittäisitte tietojärjestelmien käyttöä organisaatiossanne?

## **Vuorovaikutus**

Pyritään selvittämään kuinka sujuvasti tietojärjestelmät ja liiketoiminta kohtaavat organisaatiossa

- Kuinka IT ja liiketoiminta vuorovaikuttavat organisaatiossanne?