



**Konsta Salmela**

**DATAPOHJAINEN ARVONLUONTI ARVOKETJUSSA**

Kandidaatintutkielma  
Kauppätieteen kandidaatti  
Huhtikuu 2022

**SISÄLLYS**

<b>1</b>	<b>JOHDANTO .....</b>	<b>4</b>
<b>2</b>	<b>DATA .....</b>	<b>6</b>
<b>2.1</b>	<b>Kolmen V-kirjaimen malli .....</b>	<b>7</b>
<b>2.2</b>	<b>Big datan prosessointi .....</b>	<b>10</b>
2.2.1	Tiedonhallintaprosessi .....	11
2.2.2	Data-analytiikkaprosessi .....	11
2.2.3	Pilvipalvelut .....	12
<b>3</b>	<b>ALUSTATALOUS .....</b>	<b>14</b>
<b>3.1</b>	<b>Alustatalouden ominaispiirteet .....</b>	<b>14</b>
<b>3.2</b>	<b>Datatalous ja datapohjainen liiketoiminta.....</b>	<b>15</b>
<b>4</b>	<b>DATAPOHJAINEN ARVONLUONTI JA ARVOKETJU .....</b>	<b>17</b>
<b>4.1</b>	<b>Datan arvoketju.....</b>	<b>17</b>
4.1.1	Datan keräys.....	19
4.1.2	Tiedonluonti.....	20
4.1.3	Arvonluonti ja käyttö .....	22
<b>4.2</b>	<b>Arvonluonti rahoitus- ja vakuutuslalla .....</b>	<b>25</b>
<b>4.3</b>	<b>Datan kierrätys ja monetisointi .....</b>	<b>26</b>
<b>4.4</b>	<b>Datapohjaisen arvonluonnin haasteet .....</b>	<b>27</b>
<b>5</b>	<b>YHTEENVETO .....</b>	<b>29</b>

**KUVIOT**

<b>Kuvio 1. Kolmen V-kirjaimen malli (mukaillen Russom, 2011). .....</b>	<b>8</b>
<b>Kuvio 2. Big datan prosessit (mukaillen Gandomi &amp; Haider, 2015). .....</b>	<b>10</b>
<b>Kuvio 3. Datan arvoketju (mukaillen Faroukhi ym. 2020.).....</b>	<b>19</b>

## 1 JOHDANTO

Datan kerääminen herättää yhteiskunnallista keskustelua (Chen, 2021). Keskustelua käydään siitä, loukkaako datan kerääminen kuluttajien yksityisyydensuojaa ja onko sen hyödyntäminen liiketoiminnassa kustannustehokasta. Datapohjaisten yhtiöiden kuten Alphabet:in (Google) ja Meta:n eli Facebookin emoyhtiön markkina-arvot kuitenkin yltyvät satoihin miljardeihin dollareihin. Tästä voidaan vetää nopeasti johtopäätös. Data luo arvoa.

Digitalisaation myötä kasvanut data- ja alustatalous luo monia mahdollisuuksia datan hyödynnettävyyden luoman lisäarvon monetisoinnille (Paavola, Seppänen & Eloranta, 2021, s. 94). Dataa kerätään päivittäin valtava määrä, mutta kerätty data luo arvoa vain, jos sitä hyödynnetään oikealla tavalla (Gandomi & Haider, 2015). Raakadatalle on haastavaa määrittellä arvoa perinteisten kaavojen mukaisesti. Datan arvonluontiprosessia voidaan tarkastella arvoketjun kautta. Arvoketjun avulla havainnollistetaan sitä, miten kerättyä dataa voidaan hyödyntää, miten datan hyödyntäminen luo arvoa, sekä miten arvo voidaan muuttaa rahaksi.

Kansainvälisen kilpailun kiihtyessä yritykset pyrkivät voittamaan kilpailijansa keinolla millä hyvänsä. Datan hyödyntäminen liiketoiminnassa voi nousta merkittäväksi kilpailueduksi ja auttaa liiketoiminnan prosessien tehostamisessa, kulujen pienentämisessä, sekä myynnin kasvattamisessa. Dataa voidaan hyödyntää myös julkisella sektorilla asiakaskokemuksen parantamiseen ja arvonluonti voi näkyä esimerkiksi lyhyempinä jonotusaikoina palveluihin. Jotta dataa voidaan tehokkaasti hyödyntää, on tärkeää muodostaa ymmärrys datan ominaisuuksista sekä arvonluontiprosessista. Aiheen tutkiminen on tärkeää myös tieteellisestä näkökulmasta. Tällä hetkellä aihetta käsittelevien tieteellisten tutkimusten määrä on vähäinen. Syyt tutkimusten vähäisyydelle lienevät seurausta ilmiön tuoreudesta, eli tutkimuksia ei ole ehditty tehdä vielä tarpeeksi. Lisäksi syynä voi olla teknologian nopea kehittyminen, minkä myötä ilmiö voi muuttua nopeallakin aikataululla merkittävästi. Tieteellisen tutkimuksen myötä ilmiötä voitaisiin ymmärtää syvemmin johtaen uusiin oivalluksiin. Tutkimusten löydöksiä voitaisiin hyödyntää myös datapohjaisen liiketoiminnan kehityksessä.

Tämä tutkielma esittelee datan arvonluontiprosessin eri vaiheet pelkistetyn arvoketjun kautta. Tutkielmassa pyritään luomaan yleiskuva siitä, miten datan arvonluonti tapahtuu sen elinkaaren aikana. Tutkielma pyrkii vastaamaan tutkimuskysymykseen:

*Miten datapohjainen arvonluonti tapahtuu arvoketjun eri vaiheiden kautta?*

Tutkielma toteutetaan kirjalliskatsauksena ja tavoitteena on muodostaa käsitys siitä, mitä datalla ja datataloudella tarkoitetaan, miten dataa voidaan hyödyntää eri käyttötarkoituksissa, sekä miten hyödyntämiseen liittyvä arvonluonti tapahtuu vaiheittain. Tutkielma ei pyri määrittelemään datalle arvoa rahallisena määreenä, vaan muodostamaan yleiskuvan arvonluontiin liittyvästä prosessista, sekä sen osista. Tutkielma toteutetaan tarkastelemalla aiheeseen liittyvää tutkimuskirjallisuutta ja tieteellisiä artikkeleita, sekä vertailemalla niiden tuloksia.

Tutkielman toinen pääluku keskittyy datan ominaisuuksien ja prosessoinnin tarkasteluun. Luvussa määritellään big dataa kolmen V-kirjaimen mallin kautta ja tuodaan esille prosessointiin liittyviä toimintoja. Kolmas pääluku luo kuvan datapohjaisen liiketoiminnan perustasta, eli data- ja alustataloudesta. Luvussa tarkastellaan alustatalouden erityispiirteitä ja esitellään datatalouteen perustuvia dataa hyödyntäviä liiketoimintamalleja. Neljäs pääluku luo teoreettisen viitekehyksen arvoketjumallille, minkä lisäksi datapohjaista arvonluontia tarkastellaan vaiheittain osiin pilkotun datan arvoketjun kautta. Datapohjaisen arvonluonnin sovellusta kuvaillaan käytännön kautta rahoitus- ja vakuutuslalla, minkä lisäksi nostetaan esille datan kierrätys- ja monetisointimahdollisuudet. Luvussa luodaan katsaus myös arvonluontiin liittyviin haasteisiin. Viimeiseksi viidennessä luvussa tutkielman havainnot kootaan yhteen johtopäätösten muodossa ja pohditaan aiheita jatkotutkimuksille.

## 2 DATA

Dataa voidaan pitää yhtenä tärkeimmistä yritysten tarvitsemista raaka-aineista, mikäli ne haluavat pysyä kilpailukykyisinä tänä päivänä, sekä tulevaisuudessa. Trederin (2019, s. 2) mukaan datan hallinnointia ei tule pitää vain uutena organisaation velvollisuutena, sillä lähes kaikki, mitä liike-elämässä tehdään, pohjautuu lopulta dataan. Treder määrittelee datan olevan yhdistelmä informaatiota ja tietoa.

Data raaka-aineena ei kuitenkaan ominaisuuksiensa vuoksi ole yhtä yksinkertainen ymmärtää, kuten dataan usein verrattu öljy. Öljyn tiedetään olevan nestemäistä ja sitä voi säilöä esimerkiksi öljytynnyreissä. Dataa sen sijaan ei voi koskettaa, sitä ei jalosteta suurissa tehtaissa, eikä kuljeteta laivoilla ympäri maailmaa. Viitasen, Koivistoinen, Loikkasen ja Paajasen (2013, s.15) mielestä öljyä parempi vertauskuva datalle onkin vesi. He kuvailevat dataa raaka-aineena, mikä veden kaltaisesti kiertää digitaalisessa ekosysteemissä ja on uudelleenkäytettävissä. Heidän vertauskuvansa mukaan data voi kastelun lailla luoda kasvua liiketoiminnalle tai vaihtoehtoisesti jäädä käyttämättömänä säilöön ummehtumaan.

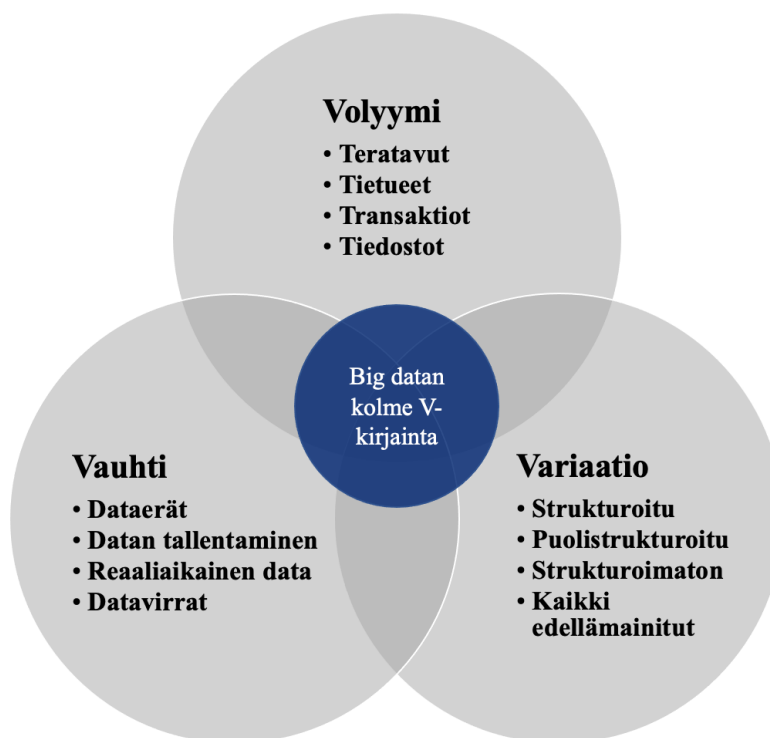
Data eroaa öljystä myös syntymisen myötä. Öljy on uusiutumaton luonnonvara, mikä tarkoittaa sitä, että sen määrä maapallolla ei kasva. Datan määrä sen sijaan on kasvanut merkittävästi viime vuosikymmenien aikana. Datan kerääminen on automatisoitunut teknologian kehittymisen myötä ja samalla datan säilyttämisen kustannukset ovat laskeneet. Tämä on johtanut siihen, että dataa voidaan kerätä ja varastoida äärimmäisen suuria määriä, eikä yksityishenkilöillä enää ole mahdollisuutta kontrolloida sitä, mitä tietoa he haluavat itsestään kerättävän. (Atzori, Iera, & Morabito, 2010.)

Tässä tutkielmassa huomio rajataan aiheen laajuuden vuoksi big dataan ja sen ominaisuuksiin. Toisena syynä on se, että aiheeseen liittyvässä kirjallisuudessa arvonluonti johdetaan usein nimenomaan big datasta. Big datalla tarkoitetaan käytännössä samaa asiaa kuin pelkällä ”data” käsitteellä, nimi viittaa vain datan suureen määrään. Tämän suurien datamääriä kuvailevan termin käyttö yleistyi vuonna 2011 (Gandomi & Haider, 2015). Samasta vuodesta lähtien aihe on räjähdysmäisesti kerännyt yleisön ja median huomiota (Alanko & Salo, 2013). Keskustelusta huolimatta

termille ei kuitenkaan ole löydetty yhteistä määritelmää. Tämä on seurausta siitä, että aiheeseen liittyvää kirjallisuutta on usealta eri tieteenalalta, minkä johdosta määritelmät vaihtelevat ja ovat jopa ristiriitaisia (Ward & Barker, 2013). McKinseyn tuottaman raportin (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh & Hung Byers, 2011) mukaan big datalla tarkoitetaan perinteisiä datasettejä suurempia tiedostoja, joita ei kokonsa vuoksi voida käsitellä perinteisten datatyökalujen avulla. Wardin ja Bakerin havainnot tukevat tätä väitettä, sillä he myös nostavat esille prosessointiin liittyviä haasteita, mitkä johtuvat termin nimen mukaisesti datan suuresta määrästä.

## 2.1 Kolmen V-kirjaimen malli

Termin eriävistä määritelmistä huolimatta big dataan liitetään yleisesti kolme eri ominaisuutta, minkä pohjalta big dataa määritellään sekä kuvaillaan. Nämä Laneyn (2001) esittelemät kolme V-kirjainta ovat ”*Volume*”, ”*Variety*”, sekä ”*Velocity*”, eli suomeksi: volyymi, variaatio ja vauhti (Alanko & Salo, 2013). Eri tahot ovat tehneet täydennyksiä kolmen V-kirjaimen rinnalle nimeämällä uusia big datan ominaisuuksia. Esimerkiksi IBM esitteli kolmen V:n rinnalle neljännen, ”*Veracity*”, eli datan oikeellisuuden, sekä Oracle jatkoi lisäämällä vielä viidennen V:n ”*Value*”, eli datan arvon (Ward & Barker, 2013). Lisäyksistä huolimatta käytetyin malli big datan määrittämiseen on yhä kolmen V-kirjaimen malli (Alanko & Salo, 2013). Tässä tutkielmassa big data määritellään kolmen V-kirjaimen mallin kautta. Kyseistä mallia käytetään sen olennaisuuden ja yksinkertaisuuden vuoksi. Big datan kolme attribuuttia esittelevä kolmen V-kirjaimen malli havainnollistetaan kuviossa 1.



**Kuvio 1. Kolmen V-kirjaimen malli (mukaillen Russom, 2011).**

Ensimmäinen ja tärkein big dataa määrittelevä attribuutti on Russomin (2011, s. 6) mukaan volyymi. Nimensä mukaisesti big data tarkoittaa määrällisesti suuria datasettejä. Big dataksi määriteltävien datasettien koot ilmoitetaan näkemyksestä riippuen yleensä joko teratavuina tai petatavuina. Raja-arvoa big dataksi määriteltävälle volyymille on kuitenkin haastavaa rajata, sillä tämä vaihtelee suhteellisesti ajan, sekä eri datatyypin myötä. Datan volyymi, mikä tänä päivänä määriteltäisiin big dataksi, ei ehkä muutaman vuoden päästä ole enää big dataa. Tämä johtuu siitä, että teknologia kehittyy jatkuvasti mahdollistaen suurempien datasettien varastoinnin (Gandomi & Haider, 2015; Russom, 2011). Kasvavaan datan volyymiin vaikuttaa myös Laneyn (2001) huomio siitä, että organisaatiot näkevät datan aineettomana hyödykkeenä, eivätkä näin ollen ole halukkaita hävittämään vanhaa dataa uuden datan edestä. Organisaatiot hankkivat mieluummin lisää säilytystilaa ja näin dataa kertyy jatkuvasti lisää.

Volyymin lisäksi big datan määrittelemiseen käytetään termiä variaatio. Variaatio liittyy myös läheisesti volyymiin. Gandomin ja Haiderin (2015) esimerkin mukaan kaksi samansuuruisia, mutta eri muodossa olevaa datasettiä voivat vaatia eri



tiedonhallinnan teknologioita. Videomuodossa olevaa dataa on erilaista analysoida, kuin esimerkiksi numeromuodossa olevaa dataa. Heidän mukaansa variaatio viittaa siis datasetin rakenteelliseen heterogeenisyyteen. Tällä hetkellä kerättävien datasettien rakenne vaihtelee paljon, sillä dataa tulee yhä useammasta eri lähteestä ja se on muodoltaan vaihtelevampaa (Alanko & Salo, 2013, s. 4; Russom, 2011, s. 7). Nykyään ainoastaan pieni osa datasta koostuu strukturoidusta datasta, eli taulukoidusta tai tietokannoissa olevista helposti haettavista tiedostoista. Loppuosa koostuu semi-strukturoidusta sekä strukturoimattomasta datasta. Nämä sisältävät esimerkiksi videotai kuvatiedostoja, sekä erilaisia teksti- ja tietokonetiedostoja (Gandomi & Haider, 2015; Russom, 2011, s. 7). Haasteeksi nousee standardoituihin tietokantoihin perustuvat tietojärjestelmät, jotka eivät kykene hyödyntämään strukturoimatonta tai semi-strukturoitua dataa (Alanko & Salo, 2013, s. 4).

Kolmas määre, eli vauhti, viittaa datan tuottamisen jatkuvasti kiihtyvään tahtiin (Alanko & Salo, 2013, s. 4). Jatkuvasti kasvavat datasetit ovat myös seurausta teknologian kehityksestä, minkä Laney mainitsi jo vuonna 2001 teoksessaan ”*3D Data Management: Controlling Data Volume, Velocity, and Variety*”. Laneyn mukaan internetissä suoritettavien transaktioiden, eli ostotapahtumien laskevien kustannusten kautta dataa pystytään keräämään yhä enemmän jokaista transaktiota kohden. Kehitys on jatkunut samansuuntaisesti, sillä nykyään yhä useammat sosiaaliset kanssakäymiset tai ostokset tapahtuvat digitaalisten laitteiden kuten älypuhelimien kautta. Tästä kehityksestä johtuen yhä enemmän dataa on entistä helpommin saatavilla. (Gandomi & Haider, 2015; Prüfer & Schottmüller, 2021.)

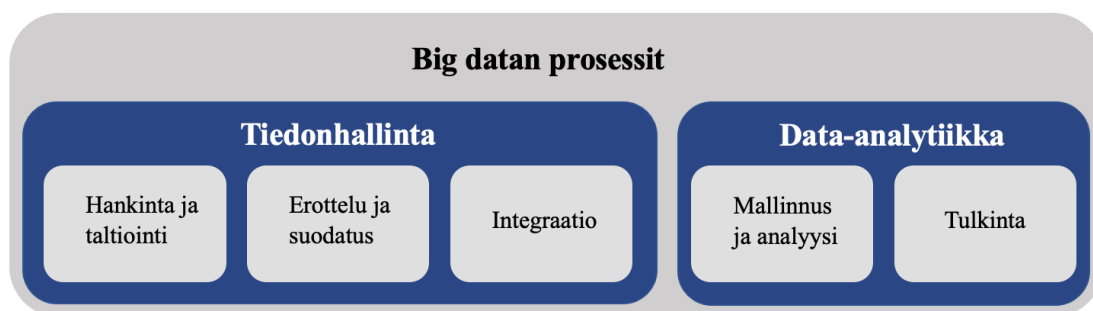
Alanko ja Salo (2013, s. 4) huomauttavat, että datan nopea tuottamistahti johtaa säilytysongelmien lisäksi myös datan analysointi- sekä hyödyntämisongelmiin. Gandomi ja Haider (2015) nostavat esille saman havainnon. Heidän mukaansa älylaitteiden suosion myötä dataa tuotetaan niin nopealla tahdilla, etteivät perinteiset datanhallintaohjelmistot pysy perässä. Tämä synnyttää kasvavan tarpeen reaaliajassa toimiville data-analytiikkaratkaisuille. Big datan analysointia varten kehitetyt ohjelmistot tarjoavat tähän ongelmaan ratkaisuja. Yksi suosituimmista ohjelmistoista on Apache Software Foundationin kehittämä Hadoop, minkä yhtenä etuna on kyky käsitellä useita eri datatyyppejä (Russom, 2011, s. 7).

## 2.2 Big datan prosessointi

Luvun 2 johdannossa ja luvussa 2.1 määriteltiin big data sen ominaisuuksien perusteella, sekä luotiin yleiskuva siitä, mitä big data on. Data itsessään ei kuitenkaan ole minkään arvoista, kuten Alanko ja Salo (2013, s. 4) huomauttavat. Heidän mukaansa datasta tulee arvokasta organisoinnin sekä analysoinnin myötä. Gandomi ja Haider (2015) ovat tehneet saman havainnon. Heidän mukaansa big datan arvo syntyy vasta sen prosessoinnin myötä, kun data-analytiikkaa hyödynnetään päätöksenteon yhteydessä.

Datasta voidaan saada uutta markkinatietoa, mikä mahdollistaa nopean reagoinnin uusiin liiketoimintamahdollisuuksiin tai muutoksiin. Kilpailuetua syntyy, mikäli tiedon avulla asioihin voidaan reagoida kilpailijoita nopeammin tai paremmin. Tiedon uuttamiseksi big dataa tulee prosessoida. Big datan prosessointi eroaa kuitenkin perinteisistä menetelmistä. Ero johtuu perinteisten prosessointiteknologioiden tehon riittämättömyydestä suurten datamassojen, eli big datan analysointiin. (Puig, Chen, Chen, Du, Li, Lu, Zhao & Zhou, 2013.)

Gandomi ja Haider (2015) jakavat big datan prosessoinnin kahteen alaprosessiin: tiedonhallintaan ja data-analytiikkaan. Tiedonhallinnan ja data-analytiikan alaprosessit koostuvat viidestä vaiheesta (Labrinidis & Jagadish, 2012). Big datan prosessointi vaiheineen esitellään kuviossa 2.



Kuvio 2. Big datan prosessit (mukaillen Gandomi & Haider, 2015).

### 2.2.1 Tiedonhallintaprosessi

Tiedonhallinnan prosessit tukevat big datan keräämistä ja säilömistä. Lisäksi tiedonhallintaa tukevien teknologioiden avulla big dataa valmistellaan analyysia varten siten, että tarvittava data on helpommin haettavissa tietokannoista käyttöä varten. Tiedonhallintaprosessiin kuuluvia toimintoja ovat datan hankinta, erottelu, suodatus ja integrointi. Tiedonhallintaprosessit ovat yhtäläisen tärkeitä data-analyysin prosessien rinnalla. Ilman tiedonhallintaa data-analyysin toteutus olisi huomattavasti haastavampaa. (Gandomi & Haider, 2015; Labrinidis & Jagadish, 2012.)

Yksi tärkeimmistä tiedonhallintaan liittyvistä toiminnoista on integraatio. Integraatiolla tarkoitetaan tiedon siirtämistä ja muuntamista tietojärjestelmien välillä. Dataa tuotetaan yhä useammasta eri lähteestä, mikä johtaa ongelmiin sen analysoinnissa. Integraation avulla useasta eri lähteestä tuleva data saadaan yhdistettyä ja suodatettua yhtenäiseksi näkemykseksi. Useimmissa tapauksissa datamäärät ovat niin suuria, että dataa on mahdollista pienentää suodattamalla siitä tärkein tieto ja koostamalla se yhteen. Kooste on yleensä riittävä tukemaan datapohjaisen päätöksenteon tarpeita. Big datan suodattamisessa kyse on siitä, että suuresta tietomäärästä pyritään ottamaan ylimääräinen turha tieto pois ja se muutetaan järjestelmälliseen yksinkertaistettuun muotoon. Suodattaminen voi tapahtua esimerkiksi koneoppimisen menetelmien avulla. (Puig ym., 2013.)

### 2.2.2 Data-analytiikkaprosessi

Toinen big datan alaprosesseista on data-analytiikka. Data-analytiikalla tarkoitetaan prosessia, missä kerättyä ja säilöttyä tietoa analysoidaan, minkä seurauksena siitä saadaan informaatiota eri käyttötarkoituksiin. Data-analyysi tapahtuu erilaisten analyysimenetelmien kautta. (Gandomi & Haider, 2015.) Data-analytiikan toimintoja ovat mallinnus, analysointi sekä tulkinta (Labrinidis & Jagadish, 2012).

Puig ym. (2013) huomauttavat, että big datan analysoinnin tulee olla perusteellista, jotta siitä olisi hyötyä päätöksenteon yhteydessä. Pintapuolinen raportointi datasta ei ole riittävää. Heidän mukaansa datasta tulee selvittää mitä nyt tapahtuu ja tämän lisäksi datan avulla täytyy kyetä ennakoimaan myös tulevaisuuden tapahtumia. Gandomi ja

Haider (2015) vahvistavat tämän havainnon. Heidän mukaansa analysointitekniikoita on useita erilaisia, mutta yksi niistä, eli ennakoiva analytiikka nousee ylitse muiden. Ennakoiva analytiikka analysoi historiallista ja nykyistä rakenteista, eli strukturoitua aineistoa, pyrkien löytämään siitä kuvioita ja havaintoja. Havaintojen avulla pyritään ennakoimaan tulevaisuuden tapahtumia. Ennakoivan analytiikan menetelmät pohjautuvat pitkälti tilastollisiin menetelmiin. Analyysissa voidaan katsoa esimerkiksi myynnin liukuvaa keskiarvoa, mistä voidaan hahmotella ensi vuoden myyntiä. Analyysin perusteella saadut havainnot tulkitaan ja tulkinnan avulla voidaan muodostaa erilaisia tulevaisuuden skenaarioita, joiden pohjalta tehdään päätöksiä liiketoimintaan liittyen. Prosessien seurauksena big datan analyysi voi luoda yritykselle arvoa ja luoda kilpailuetua.

### 2.2.3 Pilvipalvelut

Russom (2011, s. 16–17) esittää teoksessaan tärkeän big datan prosesseihin liittyvän kysymyksen: Missä big dataa kuuluisi hallita ja soveltaa? Kysymys on relevantti tänäkin päivänä, vaikkakin Russomin havainnot ovat jokseenkin aikasidonnaisia teoksen julkaisuvuoteen 2011. Russomin mukaan suosituin alusta big datan edistyneelle data-analyysille ja tiedonhallinnalle on yrityksen tietovarastot (Enterprise Data Warehouse, EDW). Yrityksen tietovarastojen heikkoudeksi nousee kuitenkin korkeat kustannukset, tehottomuus ja skaalautuvuuteen liittyvät ongelmat, kun datamäärät kasvavat eksponentiaalisesti.

Russom (2011, s. 16–17) ennakoiki kyselytuloksien perusteella uusien pilvipalveluihin perustuvien analytiikka-alustojen suosion nousevan seuraavien vuosien aikana, eli vuodesta 2011 eteenpäin. Puig ym. (2013) tekivät vuonna 2013 samankaltaisen havainnon. He ehdottivat datan suodattamiseen ja pienentämiseen käytettäväksi rinnakkaislaskentaa, missä usea tietokone prosessoi dataa. Heidän mukaansa pilvipalvelut olivat tutkimushetkellä nousemassa suureen suosioon.

Pilvipalvelut tarjoavat internetin välityksellä pääsyn tehokkaisiin muunneltaviin laskennallisiin resursseihin pienin kustannuksin. Pilvipalvelut ovat yritysten käytettävissä aina tarvittaessa (Puig ym. 2013). Yritysten ei tarvitse rakentaa ja omistaa omia datakeskuksia, vaan ne maksavat pilvipalveluista käytön mukaan.

Alanko ja Salo tarjoavat myös big datan haasteisiin vastaukseksi pilvipalvelut. He nostavat pilvipalveluiden eduksi tarpeen mukaan joustavan rajattoman tallennustilan, sekä laskentakapasiteetin. Näiden lisäksi Alankon ja Salon mielestä pilvipalvelut toimivat erinomaisena alustana datan jakamis- ja myyntitarkoituksiin. Mikäli useat datavarannot ovat samassa pilvipalvelussa, niiden jakaminen ja yhdisteleminen käy nopeasti. Edellä mainittujen asioiden lisäksi he nostavat esille pilvipalveluiden tarjoamat analytiikkapalvelut. Palvelut mahdollistavat suurien datamäärien yhdistelyn ja tehokkaan analysoinnin, mikä on big datan prosessoinnille olennaista. (Alanko & Salo, 2013, s. 7–8).

### 3 ALUSTATALOUS

Datapohjaisen arvonluontiprosessin on mahdollistanut uudenlainen moniulotteinen toimintaympäristö. Arvonluontiprosessin ohella on tärkeä ymmärtää alustatalouden ominaisuuksia, sekä alustojen kehitystä seuranneen datapohjaisen liiketoiminnan piirteitä. Tässä luvussa käydään läpi yleisellä tasolla alusta- ja datataloutta, sekä datapohjaista liiketoimintaa. Huomio on datan arvonluontiprosessin kannalta olennaisissa seikoissa.

#### 3.1 Alustatalouden ominaispiirteet

Alustataloudella tarkoitetaan tietotekniisiin järjestelmiin liittyviä alustoja, missä toimijat luovat vuorovaikutuksellisella toiminnallaan arvoa. Toiminta tapahtuu yhteisten toimintaperiaatteiden mukaan ja alustatalouden arvonluontiin liittyy vahvasti datan hyödyntäminen, sekä käyttö uusien teknologioiden avulla (Viitanen ym., 2013, s. 17). Uudet tavat arvonluontiin ja datapohjaiseen liiketoimintaan ovat seurausta digitalisaatiosta. Alustatalous ja uudet teknologiset innovaatiot muokkaavat talouden rakenteita, liiketoimintamalleja, sekä vaikuttavat siihen, miten kuluttajat toimivat markkinoilla. Muutos tuo yrityksille mukanaan haasteita sekä mahdollisuuksia. Perustasolla kyse on kuitenkin samasta kuin ennenkin; yritykset pyrkivät luomaan toiminnallaan voittoa. (Paavola ym., 2021, s. 10–12.)

Alustatalouden yhtenä perustana voidaan pitää pilvipalveluiden kehitystä. Pilvipalveluiden myötä kehittyi alustatalouden tarvitsema infrastruktuuri, minkä päälle ovat rakentuneet eri alustapohjaisten moniulotteisten markkinoiden ekosysteemit. (Kenney & Zysman, 2016.) Paavolan ym. (2021, s. 186) määritelmän mukaan alustataloudessa on kyse moniulotteisesta arvonluontijärjestelmästä. Järjestelmä yhdistää palveluiden ja tuotteiden tuottajat sekä kuluttajat. He kuvailevat alustatalouden vaihdantaa erityisesti siitä syystä, että vaihdannan kohteet ja välineet ovat joskus vaikeasti tunnistettavissa. Viitanen ym. (2013, s.15–16) ovat tehneet saman havainnon alustatalouden uudeltaisesta vaihdannasta. Heidän mukaansa digitaalisuus on luonut mahdollisuuden käyttää valuuttojen ohella myös erilaisia vaihdannan välineitä. He nimeävät vaihdannan välineiksi tiedon, datan sekä palvelut, joita voidaan vaihtaa uusilla vaihdantapaikoilla. Viitanen ym. tekstissä mainitaan

muiden lähteiden lailla alustatalouden moniulotteisuus. Heidän tulkintansa mukaan kyse on usean eri osapuolen vuorovaikutukseen perustuvasta monisuuntaisesta markkinapaikasta. Viitanen ym. huomauttavat myös alustojen eduksi skaalautuvien verkostovaikutusten mahdollisuuden.

### **3.2 Datatalous ja datapohjainen liiketoiminta**

Datatalous liittyy kiinteästi alustatalouteen ja ero käsitteiden välillä on häilyvä. Alustatalous käsitteenä on hieman datataloutta laajempi. Datataloudessa kyse on nimensä mukaisesti dataresurssien hyödyntämisestä liiketoiminnan yhteydessä. Dataa yhdistellään, jaetaan, hyödynnetään ja analysoidaan, minkä seurauksena luodaan arvoa. Datatalouden kehityssuuntaa määrittelevät tällä hetkellä suuret globaalit yritykset, kuten Google ja Amazon. (Paavola ym., 2021, s. 19.)

Alusta- ja datatalouden kasvuun liittyy datapohjaisen liiketoiminnan yleistymisen. Datapohjaiset liiketoimintamallit poikkeavat perinteisistä malleista. Datapohjaisten liiketoimintamallien rakentamisen keskeisimpinä haasteina ovat kannattavan mallin luominen sekä palveluiden kaupallistaminen. Datapohjaisiin liiketoimintamalleihin siirtyminen vaatii yritykseltä erilaisia kyvykkyyksiä sekä strategioita liittyen alustan liiketoimintalogiikkaan, arvonluontiin ja innovaatiotoimintaan, mikä nimetään yrityksen menestyksen kannalta ratkaisevan tärkeäksi. (Paavola ym., 2021, s. 93.) Viitaseen ym. (2013, s. 23–25) mukaan ratkaisevan tärkeää on innovaatiotoiminnan sijaan se, että liiketoiminnan kehityksessä syntynyt uusi oppi ja tieto kasautuvat alustalle. Näin alusta kehittyy jatkuvasti, mikä puolestaan johtaa myös uuden innovointiin.

Siirtyminen datapohjaisiin liiketoimintamalleihin voi tapahtua yritysostojen, yhteistyön tai yrityksen sisäisen kehittämisen kautta. Kannattavan datapohjaisen liiketoimintamallin luomiseen ei ole yhtä oikeaa ratkaisua. Mallin luominen ja siirtymä edellyttää usein tapauskohtaista mallien muokkaamista. Tämä johtuu eri aloilla olevien yritysten kohtaamista erilaisista haasteista. (Paavola ym., 2021, s. 114–116.)

Haasteista huolimatta Viitaseen ym. (2013, s. 14–23) mukaan sekä vuosien aikana eri lähteistä kertyneet että nykyisin luotavat datavarannot tarjoavat datapohjaiselle

liiketoiminnalle lähes rajattomasti mahdollisuuksia. Heidän mukaansa toimijat pyrkivät kehittämään uusia liiketoimintamalleja, joiden avulla edellisiä malleja voidaan täydentää tai ne voidaan jopa kokonaan korvata. Uusien liiketoimintamallien avulla on mahdollista lisätä uusia toiminnallisuuksia vanhojen arvoketjujen päälle. Uusien toiminnallisuuksien seurauksena voidaan luoda uutta arvoa.

Arvoketjumallin esitellessä Porterin (1985, s. 38) mukaan yrityksen luoma arvo mitataan liikevaihdolla. Data ja teknologian kehitys ovat kuitenkin vuosikymmenien aikana muuttaneet perinteisiä arvon mittareita. Arvonluonnin mittaaminen ei ole enää niin yksinkertaista. Paavola ym. (2021, s. 12) huomauttavat datapohjaiseen liiketoimintaan liittyvän arvonluonnin mittaamiseen sisältyvän erityisiä haasteita. Datatalouden muokatessa liiketoiminnan malleja, muokkaa se myös arvonluonnin malleja. Paavolan ym. mukaan arvonluonti voi näkyä yrityksen liikevaihdon nousun sijaan osakkeiden arvonnousuna. Arvo saadaan täten realisoitua vasta yrityksen osakkeiden myynnin yhteydessä. Tietyissä olosuhteissa menestyvän alustan liikevaihto voi jäädä datan arvoa pienemmäksi.



## 4 DATAPOHJAINEN ARVONLUONTI JA ARVOKETJU

Luvussa kaksi verrattiin dataa ja öljyä raaka-aineena ja huomioitiin niiden eroavan toisistaan eri ominaisuuksiensa myötä. Molemmat raaka-aineet luovat arvoa niille ominaisten vaiheittaisten prosessien kautta. Mikäli öljyn arvoketju pilkotaan osiin, voidaan havaita, että sen arvontuontiprosessista löytyy paljon yhtäläisyyksiä datan arvontuontiprosessin kanssa. Pelkistetysti molempien raaka-aineiden arvontuontiprosessi menee seuraavassa järjestyksessä: raaka-aineen kerääminen, prosessointi sekä käyttö.

Datapohjaisen arvontuontiprosessin tutkimiseksi arvontuontiprosessi pilkotaan osiin, eli sitä tarkastellaan arvoketjumallin kautta. Arvoketjumallin (*value chain*) esitteli ensimmäisenä Michael Porter, vuoden 1985 kirjassaan “*Competitive Advantage: Creating and Sustaining Superior Performance*”. Malli kuvaa tuotteen arvontuodostusta yrityksen eri toimintojen myötä. Mallin mukaan arvoketju voidaan jakaa prosesseihin, missä jokainen prosessin vaihe lisää tuotteelle arvoa. Arvoketjumallin avulla voidaan tunnistaa yritysten kilpailuetuja, optimoida prosesseja, sekä karsia pois turhia vaiheita, jotka eivät lisää tuotteen arvoa (Porter, 1985, s. 33–39).

### 4.1 Datan arvoketju

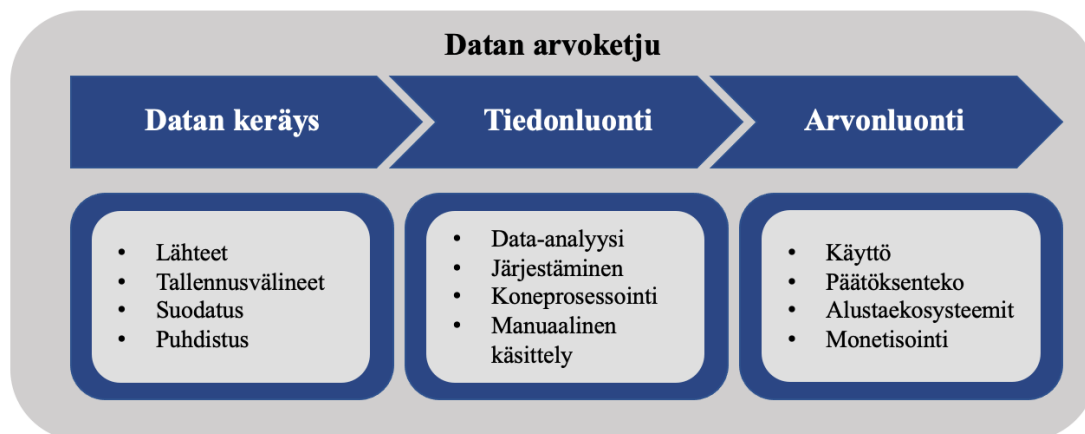
Arvoketjumallia voidaan soveltaa myös datan arvontuontiprosessin tarkastelemiseen, vaikka data itsessään ei ole fyysinen hyödyke. Faroukhi, El Alaoui, Gahi ja Amine (2020) kuitenkin huomauttavat tuoreessa julkaisussaan perinteisten arvoketjumallien olevan vanhentuneita datan arvontuonnin tarkasteluun. Ratkaisuksi tähän on arvoketjumalleista kehitelty uusia, datan arvoketjuja (Data Value Chain) kuvailevia malleja. Datan arvoketjun määritellään karkeasti olevan joukko toistettavia prosesseja, joiden myötä datasta saadaan irti arvoa sen elinkaaren aikana.

Datan arvoketjujen hallinta on hyvin tärkeää liiketoiminnan kasvun ja kilpailukyvyn edistämisen näkökulmasta. Datan arvoketjua tarkastelemalla voidaan tunnistaa, muokata ja tehostaa arvontuonnille olennaisia prosesseja, sekä ymmärtää siihen johtavat vaiheet. Arvoketjun tarkastelun avulla organisaatiot pystyvät kehittämään

toimintaansa ja luomaan arvoa kestäväällä tavalla. Arvonluonti saattaa vaatia prosessien uudelleenjärjestelyä ja vanhoista tavoista luopumista. (Faroukhi ym., 2020.)

Datan kuten minkä tahansa muun hyödykkeen arvoketju voidaan jakaa näkemyksestä riippuen useaan eri vaiheeseen. Faroukhi ym. (2020) jakavat datan arvoketjun neljään vaiheeseen. He tunnistavat vaiheiksi tuottamisvaiheen, keräysvaiheen, analyysivaiheen sekä tiedonvaihdon. Cavanillas, Curry ja Wahlster (2016, s. 32) jakavat datan arvoketjun viiteen eri vaiheeseen. Vaiheet ovat datan keräys, analysointi, kuratointi, säilytys sekä käyttö. Heidän esittämänsä vaiheet ovat järjestykseltään ristiriidassa Faroukhiin ym. vaiheiden kanssa. Voidaan huomata, että Cavanillaksen ym. mukaan analysointi tapahtuu ennen tiedon saattamista säilytykseen. Faroukhiin ym. näkemyksen mukaan analysointi suoritetaan vasta sen jälkeen, kun data on tallennettu säilytykseen keräysvaiheessa. Järjestyksen lisäksi arvoketjumallit voi poiketa myös vaiheiden määrän osalta. Arvoketju voi olla myös hyvin pienissä osissa. Lim, Kim, Kim, Heo, Kim ja Maglio (2018) luonnehtivat tietointensiivisten palveluiden datapohjaiseen arvonluontiin yhdeksänvaiheisen arvoketjun kehityksen. Nämä yhdeksän vaihetta on kuitenkin mahdollista tiivistää heidän mukaansa neljään päävaiheeseen.

Voidaan huomata, että eri lähteissä arvoketjut saattavat vaihdella järjestyksen osalta tai ne on pilkottu useaan eri osaan riippuen siitä, miten tarkasti vaiheita katsotaan. Kaikki eri lähteiden nimeämät vaiheet voidaan kuitenkin vapaasti tiivistettynä esittää kolmena päävaiheena. Olennaisuuden vuoksi tässä tutkielmassa datan arvoketjua tarkastellaan näiden kolmen päävaiheen kautta. Vaiheet ovat datan keräys, tiedonluonti sekä arvonluonti. Datan arvoketjun vaiheet havainnollistetaan kuviossa 3.



**Kuvio 3. Datan arvoketju (mukaillen Faroukhi ym. 2020.)**

Luvussa 2.2. tarkasteltiin big datan prosesseja. Prosesseissa käytiin läpi tiedonhallintaprosessia, mikä liittyy datan arvoketjun keräysvaiheeseen. Tarkastelun kohteena oli myös tiedonluontivaiheeseen liittyvä data-analytiikkaprosessi. Edellä mainitut vaiheet ja prosessit ovat oleellisia datan arvonluonnin kannalta. Ilman datan prosessointia, ei siitä saataisi arvoa. Prosessit itsessään eivät kuitenkaan tuota datalle arvoa, vaan se syntyy käytön myötä. Prosessien aikaisemmasta tarkastelusta johtuen tämän luvun huomio on erityisesti datan arvoketjun arvonluontivaiheessa. Seuraavissa alaluvuissa esitetään datan arvoketjun eri vaiheet. Tämän lisäksi nostetaan esille datapohjaiseen arvonluontiin liittyviä haasteita ja mahdollisuuksia datan monetisointiin. Lisäksi luvuissa havainnollistetaan datan hyödyntämisestä seuraavaa arvonluontia rahoitus- ja vakuutusalailla.

#### 4.1.1 Datan keräys

Datan keräyksellä tarkoitetaan niitä toimia, mitä suoritetaan, jotta data saadaan tallennettua eri lähteistä sille tarkoitettuun tallennuskohteeseen. Faroukhi ym. (2020) mukaan tallennusvälineillä on suuri merkitys arvoketjun tehokkuuteen sekä skaalautuvuuteen. Tehokkuuteen vaikuttaa tallennustilan luotettavuus, sekä tietojen saatavuus tallennuksesta data-analyysia varten. Suosittu tallennusratkaisu on pilvipalvelut, joiden toimintaa tarkasteltiin luvussa 2.2.3.

Keräysvaiheeseen kuuluvia toimintoja ovat keräys, suodatus ja puhdistus. Vaihe alkaa datalähteiden tunnistamisella. Lähteiden tunnistaminen keräysvaiheessa auttaa

tiedonluontivaiheen analyysissä ja datan käytön suunnittelussa. Data voidaan sen lähteen perusteella suunnata analysoitavaksi lopullisen käyttötarkoituksen mukaan. (Lim ym., 2018.)

Keräysvaiheessa dataa on hyvin paljon ja suuri osa siitä on vähäarvoista dataa (low-value data). Tästä johtuen datan arvoketjun keräysvaiheessa tulee keskittyä suodattamaan dataa niin, että datamäärästä ainoastaan arvokas data (high-value) jatkaa matkaa arvoketjun seuraavaan vaiheeseen. (Cavanillas ym., 2016, s. 39.) Faroukhi ym. (2020) ovat tehneet saman havainnon ja pitävät suodatusta myös tärkeänä osana datan esiprosessointia. Heidän mukaansa suodattamaton raaka data voi olla laadultaan keuhkoa ja tuottaa analysoinnissa merkityksettömiä tuloksia. Merkityksettömät tulokset eivät johda tavoiteltuun arvonluontiin, sillä tulosten tekemisestä on syntynyt kustannuksia. Merkityksettömien tulosten perusteella tehdyt päätökset voivat myös johtaa kustannuksiin.

Yhteenvedon voidaan havaita, että keräysvaiheen avulla arvonluonnin kokonaisprosessin tehostaminen on mahdollista. Varmistamalla arvokkaan datan jatkaminen seuraavaan vaiheeseen voidaan maksimoida datan arvonluontipotentiali. Cavanillas ym. (2016, s. 39) huomauttavat kuitenkin, ettei suodattaminen ole kaikille organisaatioille mielekästä. Tietyille organisaatioille suurin osa datasta voi olla potentiaalisesti arvokasta, minkä johdosta sitä ei voida karsia vielä keräysvaiheessa. Organisaatioiden tulee itse tapauskohtaisesti tarkastella ja muokata arvoketjun keräysvaihetta tarpeidensa mukaan.

#### 4.1.2 Tiedonluonti

Tiedonluontivaiheeseen sisältyy kaikki toiminnot, missä dataa käsitellään sen tallentamisen jälkeen. Yksi tärkeimmistä toiminnoista on datan analysointi. Tavoitteena on saada datasta irti tietoa, jota voidaan käyttää päätöksenteon tukena tai muuhun tiettyyn käyttötarkoitukseen. Faroukhi ym. (2020) nimeävät data-analyysivaiheen olevan datan arvoketjun tärkein vaihe. He ovat liittäneet analyysivaiheen suoraan arvonluontivaiheeseen. Heidän mukaansa analyysivaiheessa datasta saadut tiedot luovat arvoa tukemalla yrityksen operationaalista ja strategista päätöksentekoa. Tämä on osittain ristiriidassa Cavanillaksen ym. (2016) mallin

kanssa, sillä he ovat irrottaneet data-analyysivaiheen erilliseksi vaiheeksi datan arvoketjussa. He ovat kuitenkin yhtä mieltä siitä, että data-analyysi on yksi datan arvoketjun tärkeimmistä osista. He luonnehtivat, ettei arvoketjun loppuosa toimi tarkoitettulla tavalla ilman data-analyysiä. Tämä johtuu siitä, että kerätty raaka data on pääosin vailla rakennetta eri muodoissa. Data-analyysin tehtävänä on lisätä datalle rakennetta järjestelemällä sitä luettavampaan muotoon. Järjestelyn myötä dataa on helpompi hyödyntää eri teknologioiden avulla. Mitä rakenteisempaa data (structured data) on, sitä soveliaampaa se on koneprosessointia varten. Vastaavasti vailla muotoa oleva raaka data vaatii ihmisen käsittelyä. (Cavanillas ym., 2016, s. 63–83.)

Analyysitekniikoita on useita. Ne voidaan jakaa sen mukaan, miten automatisoitua analysointi on. Toiset menetelmät ovat lähes täysin automatisoituja ja toiset vaativat ihmisten osallistamista. Osakemarkkinoilla yleensä hyödynnetty analytiikka vaatii paljon ihmisen osallistamista päätöksenteon prosessin yhteydessä. (Lim ym., 2018.) Analyysitekniikat voidaan jakaa myös kategorioihin. Faroukhi ym. (2020) tunnistaa neljä analyysikategoriaa: kuvailevan, ennakoivan, ohjailevan sekä diagnostiikan. Kuvailevan analyysin kategoriassa tarkastellaan menneisyyttä ja etsitään syitä, miksi jotain on tapahtunut. Ohjailevassa analyysissä punnitaan eri vaihtoehtoja tulevia päätöksiä varten ja diagnostiikassa tarkastellaan tuottavuuteen ja tehokkuuteen vaikuttavia seikkoja.

Faroukhista ym. eroten Ohlhorst (2013, s. 3–5) tunnistaa viisi eri kategoriala, joista sama Faroukhi ym. kanssa on jo aikaisemmin luvussa 2.2.2. tarkasteltu ennakoiva analytiikka. Useissa lähteissä esille nostetussa ennakoivassa analytiikassa pyritään tunnistamaan toistuvia kuvioita, esimerkiksi regressioanalyysien myötä (Faroukhi ym., 2020). Ohlhorstin mukaan ennakoivaa analytiikkaa käytetään laajasti taloudellisella ja tieteellisellä sektorilla, missä trendien perusteella on mahdollista ennakoita tulevaa. Hänen mukaansa ennakoivan analytiikan tarkoituksena on auttaa yrityksiä tunnistamaan liiketoimintaan liittyviä riskejä sekä mahdollisuuksia. Voidaan huomioda, että tiedonluontiin liittyviä kategorioita on arvoketjun vaiheiden mukaisesti useita. Ohlhorst huomauttaakin, että hänen esittelemänsä kategoriat ovat vain pieni osa kokonaisuutta.

#### 4.1.3 Arvonluonti ja käyttö

Data tarjoaa useita mahdollisuuksia arvonluontiin. Tässä luvussa käydään läpi datan arvoketjun arvonluontivaihetta. Arvonluontivaiheessa data luo arvoa käytössä. Tavoitteena on luoda yleiskuva siitä, millaisia ovat erilaiset datapohjaisen arvonluonnin mahdollisuudet, sekä miten datan käyttö eri tarkoituksissa, kuten päätöksenteon yhteydessä, johtaa arvonluontiin.

Arvoa voidaan mitata monin eri tavoin. Arvo voi tarkoittaa puhtaasti hyödykkeen myynnistä saatavaa rahaa tai arvonluonti voi näkyä yrityksen prosessien tehostamisen kautta kulujen pienenemisenä, sekä myynnin kasvattamisena. Kuluttajalle yritys voi luoda arvoa parantamalla tuotteen ominaisuuksia, kuten valmistamalla uusista automalleista turvallisempia. Arvoa voidaan luoda saavuttamalla datan avulla haluttuja lopputulemia tai estämällä ei-toivottuja lopputulemia. Arvonluonti tapahtuu, kun datasta saatua tietoa käytetään tiettyyn käyttötarkoitukseen. (Lim ym., 2018.)

Cavanillaksen ym. (2016, s. 143) mukaan datan käyttö tuo liiketoiminnalle arvoa kahdella eri tavalla. Ensimmäiseksi datan hyödyntäminen auttaa hyödykkeen arvoketjun hallinnassa ja toiseksi se tuo arvoketjulle läpinäkyvyyttä. Vaihtelevasti Ohlhorstin näkemyksen mukaan datapohjainen arvonluonti näkyy tehokkuuden parantumisenä, uusina innovaatioina, sekä uusien liiketoimintamahdollisuuksien ja prosessien löytämisessä big datan avulla. Hän lisää, että big datan analytiikka on avainasemassa tukemassa yritysten kasvua ja kilpailukykyä. Hänen mukaansa seurauksena todennäköisesti syntyy uutta tuottavuuden kasvua, sekä kuluttajien ylijäämää (consumer surplus). (Ohlhorst, 2013, s. 22.) Faroukhin ym. (2020) näkemyksen mukaan datan arvoketjun viimeisessä vaiheessa data-analyysin tulokset jaetaan organisaation sisällä tai yhteistyökumppaneille; näin data auttaa heitä arvonluonnissa. Voidaan siis todeta, että datapohjainen arvonluonti voi tapahtua usealla eri tavalla, joita kaikkia yhdistää se, että dataa hyödynnetään käytössä.

Viitanen ym. esittelevät big datan arvonluontiin eri näkökulman. Heidän mukaansa alustatalouden tuomien alustojen myötä arvonluonti tulee muuttumaan. Tämä tarkoittaa sitä, että arvonluonti ei tapahdu enää Porterin (1985) arvoketjumallin mukaan arvoketjun vaiheita optimoimalla ja myymällä valmis tuote kuluttajille.

Viitanen ym. näkevät, että arvonluonti tapahtuu uusilla alustaekosysteemeillä. Alustaekosysteemeillä arvoa on mahdollista luoda toimijoiden välillä. Alustalle syntynyt arvo jaetaan toimijoiden kesken. Jakaminen tapahtuu osallistumisen perusteella toimijoiden yhdessä määrittelemien sääntöjen mukaan. Faroukhi ym. (2020) esittävät samankaltaisen havainnon. Heidän mukaansa arvonluonti voi lineaarisen arvoketjun sijaan tapahtua moniulotteisesti arvoverkostossa (value grid). Verkostossa arvoa on mahdollista luoda usean eri polun kautta. Huomionarvoista on kuitenkin se, että suurin osa yrityksistä ei ole siirtynyt Viitanen ym. ja Faroukhi ym. esittelemiін alustaekosysteemeihin ja verkostoihin perustuviin arvonluontimalleihin. Tästä johtuen datapohjaista arvonluontia tapahtuu edelleen pääosin hyödyntämällä dataa nimenomaan arvoketjujen optimoinnissa ja sitä kautta tuottavuuden parantamisessa. (Viitanen ym., 2013, s. 22–23.)

Dataa hyödyntävät yritykset haluavat tietenkin maksimoida datan arvon. Tätä varten he tarvitsevat Ohlhorstin mukaan suunnitelman. Ohlhorst luonnehtii big datan olevan yritykselle häiritsevä tekijä (disruptive element). Tällä hän tarkoittaa sitä, että big datan hyödyntämisestä seuraa kuluja esimerkiksi varastoinnin ja käsittelyyn tarvittavan henkilökunnan muodossa. Big datan implementointi liiketoimintaan on yritykselle ylimääräistä, sillä tuote tai palvelu voitaisiin tietyissä tilanteissa valmistaa myös ilman datan hyödyntämistä. Hänen mukaansa big datan hyödyntämiseen liittyvässä liiketoimintasuunnitelmassa tärkeässä asemassa ovat sijoitetun pääoman tuotto (Return on Investment, ROI) sekä omistamisen kokonaiskustannukset (Total Cost of Ownership, TCO). Omistamisen kokonaiskustannuksilla tarkoitetaan tässä yhteydessä datan elinkaareen liittyvien hankinta- ja käyttökustannuksien summaa. Arvonluonti maksimoidaan kasvattamalla sijoitetun pääoman tuottoa ja minimoimalla omistamisen kokonaiskustannuksia. Ohlhorst ehdottaa tämän olevan mahdollista hyödyntämällä big dataa yrityksen tarpeiden mukaisten IT-projektien yhteydessä. Tuoton ja kustannusten suhde paranee, kun uutta arvoa löydetään data-analytiikan projektiin soveltamisen kautta. (Ohlhorst, 2013, s. 22–25.)

Yksi datan tärkeimmistä käyttökohteista on toimia yrityksen päätöksenteon tukena. Brynjolfsson, Hitt ja Kim (2011) suorittivat tutkimuksen liittyen datapohjaisen päätöksenteon hyödyntämiseen yrityksissä. Otantana oli 179 suurikokoista pörssilistattua yhtiötä. He havaitsivat datapohjaisen päätöksenteon hyödyntämisen

korreloivan korkeamman tuottavuuden ja markkina-arvon kanssa. Datapohjaisen päätöksenteon ja yhtiöiden tulosten yhteys oli nähtävillä myös eri tunnusluvuissa. Datapohjaisen päätöksenteon hyödyntäminen korreloi korkeamman oman pääoman tuoton (Return on Equity, ROE) kanssa. Tutkijoiden mukaan yhtiöiden kyvykkyys datapohjaiseen päätöksentekoon voidaan nähdä yhtiön aineettomana hyödykkeenä. Kyvykkyys hyödyntää dataa päätöksenteon yhteydessä on arvostettua myös sijoittajien keskuudessa, mikä voi näkyä osakekurssien nousuna. Päätöksenteon tukemiseksi datan visualisointi on tärkeää. Cavanillas ym. ja Faroukhi ym. esittelevät molemmat visualisoinnin osana datan arvoketjumalliaan. Visualisoimalla data-analytiikasta saatuja tuloksia, on niistä helpompi vetää johtopäätöksiä sekä löytää tiettyjä kuvioita. Data-analyysin tuloksien visualisoinnilla voidaan tarkoittaa esimerkiksi trendien esittelyä. Hyvä visualisointi auttaa analyysin tuloksien laadun määrittämisessä. (Cavanillas ym., 2016, s. 152; Faroukhi ym., 2020.)

Data tarjoaa arvonluonnin mahdollisuuksia eri toimialoille. Dataa hyödynnetään paljon tuotannossa ja teollisuudessa sekä finanssialalla. Cavanillas ym. nostavat esille ennakoivan analytiikan arvonluonnin mahdollisuudet. Heidän esimerkkinsä mukaan datan hyödyntäminen luo arvoa tuotanto- ja logistiikkasektorille. Koneisiin voidaan lisätä sensoreita ja niiden tuottaman datan analysoinnin avulla voidaan ennakoida huoltotarpeita. Ongelmien ilmetessä huolto voidaan suorittaa ennen kuin kone hajoaa aiheuttaen kustannuksia korjauksen osalta sekä tuotannon seisahtumisena. Mikäli koneissa ei ilmene ongelmia, voidaan huoltoväliä vastaavasti pidentää säästäen rahaa. Viitanen ym. (2013) esittelevät saman havainnon. Heidän mukaansa tuotteisiin lisättävien komponenttien avulla luodaan tuote-palvelukokonaisuuksia. Tämä mahdollistaa myös jakeluketjujen automatisoinnin. Paavola ym. (2021) edelleen tukevat samaa havaintoa ja mainitsevat autonomisen toiminnan olevan tärkeässä asemassa tulevaisuuden lyhyen aikavälin päätöksenteossa. Autonomisessa toiminnassa järjestelmät säätelevät datan avulla itsenäisesti tekemistään. (Cavanillas ym., 2016, s.143–145; Paavola ym., 2021, s. 103; Viitanen ym., 2013, s. 23.) Ennakoivan analytiikan käytön avulla siis luodaan arvoa kustannusten säästymisen muodossa sekä prosessien automatisoinnilla.



## 4.2 Arvonluonti rahoitus- ja vakuutuslalla

Manyikan ym. (2011) mukaan rahoitus- ja vakuutuslalla voi erityisesti hyötyä big datan soveltamisesta. Tämä johtuu siitä, että mahdollisuuksia datapohjaiseen arvonluontiin on useita. Cavanillaksen ym. (2016, s. 209) havaintojen mukaan rahoitus- ja vakuutuslalla dataa on hyödynnetty jo useiden vuosien ajan. He jakavat näkemyksen Manyikan ym. (2011) kanssa ja mainitsevat, että big datan markkinan rahoituslalla arvioidaan olevan kasvussa. Kasvu johtuu organisaatioiden tarpeesta vastata rahoitusalan sääntelyyn. Sääntelyyn vastaamista voidaan edistää datan hyödyntämisen avulla.

Muita Cavanillaksen ym. (2016) mainitsemia tapoja luoda arvoa on hyödyntää markkinadataa osakkeista tai valuuttakursseista. Tavoitteena on saada suuresta datamäärästä tehokkaasti tietoa, jonka perusteella voidaan toimia. Tieto voidaan sitten yhdistää muista lähteistä saatavien tietojen kanssa. Tämä kaikki tapahtuu edistyneiden tietokonealgoritmien myötä reaaliajassa luoden hybrididatasettejä. Hybrididatasetti tarkoittaa strukturoidun ja strukturoimattoman datan yhdistelmää. Datasetit tarjoavat kattavan näkemyksen markkinoista, jota voidaan hyödyntää kaupankäynnin yhteydessä. Dataseiteistä voidaan nähdä markkinasignaaleja, joiden avulla tehdään myynti- tai ostopäätöksiä. Datan hyödyntämisen tarjoama informaatio mahdollistaa tehokkaamman kaupankäynnin, luoden sijoittajille sekä rahoituslaitoksille arvoa. (Cavanillas ym., 2016, s. 209–222.)

Digitalisaation myötä organisaatioilla on saatavilla laajasti asiakasdataa ja sitä voidaan hyödyntää eri toiminnoissa. Yksi tärkeimmistä hyödyntämiskohteista on luottoriskien hallinta. Erityisesti rahoituslaitosten liiketoiminnalle luottoriskien hallinta on tärkeää, sillä suuri osa liiketoiminnasta perustuu rahoitustoimintaan. Luottoriskillä tarkoitetaan tilannetta, missä velallinen ei maksa lainan myöntäjälle takaisin lainaamaansa summaa lainan ehtojen mukaisesti. Luottoriskin toteutuessa lainan myöntäjänä toiminut rahoituslaitos tekee tappiota. Mikäli luottoriskiä voidaan dataa hyödyntämällä hallitusti pienentää, pienenevät myös luottoriskiin liittyvät tappiot. Käänteisesti tämä tarkoittaa sitä, että arvoa luodaan pienentämällä tappioiden mahdollisuuksia. Riskienhallintaprosessiin kuuluu riskien tunnistaminen ja mittaaminen sekä strategian luominen riskien torjumiseksi. Luottoriskien hallintaan voidaan käyttää pisteytykseen

perustuvaa mallia (credit scoring model) asiakkaista. Mallissa pisteytyksen perusteella määritellään todennäköisyyttä tietyn riskin toteutumiselle. Mallin avulla mitataan kolmea eri riskiä: maksuhäiriöriskiä, luottotappioriskiä sekä altistumisriskiä. Riskien arvioinnista saatua tietoa käytetään päätöksenteon tukena lainan myöntämisen yhteydessä ja hallitaan korkean riskin asiakkaiden luottorajoja. Van Gestelin ja Baesensin (2009) mukaan rahoituslaitokset pystyvät luomaan mallit itse tai hankkimaan ne ulkoisilta toimittajilta, kuten luottotoimistoilta (credit bureau). Rahoituslaitokset suosivat omien mallien kehittämistä, mikäli heillä on tarpeeksi sisäisiä resursseja, kuten laadukasta dataa. Ulkoisten toimittajien luottopisteytysmalleihin liittyy kuitenkin tiettyjä etuja. Luottotoimistot rakentavat mallinsa perustuen useilta eri rahoituslaitoksilta saatuun dataan ja näin ollen heidän luottopisteytysmallinsa toimivat tehokkaammin ja tarjoavat tarkempaa informaatiota. Mallien toimiessa tarkemmin eri toimijoilta kerätyn suuremman datamäärän avulla, vaikuttaa se myös luottoriskien pienentymiseen eli oletettavasti rahoituslaitosten tulos paranee. (Van Gestel & Baesens, 2009, s. 38, 106–110.) Edellä mainituin menetelmin asiakasdataa hyödynnetään luottoriskien hallinnassa luoden arvoa rahoituslaitoksille.

### **4.3 Datan kierrätys ja monetisointi**

Faroukhi ym. (2020) väittävät big dataa hyödyntävien organisaatioiden saavuttavan korkeampia hyötyjä normaaleihin organisaatioihin verrattuna. Syyksi tähän he esittävät mahdollisuuden datan uudelleenkäyttöön. Data ei käytön myötä kulu. Kun data on kerran kerätty, voidaan sitä hyödyntää useasti uudelleen. Näin ollen samaa dataa voidaan käyttää eri käyttötarkoituksiin. Faroukhin ym. (2016) mukaan tämä on synnyttänyt markkinat datalle. Markkinoilla yritykset monetisoivat datansa, eli myyvät sitä muiden käyttöön korvausta vastaan. Cavanillas ym. ovat aiheesta yhtä mieltä. Heidän mukaansa mahdollisuus datan monetisointiin on luonut aiheen ympärille uusia liiketoimintamalleja. Liiketoimintamallit perustuvat siihen, että data ja sen analyysi yhdellä toimialalla voi olla arvokasta myös toisella toimialalla olevalle yritykselle. (Cavanillas ym., 2016, s. 231; Faroukhi ym., 2020.)

Najjar ja Kettinger (2013) vahvistavat havaintoa datan monetisoinnin mahdollisuudesta. Heidän mukaansa monetisoinnin ovat mahdollistaneet data-analytiikan työkalut sekä pilvipalvelut. Heidän mukaansa datan monetisoinnin

erityispiirteensä on se, että dataa voidaan myydä sen elinkaaren eri vaiheissa. Myyty data voi olla kerättyä raakaa dataa tai jo valmiiksi analysoitua ja käsiteltyä. Najjarin ja Kettingerin mukaan datalle voidaan lisätä arvoa valmistelemalla se myyntiä varten integroimalla siihen ylimääräistä relevanttia dataa. Arvoa voidaan kasvattaa myös paketoimalla data eri tarpeisiin vastaavalla tavalla. Myyntiprosessin synnyttämää kokonaisarvoa voidaan edelleen lisätä tuomalla monetisoinnin ekosysteemiin kolmansia osapuolia. Kolmannet osapuolet voivat lisätä arvoa toimimalla datan myyjien ja ostajien välittäjänä, tehostaen datan myynti- ja ostoprosessia. (Najjar & Kettinger, 2013.)

#### **4.4 Datapohjaisen arvонуonin haasteet**

Data-analytiikasta ja datan hyödyntämisestä ei kuitenkaan aina seuraa tehokkuuden parantuminen tai suuret voitot. Kuten Shah, Horne ja Capellá (2012) artikkelissaan kirjoittavat: hyvä data ei takaa hyviä päätöksiä. Heidän mukaansa data-analytiikkaan tehdyt investoinnit voivat osoittautua turhiksi, mikäli analytiikan hyödyntämistä ei onnistuta liittämään päätöksenteon yhteyteen. Paavola ym. (2021) toteavat myös, että datan hyödyntämisen kustannukset voivat kasvaa suuremmiksi, kuin sen hyödyntämisestä seuraava liikevaihto. Tällöin datan hyödyntäminen on yritykselle haitallista. Huomattavaa on, että vahvan alustaekosysteemin ja verkostovaikutusten luomiseksi osa alustayrityksistä joutuu toimimaan useita vuosia tappiollisesti. (Paavola ym., 2021, s. 16–25.)

Data-analyysin päätöksentekoon implementoinnin lisäksi erääksi datan arvонуonin haasteeksi Alanko ja Salo (2013) nostavat esille big dataan liittyvien tiedon jakamisen kanavien rajallisuuden. Merkittävä haaste on myös yksityisyyteen ja tietoturvaan liittyvät oikeudelliset seikat. Tulevaisuudessa heidän mielestään yksi suurimmista haasteista tulee olemaan oikeusturvan valvominen kerättävien tietojen jakamisen ja tallentamisen kontekstissa. Paavola ym. (2021) noteeraavat myös saman ongelman. Heidän mukaansa datan potentiaalinen hyödyntämiseksi tarvitaan suotuisaa politiikkaviitekehystä. Viitekehysten luomiseen tulisi osallistaa eri sidosryhmiä siten, että viitekehys hyödyttää kaikkia osapuolia, edistäen hyvinvointia, luottamusta sekä tietoturvallisuutta. Keskeistä on heidän mukaansa datan saatavuus, mitä voidaan

edistää esimerkiksi lainsäädännöllisiä turvallisuusintressejä edustavilla jakamismekanismeilla. (Paavola ym., 2021, s. 145–146.)

## 5 YHTEENVETO

Tutkielman tavoitteena oli tarkastella datapohjaisen arvonluonnin ominaisuuksia, erityispiirteitä sekä sovellusta käytännössä. Datapohjaista arvonluontiprosessia tarkasteltiin pilkkomalla prosessi arvoketjumallin mukaisesti vaiheisiin, minkä myötä pyrittiin vastaamaan johdannossa esiteltyyn tutkimuskysymykseen: Miten datapohjainen arvonluonti tapahtuu arvoketjun eri vaiheiden kautta? Datapohjaista arvonluontia tutkittiin aluksi määrittelemällä arvonluonnissa hyödynnettävää big dataa kolmen V-kirjaimen mallin kautta. Tutkielmassa tarkasteltiin myös datapohjaiseen arvonluontiin liittyvää toimintaympäristöä eli alustataloutta, sekä alustojen mahdollistamia uusia liiketoimintamalleja. Lopuksi tutkimuskysymykseen pyrittiin vastaamaan tarkastelemalla datapohjaisen arvonluonnin arvoketjua. Arvoketjun osien ohella datapohjaista arvonluontia tarkasteltiin rahoitus- ja vakuutusalan kautta. Tutkielmassa luotiin myös katsauksen datan monetisointiin, sekä arvonluontiin ja datan hyödyntämiseen liittyviin haasteisiin.

Tutkielman tulokset osoittavat, että datan hyödyntäminen liiketoiminnan yhteydessä voi synnyttää huomattavaa kilpailuetua ja arvoa. Tutkielma huomioi alustatalouden kehityksen avaavan uusia mahdollisuuksia datapohjaiselle liiketoiminnalle sekä arvonluonnille. Arvoketjun tarkastelu osoitti, että datapohjainen arvonluonti arvoketjussa konkretisoituu arvoketjun lopussa. Datapohjainen arvonluonti syntyy käytön yhteydessä tai monetisoinnin kautta. Tästä huolimatta arvoketjussa dataa prosessoivat aikaisemmat vaiheet ovat elintärkeässä roolissa, sillä ilman niitä arvonluonti ei onnistuisi. Arvoketjun vaiheiden tutkiminen osoitti, että arvoketjun aikaisemmillä vaiheilla on vaikutusta arvoketjun loppupään arvonluontiin. Vaiheita tehostamalla voidaan maksimoida datan arvonluontipotentiaali. Tutkielman tulosten mukaan datapohjainen arvonluonti voi tarkoittaa kilpailuedun luomista, tehokkuuden paranemista, prosessien automatisointia ja tehostamista, kustannusten pienentymistä tai suoraa datan myymisestä seuraavaa voittoa. Rahoitusallalla dataa voidaan hyödyntää osakekaupankäynnin yhteydessä ja riskienhallinnassa. Tutkielman löydösten mukaan datan hyödyntämiseen liittyy erilaisia haasteita, jotka voivat vaikuttaa lopputulokseen. Haasteiksi nousi datan hyödyntämiseen liittyvät kustannukset, jotka mahdollisesti nousevat hyötyjä suuremmiksi. Toiseksi suureksi

haasteeksi tuli esille oikeusturvaan ja lainsäädäntöön liittyvät seikat, jotka rajoittavat datan hyödyntämisen mahdollisuuksia.

Tutkielma osoitti, että tutkimuskirjallisuutta big datasta ja arvonaluonnista on suhteellisen kattavasti, vaikka kyseessä on suhteellisen tuore, kasvava ilmiö. Tutkimuskirjallisuus oli big datan ja alustatalouden määrittelemisen suhteen yhteneväistä. Kirjallisuuden mukaan big dataa määrittelevistä ominaisuuksista tärkeimmäksi nousi datan suuri määrä, mihin liittyi erilaisia prosessoinnin haasteita. Datatalouden tunnistettiin lisäksi yhteneväisesti tarjoavan merkittäviä mahdollisuuksia yrityksille, uusien arvonaluonnin keinojen ohella. Arvonaluontia ja arvoketjua käsittelevä kirjallisuus oli osin yhteneväistä ja osin ristiriitaista. Erona oli arvoketjun vaiheiden jakaminen eri suuruisiin osiin. Kirjallisuus tarjosi lisäksi näkemyksestä riippuen eri keinoja arvonaluontiin. Tutkimuskirjallisuus oli kuitenkin yhtä mieltä siitä, että data luo arvoa pääosin käytön yhteydessä.

Datan arvoketjuihin liittyvä aikaisempi tutkimuskirjallisuus keskittyy arvoketjun mallintamiseen omien vaiheidensa kautta. Tämä tutkielma luo datan lailla arvoa kokoamalla yhteen tutkimuskirjallisuudelle yhtenäiset arvoketjun vaiheet ja esittelemällä niistä tiivistetyn version. Tutkielma visualisoi lukijalle datapohjaisen arvonaluonnin ymmärrettävällä tavalla yksinkertaisen arvoketjumallin kautta.

Tulevaisuuden suhteen tutkimuskirjallisuudessa mainittiin lähes poikkeuksetta datan hyödyntämisen olevan yhä kehittyvä ilmiö ja datan potentiaalinen olevan sen nykyistä käyttöastetta suurempi. Datan kerääminen ja säilyttäminen tulee halvemmaksi teknologian kehityksen myötä, mikä osaltaan lisää sen hyödyntämisen mahdollisuuksia tulevaisuudessa. Kuten Ohlhorst (2013) mainitsee, hän on optimistinen datan hyödyntämisen tarjoamista mahdollisuuksista. Hänen mukaansa big datan analyysi vie eteenpäin lähes jokaista modernin yhteiskunnan toimialaa. Big data on jo vaikuttanut eri tieteenaloihin, finanssialaan, tuotantoon ja digitaalisiin palveluihin, mutta Ohlhorstin mukaan tämä on vasta alkua. Hän näkee big datasta saatavan paljon enemmän arvoa irti uusien innovaatioiden myötä. (Ohlhorst, 2013, s. 113–115.) Aiheeseen liittyviä jatkotutkimuskysymyksiä on paljon ilmiön tuoreudesta johtuen. Jatkotutkimukset voisivat keskittyä case-tapauksiin, joissa datapohjaista arvonaluontia tutkittaisiin yksittäisten yritysten kautta. Tämän lisäksi

jatkotutkimuksissa voitaisiin tutkia malleja datan arvonmääritykselle tarkastelemalla monetisointiin liittyvässä kaupankäynnissä datan kauppahintaan vaikuttavia tekijöitä.

## LÄHTEET

- Alanko, M., & Salo, I. (2013). Big data Suomessa. *Liikenne- Ja Viestintäministeriön Julkaisuja 25/2013*. <http://urn.fi/URN:ISBN:978-952-243-358-9>
- Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, 54(15), 2787–2805. <https://doi.org/10.1016/J.COMNET.2010.05.010>
- Brynjolfsson, E., Hitt, L., & Kim, H. (2011). Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? *International Conference on Information Systems 2011, ICIS 2011, 1*, 541–558. <https://doi.org/10.2139/SSRN.1819486>
- Cavanillas, J. M., Curry, E., & Wahlster, W. (2016). *New Horizons for a Data-Driven Economy A Roadmap for Usage and Exploitation of Big Data in Europe*. Springer.
- Chen, B. (16.9.2021). The battle for digital privacy is reshaping the internet. *The New York Times*. <https://www.nytimes.com/2021/09/16/technology/digital-privacy.html>
- Faroukhi, A. Z., El Alaoui, I., Gahi, Y., & Amine, A. (2020). Big data monetization throughout Big Data Value Chain: a comprehensive review. *Journal of Big Data*, 7(1), 1–22. <https://doi.org/10.1186/S40537-019-0281-5>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/J.IJINFOMGT.2014.10.007>
- Kenney, M. & Zysman, J. (2016) The rise of the platform economy. *Issues in Science and Technology*, 32(3), 61–69. <https://issues.org/rise-platform-economy-big-data-work/#.Yl-3riDG-Fo.link>



- Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5(12), 2032–2033. <https://doi.org/10.14778/2367502.2367572>
- Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety*. META Group. <https://studylib.net/doc/8647594/3d-data-management--controlling-data-volume--velocity--an...#>
- Lim, C., Kim, K. H., Kim, M. J., Heo, J. Y., Kim, K. J., & Maglio, P. P. (2018). From data to value: A nine-factor framework for data-based value creation in information-intensive services. *International Journal of Information Management*, 39, 121–135. <https://doi.org/10.1016/J.IJINFOMGT.2017.12.007>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. [https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI\\_big\\_data\\_exec\\_summary.pdf](https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.pdf)
- Najjar, M. S., & Kettinger, W. J. (2013). Data monetization: lessons from a retailer's journey. *MIS Quarterly Executive: A Research Journal Devoted to Improving Practice*, 12(4), 213–225.
- Ohlhorst, F. (2013). Big Data Analytics: What it is and why it matters. *Sas*, 8302(April), 1–5. [https://www.sas.com/en\\_us/insights/analytics/big-data-analytics.html](https://www.sas.com/en_us/insights/analytics/big-data-analytics.html)
- Paavola, H., Seppänen, M., & Eloranta, V. (2021). *Työ- ja elinkeinoministeriön julkaisuja • Yritykset • 2021:3 Datapohjaisen arvonluonnin strategiset vaihtoehdot*.
- Porter, M. E. (1985). Competitive advantage: creating and sustaining superior performance. In *New York* (Vol. 2). Free Press.

- Prüfer, J., & Schottmüller, C. (2021). Competing with Big Data. *The Journal of Industrial Economics*, 69(4), 967–1008. <https://doi.org/10.1111/JOIE.12259>
- Puig, M., Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: a data management perspective. *Frontiers Of Computer Science*, 7(2), 157–164. <https://doi.org/10.1007/s11704-013-3903-7>
- Russom, P. (2011). Big Data Analytics. TD WI Best Practices Report, Fourth Quarter. TDWI Research.
- Shah, S., Horne, A., & Capellá, J. (2012). *Good Data Won't Guarantee Good Decisions*. Harvard Business Review, 90(4), 23-25. <https://hbr.org/2012/04/good-data-wont-guarantee-good-decisions>
- Treder, M. (2019). What is Data? Teoksessa *Becoming a Data-Driven Organisation*, (ss. 1–39). Berlin, Heidelberg: Springer Vieweg. [https://doi.org/10.1007/978-3-662-60304-8\\_1](https://doi.org/10.1007/978-3-662-60304-8_1)
- Van Gestel, T., & Baesens, B. (2009). *Credit Risk Management: Basic Concepts: Financial Risk Components, Rating Analysis, Models, Economic and Regulatory Capital*. Oxford, UK: Oxford University Press. <https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199545117.001.0001/acprof-9780199545117>
- Viitanen, J., Koivistoinen, A., Loikkanen, V., & Paajanen, R. (2017). Digitaalisen alustatalouden tiekartasto. Helsinki: Valtioneuvoston kanslia, Työ- ja elinkeinoministeriö, Innovaatorahoituskeskus Business Finland. [https://www.businessfinland.fi/globalassets/julkaisut/alustatalouden\\_tiekartasto\\_web\\_x.pdf](https://www.businessfinland.fi/globalassets/julkaisut/alustatalouden_tiekartasto_web_x.pdf)
- Ward, J. S., & Barker, A. (2013). *Undefined By Data: A Survey of Big Data Definitions*. [Internet]. Ithaca, NY: Cornell University Library. <https://arxiv.org/abs/1309.5821v1>