

Samuli Suutari

# DATA PALVELUNA -MALLI



JYVÄSKYLÄN YLIOPISTO  
TIETOJENKÄSITTELYTIETEIDEN LAITOS  
2020

## TIIVISTELMÄ

Suutari, Samuli

Data palveluna -malli

Jyväskylä: Jyväskylän yliopisto, 2020, 26 s.

Tietojärjestelmätiede, kandidaatintutkielma

Ohjaaja: Seppänen, Ville

Data palveluna on pilvipalvelumalli, joka pohjautuu pilvilaskennan arkkitehtuuriin. Data palveluna -mallin tavoitteena on useita datalähteitä käyttäen toimittaa loppukäyttäjälle tarkoitukseen sopivaa dataa skaalautuvasti ja riippumatta loppukäyttäjän käyttämistä laitteista tai maantieteellisestä sijainnista. Tämän tutkielman tarkoituksena oli esitellä pilvilaskenta ja sen yleisimmät palvelumallit. Lisäksi tarkoituksena oli käsitellä data palveluna-mallia, sen ominaisuuksia, käyttökohteita ja yleisimpiä haasteita. Tutkielmaa voidaan käyttää pohjana mallin tutkimiseen tulevaisuudessa, ja se toteutettiin kirjallisuuskatsauksen keinoin. Tutkielmaa tehdessä kävi ilmi, että data palveluna- malli on ominaisuuksiltaan ja eduiltaan tunnetumpien pilvipalvelumallien kaltainen. Mallin käyttöön liittyy kuitenkin haasteita, ja etenkin tietosuojan ja datan laatuun liittyviin kysymyksiin tulee löytää vastaukset, ennen kuin mallin käyttö voi yleistyä muiden pilvipalvelumallien mukaisesti. Data palveluna -mallin hyödyt liittyvät sen kykyyn varastoida ja käsitellä useissa eri muodoissa olevaa ja useista eri datalähteistä tuotettua dataa, mutta samalla palveluratkaisujen toimittajien on kyettävä varmistamaan, että loppukäyttäjälle toimitettava valmis data on tietosuojastandardien mukaista. Datan tuotannon jatkuva kasvu asettaa myös palveluntarjoajille haasteita, sillä kilpailukykyisen palvelun pitää pystyä varmistamaan lopputuotteen datan laatu.

Asiasanat: pilvilaskenta, pilvipalvelut, data palveluna

## ABSTRACT

Suutari, Samuli

Data palveluna -malli

Data as a service

Jyväskylä: University of Jyväskylä, 2020, 26 p.

Information Systems Science, Bachelor's thesis

Supervisor: Seppänen, Ville

Data as a Service is a cloud computing service model based on the cloud computing architecture. The intends to combine and manage data based on several different data sources in order to provide the end user with a data product for a specific purpose that is scalable and accessible anywhere and on any device. The aim of this thesis was to present the concept of cloud computing and it's most service models. In addition, it was intended to introduce the Data as a Service model, it's features, use scenarios and the biggest challenges related to it's use. This thesis can be used as a basis for future research of the subject in the future, and it was conducted as a literature review. While conducting the thesis, it became apparent that the features and advantages of the Data as a Service model are similar to those of the other cloud computing service models. However, there are several challenges related to use of the service model before it can become more common, the most important being questions related to privacy and data quality. The advantages of the model are its capability to store, combine and manage data from several different data sources and formats. At the same time, the service providers need to be able to ensure that all privacy regulations are followed. The continuous expansion of data production is also a challenge to service providers, and in order to compete, the data quality challenges need to be addressed.

Keywords: cloud computing, cloud services, data as a service

## KUVIOT

Kuvio 1 - Pilvilaskennan arkkitehtuuri .....	10
Kuvio 2 - Elinvoimainen arvoketju .....	17

## TAULUKOT

Taulukko 1 - Datun laadun indikaattorit.....	20
--	----

## SISÄLLYS

1	JOHDANTO.....	6
2	PILVILASKENTA .....	8
2.1	Pilvilaskennan käsite.....	8
2.2	Pilvilaskennan arkkitehtuuri .....	9
2.3	Pilvilaskennan ominaisuudet.....	10
2.4	Pilvilaskennan hyödyt .....	11
2.5	Pilvilaskennan palvelumallit .....	12
2.5.1	Infrastruktuuri palveluna -malli .....	13
2.5.2	Alusta palveluna -malli .....	13
2.5.3	Sovellus palveluna -malli.....	14
3	DATA PALVELUNA.....	16
3.1	Palvelumalli.....	16
3.2	Käyttökohteet.....	17
3.3	Tietosuoja.....	18
3.4	Datan laatu.....	19
4	YHTEENVETO .....	21
5	LÄHTEET .....	23

# 1 JOHDANTO

*Pilvilaskennalla* tarkoitetaan laskentatehokkuuden ja erilaisten sovellusten tarjoamista palveluna tietoverkkojen välityksellä (Ojala & Tyrväinen, 2012). Pilvilaskennan ketteryys, skaalautuvuus ja vaivattomuus ovat mahdollistaneet maantieteellisestä sijainnista riippumattoman ja tehokkaan virtuaalisen tietoteknisen perusrakenteen synnyn, jossa kaikkia sen osia voidaan käsitellä lähtökohtaisesta palveluna (Tsai, Sun & Balasooriya, 2010). Tämän perusrakenteen ja ajatuksen pohjalle on muodostunut laaja kirjo erilaisia palvelukeskeisiä sovellutuksia aina tiedon varastoinnista kehitysalustoihin (Youseff, Butrico & Da Silva, 2008).

Pilvipohjaisesta arkkitehtuurista on sen kehityskaaren aikana esitetty useita erilaisia teorioita ja malleja sekä akateemisten, että teollisten vaikuttajien toimesta, ja useat niistä jakavat keskenään tiettyjä pääpiirteitä. Pilvilaskennan toiminta perustuu datakeskuksiin, jotka toimivat fyysisenä alustana pilvipohjaiselle arkkitehtuurille. Näiden keskusten pohjalta on rakennettu arkkitehtuuri, jonka kolmena tasona pidetään *infrastruktuuri palveluna* -mallia (eng. Infrastructure as a Service, IaaS), *alusta palveluna* -mallia (eng. Platform as a Service, PaaS) ja *sovellus palveluna* -mallia (eng. Software as a Service, SaaS) (Tsai, Sun & Balasooriya, 2010).

Näiden pilvipohjaisen arkkitehtuurin perinteisempien palvelumallien lisäksi lähivuosina pilvipohjaiseen arkkitehtuuriin on ilmestynyt useita uudempia ja vähemmän tutkittuja käsitteitä. Tässä tutkielmassa keskitytään *data palveluna* -malliin (eng. Data as a Service, DaaS), joka perinteisten palvelumallien arkkitehtuurin hierarkiassa sijoittuu samalle tasolle SaaS:n kanssa (Rajesh, Swapna & Reddy, 2012). DaaS on perusominaisuuksiltaan samankaltainen muiden aiemmin mainittujen palvelumallien kanssa. Laadukasta ja helposti hallinnoitavaa dataa voidaan tarjota käyttäjälle, milloin vain ja missä vain, ketterästi ja kustannustehokkaasti (Rajesh, Swapna & Reddy, 2012). Tulevaisuuden kannalta mielenkiintoisia ovat myös DaaS:n kohtaamat haasteet, jotka kulkevat osittain käsi kädessä massadataan (eng. Big Data) liittyvien ajankohtaisten haasteiden kanssa johtuen palvelun kautta tarjottavan datan massiivisesta määrästä (Terzo ym., 2013). Tämän lisäksi palvelukonseptiin liittyvät vahvasti datan laatuun, käyttöön,

lähteisiin ja hallintaan liittyvät kysymykset ja ongelmat ja ansaintamalleihin liittyvät vastaamattomat kysymykset (Truong & Dustdar, 2009).

Tämä tutkielma käsittelee pilvilaskentaa, ja DaaS-palvelumallia, ja tarkastelee DaaS-mallin sijoittumista pilvilaskennan arkkitehtuuriin. Tutkielma pyrkii myös esittelemään malliin liittyviä haasteita ja niiden ratkaisuehdotuksia. Tutkimuskysymys on:

- Mitkä haasteet vaikuttavat DaaS:n yleistymiseen ja käyttöön?

Tutkielma on toteutettu kirjallisuuskatsauksena, ja lähteenä on käytetty ensisijaisesti Google Scholar-, IEEE explore, ja Springer LNCS -palveluiden kautta etsityjä pilvilaskentaa, pilviarkkitehtuuriin ja data palvelumalliin liittyviä, useissa artikkeleissa viitattuja ja luotettavia julkaisuja. Tutkielmassa käytetyn kirjallisuuden löytämiseksi on käytetty muun muassa seuraavia hakusanoja ja niiden yhdistelmiä: *cloud computing, cloud architecture, data as a service daas, data quality, data privacy, daas*. Tutkielmassa käytetty kirjallisuus koostuu pääosin tietojärjestelmätieteen tutkimuksista ja konferenssijulkaisuista. Kirjallisuuskatsauksessa on pyritty käyttämään tieteellisiä julkaisuja, joiden Julkaisufoorumi-luokka on vähintään tasolla 1.

Tutkielman rakenne muodostuu kahdesta luvusta, joista ensimmäinen sisältää pilvilaskennan määrittelyn ja sen yleisimpien palvelumallien esittelyn. Toisessa luvussa esitellään DaaS-malli hyödyntäen ensimmäisen luvun pilvilaskennan esittelyä, ja tarkastellaan lisäksi palvelumallin käyttökohteita ja suurimpia haasteita. Tutkielman lopussa on yhteenveto.

## 2 PILVILASKENTA

Tässä luvussa tutustutaan pilvilaskentaan määrittelemällä yksityiskohtaisesti pilvilaskennan käsite ja pilvipalveluiden tarkemmat ominaisuudet. Tämän lisäksi luvussa esitellään pilvilaskennan perinteiset palvelumallit kerroksittain sekä luodaan katsaus uudempiin ja vähemmän tutkittuihin palvelumalleihin. Tarkoituksena on luoda pohja DaaS-mallin esittelyllä myöhemmin tässä tutkielmassa.

### 2.1 Pilvilaskennan käsite

Pilvilaskennalla tarkoitetaan sekä palveluiden tarjoamista ja toimittamista internetin välityksellä, että niitä ylläpitävien datakeskusten laitteiston ja järjestelmäsovellusten muodostamaa pilvi-infrastruktuuria (Armbrust ym., 2010). Dillon, Wu ja Chang (2010) toteavat tutkimuksessaan, että pilvilaskennalle on luotu useita eri määritelmiä, mutta päätyvät NIST:n (National Institute of Standards and Technology) määritelmään, jonka mukaan pilvilaskenta on malli, joka mahdollistaa vaivattoman ja tarpeen mukaan käytössä olevan tietoverkkoyhteyden jaettuun resurssikokoelmaan. Tämän resurssikokoelman provisiointi ja käyttöönotto on näin ollen mahdollista nopeasti ja mahdollisimman pienellä ylläpidolla (Dillon ym., 2010). Resurssi käsitteenä määritellään tässä tutkielmassa omassa luvussaan.

Pilvipalveluilla sen sijaan viitataan tässä tutkielmassa yleisesti kaikkiin pilvilaskennan ympäristössä tarjottaviin palveluihin. Pilvilaskennan yleistymisen, ja sen pilvipalveluiden suosion kasvun tärkeimpinä taustavoimina voidaan pitää muun muassa laadukkaiden internetyhteyksien määrän ja saatavuuden paranemista, langattomien yhteyksien yleistymistä, datan varastoinnin kustannusten laskua ja tietoteknisten järjestelmien kehittymistä. (Dikaiakos ym., 2009.)

Pilvi-käsite sisältää useita erilaisia pilvityyppejä, jotka mahdollistavat uniikkeja hyötyjä ja käyttötarkoituksia. *Julkisessa pilvessä* palveluntarjoajat välit-



tävät pilvipalveluita yleisesti mille tahansa loppukäyttäjälle. Julkista pilveä tarjoavat tahot hyötyvät sen käyttöönoton helppoudesta ja halpuudesta, mutta johdun tarkan kontrollin mahdollisuuden puutteesta yritykset kokevat sen usein kelpaamattomaksi liiketoiminnan harjoittamiseen. *Yksityinen pilvi* on tarkoitettu yhden yksittäisen yrityksen tai tahon käyttöön, ja sen ylläpito on mahdollista joko sisäisesti tai ulkoistettuna. Yksityisen pilven etuina on suurin mahdollinen kontrolli, joka takaa muun muassa korkean tietoturvan tason, mutta samalla menetetään tiukan kontrollin vuoksi osa pilville yleisesti tunnetuista hyödyistä. *Hybridipilvet* ovat kahden edellä mainitun pilvityypin sekoituksia, jotka yhdistämällä molempia tyyppisiä pyrkivät ratkaisemaan niitä heikkouksia, joita niiden käyttö yksittäin aiheuttaa. Käytännössä tämä tarkoittaa sitä, että osa palvelusta on käytössä yksityisenä ja tarkasti kontrolloituna, ja osa puolestaan julkisessa pilvessä pyrkimyksenä hyötyä sen eduista. Neljäntenä tyyppinä voidaan pitää *virtuaalista yksityistä pilveä*, joka on julkisen pilven pohjalla toimiva alusta. Sitä voidaan pitää kokonaisvaltaisempina pakettina, sillä se virtualisoi tarvittaessa sovellusten lisäksi myös käytettävissä olevan tietoverkon, sekä mahdollistaa käyttäjälle vapauden konfiguroida turvallisuusasetukset oman käyttötarpeensa mukaisiksi. (Zhang ym., 2010.)

Virtualisoinnilla tarkoitetaan pilvipalveluna toimitettujen resurssien dynaamista allokointia niiden palveluiden tai prosessien tarpeisiin, jotka niitä käyttökäytännöllä eniten tarvitsevat (Lombardi & Di Pietro, 2009). Käytännön esimerkkinä tästä voidaan nähdä esimerkiksi virtuaalityöasemilla sijaitsevat, tiettyä palvelua ylläpitävät palvelimet, joiden verkkoresursseja voidaan kasvattaa tai vähentää palvelimen kuormituksen mukaan. Yksi fyysinen työasema voi ylläpitää useita virtualisoituja ympäristöjä, joiden laskentatehoa ja resursseja voidaan asettaa eri kerroksissa toimivien palveluiden tarpeiden mukaan. (Lombardi & Di Pietro, 2009).

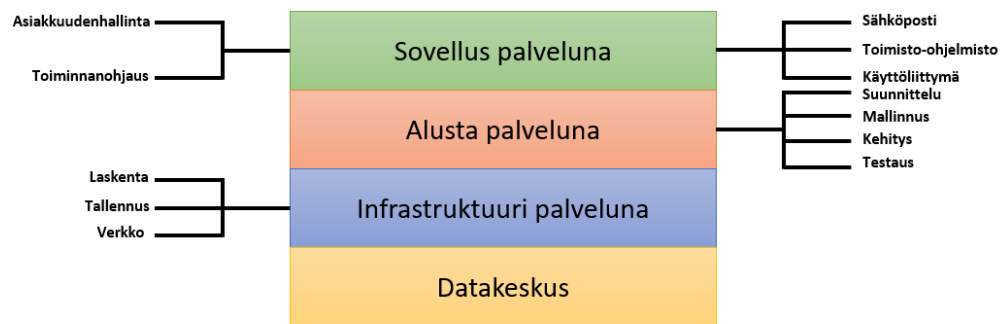
Koska useat pilvipalvelut perustuvat nimenomaan resurssien vuokraamiselle asiakkaan käyttöön, on resurssin käsite hyvä määritellä. Manvin ja Shyamin (2014) mukaan resurssi on mikä tahansa määrältään rajallinen tietokonejärjestelmän fyysinen tai virtuaalinen komponentti. Yleisenä esimerkkinä resurssista voidaan pitää infrastruktuuria tarjoavan tahon palvelinta, joka vuokrataan asiakkaan käyttöön niin, että asiakas voi asentaa palvelimelle käyttämänsä virtuaaliympäristön (Prodan & Ostermann, 2009). Muita yleisesti pilvipalveluina vuokrattuja fyysisiä resursseja ovat muun muassa suorittimet, muistit, varastot sekä tietoverkkojen osat kuten hubit ja reitittimet (Manvi & Shyam, 2014).

## 2.2 Pilvilaskennan arkkitehtuuri

Yleisesti puhuttaessa pilvilaskennan arkkitehtuurista, jaetaan se neljään eri tasoon. Alimpana tasona pidetään *laitteistotasoa*, joka luo pohjan pilven fyysisten resurssien hallinnalle. Laitteistotasoa ylläpidetään yleisesti tuhansista palvelimista koostuvissa *datakeskuksissa*. Laitteiston jälkeen seuraava taso on *infrastruk-*

*tuuritaso*, joka vastaa resurssien kokoamisesta *virtualisoinnissa* käytettäviksi kokonaisuuksiksi, ja mahdollistaa infrastruktuurin tarjoamisen palveluna. Infrastruktuuritason päälle on muodostunut *alustataso*, joka koostuu muun muassa käyttöjärjestelmistä. Sen tarkoituksena on toimia alustana pilvessä käytettäville sovelluksille ja toimia niin sanottuna välittäjätasona, joka edistää käytettävyyttä. Ylin taso pilviarkkitehtuurissa on *sovellustaso*, joka koostuu pilvisovelluksista. Kaikki arkkitehtuurin tasot ovat yhteydessä niiden kanssa kosketuspinnassa oleviin tasoihin, mutta ovat tästä huolimatta kykeneviä toimimaan yksilöinä, joka mahdollistaa niiden kehityksen omina moduuleinaan. (Zhang ym., 2010.)

Näillä tasoilla toimivista pilvipalveluista kolme parhaiten tunnettua ovat IaaS, PaaS, SaaS. Pilvilaskenta ja sen arkkitehtuuri jakavat paljon ominaisuuksia palvelukeskeisen arkkitehtuurin mallin kanssa. Palvelukeskeisen arkkitehtuurin tarkoituksena on luoda ratkaisuja, joissa palveluita kehittävät tahot voivat uudelleen käyttää jo aiemmin luomiaan sovelluksia ja niiden osia hyödykseen luodessaan uusia palveluita. Helpottaakseen näiden ratkaisuiden ja palveluiden luomista, yritykset käyttävät usein kehitysalustoissaan hyödykseen pilvilaskennan joustavuutta ja ketteryyttä. (Tsai ym., 2010.)



## Kaikki palveluna

Kuvio 1 - Pilvilaskennan arkkitehtuuri (Tsai ym., 2010)

## 2.3 Pilvilaskennan ominaisuudet

Pilvipalveluiden ominaisuuksien määritelmiä löytyy runsaasti. Dillon, Wu & Chang (2010) esittelevät yleisesti käytössä olevan NIST:n määritelmän mukaiset pilvilaskennan viisi ominaispiirrettä:

- *Tarvittaessa käytössä oleva palvelu*. Palvelu on mahdollista käynnistää automaattisesti asiakkaan tarpeen mukaan. Käytännössä tämä tarkoittaa sitä, että palveluiden käyttöönotto ei vaadi ihmisten välistä kanssakäymistä, palvelu käynnistyy niin sanotusti itsepalveluna.

- *Laaja verkkoyhteys.* Resurssit ja palvelut voidaan tarjota käyttäjälle tietoverkon välityksellä, ja niitä on mahdollista käyttää useilla eri alustoilla ja laitteilla, jotka ovat asiakkaan käytössä.
- *Resurssien yhdistäminen.* Palveluntarjoajan resurssit on kerätty yhteen, ja niitä on mahdollista tarjota useille käyttäjille samanaikaisesti esimerkiksi käyttämällä virtualisaatiota. Yhdistetyt resurssit ovat käyttäjälle näkymättömiä, eikä niiden sijainti ole julkisesti esillä.
- *Nopea elastisuus.* Asiakkaan käytössä olevien resurssien määrä on muutettavissa käyttäjän toimesta välittömästi. Asiakkaan käyttöön on mahdollista provisoida resursseja käytön mukaan, eikä asiakkaan tarvitse etukäteen varautua ylimääräiseen resurssitarpeeseen.
- *Mitattava palvelu.* Resurssien käytön määrää pystytään mittaamaan käyttäjäkohtaisesti huolimatta siitä, että resurssit ovat usein monen eri tahon käytössä samanaikaisesti. (Dillon ym., 2010.)

Gong ym. (2010) listaavat tutkimuksessaan pilvilaskennan yleisimpiä ominaisuuksia, joista he pitävät *palvelukeskeisyyttä* (eng. service oriented), *löyhyyttä* (eng. loose coupling), *vahvaa vikasietoisuutta* (eng. strong fault tolerance), pilvilaskennan liiketoimintamallia ja helppokäyttöisyyttä niistä tärkeimpinä. Palvelukeskeisyydellä tarkoitetaan palvelun käytön esteettömyyttä, jonka myötä käyttäjän on mahdollista käyttää pilvipalveluita tuntematta sen arkkitehtuuria. Löyhyys mahdollistaa sen, että pilvessä olevat infrastruktuurit ovat erillään toisistaan eivätkä niiden sisällä tapahtuvat toiminnot vaikuta ympärillä oleviin osiin. Vahva vikasietoisuus liittyy edellä mainittuun löyhyyteen siten, että erillään toimivien pilven osien sisällä tapahtuvat vikatilanteet eivät yleensä vaikuta niitä ympäröivään infrastruktuuriin. Pilvijärjestelmän ominaisuuksiin kuuluu myös mahdollisuus palata niin sanottuun palautuspisteeseen vikatilanteen sattuessa. Pilvilaskennan liiketoimintamallit perustuvat hyvin usein siihen, että käyttäjät maksavat pilvipalveluista käytön mukaista hintaa, vaikkakin useiden tunnettujen pilvipalveluiden kuten perusominaisuuksia tukevat versiot ovat käyttäjälle ilmaisia. Viimeinen niin sanotuista tärkeistä ominaisuuksista on helppokäyttöisyys. Pilvipalvelut ovat järjestään helppokäyttöisiä, sillä niiden tarkoitus on olla mahdollisimman käyttäjäystävällisiä vähentääkseen käyttöönoton kynnystä. (Gong ym., 2010.)

## 2.4 Pilvilaskennan hyödyt

Aiemmin mainittujen tekijöiden lisäksi pilvilaskennan yleistymiselle on löydetävissä sekä taloudellisia, että teknisiä syitä (Aymerich ym., 2008). Selvyiden vuoksi on tässä vaiheessa hyvä erottaa toisistaan pilvilaskennan palveluntarjoajien kaksi eri tasoa. Pilvi-arkkitehtuurissa niin sanotusti ylempänä tahona ovat

infrastruktuurintarjoajat, jotka tarjoavat pilvi-infrastruktuureja palveluntarjoajille, jotka puolestaan toimittavat näiden infrastruktuurien pohjalle rakennettuja palveluita loppukäyttäjille (Zhang ym., 2010).

Pilvi-infrastruktuureja ylläpitävät ja laskentatehoa tarjoavat yritykset hyötyvät arkkitehtuurien skaalautuvuudesta ja siitä, että ne mahdollistavat esimerkiksi massiivisten datamäärien säilytyksen ja hallinnan (Grossman, 2009). Pilviarkkitehtuurit selviävät tehokkaan skaalautuvuutensa takia hetkittäisistä rasiustilanteista ilman ylimääräistä laitteistoa (Aymerich, 2008). Skaalautuvuus mahdollistaa myös uusien yritysten tulon markkinoille, sillä käytössä olevien resurssien määrää on helppo kasvattaa toiminnan laajentuessa (Grossman, 2009).

Palveluntarjoajien näkökulmasta pilvipalveluiden arvoa lisäävät usein myös niiden helppokäyttöisyys ja saavutettavuus maantieteellisestä sijainnista riippumatta. Pilvipalveluina tarjotut palvelut ovat usein web-pohjaisia, joka mahdollistaa niiden käytön lähes kaikilla laitteilla, jotka on mahdollista liittää internettiin, ja usein myös langattomasti. Teknisten ongelmien luomat riskit ovat usein helpommin ja ammattitaitoisemmin hoidettavissa pilvipalveluiden infrastruktuuria ylläpitävän tahon puolesta, jolloin palveluntarjoaja niin sanotusti ulkoistaa tekniset riskinsä. Tämän lisäksi palveluntarjoajat hyötyvät muun muassa pilvipalveluille ominaisesta mahdollisuudesta sijoittaa vähemmän kuormitettujen palveluiden laskentaresursseja niihin palveluihin, jotka niitä sillä hetkellä eniten vaativat. (Zhang ym. 2010).

Pilvipalveluiden käytön etuja voidaan tarkastella myös loppukäyttäjän näkökulmasta. Käyttöönoton helppous ja kustannustehokkuus tulevat tärkeinä ominaisuuksina esiin etenkin pienten ja keskisuurten yritysten kohdalla, joille esimerkiksi mittavat perinteisin keinoin tehdyt ohjelmistohankinnat saattavat olla suuri haaste toiminnan alkuvaiheessa (Kourik, 2011). Pilvipalveluina hankitut sovellukset ovat käytettävissä ilman, että loppukäyttäjä on sidottu tietyn tyyppiseen laitteeseen ja niiden päivittäminen jokaiselle laitteelle onnistuu käytön yhteydessä. (Aymerich, 2008). Taloudellisuuden kannalta tärkeää on se, että yritykset voivat säästää huomattavasti vuokraamalla aktiiviseen käyttöönsä vain sillä hetkellä tarvitsemansa resurssit siitä huolimatta, että tarvittaessa niiden määrää on mahdollista kasvattaa milloin tahansa (Kourik, 2011). Kuten palveluntarjoajat, myös loppukäyttäjät hyötyvät siitä, että mahdolliset viat ja tekniset ongelmat ovat palveluita tarjoavan tahon vastuulla, eivätkä esimerkiksi mahdolliset sovelusten yhteensopivuusongelmat ole käyttäjän vastuulla (Aymerich, 2008).

## 2.5 Pilvilaskennan palvelumallit

Tässä kappaleessa käsitellään pilvilaskennan perinteiset palvelumallit. Tarkoituksena on selvittää jokaiselle mallille tärkeimpiä ominaispiirteitä aiemmin esiteltujen pilvipalveluiden yleisten ominaisuuksien lisäksi. Tavoitteena on myös luoda pohja palvelumallien väliselle vertailulle myöhemmin tässä tutkielmassa.

### 2.5.1 Infrastrukturi palveluna -malli

Pilviarkkitehtuurin mallien kerroksittaisessa luokittelussa IaaS on niin sanotusti pohjataso, jonka pohjalle muut palvelumallit rakentuvat. Sen yleisin käyttötarkoitus on virtuaalisten resurssien vuokraaminen, joka käytännössä tarkoittaa usein asiakkaan käyttöön räätälöityjen virtuaalikoneiden ylläpitoa (Villegas ym., 2012). Toisin sanoen tarkoituksena on tarjota asiakkaan tarvitsema määrä laskentatehoa ja perustason palveluita käytettävän ympäristön luomiseksi (Manvi & Shyam, 2014). Näiden palveluiden perustana toimivat usein datakeskukset, joiden kautta virtuaalisia resursseja ja laitteistoa vuokrataan asiakkaiden käyttöön (Prodan & Ostermann, 2009).

IaaS mahdollistaa resurssien vuokraamisen usealle eri käyttäjälle samanaikaisesti ja niin, että yhden tahon käytöstä vapauttamat resurssit voidaan helposti ja nopea siirtää tarvittaessa vastaamaan toisen käyttäjän tarpeita (Villegas ym., 2012). Fyysisiä resursseja ohjaamaan voidaan myös vuokrata niin sanottuja loogisia resursseja, kuten käyttöjärjestelmiä ja protokollia, jotka usein tukevat asiakkaan tarpeita infrastruktuurin tyyppin suhteen (Manvi & Shyam, 2014). Kaupallisista IaaS-palveluntarjoajista suurimpia ovat muun muassa Amazon, Rackspace, Microsoft (Jackson ym., 2010), IBM ja Google (Chang ym., 2010). Amazon Web Services (AWS) oli vuonna 2015 tehdyn tutkimuksen mukaan palveluntarjoajien markkinajohtaja, jonka asiakkaihin sisältyy muun muassa Yhdysvaltojen hallitus. (Gartner, 2015). AWS tarjoaa käyttäjilleen myös PaaS-palveluita sekä markkina- paikan, joka mahdollistaa lisäpalveluiden ja kolmannen osapuolen palveluiden hankintaan (Gartner, 2015).

Toosi ym. (2011) mukaan puhuttaessa infrastruktuurin tarjoamisesta palveluna on pilvilaskennan mahdollistama usean asiakkaan palveleminen samoilla resursseilla yhdenaikaisesti yksi sen pääominaisuuksista. Pystyäkseen maksimoimaan voittonsa on palveluntarjoajan pystyttävä varmistamaan palveluiden laatu sen lisäksi, että palvelu on oltava käytettävissä mahdollisimman monella asiakkaalla. Onnistuakseen tässä, IaaS-palveluntarjoajat joutuvat huolellisesti suunnittelemaan resurssienhallintansa. Yksi yleisimmistä strategioista on pitää ennalta päätettyä määrää vapaita resursseja koko ajan valmiudessa. Joissain tapauksissa palveluntarjoajan kannalta kannattavinta on hankkia nämä niin sanotut valmiudessa olevat resurssit kilpailijaltaan. (Toosi ym., 2011.)

### 2.5.2 Alusta palveluna -malli

Palvelumallien arkkitehtuurissa infrastruktuuritason päälle sijoittuu niin kutsuttu alustataso eli PaaS. Beimborn ym. (2011) mukaan PaaS on usein se taso, joka jää käyttäjän kannalta pimentoon johtuen sen roolista mahdollistavana alustana palveluina toimitetuille sovelluksille. Sen käyttäjäkunta koostuu kehittäjistä, joiden on mahdollista palvelualustaa ja sen joustavuutta hyväksi käyttäen kehittää ja suorittaa sovelluksiaan (Zeng & Xu, 2010). Etuna perinteiseen kehitystyöhön verrattuna voidaan pitää sitä, että kaikki sovelluskehityksen vaiheet suunnittelusta ylläpitoon asti voidaan hoitaa alustapalvelun tarjoaman sovelluspaketin

kautta niin, että käyttäjän ei tarvitse itse hankkia jokaista työkalua erikseen (Jadeja & Modi, 2012). Lisäksi Zeng ja Xu (2010) listaavat tutkimuksessaan PaaS:n etuja:

- Alustalla kehitettyjä sovelluksia on mahdollista käyttää ilman niiden asentamista tai lataamista.
- Kehittäjien välinen yhteistyö helpottuu.
- Muita tarpeellisia web-pohjaisia palveluita on tarvittaessa mahdollista integroida alustaan.
- Alusta on skaalautuva, luotettava ja turvallinen.
- Kehitetyn palvelun käyttäjien toimintoja on mahdollista seurata ja analysoida.
- Alustapalvelusta maksetaan käytön mukaan. (Zeng & Xu, 2010)

PaaS:n niin sanottu pääkomponentti on *ajonaikainen käyttöjärjestelmä* (eng. Application Runtime Environment, ARE), jonka tehtävänä on tarjota kehittäjälle alustalle tyypilliset ominaisuudet kuten skaalautuvuuden ja tietoturvan. Tämän lisäksi alusta sisältää usein myös *ohjelmistoympäristön* (eng. Integrated Development Environment, IDE). IDE sisältää tuen yleisempiin ohjelmointikieliin ja tarjoaa kehittäjän käyttöön mittavan määrän työkaluja sekä tallennuspalveluita. Kehittäjien kannalta yksi suurista alustapalveluiden eduista on myös se, että osa alustoista tarjoaa sovelluskehitystyökalujen lisäksi kanavan myös muita lisäpalveluita. Hyvänä esimerkkinä tästä on Google App Marketplace, joka tarjoaa kehittäjälle niin sanottujen runkopalveluiden lisäksi myös kanavat markkinointiin ja myyntiin. (Beimborn ym., 2010).

### 2.5.3 Sovellus palveluna -malli

Pilvipalveluna toimitettu sovellus palveluna on nopeasti kasvava malli, jonka suurimpana erona perinteiseen sovellusten lisensointimalliin voidaan pitää sitä, että asiakas ostaa SaaS-tuottajalta ns. tilauksen sovelluksen käyttöön normaalin, sovelluspaketin lisenssin sijaan (Choudhary, 2007). Toisin sanoen asiakas saa käyttöönsä selainpohjaisen, usealla eri alustalla käytettävissä olevan tuotteen, joka ei vaadi esimerkiksi yksittäiseen ohjelmistopakettiin sidonnaista lisenssiä toimiakseen. Tällaisia sovellusratkaisuja kutsutaan myös web-pohjaisiksi sovelluksiksi. Sun ym. (2007) mukaan SaaS-tuotteet kattavat yleisimmät yritysten sovellustarpeet esimerkiksi *henkilöstöhallinnon järjestelmistä* (eng. Human Resource Management, HRM) *asiakkuudenhallinnan sovelluksiin* (eng. Customer Relationship Management, CRM). Yksi SaaS:lle ominaisista piirteistä on se, että pilvipalveluna tarjottu ohjelmistotuote on yleensä web-pohjainen, ja palvelun resurssointi ja ylläpito tapahtuu palvelua tarjoavan tahon palvelimella (Godse & Mulik, 2009). SaaS-pohjaisille ratkaisuille on yleistä se, että ne perinteisestä lisenssipohjaisesta mallista poiketen päivittyvät säännöllisesti ja usein, ja uudet versiot ovat

heti asiakkaiden käytettävissä sen sijaan että sovelluspäivitykset olisivat suuria, ja ne julkaistaisiin harvoin kuten perinteisille lisensoituille sovelluksille on ominaista. (Choudhary, 2007).

SaaS-ratkaisut ovat usein ns. standardimallisia, eli toimittajan tarjoama ratkaisu on asiakkaasta riippumatta perusominaisuuksiltaan hyvin samanlainen. Web-pohjaiset tuotteet tarjoavat tästä syystä usein asiakkaalle mittavan määrän muokkausmahdollisuuksia, joka puolestaan mahdollistaa ratkaisun konfiguroinnin asiakkaan toimesta hyvin erilaisiin liiketoimintamalleihin ja liiketoiminta-alueisiin sopiviksi (Sun, Zhang, Guo, Sun & Su, 2008). Tämä on puolestaan mahdollistanut hyvin kattavan, niin kutsutun. partner-liiketoiminnan muodostumisen, jossa asiakkaiden keskuudessa yleistyneiden SaaS-ratkaisujen toimittajat ovat mahdollistaneet esimerkiksi erilaisten konsulttiyritysten toiminnan eräänlaisena korkeamman tason palvelu- tai tukiyhteistyökumppanina tarjoamalla tehokkaampia ja monipuolisempia työkaluja SaaS-ratkaisun konfigurointiin ja ylläpitoon. Esimerkkeinä tällaisista tuotteista ovat muun muassa Salesforce (CRM) ja Workday (HCM).

Kuten muissakin tässä mainituissa pilvipalvelumalleissa, voidaan SaaS:n etuna pitää sen skaalautuvuutta. SaaS:n kustannusrakenne muodostuu usein sekä aikamääreisestä tilausmallista, kuten esimerkiksi vuosittaisesta käyttömaksusta, sekä kertaluonteisesta käyttöönottomaksusta (Godse & Mulik, 2009). Choudhary (2007) esittää tutkimuksessaan mallin, jonka mukaan web-pohjaisten sovellusten laatu on perinteisten lisensoitujen sovellusten laatua korkeampi joutuen siitä, että toimittaja sijoittaa keskimäärin enemmän resursseja niiden jatkuvaan kehitykseen ja laadunhallintaan, mutta samalla toimittajan rahallinen hyöty kasvaa suhteessa asiakkaiden määrän kasvaessa.

### 3 DATA PALVELUNA

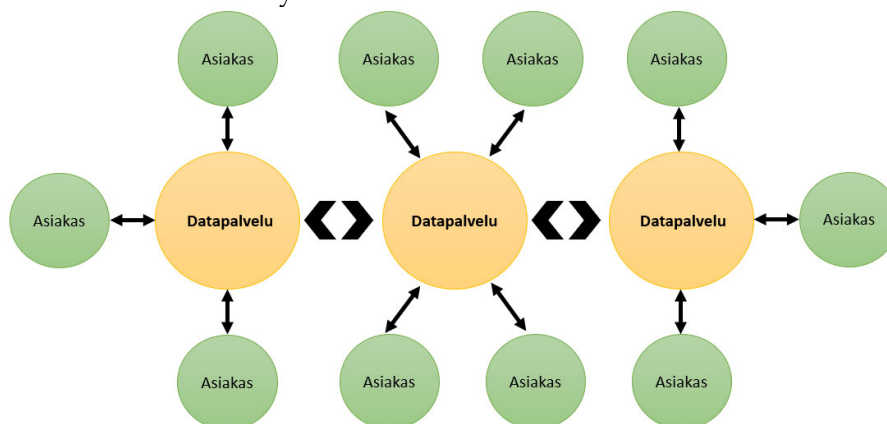
Tässä luvussa käsitellään data palveluna -mallia, palvelumallin käyttökohteita ja siihen liittyviä haasteita, sekä niihin kirjallisuudessa esitetyjä ratkaisuja. Mallin esittelyssä käytetään pohjana tämän tutkielman pilvilaskenta-lukua, sillä DaaS on perusominaisuuksiltaan muiden pilvilaskennan arkkitehtuurin palvelumallien kaltainen.

#### 3.1 Palvelumalli

Data palveluna on pilvipalvelumalli, jossa dataa myyvä tai vuokraava palveluntarjoaja toimittaa asiakkaalle käyttötarpeeseen räätälöityä, usein useista eri lähteistä tuotua dataa (Søilen, 2016). DaaS jakaa muiden, aikaisemmin yleistyneiden pilvipalveluiden positiiviset ominaisuudet. Toisin kuin muut tunnetummat pilvipalvelumallit, DaaS:n arvo on sen tuottamassa datassa, eikä se välttämättä vaadi rinnalle esimerkiksi olemassa olevaa alustaa tai sovellusta (Terzo ym., 2013). Tästä huolimatta DaaS-mallin ominaisuudet voidaan tietyissä tapauksissa mahdollista yhdistää esimerkiksi SaaS-pohjaisiin sovellusratkaisuihin. Poiketen perinteisistä lisensoiduista sovelluksista, jotka myytiin aiemmin osana sovelluspakettina sisältäen käyttöön tarvittavan datan, mahdollistavat jatkuvasti yleistyneet pilvipalveluna toimitetut sovellukset datan käsiteltäväksi tuomisen sovelluksen ulkopuolisista, pilvessä sijaitsevista lähteistä (Rajesh, Swapna & Reddy, 2012). Vu ym. (2012) mukaan dataa käyttävän tahon ei tarvitse enää säilöä massadataa omassa tietokannassaan tai omalla palvelimellaan, vaan dataa voidaan tarvittaessa hakea sillä hetkellä sitä hyödyntävien sovelluksien käyttöön *ohjelmointirajapintoja* (eng. Application Programming Interface, API) käyttäen. Palveluna toimitettu data on usein käyttövalmista, ja tiettyjä liiketoimintaan liittyviä päätöksiä tukevaa, eikä sen raaka versio tai muokkausprosessi ole asiakkaalle näkyvä (Søilen, 2016). Rajesh, Swapna & Reddy (2012) määrittelevät palveluna



toimitetun datan arvoketjun elinvoimaiseksi silloin kun asiakkaita on useita, datansiirto on jatkuvaa, lopputuote on käyttäjälle valmiiksi arvokasta ja tuotetut palveluratkaisut voivat täydentää toisiaan.



Kuvio 2 Elinvoimainen arvoketju (Rajesh, Swapna & Reddy, 2012)

### 3.2 Käyttökohteet

Palveluna toimitetun datan käytöllä voidaan yksinkertaisimmillaan tarkoittaa ennalta määritetyn, pilvi-infrastruktuurissa sijaitsevan käyttövalmiin datakoelman analysointia rajapintaa hyväksi käyttäen (Zheng, Zhu & Lyu, 2013). Tbahriti ym. (2011) mukaan ajoittain pelkkä tarvittavan ja tiettyä käyttötarkoitusta varten toimitetun datan saavuttaminen ilman laitteiden tai sijainnin rajoitteita voi itsessään olla asiakkaalle hyödyllistä, mutta usein lisäarvon tuottamista varten datan jatkokäsittely vaatii erillisten palveluiden, sovellusten tai prosessien hyödyntämistä. Esimerkkinä tästä voidaan pitää vertailudatan käyttöä liiketoimintaprosessien kehityksessä, jossa kaupallinen organisaatio ostaa pilvipalveluna toimitetun sovelluksen tarjoajalta oman organisaationsa liiketoiminta-alueelta vastaavien, samaa sovellusta käyttävien tahojen vertailukelpoista ja anonymisoitua dataa. Vertailuanalyysin tarkoituksena on vastata ennalta määritettyyn kysymykseen, joka asettaa vastakkain erilaisia vastaavia prosesseja tai lopputuotteita (Folkerts ym., 2012). Tosielämän tilanteessa yrityksellä voi olla esimerkiksi tarve vertailla pilvipalveluna toimitetun sovelluksen sisällä mallinnetun liiketoimintaprosessin tehokkuutta vertailukelpoisen organisaation vastaavaan prosessiin, ja mahdollisuuksien mukaan tunnistaa esimerkiksi optimoinnin kohteita.

Yksi palveluna toimitetun ja datan mahdollisista käyttökohteista on Internet of Things (IoT) ja sen ratkaisut. IoT-ratkaisujen tehokkuus perustuu keskenään yhteydessä oleviin laitteisiin, niiden kykyyn sulautua olemassaoleviin tietoverkkoihin sekä verkkoon liitettyjen laitteiden tuottamaan sensoridataan, jonka perusteella kokonaisuutta pystytään hallitsemaan (Gubbi, Buyya, Marusic & Palaniswami, 2013). Yhtenä IoT-ratkaisujen suurimmista haasteista pidetään

niiden integraatiota, sillä niiden osana olevat laitteet tuottavat usein dataa eri formaateissa, jolloin niihin liittyvät palvelut eivät oletusarvoisesti keskustele keskenään (Barros, Estrella, Prates & Bruschi, 2018). Myös IoT-ratkaisujen skaalan ja tietoverkkoon liitettyjen laitteiden määrän kasvaessa niiden tuottaman sensoridatan määrä kasvaa, jolloin perinteisten datapalveluiden sijaan DaaS tarjoaa skaalautuvamman ratkaisun datan säilytykseen, hallintaan ja toimittamiseen (Badidi, Routaib & El Koutbi, 2016). DaaS-palvelut tarjoavat lisäksi samalla datansäilytysalustalla toimivia ratkaisuja esimerkiksi eri sensoreista lähtöisin olevien keskenään erilaisten dataformaattien harmonisointiin, jolloin loppukäyttäjälle voidaan tarjota täysin hänen palveluunsa valmis dataratkaisu (Barros ym., 2018). Toisin kuin perinteisiä tietokantoja käyttäen, palveluntarjoaja pystyy myös allokoimaan datan säilöntään ja käsittelyyn tarvittavia resursseja IoT-ratkaisujen käyttöasteen ja niiden tuottaman datan mukaan, jolloin palveluiden käyttämät resurssit voidaan optimoida vapauttaen niitä käyttöasteen laskiessa muiden ratkaisujen käyttöön (Badidi, Routaib & El Koutbi, 2016).

### 3.3 Tietosuoja

Huolimatta siitä, että DaaS:n tietoturvaan ja tietosuojaan liittyvät ongelmat ja niiden ratkaisuehdotukset muodostavat huomattavan suuren osan palvelumallin olemassa olevasta kirjallisuudesta, ovat ne mallin nykytilanteen ymmärtämisen kannalta niin merkittäviä, että niiden käsittely tässä kirjallisuuskatsauksessa on mielekästä. Pystyäkseen tarjoamaan vaivattomia DaaS-ratkaisuita, on palveluntarjoajien pystyttävä ratkaisemaan mallin tietosuojaan liittyvät ongelmat (Barhamgi ym., 2011). Pelkästään relevantin datan saavutettavuus palvelumallin mukaisesti tuottaa ajoittain loppukäyttäjälle arvoa, mutta usein lisäarvoa tuottavan lopputuotteen muodostaminen sisältää useiden datalähteiden tuottaman datan yhtenäistämistä ja käsittelyä, jolloin prosessoitava kokonaisuus voi sisältää myös tietosuojattua dataa (Tbahriti ym., 2011).

Palveluna toimitetun datan käsittely vaatii usein dataa analysoivien ja muokkaavien työkalujen käyttöä, joista yhtenä tärkeänä pidetään anonymisointityökaluja (Truong & Dustdar, 2010). Anonymisointiprosessin tarkoituksena on muokata esimerkiksi arkaluontoista dataa tai niitä datan sisältämiä tietoja, jotka voivat johtaa yksittäisen henkilön tai entiteetin tunnistamiseen, mutta toimiakseen tehokkaasti se vaatii taustalle anonymisoinnin kohteita datassa tunnistavia prosesseja. Tbahriti, Ghedira, Medjahed ja Mrissa (2013) esittelevät julkaisussaan DaaS-rakenteeseen mallin, jonka tavoitteena on lisätä datankäsittelyprosessiin selkeät tietosuojakäytäntö- ja vaatimukset, jotka mahdollistavat tietosuojaan liittyvien ongelmien ratkaisemisen prosessin aikana. Tietosuojaan liittyvien ongelmien ratkaisemisen kannalta on tärkeää määritellä datalähteen omistajan ja DaaS-palveluntarjoajan roolien ja vastuiden erot, sekä tunnistaa datan omistajan ja käsittelijän tietosuojakäytäntöjen tarpeet (Mrissa, Tbahriti & Truong, 2010). Tutkimuksessaan Barhamgi ym. (2011) esittävät viitekehyksen, jonka periaat-

teena pidetään sitä, että ennen eri lähteistä tuodun datan automaattisen käsittelyprosessin jatkokehittämistä tehokkaammaksi, tulee myös käytetyille datalähteille määritellä yksilölliset tietosuojakäytännöt. DaaS-mallin yleistymisen kannalta voidaan pitää tärkeänä, että tietosuojaan liittyviä haasteita on käsitelty etenkin palvelumalliin liittyvässä varhaisessa kirjallisuudessa. Vuonna 2018 voimaan tullut EU:n yleinen tietosuoja-asetus selkeytti datan omistajan ja käsittelijän vastuuta, mutta tätä edeltäviä tietosuojakäytänteitä voidaan pitää tärkeinä palveluna toimitetun datan käytön yleisyyden takia jo ennen EU:n asetuksen voimaantuloa.

### 3.4 Datan laatu

Johtuen palveluna toimitetun datan määrästä, heterogeenisyydestä ja loppukäyttäjien korkeista laatuvaatimuksista, on DaaS-palveluntarjoajien pystyttävä sekä varmistamaan, että osoittamaan toimitettavan datan laatu (Badidi, Routaib & El Koutbi, 2016). Osa palveluna toimitetusta datasta on esimerkiksi sitä käyttävän sovelluksen vaatimusten takia strukturoimatonta, jolloin datan laatuun liittyvät ongelmat korostuvat entisestään (Pringle, Baer & Brown, 2014). Etenkin strukturoimattoman datan kohdalla datalähteeseen ja datan käsittelyyn liittyvä metadata on käyttäjälle tärkeä työkalu datan laadun määrittelyyn. DaaS-palveluiden käyttäjien haastatteluun perustuvassa tutkimuksessaan Søylen (2016) määrittelee datan laatuun ja luotettavuuteen liittyvät haasteet palvelumallin suurimmiksi ongelmiksi tietosuojan ja tietoturvan ohella. Palveluna toimitetun datan loppukäyttäjät ovat usein yritykset, jotka käyttävät dataa tukena toimintaansa liittyvissä päätöksissä, jolloin huonolaatuinen data voi johtaa liiketoiminnan kehityksen kannalta haitallisiin ratkaisuihin (Badidi, Routaib & El Koutbi, 2016).

DaaS-malli mahdollistaa perinteisistä tietokantaratkaisuista poiketen datan laadun keskitetyn hallinnan, jonka avulla esimerkiksi useista datalähteistä tuotavaa dataa on mahdollista yhtenäistää laadukkaan lopputuloksen saavuttamiseksi (Segura, Cuadrado & de Lara, 2014). Rajesh, Swapna & Reddy (2012) mukaan datan validointi ja laadunvarmistus ovat osa elinvoimaista datan arvoketjua. Palveluntarjoajaa valitessa kriteerinä käytetään usein arviota palvelukokonaisuudesta, mutta eri ratkaisujen mielekäs vertailu vaatii myös datan laadun yksityiskohtiin liittyvää tarkempaa määrittelyä (Truong & Dustdar 2009). Vuon. (2009) mukaan palveluna toimitettavalle datalle on tyypillistä, että eri palveluntarjoajat voivat tarjota vastaavia ratkaisuita tai jopa samaa täysin samaa dataa, mutta sekä palvelun että datan hinnassa ja laadussa voi olla eroja. DaaS-palveluiden hinnoittelun ja luokittelun kriteerinä käytetään usein palvelukokonaisuuden laatuluokittelua, joka ei ota kantaa itse datan sisältöön tai laatuun (Truong & Dustdar, 2009).

Vaikka palvelun ja datan laatua on käsitelty palvelumalliin liittyvässä kirjallisuudessa kattavasti, ei datan laadulle ole määritelty yhtä yhtenäistä mittaria, vaan datan laadun luotettava määrittely vaatii useamman laatua määrittävän mittayksikön käyttöä (Truong ym., 2011). Artikkelissaan Truong & Dustdar (2009)

määrittelevät palvelumallin keskeisimmäksi tekijäksi datan, jolloin asiakkaan olisi tärkeää pystyä erottamaan datan ja palvelun laadun kriteerit toisistaan voidakseen valita omaan käyttötarkoitukseensa sopivimman ratkaisun tai toimittajan. Kuten taulukosta 1 käy ilmi, Badidi, Routaib & El Koutbi (2016) esittelevät tutkimuksessaan kuusi indikaattoria, joita käyttämällä datan laatua voidaan arvioida

Taulukko 1 - Datan laadun indikaattorit (Badidi, Routaib & El Koutbi, 2016)

<b>Indikaattori</b>	<b>Määritelmä</b>
Tarkkuus	Toimitetun tietojoukon laajuus suhteessa kohdekokonaisuuteen
Avaruudellinen resoluutio	Datan käyttökohdealueen määrittelyn tarkkuus
Ajallinen resoluutio	Datayksikön oikeudenmukaisuuden ajallinen jakso
Tuoreus	Datan ikä sen muodostushetkestä toimitushetkeen
Virheettömyys	Datan paikkansapitävyys
Johdonmukaisuus	Datan instanssien vastaavuus tietojoukon eri osissa
Täydellisyys	Tarvittavien datakomponenttien olemassaolo

## 4 YHTEENVETO

Tässä kandidaatintutkielmassa on käsitelty kirjallisuuskatsauksen keinoin data palveluna -malli, joka on yksi uudemmissa palvelukeskeiseen arkkitehtuuriin perustuvista pilvilaskennan malleista. Tutkielmassa on myös selitetty pilvilaskennan käsite ja ominaisuudet, joiden ymmärtäminen on tärkeää DaaS-mallin käsittelyn kannalta. Datan tuotanto kasvaa jatkuvasti, ja yksi massadatan haasteista liittyy sen säilytykseen, hallintaan, harmonisointiin ja saavutettavuuteen. Tästä huolimatta datapalveluiden datan hallinnan malleja ei ole laajasti tutkittu, vaikka datan käyttö päätösten ja liiketoiminnan kehittämisen apuna on yleistynyt. Tämän tutkimuksen tarkoituksena on ollut esitellä data palveluna -malli, sen yleisimmät käyttökohteet sekä tuoda esiin siihen liittyviä haasteita. Apuna tutkimuksen tekemisessä on käytetty tutkimuskysymyksiä:

- Mikä data palveluna -malli on?
- Mitä pilvilaskenta on?
- Mitkä haasteet vaikuttavat DaaS:n yleistymiseen ja käyttöön?

Vaikka data palveluna -mallia ei ole tutkittu niin laajasti kuin muita pilvipalvelumalleja, löytyy aiheesta tarpeeksi aiempaa julkaistua tietoa, että kirjallisuuskatsauksen tekeminen on ollut mielekästä. Vu ym. (2012) mukaan huolimatta siitä, että DaaS:lla on jo omat markkinansa, ei käsitteestä ole muodostettu yksiselitteistä ja hyvin rakennettua mallia. Aiemmin mainitut massadatan haasteet tekevät DaaS:sta ajankohtaisen aiheen myös siksi, että se tarjoaa mahdollisia ratkaisuja datan käsittely- ja varastointiongelmiin, jotka tulee ratkaista, että sitä voidaan tulevaisuudessakin käyttää teknologisten innovaatioiden kehitykseen (Rajesh, Swapna & Reddy, 2012).

Aiheen tutkimuksen tarkoituksena on ollut esitellä data palveluna -malli, sen ominaisuudet, jotka perustuvat pilvipalveluarkkitehtuurin muihin ratkaisuihin, sekä mallin käyttöön ja yleistymiseen liittyvät haasteet. DaaS-mallin tutkiminen on ollut tärkeää, sillä poiketen muista pilvilaskennan jo aiemmin vakiintuneista malleista, kuten IaaS, PaaS ja SaaS, aihetta ei ole vielä käsitelty tarpeeksi

laajasti tieteellisen tutkimuksen keinoin, ja aiempi tutkimus painottuu suurelta osin DaaS:n mahdollisiin tietosuojongelmiin, joiden tarkempi tutkiminen on ollut yksi palvelumallin käytön yleistymiseen edellytyksistä. Tietosuojaan liittyvien ongelmien lisäksi on ollut tärkeää tutkia muita DaaS:n käytön haasteita, ja esitellä etenkin datan laadun hallintaan liittyviä kysymyksiä ja ratkaisuehdotuksia, sillä datapalveluiden arvioiminen ilman laadun arvioinnin malleja ei ole mielekästä. Datan laatuun liittyvät kysymykset ovat etenkin loppukäyttäjälle oleellisia palveluntarjoajia vertaillessa. Lisäksi on ollut oleellista tutkia pilvilaskennan arkkitehtuuria ja yleisimpiä pilvipalvelumalleja, sillä DaaS on perusominaisuuksiltaan muiden pilvipalvelumallien kaltainen ja jakaa niiden positiiviset ominaisuudet.

Tutkielman rajoitteena voidaan pitää data palveluna -mallin tutkimusten rajattua määrää. Siinä missä muita pilvipalvelumalleja on tutkittu laajasti niiden yleistyessä, DaaS:n liittyvät tutkimus sijoittuu pääosin 2010-luvun alkuun, ja perustuu usein mallin tietosuojongelmiin. Datapalveluiden käyttö on myös lähivuosina lisääntynyt niin nopeasti, ja DaaS-ratkaisut ovat usein osa palveluntarjoajien laajempia palvelukokonaisuuksia, ja DaaS-malliin liittyvä tutkimus on sulautunut yleisempään pilvipalveluratkaisuja esittelevään tutkimukseen. Tulevaisuuden kannalta mielenkiintoisia tutkimuksen aiheita ovat olemassa olevasta kirjallisuudesta huolimatta etenkin datan laatuun liittyvät kysymykset, sillä dataa käytetään jatkuvasti enemmän liiketoiminnan tukena. Nykyinen kirjallisuus ei myöskään esittele DaaS-palveluita siitä näkökulmasta, etenkin suuret palveluntarjoajat ovat avanneet yhä useammin datapalvelunsa loppukäyttäjälle. Tämä poikkeaa huomattavasti mallin varhaisesta kirjallisuudesta, sekä esimerkiksi tuottavan DaaS-palvelun määrittelystä aiemmissä tutkimuksissa. Olisi näin ollen mielenkiintoista tutkia sitä, miten DaaS-palveluiden kenttä kehittyy, kun palveluntarjoajat eivät enää ainoastaan tarjoa loppukäyttäjille valmiita ratkaisuja, vaan käyttäjät saavat lisäksi käyttöön alustan, jossa palveluiden yksilöllistäminen on huomattavasti helpompaa.

## 5 LÄHTEET

- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., ... & Zaharia, M. (2010). A view of cloud computing. *Communications of the ACM*, 53(4), 50-58.
- Aymerich, F. M., Fenu, G., & Surcis, S. (2008, August). An approach to a cloud computing network. In *2008 First International Conference on the Applications of Digital Information and Web Technologies (ICADIWT)* (pp. 113-118). IEEE.
- Badidi, E., Routaib, H., & El Koutbi, M. (2016, May). Towards data-as-a-service provisioning with high-quality data. In *International Symposium on Ubiquitous Networking* (pp. 611-623). Springer, Singapore.
- Barhamgi, M., Benslimane, D., Ghedira, C., Tbahriti, S. E., & Mrissa, M. (2011, July). A framework for building privacy-conscious daas service mashups. In *2011 IEEE International Conference on Web Services* (pp. 323-330). IEEE.
- Barros, V. A., Estrella, J. C., Prates, L. B., & Bruschi, S. M. (2018, October). An IoT-daas approach for the interoperability of heterogeneous sensor data sources. In *Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems* (pp. 275-279).
- Beimborn, D., Miletzki, T., & Wenzel, S. (2011). Platform as a service (PaaS). *Business & Information Systems Engineering*, 3(6), 381-384.
- Chang, V., Bacigalupo, D., Wills, G., & De Roure, D. (2010, May). A categorisation of cloud computing business models. In *Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing* (pp. 509-512). IEEE Computer Society
- Choudhary, V. (2007, January). Software as a service: Implications for investment in software development. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)* (pp. 209a-209a). IEEE.
- Gong, C., Liu, J., Zhang, Q., Chen, H., & Gong, Z. (2010, September). The characteristics of cloud computing. In *Parallel Processing Workshops (ICPPW), 2010 39th International Conference on* (pp. 275-279). IEEE.
- Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7), 1645-1660.

- Dikaiakos, M. D., Katsaros, D., Mehra, P., Pallis, G., & Vakali, A. (2009). Cloud computing: Distributed internet computing for IT and scientific research. *Internet Computing*, IEEE, 13(5), 10-13.
- Dillon, T., Wu, C., & Chang, E. (2010, April). Cloud computing: issues and challenges. In *Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on* (pp. 27-33). IEEE.
- Folkerts, E., Alexandrov, A., Sachs, K., Iosup, A., Markl, V., & Tosun, C. (2012, August). Benchmarking in the cloud: What it should, can, and cannot be. In *Technology Conference on Performance Evaluation and Benchmarking* (pp. 173-188). Springer, Berlin, Heidelberg.
- Godse, M., & Mulik, S. (2009, September). An approach for selecting software-as-a-service (SaaS) product. In *2009 IEEE International Conference on Cloud Computing* (pp. 155-158). IEEE.
- Grossman, R. L. (2009). The case for cloud computing. *IT professional*, 11(2), 23-27.
- Jackson, K. R., Ramakrishnan, L., Muriki, K., Canon, S., Cholia, S., Shalf, J., ... & Wright, N. J. (2010, November). Performance analysis of high performance computing applications on the amazon web services cloud. In *Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on* (pp. 159-168). IEEE
- Jadeja, Y., & Modi, K. (2012, March). Cloud computing-concepts, architecture and challenges. In *Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on* (pp. 877-880). IEEE
- Kourik, J. L. (2011, April). For small and medium size enterprises (SME) deliberating cloud computing: a proposed approach. In *Proceedings of the European Computing Conference* (Vol. 13, pp. 216-221).
- Lombardi, F., & Di Pietro, R. (2011). Secure virtualization for cloud computing. *Journal of network and computer applications*, 34(4), 1113-1122.
- Magic Quadrant for Cloud Infrastructure as a Service, Worldwide. *Haettu 25.4.2016 osoitteesta* <https://www.gartner.com/doc/reprints?id=1-2G2O5FC&ct=150519>
- Manvi, S. S., & Shyam, G. K. (2014). Resource management for Infrastructure as a Service (IaaS) In *cloud computing: A survey. Journal of Network and Computer Applications*, 41, 424-440.
- Mrissa, M., Tbahriti, S. E., & Truong, H. L. (2010, December). Privacy model and annotation for DaaS. In *2010 Eighth IEEE European Conference on Web Services* (pp. 3-10). IEEE.
- Ojala, A., & Tyrväinen, P. (2011) "Developing cloud business models: A case study on cloud gaming." *IEEE software* 28.4 (2011): 42.
- Pringle, T., Baer, T., & Brown, G. (2014, July). Data-as-a-service: the Next Step in the As-a-service Journey. In *Oracle Open World Conference*.
- Prodan, R., & Ostermann, S. (2009, October). A survey and taxonomy of infrastructure as a service and web hosting cloud providers. In *Grid Computing, 2009 10th IEEE/ACM International Conference on* (pp. 17-25). IEEE.
- Rajesh, S., Swapna, S., & Reddy, P. S. (2012). Data as a Service (Daas) in Cloud Computing. *Global Journal of Computer Science and Technology*, 12(11-B)



- Segura, Á. M., Cuadrado, J. S., & de Lara, J. (2014). ODaaS: Towards the model-driven engineering of open data applications as data services. In *2014 IEEE 18th International Enterprise Distributed Object Computing Conference Workshops and Demonstrations* (pp. 335-339). IEEE.
- Shu-Qing, Z. E. N. G., & Jie-Bin, X. (2010, October). The improvement of PaaS platform. In *Networking and Distributed Computing (ICNDC), 2010 First International Conference on* (pp. 156-159). IEEE
- Søilen, K. S. (2016). Users' perceptions of Data as a Service (DaaS). *Journal of Intelligence Studies in Business*, 6(2).
- Sun, W., Zhang, K., Chen, S. K., Zhang, X., & Liang, H. (2007, September). Software as a service: An integration perspective. In *International Conference on Service-Oriented Computing* (pp. 558-569). Springer, Berlin, Heidelberg.
- Sun, W., Zhang, X., Guo, C. J., Sun, P., & Su, H. (2008, September). Software as a service: Configuration and customization perspectives. In *2008 IEEE congress on services part II (services-2 2008)* (pp. 18-25). IEEE.
- Subashini, S., & Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing. *Journal of network and computer applications*, 34(1), 1-11
- Terzo, O., Ruiu, P., Bucci, E., & Xhafa, F. (2013, July). Data as a service (DaaS) for sharing and processing of large data collections in the cloud. In *Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on* (pp. 475-480). IEEE.
- Tbahriti, S. E., Mrissa, M., Medjahed, B., Ghedira, C., Barhamgi, M., & Fayn, J. (2011, August). Privacy-aware DaaS services composition. In *International Conference on Database and Expert Systems Applications* (pp. 202-216). Springer, Berlin, Heidelberg.
- Tbahriti, S. E., Ghedira, C., Medjahed, B., & Mrissa, M. (2013). Privacy-enhanced web service composition. *IEEE Transactions on Services Computing*, 7(2), 210-222.
- Toosi, A. N., Calheiros, R. N., Thulasiram, R. K., & Buyya, R. (2011, September). Resource provisioning policies to increase IaaS provider's profit in a federated cloud environment. In *High Performance Computing and Communications (HPCC), 2011 IEEE 13th International Conference on* (pp. 279-287). IEEE
- Truong, H. L., & Dustdar, S. (2009, December). On analyzing and specifying concerns for data as a service. In *Services Computing Conference, 2009. APSCC 2009. IEEE Asia-Pacific* (pp. 87-94). IEEE.
- Truong, H. L., & Dustdar, S. (2010, December). On evaluating and publishing data concerns for data as a service. In *2010 IEEE Asia-Pacific Services Computing Conference* (pp. 363-370). IEEE.
- Truong, H. L., Gangadharan, G. R., Comerio, M., Dustdar, S., & De Paoli, F. (2011). On analyzing and developing data contracts in cloud-based data marketplaces. In *2011 IEEE Asia-Pacific Services Computing Conference* (pp. 174-181). IEEE.

- Tsai, W. T., Sun, X., & Balasooriya, J. (2010, April). Service-oriented cloud computing architecture. In *Information Technology: New Generations (ITNG), 2010 Seventh International Conference on* (pp. 684-689). IEEE.
- Villegas, D., Antoniou, A., Sadjadi, S. M., & Iosup, A. (2012, May). An analysis of provisioning and allocation policies for infrastructure-as-a-service clouds. In *Cluster, Cloud and Grid Computing (CCGrid), 2012 12th IEEE/ACM International Symposium on* (pp. 612-619). IEEE.
- Vu, Q. H., Pham, T. V., Truong, H. L., Dustdar, S., & Asal, R. (2012, March). Demods: A description model for data-as-a-service. In *Advanced Information Networking and Applications (AINA), 2012 IEEE 26th International Conference on* (pp. 605-612). IEEE.
- Youseff, L., Butrico, M., & Da Silva, D. (2008, November). Toward a unified ontology of cloud computing. In *Grid Computing Environments Workshop, 2008. GCE'08* (pp. 1-10). IEEE.
- Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: state-of-the-art and research challenges. *Journal of internet services and applications*, 1(1), 7-18.
- Zheng, Z., Zhu, J., & Lyu, M. R. (2013, June). Service-generated big data and big data-as-a-service: an overview. In *2013 IEEE international congress on Big Data* (pp. 403-410). IEEE.