

Jari Forsten

3D-KIIHDYTETYT PILVIPALVELUT

Tieto- ja viestintäteknikan koulutusohjelma

2020

3D-kiihdytetyt pilvipalvelut

Forsten, Jari
Satakunnan ammattikorkeakoulu
Tieto- ja viestintätekniikan koulutusohjelma
Toukokuu 2020
Sivumäärä: 33
Liitteitä:

Asiasanat: pilvipalvelut, virtualisointi, 3D-kiihdytys, verkkopelaaminen

Tässä tutkimuksessa tarkasteltiin työn tilaajan, Lounea Palveluiden mahdollisuutta ja kannattavuutta aloittaa 3D-kiihdytettyjen pilvipalveluiden tarjoamista yrityksille ja yksityisille. Työssä käytiin läpi pilvipalveluiden kehitystä ja eri muotoja, tutkimuksen painottuessa pelaamiseen tarkoitettuihin palveluihin.

Tutkimuksessa vertailtiin neljää kaupallista pilvipelaamiseen tarkoitettua palvelua, niiden hintoja, pelikirjastoja ja laitevaatimuksia sekä niille ominaisia ominaisuuksia. Grafiikkakortteja vertaillaessa tutkittiin tarkemmin Nvidian ja AMD:n tarjoamia malleja ja niiden ominaisuuksia edellisen sukupolven arkkitehtuureista uusimpiin. Samalla analysoitiin myös näiden kahden yrityksen tarjoamia, palvelimille suunnattuja ohjelmistoja ja niiden lisenssihintoja.

Testausvaiheessa vertailtiin pöytäkoneen, kahden eritasoisen virtuaalikoneen ja kaupallisen pelipalvelun eroavaisuuksia toiminnallisesti ja graafisesti testausohjelman ja kahden pelin kanssa. Lopuksi vielä testattiin raskaamman sarjan 3D-suunnitteluun tarkoitettua ohjelmaa sekä pöytäkoneella että toisella virtuaalikoneista.

3D-accelerated cloud computing

Forsten, Jari

Satakunta University of Applied Sciences

Degree Programme in Information and Communication technology

May 2020

Number of pages: 33

Appendices:

Keywords: cloud computing, virtualizing, accelerated 3D

Lounea Palvelut has been interested in providing 3D accelerated cloud services for enterprises and private users in a profitable way. Different commercial gaming services were examined for pricing, libraries and other key functions in this thesis.

Two major graphics card manufacturers Nvidia and AMD were compared for their server line of products, reaching from former generations to current ones. Services and licenses related to those products were also compared in detail.

Differences in graphical capabilities between classic gaming desktop, couple variants of 3D accelerated virtual desktops and a commercial gaming service were analysed during the testing phase. For this purpose, a graphical benching tool, two games and industrial level 3D design program were applied.

SISÄLLYS

1	JOHDANTO.....	5
2	PILVIPALVELUN MÄÄRITELMÄ.....	6
2.1	Pilvipalvelun tärkeimmät ominaisuudet	6
2.2	Palvelumallit	7
2.3	Toteutustavat.....	9
3	3D-KIIHDYTETTY PILVIPALVELU	10
4	PELAAMINEN PILVIPALVELUSSA	12
4.1	Haasteet.....	12
4.2	Palvelut	13
4.2.1	Vertailu	14
5	GRAFIKKAKORTIT	17
5.1	Nvidia.....	17
5.1.1	Pascal	17
5.1.2	Volta	18
5.1.3	Turing	19
5.2	AMD	20
5.2.1	Firepro S-sarja	20
5.2.2	Instinct	20
5.3	Ohjelmistot ja lisenssit.....	21
5.3.1	NVIDIA	21
5.3.2	AMD	24
6	TESTAAMINEN.....	25
	Testikoneet	25
6.1	Penkitys.....	26
6.2	Pelitestaus	27
6.2.1	Counter Strike: Global Offensive.....	28
6.2.2	The Witcher 3: Wild Hunt.....	31
7	LOPUKSI.....	33
	LÄHTEET.....	34
	LIITTEET	

1 JOHDANTO

Pilvipalvelut ovat olleet jo pitkään jatkuvassa kasvussa oleva palvelumalli, joka säästää asiakkaalta laite- ja tilakustannuksia, tarjoamalla jo valmiina olevan infrastruktuurin. Nykyään suurin osa päivittäin käyttämistämme palveluista toimii pilvessä. Myös peliteollisuus ja muiden graafisten ohjelmistojen valmistajat ovat havahtuneet tähän jatkuvassa nousussa olevaan trendiin ja monet yritykset ovatkin jo lähteneet tarjoamaan jonkin asteisia pelaamiseen ja 3D-suunnitteluun liittyviä pilvipalveluita.

Pelaamiseen tarkoitettujen pilvipalveluiden historian voidaan sanoa alkaneen vuonna 2003, OnLive-nimisen yrityksen aloittaessa toimintansa. Tällöin verkon nopeudet olivat vielä suhteellisen hitaita, jolloin kuvanlaatu ei aina ollut paras mahdollinen. Verkonopeuksien ja niiden luotettavuuden nousu on tehnyt palvelun perustamisesta kannattavampaa ja halvempaa. Myös suuret toimijat ovat huomanneet kyseisen palvelumallin potentiaalin ja ovat lähteneet kehittämään omia versioitaan.

Työn tilaaja Lounea Palvelut Oy on lounaisessa Suomessa toimiva, teleliiketoimintaa harjoittava yritys. Yrityksen palveluihin kuuluu jo monia pilvipalveluita ja mahdollisesti tulevaisuudessa palvelut laajenevat koskemaan 3D-kiihdytettyjä pilvipalveluita.

Tässä työssä käsittelen yleisesti pilvipalveluita, niiden määritelmiä ja toteutustapoja. Tutkimuksen painopiste tulee olemaan pelaamisessa, mutta käsittelee myös osittain muitakin graafisia palveluita.

2 PILVIPALVELUN MÄÄRITELMÄ

Pilvipalvelu on malli, jossa tarjotaan esimerkiksi lähiverkoille, palvelimille, verkkokiintolevyille, sovelluksille ja palveluille kaikkialla toimiva ja kätevä on-demand pääsy jaettuun verkkoon, jossa resurssit on helppoa varata ja vapauttaa. Se koostuu viidestä oleellisesta ominaisuudesta, kolmesta palvelumallista ja neljästä käyttöönottomallista (Mell P, Grace T. 2012).

2.1 Pilvipalvelun tärkeimmät ominaisuudet

- *Itsepalvelu (On-demand self-service)*. Henkilö voi lokaatiostaan riippumatta varata itselleen palveluita, esimerkiksi palvelimilta tai verkkokiintolevyiltä, ilman tarvetta yhteydenpitoon palveluntarjoajien kanssa.
- *Laajat verkkoliitännät (Broad network access)*. Palvelut ovat saatavilla tietoverkon kautta erilaisilla päätelaitteilla, kuten tabletti, puhelin ja tietokone.
- *Resurssien yhdistettävyys (Resource pooling)*. Palveluntarjoajan resurssit on yhdistetty palvelemaan monia asiakkaita ja niitä voidaan dynaamisesti ohjata asiakkaiden tarpeiden mukaan.
- *Joustavuus (Rapid elasticity)*. Virtualisoinnin ja automaation ansiosta voidaan yksittäisen palvelun kuluttamia resursseja tarvittaessa skaalata.
- *Monitoroitu palvelu (Measured service)*. Palvelut optimoivat ja hallitsevat resurssien käyttöä automaattisesti käyttämällä mittausta, jonka mukaan asiakas maksaa palvelun käytöstä. Resurssien käyttöä voi seurata, hallita ja raportoida, joka luo läpinäkyvyyttä asiakkaan ja palveluntarjoajan välille (Mell P, Grace T. 2012).

2.2 Palvelumallit

- *SaaS (Software as a Service)*. Palveluntarjoaja vastaa tarvittavasta infrastruktuurista, virtuaalipalvelimista ja ohjelmistoista. Asiakas ei pääse hallinnoimaan infrastruktuuria, jonka päällä tilattu palvelu toimii, sillä vastuu niistä kuuluu palveluntarjoajalle.



Kuva 1. SaaS-palvelu. O7services.2019.

- *PaaS (Platform as a Service)*. Palveluntarjoaja vastaa infrastruktuurista, virtuaalipalvelimista ja niiden päälle asennetusta asiakkaan valitsemasta alustasta. Asiakas ei hallitse alustan alla pyörivää infrastruktuuria mutta saa hyödyntää alustaa parhaalla näkemällään tavalla.



Kuva 2. PaaS-palvelu. O7services.2019.

- *IaaS (Infrastructure as a Service)*. Palveluntarjoaja tarjoaa asiakkaalle fyysisen infrastruktuurin, jonka päälle asiakas rakentaa oman palvelunsa (Mell P, Grace T. 2012)



Kuva 3. IaaS-palvelu. O7services.2019.

- *DaaS (Desktop as a Service)*. Uusin palvelumuoto, jossa virtuaalikoneet sijaitsevat palveluntarjoajan palvelimilla ja asiakas käyttää niitä joko erillisen ohjelmiston tai verkkoselaimen kautta. Siinä palveluntarjoaja hoitaa VDI-ratkaisun (Virtual Desktop Infrastructure) ylläpidolliset toimet, kuten käyttöönoton, päivitykset tietoturvan, tallennuksen ja varmuuskopioinnin. Asiakas hallinnoi sovelluksiaan ja työpöytiä (Rouse. 2017, Citrix).

2.3 Toteutustavat

- *Yksityinen pilvi (Private cloud)*. Yksityinen pilvi on tarkoitettu yksittäisen organisaation käyttöön. Sen voi omistaa ja sitä voi hallita itse käyttävä organisaatio, kolmas osapuoli tai jokin niiden yhdistelmä.
- *Yhteisöpilvi (Community cloud)*. Yhteisöpilvi on tarkoitettu yhteisöille, joiden välillä on suurempi luottamus kuin julkisessa pilvessä.
- *Julkinen pilvi (Public cloud)*. Kun puhutaan pilvipalvelusta, puhutaan yleensä julkisen pilven palveluista. Resurssit on jaettu suuren yleisön käyttöön. Julkisen pilven palveluntarjoajia ovat esimerkiksi Amazon ja Microsoft.
- *Hybridi-pilvi*. Muodostuu kahdesta tai useammasta yllä mainitusta ympäristöstä, esimerkiksi SaaS ja IaaS (Mell P, Grace T. 2012).

3 3D-KIIHDYTETTY PILVIPALVELU

3D-kiihdytetty pilvipalvelu antaa asiakkaalle joustavan mahdollisuuden käyttää graafisesti raskaitakin sovelluksia jopa puhelimella tai tabletilla, sillä kaikki raskas laskeinta ja renderöinti tapahtuu palveluntarjoajan laitteissa ja video lähetetään verkon yli asiakkaan päätelaitteeseen. Tämä helpottaa varsinkin aloittavien yritysten alkutaivalta poistamalla kustannuksista tehokkaan ja kalliin laitteiston hankinnan, jolloin rahaa jää muuhun.

Siirrettäessä palvelut verkkoon myös tilan tarve yrityksiltä vähenee. Missä ennen tarvittiin tilaa vievää pöytäkoneita, riittää nyt kannettava tietokone, tabletti tai puhelin. Palveluntarjoajalta palvelun pystyttäminen vaatii näytönohjaimella varustetun palvelinkaluston hankkimista. Lisäksi ostohinnan päälle tulee vielä erinäisiä lisenssikustannuksia, joita käydään myöhemmin läpi. 3D-kiihdytystä voidaan soveltaa lukemattomilla eri aloilla, kuten mainostoimistoissa, konepajoissa, arkkitehtiyrityksissä ja yksityishenkilöiden käytössä. Seuraavaksi käydään läpi syitä 3D-kiihdytetyn palvelimen valintaan perinteisen palvelimen sijasta.

Grafiikkakiihdytys on ennen toteutettu prosessoreilla, jolloin laitteiden tarve oli verrattuna grafiikkakortin sisältäviin palvelimiin hyvin suuri. Otettaessa GPU-kiihdytys käyttöön, tarvitaan vähemmän laitteita ja tilaa. Seuraavissa laskutoimituksissa käytän Nvidian tekemiä esimerkkejä vain prosessoreja sisältävien ja grafiikkakortin sisältävien palvelimien eroista.

Oletetaan että on olemassa kaksi datakeskusta. Toinen koostuu pelkästään prosessoreita omaavista palvelimista ja toinen sekoituksesta vain prosessoreilla toimivista ja grafiikkakortin omaavista palvelimista. Jokainen palvelin on ns. dual-socket eli palvelimessa on kaksi prosessoria. Grafiikkakortin sisältävissä palvelimissa on myös kaksi Tesla V100 kiihdytintä. Kummallekin datakeskukselle lasketaan 70% grafiikkakorttia tukevaa työtä.

Jos asetetaan kummallekin yllämainituista datakeskuksista 1000 yksikköä työtä, tarvitaan 1000 pelkän prosessorin sisältävää palvelinta hoitamaan työt päivässä, koska 700 työtä on mahdollista suorittaa grafiikkakortilla, voidaan ne asettaa kiihdytetyille

palvelimille, jolloin kiihdyttämättömille palvelimille jää 300 työtä. Oletetaan että grafiikkakortti on 20 kertaa nopeampi kuin kiihdyttämätön palvelin, tarvitaan vain 35 kiihdytettyä palvelinta ja 300 kiihdyttämätöntä palvelinta. Täten käyttämällä kiihdytettyjä palvelimia tarvitaan 665 laitetta vähemmän.

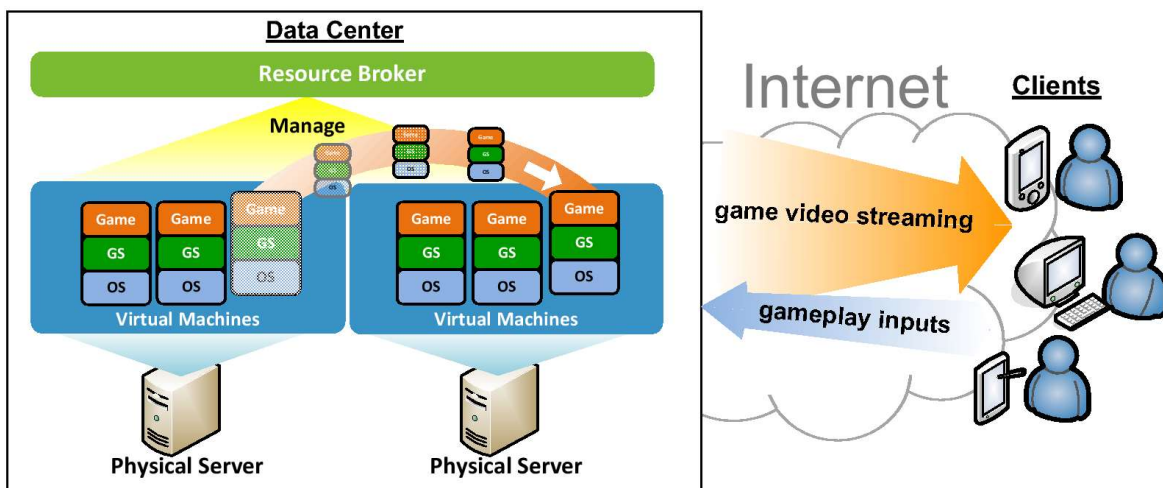
Hankintahintoja vertailtaessa huomataan grafiikkakorttien tuovan palvelimelle huomattavan määrän lisähintaa, tulee se lopulta halvemmaksi kuin tavallisten palvelimien hankinta. Seuraava taulukko perustuu edeltävään laskutoimitukseen, vaikka on kuitenkin otettava huomioon, että seuraava kaavio sisältää vain hankintakustannukset ilman lisenssejä (Nvidia 2018).

ACQUISITION COSTS	THE CPU-ONLY DATA	THE ACCELERATED DATA CENTER
CPU Nodes	\$9,000 x 1000 Nodes	\$9,000 x 300 Nodes
Tesla V100 Nodes	—	\$44,000 x 35 Nodes
Servers	\$9.0M	\$4.2M
Storage	\$3.0M	\$3.0M
Software and Services	\$1.5M	\$.9M
Total Acquisition Cost	\$13.5M	\$8.1M

Taulukko 1. Hankintahinnat. Nvidia 2018.

4 PELAAMINEN PILVIPALVELUSSA

Pilvipelaamisen merkittävimpiä pioneereina voidaan pitää Onliveä ja Gaikaita, jotka aikanaan keräsivät miljoonia käyttäjiä omille palveluilleen. Voidaan sanoa, että Sonyn ostettua edellä mainitut palvelut, alkoi pilvipelaamisen todellinen kehitys. Seuraavassa kuvassa on selitetty toimintaperiaate, jolla pelaaminen pilvessä toimii. Tätä mallia voidaan käyttää myös muun graafisen toiminnan selittämiseen.



Kuva 4. Palvelun toimintaperiaate. Hong H, Chen D, Huang C, Chen K, Hsu C. 2019.

4.1 Haasteet

Verkon yli pelaamisessa nousevat suurimmiksi haasteiksi syöttöviive (input lag) ja kuvan laatu. Välimatkan kasvaessa serveriin nähden, alkaa latenssi nousta ja syöttöviive pahentua, kuitenkin nykyään kovaa vauhtia yleistyvät kuituliittymät ja muut pieniviiveiset yhteystyypit poistavat tämän ongelman melkein kokonaan. Kuvanlaadun takaamiseksi täytyy palvelimella olla enkooderi, joka kykenee hyvään kuvanlaatuun myös pienillä siirtonopeuksilla (bitrate). Lisäksi pelamistarkoitukseen tulevassa prosessorissa tulisi olla mieluummin mahdollisimman korkea ytimen kellotaajuus (core clock) kuin ytimien määrä (core count). Kuitenkin ytimiä pitäisi olla vähintään neljä, sillä monet pelit on optimoitu kyseiselle määrälle.

Laadukkaan videon lähetyksen pienilläkin nopeuksilla vaatii tehokkaan encoderin ja decoderin, jotka aiheuttavat mahdollisimman vähän kuvapakettien kasautumista ja kuvanlaadun heikkenemistä. Myös asiakaspään laitteen kuormitus täytyisi pitää mahdollisimman alhaisena.

Lisäksi suurten toimijoiden, kuten Amazonin kilpailuun mukaan tuleminen ja tarvittavan palvelinkaluston kustannukset hankaloittavat pienempien yritysten markkinoille pääsyä ja menestymistä. (Chris. 2018, Mäntylä. 2019).

4.2 Palvelut

Pelaamiseen tarkoitettut, verkon välityksellä toimivat palvelut on jaettu kolmeen luokkaan. Peer-to-Peer palveluihin, suoratoisto palveluihin ja progressiivisen latauksen palveluihin.

Peer-to-Peer-palvelut yhdistävät kaksi laitetta suoraan toisiinsa verkon välityksellä. Peli itsessään asennetaan ja se ajetaan vain toisella laitteella, jonka jälkeen video lähetetään toiselle laitteelle. Näitä palveluita ovat esimerkiksi Steamin Remote Play ja Parsec,

Suoratoistopalvelut omaavat oman pelikirjaston, jonka saa käyttöön kuukausimaksua vastaan. Suomessa tällä hetkellä toimivista tuottavista palveluista tunnetuimpia ovat Googlen Stadia, Vortex, PlayStation Now. Myös Microsoft on lähdössä kilpailuun mukaan omalla, vielä rajatussa kokeilussa olevalla xCloud-palvelullaan.

Progressiivisen latauksen palveluissa vain pieni osa pelistä tarvitaan pelaamisen aloittamiseen. Loput tiedostoista latautuvat taustalla. Tämä palvelu auttaa pääsemään nopeasti peliin pienemmälläkin verkkonopeudella. Tällaisia palveluita ovat esimerkiksi Blizzardin Battle.net.

Lisäksi on vielä palveluita, jotka tarjoavat virtuaalikoneen, jolle asiakas voi asentaa haluamansa pelit omasta kirjastostaan, joko rajoituksetta tai palveluntarjoajan asettamilla rajauksilla. Tiedossa on myös muutamia pelitaloja, jotka estävät peliensä pelaamisen virtuaalikoneella. Tällaisia palveluita tarjoaa esimerkiksi Geforce Now ja Paperspace.

4.2.1 Vertailu

Seuraavassa kaaviossa käydään läpi muutaman palvelun hinnastoa, mitä kyseisellä maksulla saa sekä minimivaatimukset asiakkaan laitteille. Valikoin nämä palvelut vertailtaviksi, sillä ne ovat mielestäni hyvät edustajat omaan palvelumalliinsa. Käytän esimerkkeinä vain suoratoistopalveluihin ja virtuaalikonepalveluihin kuuluvaa palvelua, sillä kaksi muuta palvelua ovat yleensä saatavilla ilmaiseksi. Tiedot on kerätty palveluiden verkkosivuilta. Ensin kuitenkin käyn läpi perustietoa vertailtavista palveluista.

Google Stadia. Stadia julkaistiin marraskuussa 2019. Se toimii Debian Linux palvelimilla ja käyttää Vulkania ohjelmointirajapintanaan. Sen toimintaperiaatteena on, että asiakas ostaa pelin palvelun kautta ja on sen jälkeen oikeutettu pelaamaan kyseistä peliä missä vain. Oikeus kuitenkin raukeaa, jos asiakas irtisanoo Stadia-jäsenyytensä. Julkaisuhetkellä palvelussa oli valittavana 22 peliä, jonka jälkeen kirjasto on kasvanut 42:n ja lisää on ilmoitettu tulevan. Julkaisun jälkeen Google on lisännyt palveluun mahdollisuuden käyttää yleisimpiä peliohjaimia pelattaessa tietokoneella tai kannettavilla älylaitteilla. Television kautta käytettäessä palvelu vaatii Stadian oman ohjaimen ja Chromecast ultran, jotka kuuluvat pakettiin. Stadialla on kaksi jäsenyysvaihtoehtoa. Pro-jäsen, jonka hinta on 9,99€/kk. Maksulla saa kuukausittaisen maksuttoman pelin ja on oikeutettu jäsentarjouksiin. Pro-jäsen pystyy myös pelaamaan jopa 4K resoluutiolla, kun ilmaisjäsen on lukittu 1080p tarkkuuteen (Google, Dov Z).

Playstation now. PS Now julkaistiin Euroopassa vuonna 2015 ja Suomessa 2019. Se tarjoaa asiakkailleen suuren määrän jopa PlayStationille yksinoikeutettuja pelejä pelattavaksi PlayStation 4 konsolilla tai tietokoneella. Pelejä ei tarvitse erikseen ostaa eikä ladata, vaan ne suoratoistetaan suoraan palvelimelta verkon yli. Palveluun on saatavissa seitsemän päivän ilmainen kokeilujakso, jonka jälkeen hinta on 9,99€/kk, 24,95€/3kk tai 59,95€/12k (Sony).

Vortex. Vortex toimii samalla periaatteella kuin PS Now. Sen kirjastoon kuuluu pelejä eri palveluista, kuten Steam, Electronic Artsin Origin ja Ubisoftin Uplay. Palvelua voi käyttää Androidilla, Windowsilla tai macOS:lla. Valittavana on kolme jäsenyystasoa: Basic 9,99€/kk, joka toimii HD-tasoisena ja sisältää 50 tuntia peliaikaa kuukaudessa, sekä 97 pelin kirjaston. Pro-taso maksaa 18,99€/kk, sisältää kahdeksankymmentä

tuntia peliaikaa kuukaudessa ja 178 pelin kirjaston. Ultra-taso maksaa 27,99€/kk, sisältäen 140 tuntia peliaikaa ja 178 pelin kirjaston (Vortex).

Geforce Now. Geforce Now käyttää muista esitellyistä palveluista poiketen käyttäjän omaa pelikirjastoa. Tällä hetkellä valittavana on kaksia jäsenyystasoa: Ilmainen, joka sisältää tunnin peliaikaa istuntoa kohti ja perustajajäsen tason, joka maksaa 5,49/kk/12kk ja sisältää 90 päivän ilmaisen kokeilujakson. Perustajajäsen saa etuoikeuden peleihin pääsyssä, jopa kuuden tunnin peli-istunnot ja RTX-ominaisuuden (Ray Tracing Platform) (Nvidia).

Palvelut	Stadia	Play Station Now	Vortex	Geforce Now
Hinta /kk	Pro: 9,99€. Basic: ilmainen	9,95€	Basic: 9.99 Pro: 18.99 Ultra: 27.99	Perustajajäsen: 5,49€ Ilmainen
Kirjasto	42	700+	179	Käyttäjän oma pelikirjasto
Luvattu tarkkuus	4K ja HDR	4K	HD	?
Luvattu virkistystaajuus	60 fps	?	60 fps	?
Laitevaatimukset (minimi)	Ei ole. Toimii Chromecast Ultran kautta televisiossa, selaimen kautta tietokoneessa. Chrome OS älypuhelimella / tabletilla.	Win7 (SP1), 8.1 tai 10 Core i3 2.0GHz 300Mt vapaata tilaa 2 Gt RAM Äänikortti USB-portti	Android 4.4 Windows 7, 8, 10 64-bit Mac v.10.13	2.0 GHz:n kaksiytiminen X86-prosessori. Direct X11 Nvidia 600-sarja AMD Radeon HD3000 -sarja Intel HD Graphics 2000-sarja
Mukana tulevat lisälaitteet	Chromecast Ultra ja Stadia-ohjain			
Lisämaksut	129€, joka pitää sisällään Chromecast Ultran, Stadia ohjaimen ja 3kk Stadia Pro-tilausta.			

Taulukko 2. Google, Sony, Nvidia 2020.

5 GRAFIKKAKORTIT

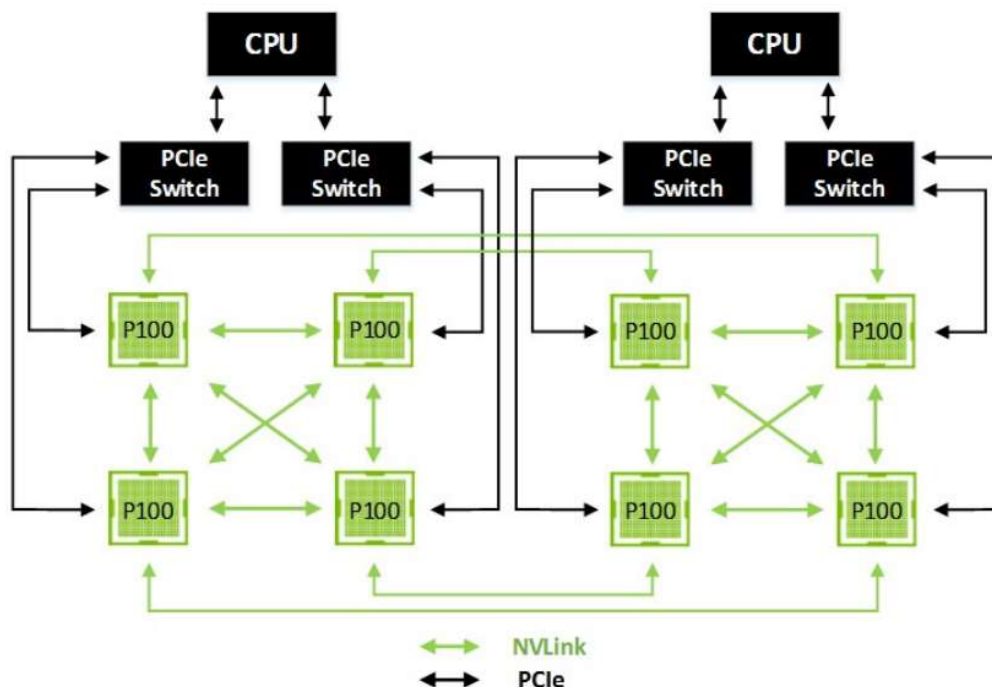
Vaikka palvelinvalmistajia onkin useita, niin grafiikkakortit ovat pääosin joko Nvidian tai AMD:n valmistamia.

5.1 Nvidia

Nvidialla on valikoimassaan kolme GPU-arkkitehtuuria, Pascal, Volta ja Turing. Seuraavaksi käydään hieman läpi näitä arkkitehtuureja ja niiden tärkeimpiä ominaisuuksia.

5.1.1 Pascal

Pascal on vanhaa sukupolvea edustava arkkitehtuuri, joka julkaistiin vuonna 2016. Se korvasi vanhan Maxwell arkkitehtuurin. Sen yksi tärkeimmistä ominaisuuksia on NVLink joka mahdollistaa jopa 160 Gb/s kaksisuuntaisen yhteyden kahden grafiikkakortin välillä. Kuvassa viisi esitellään miten NVLink yhdistää grafiikkakortteja.



Kuva 5 NVLinkin toiminta. Nvidia 2017.

Pascal-arkkitehtuuri käytti ensimmäisenä maailmassa HBM2 (High Bandwidth Memory) muistia, joka tarjoaa jopa kolminkertaisen muistiväylän edeltäjänsä nähden. Koska HBM2 on pinomuisti ja se sijaitsee grafiikkakortin sisällä, toisin kuin vanha GDDR5 (graphics double data rate type five synchronous dynamic random-access memory), vähentää se tilan tarvetta palvelimen sisältä.

5.1.2 Volta

Volta edustaa Nvidian grafiikkakorttien uudempaa sukupolvea ja se korvaa vanhan Pascal arkkitehtuurin. Tesla V100 SXM2 lupaa 7.8 teraFLOP:a sekunnissa (Floating point Operations Per Second), mikä on yli viisinkertainen määrä Pascaliin verrattuna. Suuriin muutoksia Pascaliin verrattuna on CUDA-ydinten, joita Voltassa on 5120 ja Tesla V100:ssa 3840, sekä Tensor-ydinten käyttö CUDA:n rinnalla.

NVLink on myös päivitetty vanhasta versiosta. Se lupaa kaksinkertaisen, eli noin 320 Gb/s kaksisuuntaisen yhteyden.

5.1.3 Turing

Turing on Nvidian uusin ja kehittynein grafiikkakorttien sukupolvi. Suurimmat muutokset edeltäjäänsä ovat muistityypin muuttuminen GDDR6: n ja uusien RT-ydinten lisääminen. RT-core mahdollistaa Ray Tracing- mekaniikan käytön, joka tarkoittaa käytännössä valonsäteiden seuraamista lähteestä kameraan ja samalla simuloimalla muutoksia, jotka aiheutuvat säteen osuessa objekteihin. Seuraava kuva on otettu Nvidian sivuilta ja se havainnollistaa säteenseurannan käyttöä.



Kuva 5. Ray tracing. Nvidia 2018.

5.2 AMD

AMD on hiljalleen valtaamassa konsolimarkkinat ja laajenevan osan pc-markkinoista Nvidialta uusilla innovaatioillaan. Palvelinkäyttöön yrityksellä on yksi grafiikkakorttimalli. Yhtiö on myös päättänyt estää kuluttajille suunnattujen korttiensa ajurien asentamisen Windowsin palvelinkäyttöjärjestelmiin.

5.2.1 Firepro S-sarja

FirePro S-sarja on vuonna 2016 julkaistu palvelingrafiikkakorttisarja, jonka myöhemmin korvasi Radeon Instinct. Sen lippulaivamallissa S7150x2:a on 16GB GDDR5 muistia ja se lupaa 7.54 teraFLOP:a sekunnissa. Yhtenä ominaisuutena kortissa on AMD Multiuser GPU teknologia, joka mahdollistaa tarjota käyttäjälle ISV-sertifioitun virtuaalikoneen. Lisäksi kortissa on erilaisia virrankulutusta vähentäviä ratkaisuja. (AMD 2016).

5.2.2 Instinct

Radeon Instinct on FirePro S-sarjan seuraaja, joka käyttää Vega arkkitehtuuria ja on AMD:n uusin puhtaasti palvelinkäyttöön tarkoitettu grafiikkakortti. Se on suurimaksi osaksi suunniteltu syväoppimiseen ja muihin HPC (High-Performance Computing) tehtäviin.

Ominaisuuksina Instinctillä on jopa 26.5 TeraFLOP:a sekunnissa ja se käyttää Nvidian tapaan HBM2 muistia. Kortissa on 3840 CUDA-corea, joita AMD nimittää Stream Processoreiksi. Instinct käyttää myös Infinity Fabrik linkkiä, joka toimii samaan tapaan kuin Nvidian NVLink, yhdistäen GPU clustereita jopa 92 GB/s nopeudella (AMD 2020).

5.3 Ohjelmistot ja lisenssit

Seuraavaksi käydään läpi kummankin edellä esitellyn yrityksen palveluita, lisenssejä ja niiden hintoja.

5.3.1 NVIDIA

Nvidian vGPU ohjelmistot tuovat eri tasoisia palveluita erilaisiin tarpeisiin. Erilaisia ohjelmia on neljä.

GRID vPC joka tuo asiakkaalle virtualisoidun työpöydän kaikilla pc:n toiminnoilla.

*Quadro vDWS on paranneltu GRID vPC, joka on suunnattu 3D-suunnittelijoille ja si-
sällöntuottajille. Palvelu toimii millä tahansa laitteella.*

GRID vApps tuo ohjelmat virtualisoituna palveluna

NVIDIA vComputeServer on palvelu, joka sopii hyvin laskentaraskaisiin suoritteisiin, kuten tekoäly ja syväoppiminen.

Seuraavassa taulukossa verrataan yllämainittuja palveluita

Feature	GRID vApps	GRID vPC	Quadro vDWS	vComputeServer
License Entitlement				
Concurrent User (CCU)	✓	✓	✓	
Per GPU				✓ ¹
Capability Entitlement				
Desktop Virtualization		✓	✓	
RDSH App Hosting	✓	✓	✓ ²	
RDSH Desktop Hosting	✓	✓	✓ ²	
Compute Virtualization			✓	✓
Windows Guest OS	✓	✓	✓	
Linux Guest OS		✓	✓	✓
Maximum Displays	1 ³	4	4	1
Maximum Resolution ⁴	1280*1024	5120*2880 (5K)	7680*4320 (8K)	4096*2160 (4K)
NVIDIA Quadro Software Features			✓	
CUDA & OpenCL Supported			✓ ⁵	✓
ECC & Page Retirement			✓	✓ ⁶
Multi-GPU			✓	✓
NVLINK			✓	✓
GPU Pass-through Supported ⁷	✓		✓	✓
Bare Metal Supported ⁸	✓		✓	
vGPU Profiles Supported ⁹				
512 MB		✓	✓	
1 GB	✓	✓	✓	
2 GB	✓	✓	✓	
3 GB	✓		✓	
4 GB	✓		✓	✓
6 GB	✓		✓	✓
8 GB	✓		✓	✓
12 GB	✓		✓	✓
16 GB	✓		✓	✓
24 GB	✓		✓	✓
32 GB	✓		✓	✓
48 GB			✓	✓

Taulukko 3. Palveluiden vertailu. Nvidia 2019

Lisenssit yllä mainittuihin palveluihin voivat olla joko vuotuisia tai kertaluonteisia. Vuotuisiin lisensseihin kuuluu sisällytettynä Nvidian SUMS (Support Updates and Maintenance Subscription), kerralla maksettaviin lisensseihin kyseinen palvelu täytyy maksaa erikseen. Esimerkit lisenssien hinnoista alla olevissa kaavioissa.

Enterprise Annual Subscription Pricing	
GRID Virtual Applications	\$10 per CCU subscription
GRID Virtual PC	\$50 per CCU subscription
	4 year subscription - \$43.75 per CCU 5 year subscription - \$40 per CCU
Quadro Virtual Data Center Workstation	\$250 per CCU subscription

Taulukko 4. Vuotuiset lisenssimaksut. Nvidia 2019

Enterprise Perpetual Licensing + SUMS Pricing	
GRID Virtual Applications	\$20 per CCU perpetual license
	\$5 SUMS per year
GRID Virtual PC	\$100 per CCU perpetual license
	\$25 SUMS per year
Quadro Virtual Data Center Workstation	\$450 per CCU perpetual license
	\$100 SUMS per year

Taulukko 5. Kertaluonteinen lisenssimaksu. Nvidia 2019

5.3.2 AMD

AMD on ottanut erilaisen lähestymistavan virtualisointiin yhdistämällä voimansa VMwaren kanssa. MxGPU teknologia ja Radeon Pro-ohjelmisto mahdollistavat AMD-grafiikkakorttien ohjaamisen VMwaren vSpheren kautta (AMD 2020).

Ohjelmiston avulla voidaan virtuaalikoneet jakaa neljään eri kategoriaan, niiden teho- ja hontarpeiden mukaan.

User Type/Performance	vCPU cores	System memory (GB)	# of enabled MxGPUs	Frame buffer size (MB) ¹
<i>Workstation (High-performance)</i>	4	8	2	3840
<i>Power (Professional)</i>	2	4	4	1920
<i>Knowledge (Enhanced)</i>	2	3	8	960
	2	3	10	768
<i>Task (Standard)²</i>	1	2	16	480
<i>Note: All guest memory must be locked/reserved.</i>				
1. Amount of memory available to each VF.				
2. ESXi limits the number of VFs per GPU to 15.				

Taulukko 6. Työasemien jaottelu. AMD 2020.

Palvelun käyttö vaatii VMwaren vSphere Enterprise Plus paketin. Koska toimittajalla on jo käytössään VMwaren palveluita niin en nähnyt tarpeelliseksi tutkia sen tarkemmin eri lisensointivaihtoehtoja.

6 TESTAAMINEN

Aloitin testaamisen Satakunnan ammattikorkeakoulun palvelimelta saamallani virtuaalikoneella, jolloin törmäsin ensimmäiseen ongelmaan. Hiiren toiminta peleissä oli täysin olematonta. Erinäisten keskustelupalstojen mukaan ongelmana on, että RDP-yhteys (Remote Desktop) ottaa vastaan hiiren absoluuttisen paikan, kun taas suuri osa peleistä esim. FPS (First Person Shooter) käyttävät hyväkseen hiiren relatiivista paikkatietoa. Pelit, kuten Cities Skylines ja RTS (Real Time Strategy) pelit, jotka käyttävät absoluuttista paikannusta toimivat hiiren kannalta hyvin.

Toinen tapa, jota käytin testaamiseen oli kolmannen osapuolen tarjoama tunnelipalvelu (Parsec) kahden oman koneeni välillä sekä Testikone 2:n ja henkilökohtaisen pöytäkoneeni välillä. Se toimi yllättävän hyvin vaikka puhelimella jaetun verkon nopeus vaihteli 5 – 27 Mbps välillä. Nopeissa liikkeissä oli havaittavissa merkittävää pikselöitymistä, verkkonopeuden ollessa minimissä. Pelaamisen lisäksi testasin Autodeskin nimistä 3D-suunnitteluun tarkoitettua ohjelmaa. Ohjelma on ammattilaiskäyttöön tarkoitettu ja graafisesti melko vaativa, kuitenkin suurta eroa ei eri testilaitteiden välillä ollut havaittavissa.

Testikoneet

Testaamisessa vertailukohteena käytin henkilökohtaista pöytäkoneettani, jonka speksit ovat:

CPU: AMD Ryzen 7 3700X 4.0 GHz (8core)

GPU: AMD Radeon RX 5700 8176 MB GDDR6

RAM: 16GB

Testikone 1

SAMK:n palvelimelta saatu virtuaalikone:

CPU: Intel Xeon Gold 6144 3.5GHz (2core)

GPU: NVIDIA Tesla P40 (2 core)

RAM: 16GB

Testikone 2

Lounean palvelimelta saatu virtuaalikone:

CPU: AMD EPYC 7702P 64-Core Processor 2.0GHz (1core)

GPU: NVIDIA Tesla T4

RAM: 8GB

Nvidia GeForce NOW

CPU: Intel CC150 CPU 3.50GHz

GPU: NVIDIA RTX T10-8

RAM: 7GB

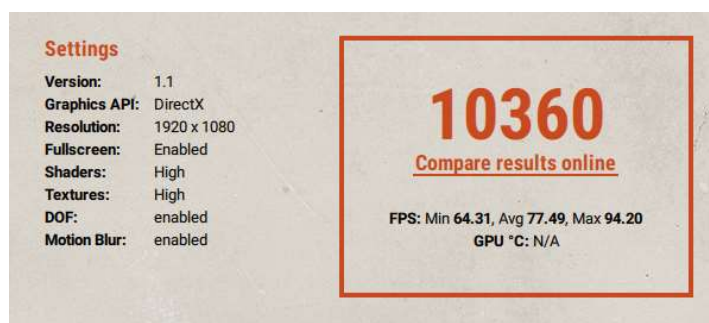
Verkkonopeutena toimi Lounean 1000/100Mbps kuituliittymä.

Ohjelmistot, joita käytin testeihin ovat MSI Afterburner, joka tallentaa tietoja kuvantaajuuden arvoista, Superposition niminen testipenkkiohjelma ja Parsec, jota käytin ottamaan yhteyttä Testikone 2:n

6.1 Penkitys

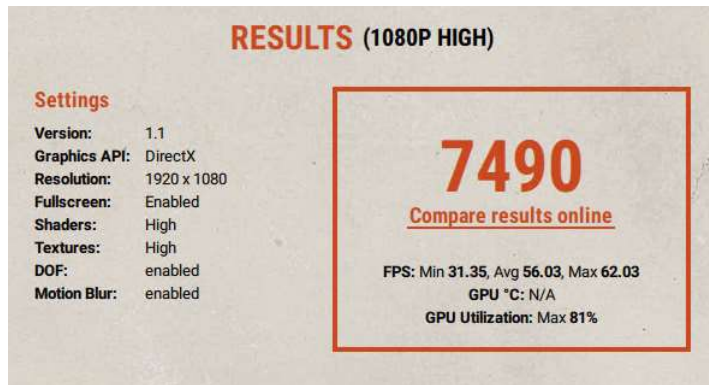
Ensimmäisenä testinä jokaiselle koneelle tein Superposition-nimisen testiajon, joka testaa koneen graafista suorituskykyä. Kaikki testit tehtiin samoilla ohjelmallisilla säädöillä. Tulokset alla.

Vertailukone:



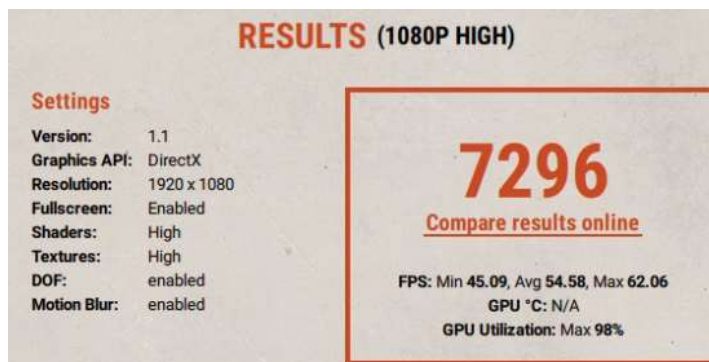
Kuva 6. Vertailukoneen testitulokset: Superposition

Testikone 1:



Kuva 7. Testikone 1:n testitulokset: Superposition

Testikone 2:



Kuva 8. Testikone 2:n testitulokset: Superposition

6.2 Pelitestausta

Seuraavaksi testasin koneita pelikäytössä kolmen eri pelin kanssa. Testikohteiksi valikoituivat Witcher 3: Wild Hunt ja Counter-Strike: Global Offensive, sillä ne löytyivät omasta pelikirjastostani sekä GeForce NOW:n valikoimista. Lisäksi ne ovat suosittuja ja täten hyviä testipenkkejä koneille. Tässä kohtaa testausta valitettavasti Testikone 1 ei ollut enää käytettävissä mutta sen tilalle testaukseen tuli GeForce NOW-palvelun ilmainen ja rajoitettu versio. GeForce NOW palveluun ei ollut mahdollista asentaa testiohjelmaa, joten kuvataajuudet ovat itse seuraamiani ja laskemiani Steamian kuvantaa-juusseurantalisisäosan avulla. Lisäksi Testikone 2:een otan yhteyden Parsec-ohjelman kautta, sillä se tukee suoraa hiiren syötteitä. Tämä ei kuitenkaan ole paras tapa yhdistää koneita sillä testatessa huomasin äkillisiä kuvanlaadun laskuja. Tuloksissa käyn läpi testipenkin tuloksen ja sen jälkeen omia kokemuksiani testauksen kulusta.

6.2.1 Counter Strike: Global Offensive

Grafiikka-asetukset:

Default High

Vertailukone:

Average framerate: 326.0 FPS

Minimum framerate: 117.3 FPS

Maximum framerate: 380.9 FPS

1% low framerate: 148.1 FPS

0.1% low framerate: 29.1 FPS

Peli toimi moitteettomasti eikä kuvavirheitä tai syöttöviivettä ollut.

Testikone 2:

Average framerate: 61.8 FPS

Minimum framerate: 57.5 FPS

Maximum framerate: 61.9 FPS

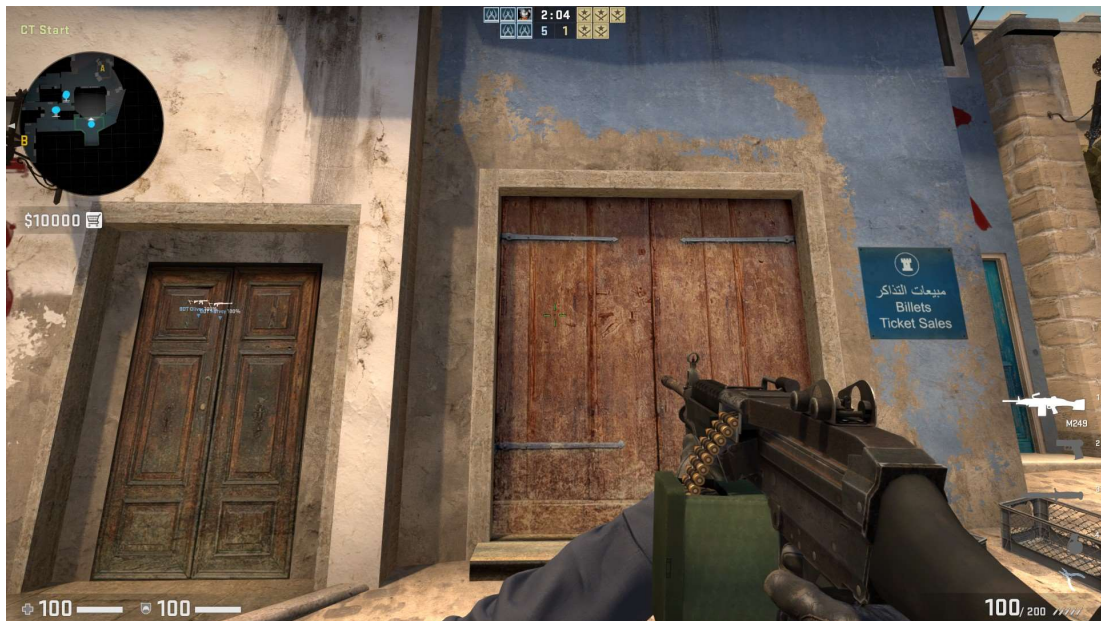
1% low framerate: 58.5 FPS

0.1% low framerate: 25.4 FPS

Nopeissa liikkeissä ja äkillisissä kuvamuutoksissa kuten valokranaatin kohdalla pientä pikselöitymistä joka meni nopeasti ohi ja johtui luultavimmin enemmän Parsecista kuin itse koneesta. Syöttöviivettä ei huomattavissa.

GeForce NOW: Kuvataajuus pysyi pitkälti yli 180 fps. Syöttöviive oli olemattoman pieni.

Eri testikohteissa ei ollut suurta eroa grafiikan laadun kanssa kuten seuraavista kuvista näkyy.



Kuva 9 Kuvakaappaus Testikone 2



Kuva 10 Kuvakaappaus GeForce NOW



Kuva 11 Kuvakaappaus Vertailukone

6.2.2 The Witcher 3: Wild Hunt

Peli on graafisesti raskaampi kuin Counter Strike, joten odotin suurempia eroja koneiden välillä. Kaikki testikoneet kuitenkin toimivat kiitettävästi, vaikkakin joitain ongelmia nousi esiin. Valitettavasti Testikone 2:sta päättyi Nvidian testilisenssi enkä täten saanut kuvakaappausta.

Grafiikka-asetukset:

Default High

Vertailukone:

Average framerate: 94.2 FPS

Minimum framerate: 82.6 FPS

Maximum framerate: 110.9 FPS

1% low framerate: 69.6 FPS

0.1% low framerate: 52.2 FPS

Testikone 2:

Hiiren ja näppäimistön kanssa pelattaessa syöttöviive oli huomattava, mutta ohjainta käytettäessä se muuttui olemattomaksi. Kuvataajuus vaihteli suuresti aiheuttaen pikselöitymistä varsinkin nopeissa liikkeissä.

Average framerate: 77.4 FPS

Minimum framerate: 76.4 FPS

Maximum framerate: 77.6 FPS

1% low framerate: 69.3 FPS

0.1% low framerate: 68.0 FPS

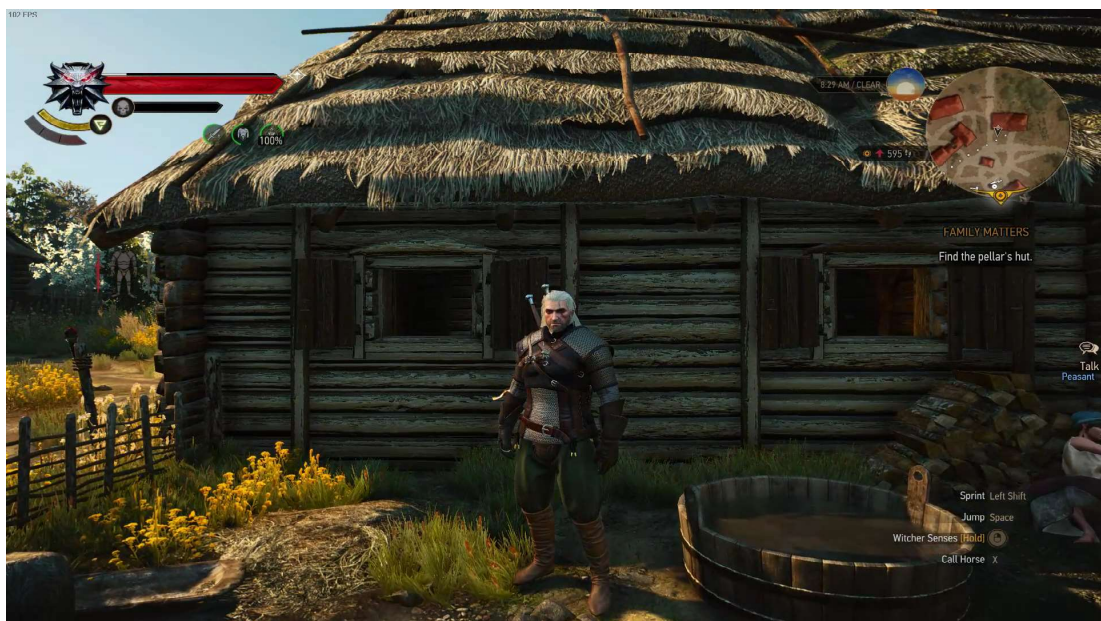
GeForce NOW:

Syöttöviive oli pieni ja kuvataajuus pysyi hyvin yli 70. Paikoittaista pikselöitymistä oli havaittavissa.

Kuvavertailussa ei suuria eroja näkynyt.



Kuva 12. Kuvakaappaus Vertailukone.



Kuva 13. Kuvakaappaus Nvidia

7 LOPUKSI

Tässä työssä tutkin pilvipalveluita yleensä ja tarkennettuna niihin liittyvää 3D-kiihdytystä. Tein tämän käymällä ensin läpi yleistä teoriaa, jonka jälkeen vertailin eri kaupallisia pilvipelaamiseen tarkoitettuja palveluita, niiden ominaisuuksia ja eroja. Vertasin myös keskenään kahta suurinta grafiikkakorttien valmistajaa Nvidiaa ja AMD:tä, niiden tarjoamia grafiikkakortteja ja palveluita.

Testauksessa käytin Satakunnan ammattikorkeakoululta ja Lounea Palveluilta saamiini virtuaalikoneita, joihin otin yhteyden Parsec-nimisellä ohjelmalla sekä Nvidian kaupallista pilvipelaamiseen tarkoitettua palvelua ja vertasin niiden graafisia ominaisuuksia pelaamiseen tarkoitettuun pöytäkoneeseeni. Ohjelmina käytin Parsecin lisäksi Superposition-nimistä testipenkkiä, jolla sain tietooni koneiden maksimaalisen graafisen tehon, MSI Afterburneria, joka seuraa ja tallentaa lokiin virkistystaajuuden, jolloin koneiden erot pelaamisessa tulivat ilmi. Testipeleinä toimivat Counter-Strike Global Offensive sekä The Witcher 3: Wild Hunt. Testeissä tuli ilmi, kunhan verkko on tarpeeksi nopea ja luotettava, toimivat virtuaalikoneet pelaamisessa yllättävänkin hyvin ja hävisivätkin vain niukasti kaupalliselle palvelulle ja perinteiselle pelikoneelle.

Jos verrataan keskiluokan pelitietokonetta, joka on hintaluokaltaan n.2000€ ja omalla kohdallani koneen elinkaari on noin viisi vuotta, tulee kuukausihinnaksi noin 33€. Viiden vuoden jälkeen, ellei aikaisemmin, koko kone tai osia siitä täytyy päivittää, jolloin kustannuksia tulee lisää. Virtuaalikoneella nämä kustannukset pysyvät kiinteinä eikä yllättäviä menoeria tule, sillä palveluntarjoaja vastaa palvelun toimivuudesta ja päivittämisestä. Tämä tekee mielestäni virtuaalikoneen varteenotettavaksi vaihtoehdoksi sekä yksityisille, että yrityspuolelle, jolla laitteistohinnat ovat vielä huomattavasti korkeammat.

Päätelmänä tästä tutkimuksesta voin sanoa, että mielestäni 3D-kiihdytettyjen virtuaalikoneiden tarjoaminen markkinoille olisi kannattavaa, varsinkin valmiin arkkitehtuurin ollessa olemassa. Vaikka tässä työssä painotus olikin yksityispuolen pelaamisella, on se kuitenkin yksin pieni ja epävakaa markkina-alue. Siksi mielestäni suurempi painotus kannattaakin suunnata yritysten tarpeisiin.

LÄHTEET

AMD. Radeon Instinct™ Accelerators 2020. Viitattu 22.4.2020.

<https://www.amd.com/en/graphics/servers-radeon-instinct-mi>

AMD. AMD FirePro™ S-Series for Virtualization 2016. Viitattu 7.5.2020.

<https://www.amd.com/system/files/documents/firepro-s-series-datasheet.pdf>

AMD. Radeon™ Pro for Virtualization 2020. Viitattu 7.5.2020.

<https://www.amd.com/en/graphics/workstation-virtual-graphics>

AMD. AMD MxGPU and VMware 2020. Viitattu 7.5.2020. https://drivers.amd.com/relnotes/amd_mxgpu_deploymentguide_vmware.pdf

Chris. The Truth About Cloud Gaming - Performance & Input Lag Analysis 2018.

Viitattu 22.11.2019. https://www.youtube.com/watch?v=eY_zjGAXs_8

Citrix. Mitä työpöytä palveluna (DaaS) tarkoittaa? Viitattu 18.5.2020.

<https://www.citrix.com/fi-fi/glossary/what-is-desktop-as-a-service-daas.html>

Dov Z. Welcome to Stadia 2019. Viitattu 28.1.2020. <https://stadia.dev/blog/welcome-to-stadia>.

Google. Stadian sivut 2020. Viitattu 22.1.2020. <https://store.google.com/fi/product/stadia>

Google. Stadia FAQ 2019. Viitattu 28.1.2020. <https://support.google.com/stadia/answer/9338946?hl=en>

Hong H, Chen D, Huang C, Chen K, Hsu C. Placing Virtual Machines to Optimize

Cloud Gaming Experience 2019. Viitattu 12.2.2020. https://www.iis.sinica.edu.tw/~swc/pub/virtual_machine_allocation_for_cloud_gaming.html

Mell P, Grace T. The NIST Definition of Cloud Computing 2011. Viitattu

11.10.2019. <https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf>

Mäntylä J-M. Unohda konsolit ja pc – pian ohjain kytketään suoraan televisioon tai älypuhelimeen. Viitattu 14.1.2020. <https://yle.fi/uutiset/3-10998455?fbclid=IwAR18TkWDdu5SKTk2Y6yTBoV7glPCs-MhetGtBD1fiMIUVg4tatQ96KIE5Uk&utm>

<https://yle.fi/uutiset/3-10998455?fbclid=IwAR18TkWDdu5SKTk2Y6yTBoV7glPCs-MhetGtBD1fiMIUVg4tatQ96KIE5Uk&utm>

Nvidia. Gforce Now-sivut 2020. Viitattu 22.1.2020. <https://www.nvidia.com/fi-fi/geforce/products/geforce-now/>

Nvidia. Nvidia Tesla P100. 2018. Viitattu 30.3.2020. <https://www.nvidia.com/en-us/data-center/resources/pascal-architecture-whitepaper/>

Nvidia. ACCELERATED COMPUTING AND THE DEMOCRATIZATION OF SUPERCOMPUTING. 2018. Viitattu 6.4.2020. <https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/tesla-product-literature/sc18-tesla-democratization-tech-overview-r4-web.pdf>

Nvidia. Tesla V100. 2018. Viitattu 30.3.2020. <https://images.nvidia.com/content/technologies/volta/pdf/tesla-volta-v100-datasheet-letter-fnl-web.pdf>

Nvidia. NVIDIA VIRTUAL GPU PACKAGING, PRICING AND LICENSING. 2019. Viitattu 20.4.2020. <https://images.nvidia.com/content/grid/pdf/Virtual-GPU-Packaging-and-Licensing-Guide.pdf>

Nvidia. NVIDIA TURING GPU ARCHITECTURE 2018. Viitattu 19.04.2020. <https://www.nvidia.com/content/dam/en-zz/Solutions/design-visualization/technologies/turing-architecture/NVIDIA-Turing-Architecture-Whitepaper.pdf>

O7services. IAAS SAAS PAAS: Difference With Example And How To Choose 2019. Viitattu 11.10.2019. <https://o7services.com/blog/2019/01/03/iaas-saas-pass/>

Rouse M. Desktop as a Service (DaaS) 2017. Viitattu 29.1.2020. <https://searchvirtualdesktop.techtarget.com/definition/desktop-as-a-service-DaaS>

Sony. PlayStation Now-sivut 2020. Viitattu 22.1.2020. <https://www.playstation.com/fi-fi/explore/playstation-now/>

