

Tieteellisten laskentaresurssien ja niiden oheispalvelujen kokonaisarkkitehtuuri

Tieteellisten laskentaresurssien ja niiden oheispalvelujen kokonaisarkkitehtuuri	1
1 Yleistä	3
1.1 Laskennallinen tiede ja tieteellinen laskenta.....	3
1.2 Rajaukset ja reunaehdot	4
1.3 Nykytila.....	4
1.3.1 Laskennallinen tiede Suomessa	4
1.3.2 Tieteellisen laskennan käyttäjät	6
1.3.3 Resurssien ja käyttäjien (hallintoalueiden) väliset raja-aidat Suomessa.....	9
1.3.4 Laskentaresurssit.....	10
1.4 Tavoittila	14
2 Tieteellisen laskennan toiminta-arkkitehtuuri	16
2.1 Strategia	16
2.1.1 Tieteellisen laskennan strategia	17
2.2 Toiminnan haasteet ja tavoitteet	18
2.2.1 Keskitettyjen resurssien aidosti kansallinen käyttö.....	18
2.2.2 CSC:n in house -asema	18
2.2.3 Teknologian kehityksen aiheuttamat haasteet.....	18
2.2.4 Laitteistohankinnat ja niiden rahoitus	19
2.2.5 Kansainvälisen käytön ja tukipalveluiden rahoitus.....	20
2.2.6 Kokonaisuoptimointi ja kustannustietoisuus.....	21
2.2.7 Tieteellinen laskenta teollisuudessa	21
2.3 Palvelut	22
2.3.1 Laskentaresurssit.....	22
2.3.2 Virtualisoidut laskentaresurssit	23
2.3.3 Korotetun tietoturvatason laskenta.....	24
2.3.4 Käyttäjätuki	24
2.3.5 Koulutus	25
2.3.6 Ohjelmisto- ja tietokantapalvelut	25
2.3.7 Kriittiset palvelut muualta.....	26
2.4 Toimijat ja sidosryhmät	26
2.4.1 Palveluiden tuottajat.....	26
2.4.2 Loppukäyttäjät	27
2.4.3 Rahoittajat	27
2.4.4 Ohjaus	28
2.5 Organisaatio	28
2.5.1 CSC	29
Yliopistot	30
2.5.2 Ammattikorkeakoulut	30
2.5.3 Tutkimuslaitokset.....	30
2.6 Prosessit	31
2.6.1 Resurssien jako	31
2.6.2 Käyttäjähallinta	31
2.6.3 Resurssien ylläpito ja uudistaminen.....	33
3 Tieteellisen laskennan tietoarkkitehtuuri.....	33

3.1	Käsitteitä	34
4	Arkkitehtuuriperiaatteet	37
4.1	Sidosarkkitehtuurit	37
4.2	Yleiset arkkitehtuuriperiaatteet	38
4.3	Toiminnalliset arkkitehtuuriperiaatteet	39
4.4	Tietoturva- ja tietosuojaperiaatteet	39
5	Liitteet	39
5.1	Lait ja säädökset	39
5.2	CSC:n tämän hetkisten koneiden tekniset yksityiskohdat	39

LUONNOS

Tiivistelmä

Johdanto

Tieteellinen laskenta on yksi Suomen tieteen suurista vahvuusalueista. Sen monella osa-alueella on Suomessa jo kauan ollut aktiivisia tutkimusryhmiä, jotka ovat hyvin korkeasti arvostettuja maailmalla. Alalla on ollut useita Akatemian huippuyksiköitä ja akatemiaprofessoreita. Pitkään jatkunut tutkimus on myös luonut vahvan kansallisen, yhteistyön kautta toisiaan tukevan tutkimusyhteisön joillekin osa-alueille. Tämän on mahdollistanut korkeatasoisten tutkijoiden 1970-luvulta alkaen luoma hyvä tutkimusilmapiiri sekä CSC:n tarjoama kansainvälisestikin erittäin korkea määrä laskentakapasiteettia ryhmää kohden.

Lisääntyvän digitalisaation myötä tieteellinen laskenta laajenee perinteisten laskennallisten tieteiden alalta kovaa vauhtia muihin tieteisiin ja esimerkiksi dataintensiivisen laskennan menetelmiä sovelletaan jo rutiininomaisesti mm. kielitieteisiin ja biologiaan.

Tieteellisten laskentaresurssien ja niiden oheispalveluiden kokonaisarkkitehtuuri kuvaa suomalaisten laskennallisen tieteen ja tutkimuksen käytössä olevaa laskenta- ja tallennuskapasiteetista, ohjelmistoista, tietokannoista ja niihin liittyvästä käyttäjätuesta ja -koulutuksesta muodostuvaa kokonaisuutta (jota jatkossa kutsutaan yksinkertaisemmin tieteelliseksi laskennaksi).

Kartturi-mallin mukainen kokonaisarkkitehtuuri käsittää toiminta-, tieto-, tietojärjestelmä- ja teknologia-arkkitehtuurien yksityiskohtaiset kuvaukset. Tätä dokumenttia laadittaessa kokonaisarkkitehtuuriperiaatteita on kuitenkin noudatettu vain löyhästi ja keskittyen toiminta-arkkitehtuuriin. Tavoitteena on ollut selkeä ja käytännönläheinen lopputulos, jota voidaan ajatella myös epämuodollisesti tieteellisen laskennan kokonaiskuvana.

Koska suomalaiset tieteellisen laskennan toimijat muodostavat itsenäisiä kokonaisuuksia ilman yhteistä hallintoa, arkkitehtuuri on luonteeltaan ohjaava eikä sitova.

Kokonaisarkkitehtuurin keskeisimmät osakokonaisuudet ovat nyky- ja tavoitetilan kuvaukset sekä tieteellisen laskennan strategia.

1 Yleistä

1.1 Laskennallinen tiede ja tieteellinen laskenta

Laskennallinen tiede on tietokoneiden kehityksen myötä vakiintunut kolmanneksi paradigmaksi perinteisten kokeellisten ja teoreettisten menetelmien rinnalle. Laskennallisen tieteen prosessi voidaan karkeasti jakaa neljään vaiheeseen:

- 1) Tutkittavasta ilmiöstä tehdään matemaattinen malli.
- 2) Matemaattiselle mallille etsitään numeerinen ratkaisumenetelmä, jota kutsutaan usein algoritmiksi.
- 3) Numeerinen algoritmi ohjelmoidaan ja ajetaan tietokoneella. Tätä kutsutaan usein simuloimiseksi.

- 4) Tulosten tulkinta ja havainnollistaminen, mikä tyypillisesti vaatii lisää laskentaa tietokoneella.

Erityisesti kohdat 2) ja 3) karakterisoivat laskennallista tiedettä. Laskennallisen tieteen ja tietokonesimulaatioiden avulla voidaan tarkastella ilmiöitä, joiden tutkiminen kokeellisesti on liian hidasta, kallista, vaarallista tai yksinkertaisesti mahdotonta. Puhtaasti teoreettinen lähestymistapa muuttuu laskennalliseksi tieteeksi, kun käytettävät matemaattiset mallit tulevat niin monimutkaisiksi, että niitä voidaan analysoida vain numeerisesti.

Tieteellinen laskenta käsittää pääasiassa edellä olevan listan kohdan 3), mutta käytettävissä olevan tietokoneen tekniset ominaisuudet voivat vaikuttaa kohtaan 2) eli algoritmin valintaan.

Edellä esitetyn laskennallisen tieteen mallin rinnalle on noussut neljänneksikin paradigmaksi kutsuttu dataintensiivisen tieteen malli. Neljäs paradigma yhdistää kokeellisen tieteen havaintoihin ja niistä kerättyyn dataan perustuvan tutkimuksen ja laskennallisen tieteen tavan käyttää tietokonekapasiteettia tutkimuksessa. Neljännen paradigman mukaista laskentaa kutsutaan usein dataintensiiviseksi laskennaksi.

Tässä dokumentissa tarkastellaan tieteellisten laskentaresurssien ja niiden oheispalvelujen kokonaisarkkitehtuuria tai -kuvaa, jonka keskeisimmät osakokonaisuudet ovat:

- 1) Laskentakapasiteetti käyttöympäristöineen.
- 2) Oheispalvelut kuten tallennuskapasiteetti, ohjelmistot, aineistot ja tietokannat jne.
- 3) Laskentakapasiteetin käyttöön ja ohjelmointiin liittyvä käyttäjätuki.
- 4) Edellisiin liittyvä koulutus.

Jatkossa tieteellisiin laskentaresursseihin ja niiden oheispalveluihin viitataan yksinkertaisten termillä tieteellinen laskenta.

1.2 Rajaukset ja reunaehdot

- Arkkitehtuuri ei ole sitova vaan ohjaava.
- Toimijaryhmät ovat mukana edustajien kautta, ja yksittäisen toimijan näkemys saattaa poiketa työryhmän suosituksista.
- Yrityksissä tehtävää tieteellistä laskentaa ei käsitelty yksityiskohtaisesti.
- Arkkitehtuuriperiaatteita on noudatettu löyhästi menemättä liian syvälle yksityiskohtiin; tavoitteena on mahdollisimman selkeä kokonaiskuva.
- Tarkastelujakso: supertietokoneiden käyttöikä on noin 3 vuotta ja EU-infrastruktuurien rahoitusinstrumenttien näkymä H2020-ohjelman kautta on noin viiden vuoden päässä, joten tarkastelujaksossa on pyritty keskittymään seuraavaan 5 vuoteen.

1.3 Nykytila

1.3.1 Laskennallinen tiede Suomessa

Laskennallinen tiede (tieteellinen laskenta) perustuu tieteellisten ongelmien ratkaisemiseen matemaattisten mallien ja kvantitatiivisten analyysimenetelmien avulla, hyödyntäen tehokkaasti tietokoneiden antamaa lisäarvoa. Olennaisesti laskennallinen tiede integroituu

parhaillaan kaikkialle – esimerkiksi kännyköihin, paikantamiseen, terveydenhoitoon ja energiantuotantoon, ja sen ohessa mitä luontevimmin uuden tieteellisen tiedon tuottamiseen.

Suomessa laskennallinen tiede kattaa oleellisesti kaikki tieteen alat (mm. fysikaaliset ja kemialliset tieteet, biologiset ja lääketieteet, teknisten tieteiden laajan kirjon, yhteiskuntatieteet mukaan lukien taloustieteet, ja menetelmien kehittämisen). Laskennalliselle tieteelle on vahvasti ominaista sen poikki- ja monitieteinen luonne. Yhä kasvavassa määrin laskennallisen tieteen tutkimus integroidaan yhteen kokeellisen tieteen kanssa, niiden molempien vaikutusta kasvattaen. Samassa hengessä laskennallinen tiede on Suomessa saanut merkittävän aseman eri tieteenalojen välisten rajapintojen kaatajana, sillä hyvin usein laskennallisen tieteen malleja ja menetelmiä käytetään poikkitieteellisten ongelmien ratkaisemiseen. Esimerkkeinä tästä suuntautumisesta voidaan mainita laskennallisen tieteen alalta mm. ilmakehän tieteet (Helsinki), biologinen fysiikka (Tampere, Oulu), nanotieteet (Espoo, Helsinki, Jyväskylä) ja bioinformatiikka (Helsinki, Kuopio, Tampere).

Mainittakoon lisäksi että Suomen Akatemian vuonna 2012 tekemässä fysiikan tieteenala-arvioinnissa todetaan laskennallisen tieteen olevan Suomessa erittäin korkeatasoista¹. Alalla on myös ollut useita Akatemian huippuyksiköitä ja akatemiaprofessoreita².

Kun katsotaan alan kehitystä viimeisen 10 vuoden ajalta, on huomion arvoista, että johtavien tiedelehtien mielenkiinto laskennallista tiedettä kohtaan on kasvanut systemaattisesti: laskennallista tiedettä sisältävien tutkimusartikkelien suhteellinen osuus kaikista julkaisuista on tieteen huippulehdissä kasvanut noin kaksinkertaiseksi edellisen vuosikymmenen aikana. Tämän trendin voidaan perustellusti arvioida kasvavan – etenkin Suomessa – koska laskennallinen tiede on saanut mm. tuomansa lisäarvon ja suhteellisen edullisuuden ansiosta kasvavan suosion kaikissa suomalaisissa tiedeorganisaatioissa, ja se näkyy konkreettisesti etenkin huipputasoisen tutkimuksen edistäjänä. Verrattuna muihin EU:n jäsenmaihin Suomessa onkin ominaista keskitettyjen (CSC:n) laskentapalvelujen suuri merkitys akateemiselle tutkimukselle. Tästä on seurannut tilanne, että vaikka monella muulla tutkimuksen alueella suomalaisten akateemisten ryhmien mahdollisuudet ovat teknisessä mielessä rajoitetut, on laskennallisissa tieteissä pystytty saavuttamaan kansainvälisesti katsottuna vahva kilpailuasema ilman, että kustannukset tutkijoille olisivat olleet kohtuuttomia. Kansainvälisellä tasolla haettavana olevat tieteellisen laskennan resurssit tukevat tätä kehitystä.

Toinen myönteinen seuraus CSC:n roolista on ollut laskennallisten tieteiden suurempi integraatio muuhun tutkimukseen. Koska keskitetty laskentaresurssien ylläpito ja tieteellisten ohjelmistojen sekä menetelmien asiantuntijatuki mahdollistavat suurteholaskennan käyttämisen jo yksittäisten tutkijoiden tasolla, ei uusien menetelmien käyttöönotto ole ollut ylivoimaista. Tätä taustaa vasten on helposti ymmärrettävissä, että Suomen Akatemian valitsemista tutkimuksen huippuyksiköistä lukuisissa³ on merkittävä laskennallisen tieteen osuus, joka tukee muuta, useimmiten kokeelliseen tutkimukseen pohjautuvaa tutkimusta.

Kattavaa arviota laskennallisen tieteen osuudesta suhteessa kaikkeen suomalaiseen tutkimukseen ei liene vielä tehty, mutta suuntaa antavana arviona voidaan pitää yksittäisten

¹ "Physics Research in Finland 2007-2011 - Evaluation Report", Academy of Finland 08/12, http://www.aka.fi/Tiedostot/Tiedostot/Julkaisut/8_12_Physics%20Research.pdf

² Huippuyksiköt COMP ja CMS, Akatemiaprofessorit Nieminen, Pyykkö.

³ Muun muassa *Center of Excellence in Biomembrane Research*, *Center of Excellence in Atmospheric Science*, *Center of Excellence in Experimental and Computational Developmental Biology Research*, *Finnish CoE in Cancer Genetics Research*, *Finnish CoE in Computational Nanoscience*, *Finnish CoE in Computational Interface Research*, *Finnish CoE in Inverse Problems Research* jne.

yliopistojen tilanteita. Esimerkiksi Tampereen teknillisessä yliopistossa laskennallisen tieteen osuus kaikesta sen tutkimuksesta on karkeasti luokkaa 20–25 %, mitä voidaan pitää merkittävänä osuutena ja korkean vaikutuksen tuottajana. Ottaen huomioon tieteen kustannuspaineet, vaikutuksen kasvuun kohdistuvat tarpeet ja laskennallisen tieteen kustannustehokkuuden, laskennallisen tieteen suhteellisen roolin kasvua osana Suomen tieteen tulevaisuutta ja yritystoimintaa voi pitää luontevana kehityskulkuna.

Yleisesti laskennallisen tieteen asema Suomessa on tällä hetkellä vahva, suorastaan erinomainen. Tämän tilanteen taustalla on se erinomainen tuki, jota etenkin CSC on tukiorganisaationa laskennalliselle tieteelle ja sen kehitykselle antanut. Jotta laskennallinen tiede yhtenä suomalaisen tieteen keihäänkärkiä ja korkean vaikutuksen tuottajana kehittyisi myös jatkossa, on CSC:n toiminnan tukeminen ja edelleen vahvistaminen välttämätöntä.

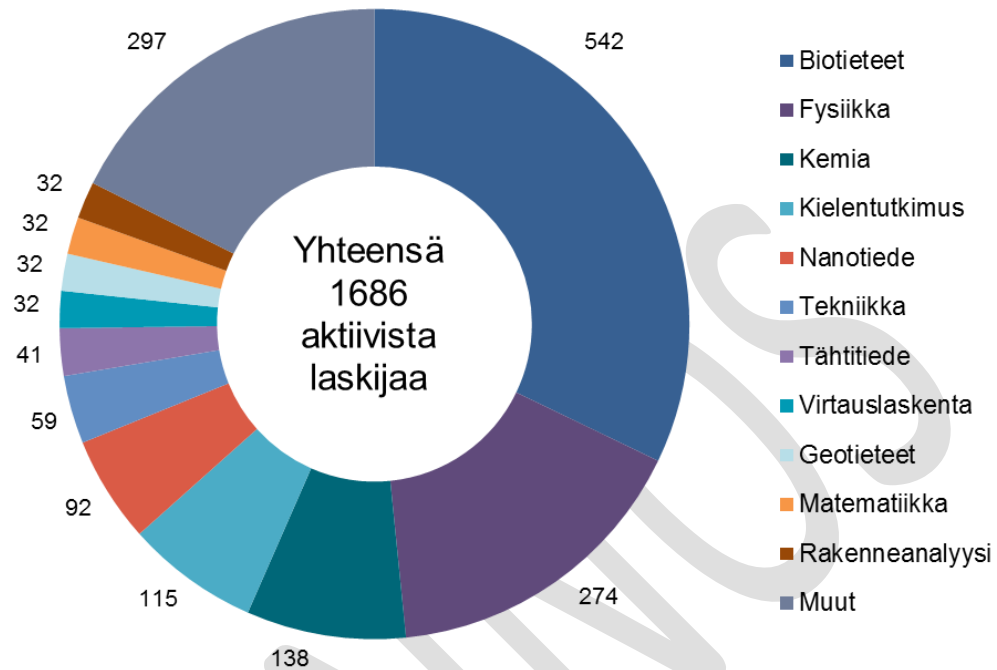
Dataintensiivisen laskennan yleistymisen tukee omalta osaltaan laskennallisen tieteen kasvua.

1.3.2 Tieteellisen laskennan käyttäjät

Yliopistot, ammattikorkeakoulut ja tutkimuslaitokset tuottavat laskentaresursseja lähinnä sisäiseen käyttöön, josta ei kattavia tilastoja ole saatavilla.

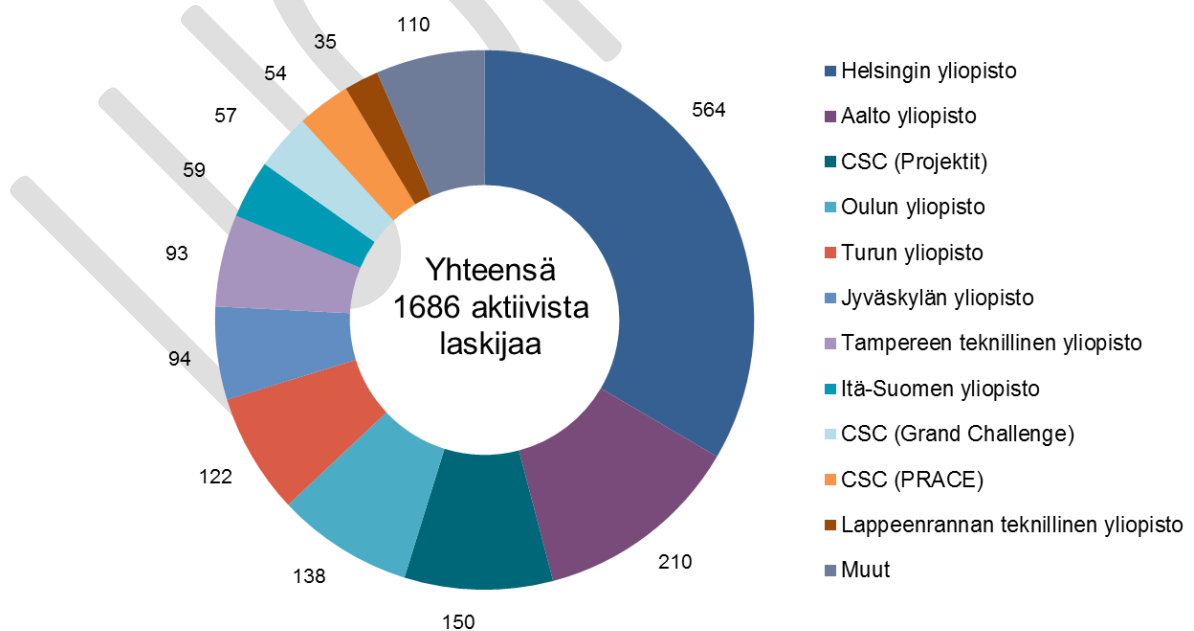
CSC:n keskitettyjä tieteellisen laskennan resursseja käytetään pääasiassa yliopistoissa, ja pienessä määrin myös ammattikorkeakouluissa ja tutkimuslaitoksissa. Käyttö jakaantuu monen tieteenalan kesken. Lukumääräisesti suurin käyttäjäryhmä on biotieteen tutkijat ennen fyysikoita, kemistejä ja kielentutkimusta (kuva 1). Eniten resursseja puolestaan käyttävät fyysikot ennen nanotieteen, kemian ja biotieteiden tutkijoita (kuva 3). Organisaatioittain suurimmat käyttäjät ovat lukumääräisesti mitattuna Helsingin yliopisto, Aalto-yliopisto ja Oulun yliopisto ja resurssien käytön mukaan Tampereen teknillinen yliopisto, Aalto-yliopisto ja Helsingin yliopisto (kuvat 2 ja 4).

Tietokoneressurssien käyttäjät tiedealoittain 2014



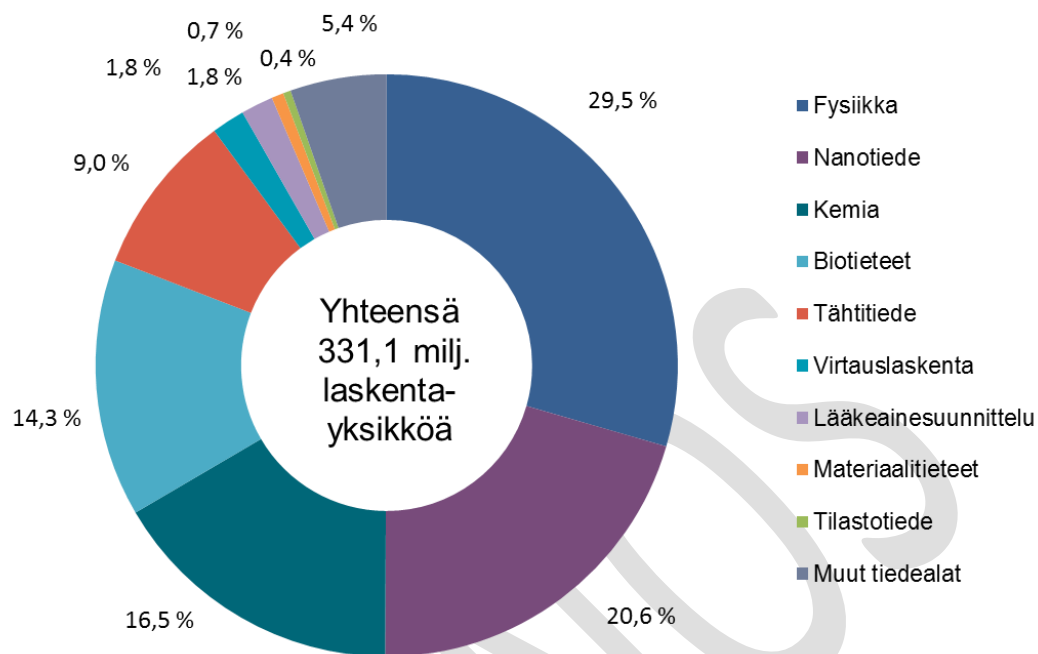
Kuva 1. CSC:n laskenta-asiakkaat tiedealoittain 2014.

Tietokoneressurssien käyttäjät organisaatioittain 2014



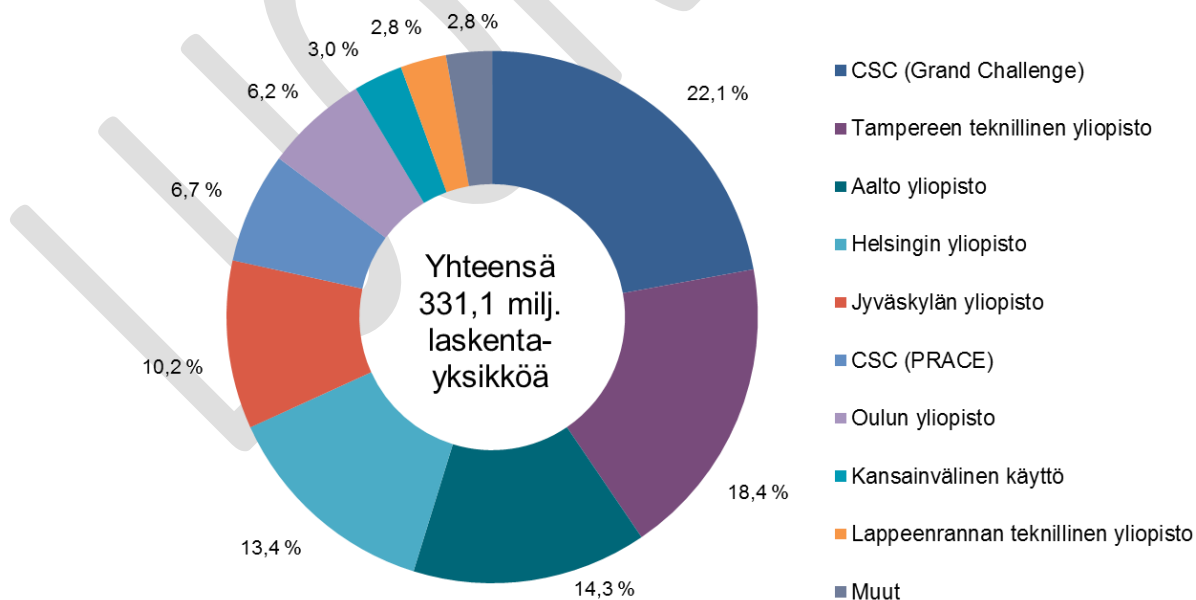
Kuva 2. CSC:n laskenta-asiakkaat organisaatioittain 2014.

Tietokoneressurssien käyttö tiedealoittain 2014



Kuva 3. CSC:n laskentaresurssien (yksi Sisun tai Taidon coreh = kaksi laskutusyksikköä) käyttö tiedealoittain 2014.

Tietokoneressurssien käyttö organisaatioittain 2014



Kuva 4. CSC:n laskentaresurssien (yksi Sisun tai Taidon coreh = kaksi laskutusyksikköä) käyttö organisaatioittain 2014.

Kokonaisuutena tarkasteltuna yliopistojen omien resurssien käyttö noudattaa luultavasti samoja linjoja.

Tutkimuslaitoksista eniten laskenta-aikaa käytetään Ilmatieteen laitoksella, jossa käyttö jakaantuu operatiivisen säänennustamisen ja tutkimuksen kesken. Ilmatieteen laitos on myös käyttänyt huomattavan määrän PRACE Tier-0 -kapasiteettia avaruussään tutkimukseen liittyen.

Tyypillisiä tieteellisen laskennan sovelluskohteita ovat muun muassa molekyyliidynamiikka, elektronirakenteen tutkiminen, plasmafysiikka, ilmastomallinnus, virtauslaskenta ja hiukkasfysiikka.

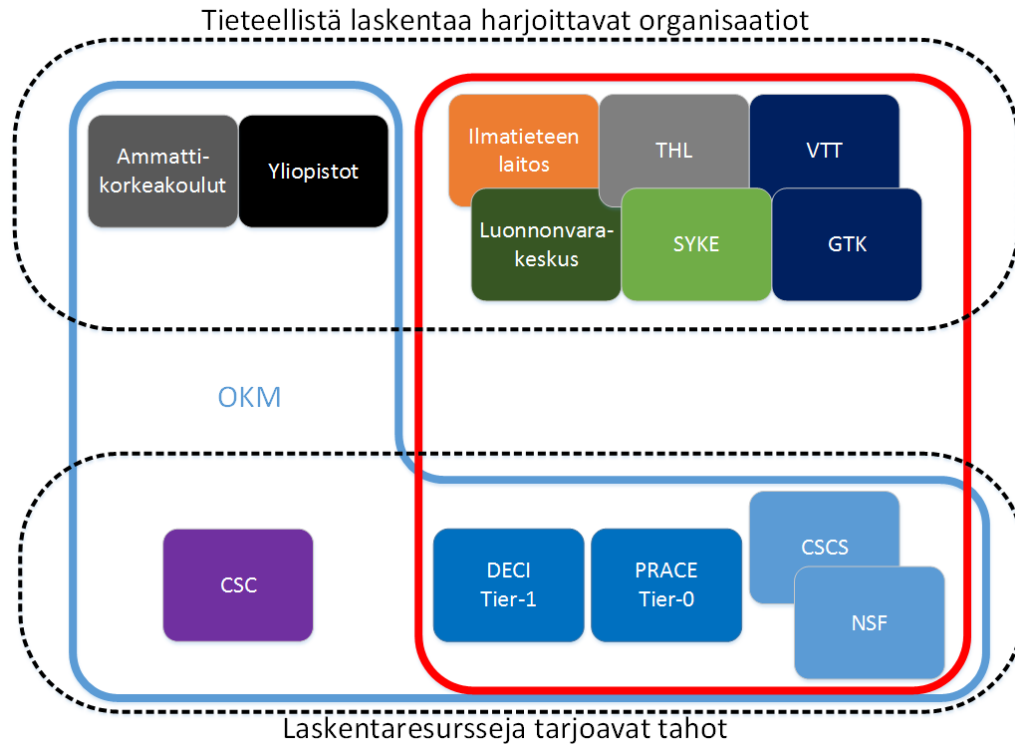
On kuitenkin huomattava, että eri sovellukset tarvitsevat erityyppisiä resursseja: optimaaliset resurssit ovat keskenään hyvin erilaisia CPU- ja muisti-intensiivisille sovelluksille, kun taas jotkut tieteenalat tarvitsevat ennen kaikkea tehokkaita tallennus- ja/tai I/O-resursseja. On lisäksi tieteenaloja, joissa reaaliaikainen tiedon yhdistäminen verkon yli on kiperin ongelma (esim. long baseline interferometry radioastronomiassa).

2014 keväällä avattu cPouta Infrastructure as a Service toi uuden tavan käyttää CSC:n laskenta-resursseja. Käyttäjät voivat valita haluamansa käyttöjärjestelmän ja asentaa ohjelmistoja täysin pääkäyttäjän oikeuksin. Virtuaalikoneisiin voi liittää tallennustilaa laskentaa varten. Projekteista noin puolet tulee biotieteistä, mihin yksi syy on alan ohjelmistokokonaisuuksien ja menettelemien nopea kehitys ja lukumäärä, johon CSC:n perinteisellä yleiskäyttöisellä laskentaklusteri ympäristöllä ei ole pystytty vastaamaan.

Kattava kuvaus laskennallisen tieteen suurteholaskentatarpeista Euroopassa on esitetty viitteessä (PRACE Scientific Case).

1.3.3 Resurssien ja käyttäjien (hallintoalueiden) väliset raja-aidat Suomessa

CSC:n keskitettyjen laskentaresurssien veloitukseton käyttö on mahdollista vain opetusministeriön toimialalle eli yliopistoille ja ammattikorkeakouluille. Kansainvälisten resurssien, kuten PRACEn, käyttö on maksutonta kaikille avointa tutkimusta tekeville toimijoille tutkimuslaitokset mukaan lukien, mutta hakemusten on läpäistävä tieteellinen ja tekninen arviointi. PRACE-käyttö tulee kuitenkin jatkossa edellyttämään lisäksi kansallisen tason kontribuutiota infrastruktuurin operointikuluihin. Tilanne on esitetty kuvassa 5.



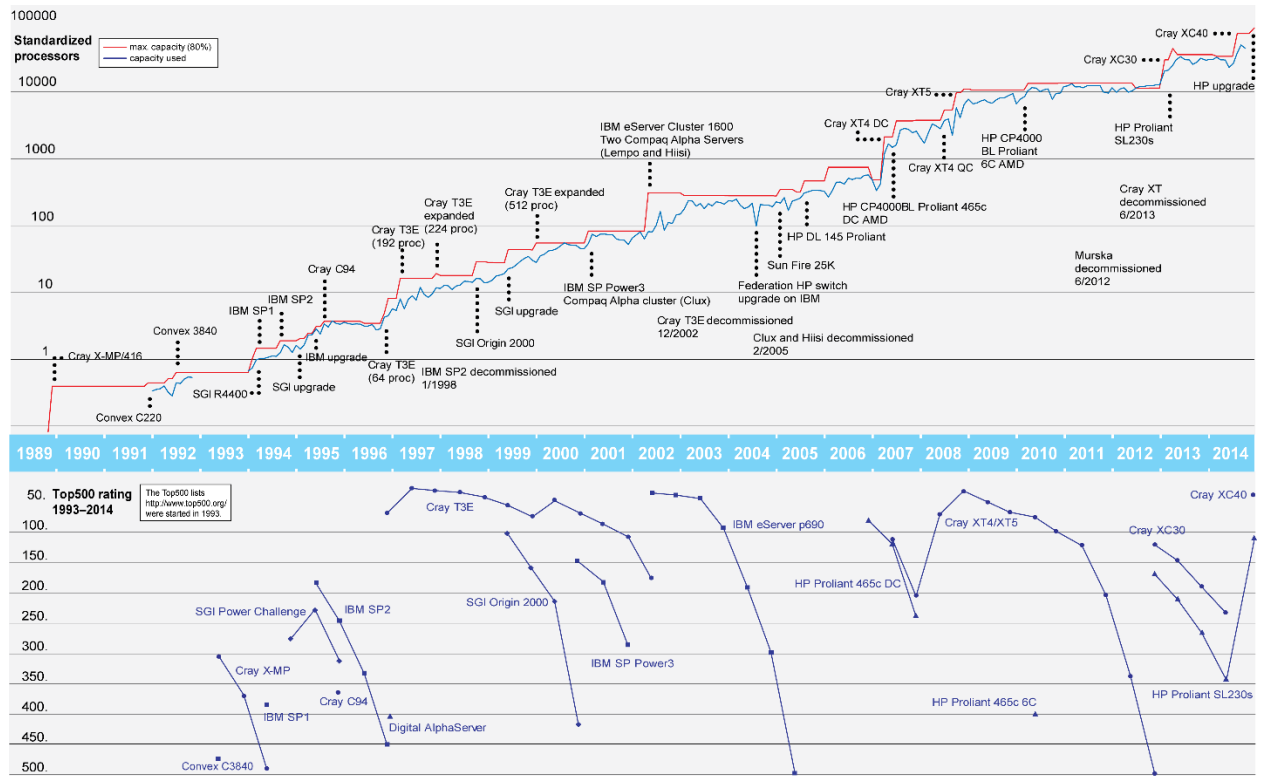
Kuva 5. Tieteellisen laskennan raja-aidat Suomessa.

Kansainvälinen käyttö (suomalaisten tutkijoiden kansainvälinen käyttö ja vice versa) ja yhteistyö kasvaa tulevaisuudessa runsaasti Euroopan yhteisten strategisten tutkimusinfrastruktuurien (ESFRI) edistymisen ja kansainvälistä yhteiskäyttöisyyttä ja uusia tieteellisen laskennan tarvitsemia palveluita valmistelevien H2020-hankkeiden kuten EUDAT ja EGI myötä.

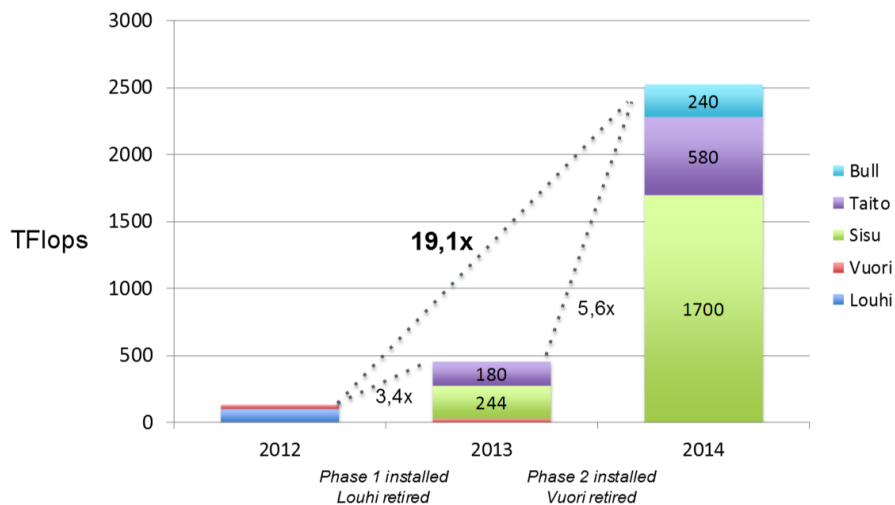
1.3.4 Laskentaresurssit

CSC:n laskentaresurssit

CSC:n laskentaresurssien kasvu vuodesta 1989 lähtien on esitetty kuvassa 6. Vuonna 2014 suoritettu päivitys moninkertaisti kokonaislaskentakapasiteetin aikaisempiin vuosiin verrattuna (kuva 7). Kasvu on pitkällä aikavälillä tarkasteltuna ollut eksponentiaalista, ja kapasiteetti on kaksinkertaistunut keskimäärin alle kahden vuoden välein. Asiakkaat ovat pystyneet hyödyntämään tarjotut resurssit tehokkaasti, ja koneiden käyttöaste on tyypillisesti ollut erittäin korkea ja elinkaaren loppupäässä jonotusajat pitkiä. Koneiden elinkaaren alkupäässä niiden suorituskyky on ollut kansainvälisellä superkoneiden TOP500-listalla viidenkymmenen paremmalla puolella ja tyypillisesti Euroopan ja Pohjoismaiden nopeimpien järjestelmien joukossa.

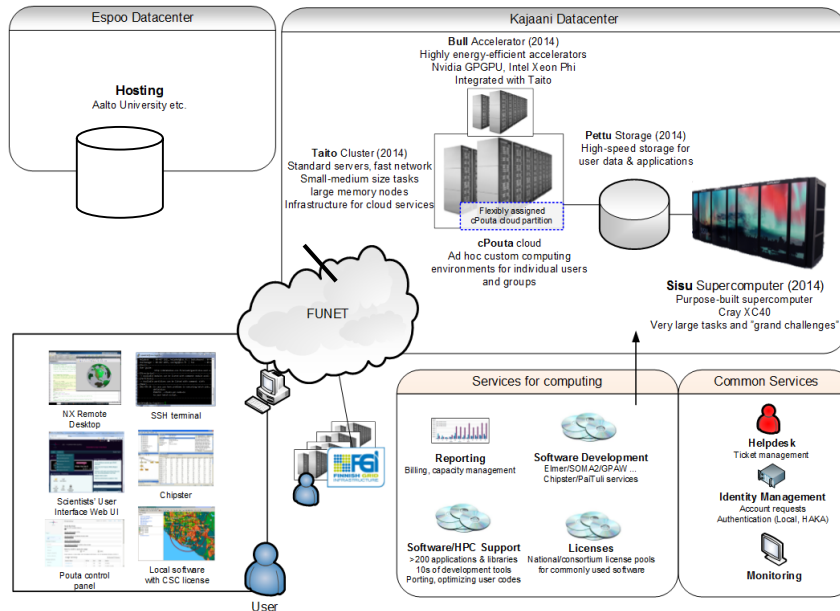


Kuva 6. CSC:n laskentakapasiteetin ja käytön kehitys 1989–2014.



Kuva 7. CSC:n laskentakapasiteetin kasvu 2012–2014.

CSC:n tarjoama suurteholaskentakapasiteetti käsittää tällä hetkellä supertietokone Sisun ja superklusteri Taidon. Erikoisproessoreita ja pilvikapasiteettia (cPouta) tarjotaan osana Taitoa. Lisäksi CSC tarjoaa laskentakapasiteettia sensitiivisen datan analysointiin. Organisaatioasiakkaille ja vaativiin tietoturvaan liittyvää pilvipalvelua (ePouta) pystytetään keväällä 2014 Espoon datakeskukseen. ePouta korvaa 2010 lähtien käynnissä olleen Biomedinfra-pilvipalvelun.



Kuva 8. CSC:n laskentaympäristö 2015 (suunnitelma).

CSC on profiloitunut Sisun isoja rinnakkaisajvoja varten: normaalin käytön puitteissa suurin sallittu työn koko on 9600 ydintä eli suunnilleen neljäsosa koko koneesta. Niin sanotuissa Grand Challenge -hauissa voidaan myöntää oikeudet isompiin ajoihin. Isojen ajojen suosiminen laskee väistämättä koneen käyttöastetta jonkin verran. Pienemmät rinnakkaistyöt ja peräkkäistyöt ajetaan Taidossa, jossa on myös suurempi valikoima CSC:n asentamia ja tukemia tieteellisen laskennan sovelluksia. Sisussa ajetaan lähinnä käyttäjien omia, tieteellisten yhteisöjen kehittämää tai avoimeen lähdekoodiin perustuvia sovelluksia.

CSC:n tämän hetkisten koneiden tekniset yksityiskohdat on kuvattu liitteessä 2.

Sisussa normaalit resurssimyönnot ovat muutamia miljoonia core-tunteja (coreh) ja Grand Challenge -projektille voidaan myöntää suunnilleen 10 miljoonaa coreh. (Sisussa vuodessa teoreettinen maksimi 350 M coreh 100 % käytöllä.)

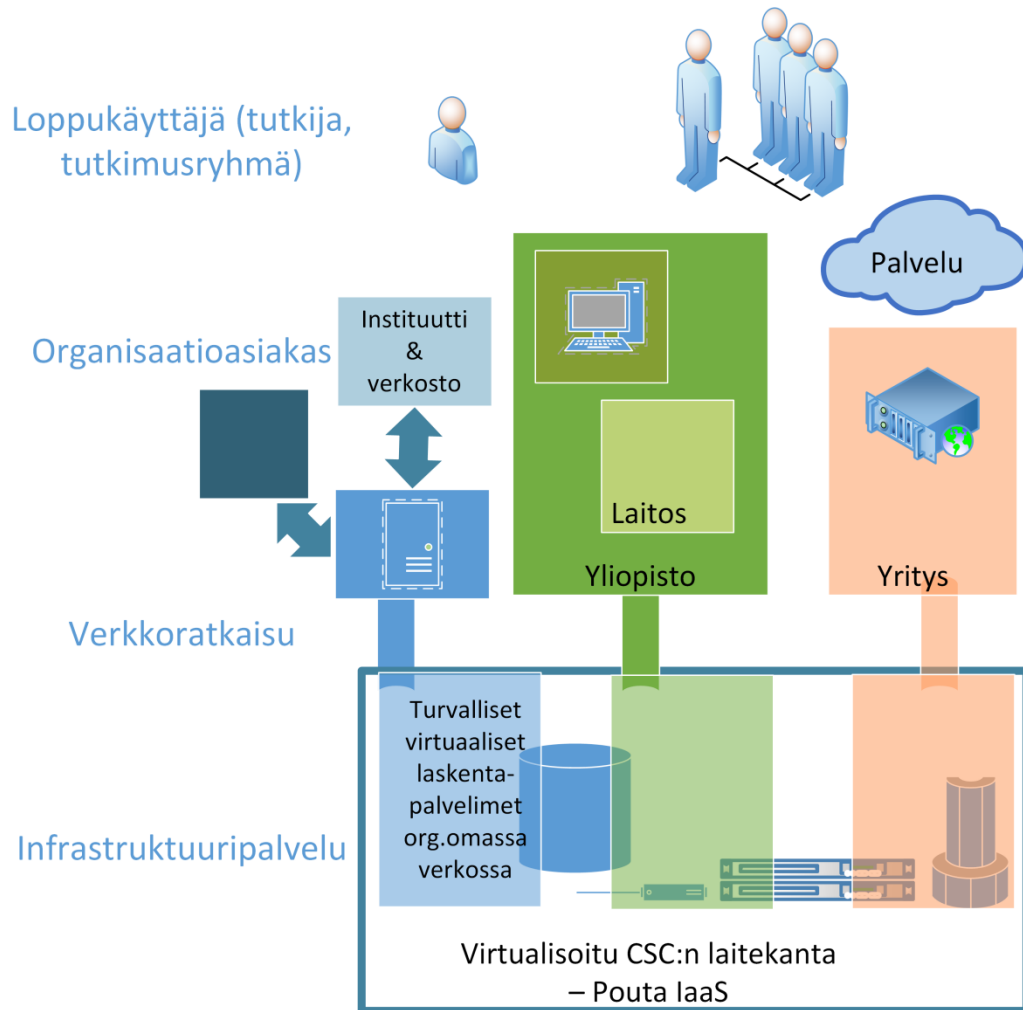
Laskentaresurssit yliopistoissa, ammattikorkeakouluissa ja tutkimuslaitoksissa

Yliopistot ovat tuottaneet omaa laskentakapasiteettia 1960-luvulta lähtien, jolloin Fysiikan laskentatoimisto perustettiin Helsingin yliopistossa. 1990-luvun alulle asti (kun internet ei ollut vielä laajasti käytettävissä) suurin osa laskentaa tehtiin paikallisilla ratkaisulla. 1990-luvulta lähtien CSC:llä on ollut yllä kuvatun mukaisesti ratkaiseva rooli suurimman tehon vaativassa laskennassa, mutta yliopistoilla, ammattikorkeakouluilla ja tutkimuslaitoksilla on edelleen ollut useita paikallisia ratkaisuja.

Yliopistojen, ammattikorkeakoulujen ja tutkimuslaitosten Tier-2-laitteistot ovat useimmiten klustereita. Paikalliset resurssit ovat tärkeä ja oleellinen osa tieteellisen laskennan hierarkiaa; niillä voidaan räätälöidysti vastata tarpeisiin, joiden huomioon ottaminen kansallisella tasolla on vaikeaa tai epätarkoituksenmukaista. Esimerkkeinä tällaisista erikoisvaatimuksista ovat hyvin pitkät ajoajat ja äärimmäisen korkea luotettavuus.

Kun Grid-teknologia, jolla laskentaresurseja voidaan käyttää ja hallita hajautetusti, tuli toimintakäyttöön 2000-luvun alussa, yliopistojen paikallisia klustereita alettiin yhdistää

toisiinsa ja CSC:n kapasiteettiin ensin Materials Grid (MGrid), ja niitä seuraavissa Finnish Grid Infrastructure (FGI) ja Finnish Grid and Cloud Infrastructure -konsortioissa (FGCI). Tällä tavalla saavutettiin merkittäviä synergiaetuja klusterien laskentaresurssien jakamisen ja yhteiskäyttöisyyden suhteen sekä niiden teknisen hallinnan aloilla, kuitenkin samalla säilyttäen paikallisten klusterien edut. Resurssit ovat toimineet myös uusien Grid- ja Cloud-teknologioiden kehitys- ja testausalustana.



Kuva 9. IaaS-pilvipalvelumalli organisaatioille.

CSC on vuosina 2011–2014 kehittänyt organisaatioille suunnattua virtualisointiin pohjautuvaa tieteellistä laskentaratkaisua (Kuva 9). Palvelumalli tuotteistetaan osana Pouta IaaS -pilvipalvelua (ePouta) ja se tulee soveltumaan myös sensitiivisten aineistojen analysointiin. Organisaation sisällä kapasiteetin jakelusta loppukäyttäjille vastaa paikallinen IT-asiantuntija tai -osasto.

Yhteispohjoismaiset resurssit: NeIC

Nordic e-Infrastructure Collaboration NeIC on pohjoismainen hajautettu e-infrastruktuuri⁴. Sen tehtävänä on kehittää ja tuottaa korkeatasoisia laskenta- ja tallennuspalveluja pohjoismaisille

⁴ <http://neic.nordforsk.org/>

tutkijoille erityisesti sellaisilla tutkimusalueilla, joilla pohjoismaisella yhteistyöllä saavutetaan lisäarvoa. NeIC:n taustalla on yhteispohjoismaiset tarpeet hiukkasfysiikan Tier-1-tason keskuksen luomiseksi, mutta toimintaa on sittemmin pyritty laajentamaan myös muille tieteenaloille. NeIC toimii NordForskin alaisuudessa. NeIC on hajautettu organisaatio, jossa itse NeIC vastaa palveluiden kehityksestä ja tuesta, mutta itse laskenta- ja tallennusresurssit ovat kansallisissa keskuksissa. NeIC on siten komplementaarinen perinteiselle suurteholaskennalle.

Kansainväliset resurssit: PRACE ja muut vastaavat resurssit

Jos kansalliset resurssit eivät riitä, tutkija voi hakea laskenta-aikaa PRACElta. Tällä hetkellä PRACEn kuuluu kuusi Tier-0-luokan konetta Saksassa, Ranskassa, Espanjassa ja Italiassa, joista voi saada käyttöönsä kymmeniä miljoonia core-tunteja. Kaikki hakemukset käyvät läpi tiukan tieteellisen ja teknisen evaluaation. Teknisiin vaatimuksiin kuuluu käytettävän ohjelmiston erittäin hyvä rinnakkaissuorituskyky: skaalautuvuuden on ulotuttava tyypillisesti vähintään 10 000 ytimeen asti. Suomalaiset tutkijat ovat menestyneet hyvin PRACE-hauissa: kokonaissaldo vuodesta 2011 lähtien on pitkälti yli 200 miljoonaa core-tuntia.

PRACEn yhteydessä osa mukana olevista keskuksista on tarjonnut tutkijoille Tier-1-aikaa vaihdantaperiaatteella DECI-ohjelmassa. CSC on osallistunut aktiivisesti tämän toiminnan suunnitteluun ja toteuttamiseen, joka on jossain mielessä jatkoa aikaisemmille EU-rahoitteisille DEISA-projekteille.

PRACEn käyttö on toistaiseksi ollut käyttäjille maksutonta, mutta infrastruktuurin seuraavassa vaiheessa keväällä 2015 Tier-0-resurssien suunnitellaan muuttuvan maksullisiksi. Kokonaisarkkitehtuuria laadittaessa rahoitusmalli, jäsenten maksuosuudet ja samalla koko infrastruktuurin tulevaisuus ovat jossain määrin epäselviä. Jos neuvotteluissa saavutetaan kaikkien osapuolien hyväksymä malli toiminnan jatkamiselle, CSC pyrkii hakemaan PRACE-käyttöä varten Suomen Akatemian FIRI-rahoitusta.

Koska Tier-0-tason kapasiteetti on osoittautunut tärkeäksi, sen saatavuus on turvattava PRACEn tulevaisuudesta riippumatta. Vaihtoehtoisia ratkaisuja on useita: avoimiin kansainvälisiin hakuihin osallistuminen, resurssien ostaminen tai vaihtaminen sisäkeskuksilta, eurooppalaisen tai pohjoismaisen konsortion muodostaminen jne.

1.4 Tavoitetila

Yleiset tavoitteet:

- Tieteellisen laskennan kokonaisuus tukee kilpailukykyistä ja kansainvälisesti houkuttelevaa tutkimusjärjestelmää.
- Tieteellisen laskennan tulee mahdollistaa uusia tieteellisiä avauksia.
- Tieteellinen laskenta organisoidaan mahdollisimman kustannustehokkaasti.

Yksityiskohtaisemmat tavoitteet, jotka pitäisi saada toteutetuksi tai turvatuksi seuraavan viiden vuoden kuluessa:

Toiminta on kansainvälistä ja verkostoitunutta

- Suomen laskennallinen tiede houkuttelee kansainvälistä yhteistyötä.

- CSC:n palvelut ja kapasiteetti toimivat yliopistojen kilpailuetuna kansainvälisiä huippututkijoita Suomeen houkuteltaessa.
- Mahdollisimman saumaton liittyminen kansainväliseen Tier-0 - ja grid-kapasiteettiin.
- Tiedonvaihto kansainvälisten toimijoiden kanssa.
- Kansainvälisten ja kansallisten data-aineistojen saatavuus laskentaympäristössä.

Riittävästi kapasiteettia

- Kansallinen kapasiteetti on riittävä kilpailukykyisen ja kansainvälisesti houkuttelevan tutkimusjärjestelmän ylläpitämiseksi.
- Mahdollisuus riittävän suuriin rinnakkaisajoihin, jonotusajat ovat kohtuulliset, riittävä kokonaiskapasiteetti (throughput, työt/aikayksikkö).
- Mahdollisuus suurien data-aineistojen (yli 100 TB) käsittelylle osana laskentaa.
- Aikakriittiselle laskennalle täytyy turvata riittävä palvelutaso (tarkoittaa lähes reaaliaikaisesti prosessoitavaa mittalaitedataa tai operatiivista laskentaa, kuten esim. hiukkasfysiikan laskentaa, meteorologiaa tai muuta tiedettä, jossa tarvitaan keskeytymätöntä 24/7 laskentaa).
- Mahdollisuus sujuvaan siirtymiseen Tier-0-tasolle.

Tarkoituksenmukaiset arkkitehtuurit

- Tehokkaalla kytkentäverkolla varustettu superkone raskaaseen rinnakkaislaskentaan.
- Klusterit peräkkäis- ja pienille rinnakkaistöille sekä dataintensiiviseen laskentaan.
- Suuri kaistanleveys ja matala latenssi laskennan skaalattaviin levyjärjestelmiin
- Suurimuistisia solmuja.
- Virtualisoinnin laajentaminen tiedealoille räätälöityjen sovellusympäristöjen tukemiseksi.
- Tarkoituksenmukainen tietoturva.
- Erikoisproessoreita niitä hyödyntämään pystyviä sovelluksia ja äärimmäistä tehoa vaativaa laskentaa varten.
- Sensitiivisten tutkimusaineiston laskentaan soveltuva, korotetun tietoturvatason järjestelmä.
- Nousevat dataintensiivisen tieteen tarpeet otetaan huomioon hyödyntäen synergioita olemassa oleviin tieteellisen laskennan palveluihin ja infrastruktuureihin.

Korkeatasoiset tukipalvelut: laskentaympäristö, menetelmät, ohjelmointi, ohjelmistot, tiede

- Varmistetaan laitteistojen tehokas hyödyntäminen.
- Tuetaan suomalaisen tieteellisen laskennan ohjelmoinnin ja laskennallisen tieteen sekä yhteiskunnan kansainvälistä kilpailukykyä.

Korkeatasoinen tieteellisen laskennan koulutus

- Varmistetaan laitteistojen tehokas hyödyntäminen.
- Tuetaan suomalaisen tieteellisen laskennan ohjelmoinnin ja laskennallisen tieteen sekä yhteiskunnan kansainvälistä kilpailukykyä.

Riittävä, tehokas ja tarkoituksenmukainen tallennuskapasiteetti

- Nopeaa koneiden ”lähellä” olevaa tallennuskapasiteettia (computable storage).
- Arkistointi, pitkäaikaissäilytys ja jakaminen.

Tarkoituksenmukainen hierarkia ja yhteistyö paikallisten ja keskitettyjen resurssien välillä

- Vältetään päällekkäisyyksiä.
- Keskitetyt ja paikalliset resurssit täydentävät toisiaan.

- Laajan käyttäjäkunnan palvelut keskitetyksi.
- Rajoitetun käyttäjäkunnan (erikois)palvelut koordinoitusti.
- Avoimuus uusia teknologioita kokeiltaan.

Infrastruktuuri on toteutettu kustannustehokkaasti

- Kokonaiskustannukset lasketaan yhteisesti sovittujen periaatteiden mukaisesti.
- Pyritään kokonaistaloudellisesti edullisimpaan ratkaisuun.
- Yhteistyö hankinnoissa.
- Turvattu rahoitus ja rahoitusmalli (jaksotus jne.), joka mahdollistaa pitkän tähtäimen suunnittelun ja tukee kustannustehokkuutta.

Keskitetyt resurssit ovat aidosti kansalliset, ei keinotekoisia raja-aitoja

- CSC:n palvelut ovat maksuttomia julkiselle sektorille.
- Kaikki ministeriöt osallistuvat rahoitukseen.

Yhteinen ohjaus

- Suomen tieteellisen laskennan kokonaisuudella on yhteinen ohjaus.
- Kaikki toimijat edustettuina.
- Kokonaisuuden koordinaatio.

2 Tieteellisen laskennan toiminta-arkkitehtuuri

2.1 Strategia

Tieteellinen laskenta on yliopistoissa ja ammattikorkeakouluissa sekä tutkimuslaitoksissa lähinnä työkalun asemassa ja vain välikappale tieteellisten tulosten tai esimerkiksi sääennusteiden tuottamisessa. Tämän vuoksi tieteellinen laskenta itsessään ei yleensä esiinny näiden toimijoiden strategioissa toisin kuin laskennallinen tiede.

Tieteellisen laskennan palvelujen tuottaminen on keskeinen osa CSC:n toimintaa, ja se esiintyy myös vahvasti CSC:n strategiassa, kuten:

- Korkeakoulut hyödyntävät opetus- ja kulttuuriministeriön tarjoamia ja CSC:n tuottamia palveluita kilpailuetunaan, esimerkiksi menestyvien tutkijoiden houkuttelemiseksi Suomeen.
- CSC kehittää palveluita ja laitteistoympäristöään proaktiivisesti niin, että suomalaisilla korkeakouluilla on käytössään globaalisti ensimmäisten joukossa kansainvälisesti kilpailukykyinen, uusimpia teknologioita hyödyntävä ja helposti saavutettava tutkimuksen tietotekniikkaympäristö.
- CSC mahdollistaa tutkijoiden huippusuoritukset tekemällä kiinteää yhteistyötä kansallisen tietokartan tutkimusinfrastruktuurien kanssa.
- CSC:n tutkimusinfrastruktuuri mukautuu uusien tieteenalojen tarpeisiin ja yhdistyy luontevasti tutkijan arkeen.
- CSC toimii sillanrakentajana kansallisten ja kansainvälisten tutkimusinfrastruktuurien välillä ja laajentaa kansallista tarjontaa eurooppalaisella yhteistyöllä. CSC:n kansainvälisen yhteistyön hedelmät tuovat suoraa hyötyä tutkimusryhmille.
- CSC yhdistää kansallisia ja kansainvälisiä tutkimusinfrastruktuureja ekosysteemeiksi rakentamalla monitieteisyyttä, avoimuutta ja yhteentoimivuutta tukevia palveluita.

Lisäksi tieteellisen laskennan strategisia linjauksia tehdään siitä vastaavan CSC:n pääyksikön vuosittaisissa toimintasuunnitelmissa.

Seuraavassa esitetään ohjeellinen strategia Suomen tieteellisen laskennan kokonaisuudelle.

2.1.1 Tieteellisen laskennan strategia

Tieteellisen laskennan hyödyntäminen lisääntyy kaikilla tieteenaloilla ja parantaa suomalaisen yhteiskunnan kilpailukykyä. Tieteellisen laskennan yhteiskunnallinen merkittävyys tunnistetaan yleisesti.

Suomalaisilla laskennallisen tieteen tutkijoilla on käytettävissään kansainvälisesti kilpailukykyiset tehokkaat, tarkoituksen- ja ajanmukaiset kustannustehokkaasti toteutetut laskentaresurssit ja niihin liittyvät palvelut. Korkeatasoinen tuki ja koulutus takaavat laitteistojen tehokkaan hyödyntämisen. Keskitetyt resurssit ovat helposti kaikkien käytettävissä.

Käyttäjäkunnan laajentaminen

Uusia käyttäjiä ja sovellusaloja tavoitellaan tiedotuksella, mahdollisimman helppokäyttöisillä palveluilla, kattavalla tuella ja yhteistyöprojekteilla. Teknologia- ja menetelmäkehityksen avaamat mahdollisuudet siirretään tutkimusyhteisöjen käyttöön nopeasti uusina palveluina. Dataintensiivisen laskennan palveluilla voidaan saavuttaa käyttäjiä uusilta aloilta, kuten esimerkiksi bio- ja lääketieteistä.

Tehokkuus ja tarkoituksenmukaisuus

Laitteistot eivät aseta rajoituksia tieteelle, ja tarjottavat arkkitehtuurit vastaavat käyttäjien tarpeita. Tarjolla on superkonekapasiteettia massiiviseen rinnakkaislaskentaan, massiivista tallennuskapasiteettia datan prosessointia varten, klustereita pienille rinnakkaisajoilla ja peräkkäistöille, suurimuistisia noodeja, pilvikapasiteettia, korotetun tietoturvatason kapasiteettia sensitiivisten aineistojen analysointiin, erikoisprosessoreita ja nopeaa tallennuskapasiteettia. Tarkoituksenmukaiset palvelut virtualisoidaan monikäyttöisyyden lisäämiseksi.

Kustannustehokkuus

Tarjoamalla raskain kapasiteetti keskitetysti ja alemman tason (Tier-2) kapasiteetti hajautetusti vältetään päällekkäisyydet ja saavutetaan kokonaisuuden kannalta tarkoituksenmukaisin ratkaisu. Kapasiteetti, jota ei kannata toteuttaa kansallisella tasolla, hankitaan kansainvälisesti, esim. PRACElta.

Kalliit erikoisratkaisut varataan vain niitä tarvitseville töille ja työt ohjataan sopivimpaan arkkitehtuuriin. Tavoitteena on hyvä käyttöaste ja töiden nopeat läpimenoajat.

Suositaan energiatehokkaita arkkitehtuureja ja konesaleja. Laitteistot uusitaan riittävän usein. Kaikki toimijat laskevat kustannukset samojen yhteisesti sovittujen periaatteiden mukaisesti.

Ajanmukaisuus

Tutkijoille tarjotaan kansainvälisesti kilpailukykyistä laskentakapasiteettia. Turvaton ja jatkuva rahoituksen turvin laitteistojen uusiminen on mahdollista suunnitellusti ja pitkäjänteisesti. Teknologia seurannan ja -testauksen avulla pystytään tekemään oikeita valintoja laitteistoja hankittaessa.

Tuki ja koulutus

Tukihenkilöstön asiantuntemusta kehitetään jatkuvasti ja tieto siirretään käyttäjille hyvin järjestetyn koulutuksen ja selkeän dokumentaation avulla. Käyttäjätukea annetaan sekä keskitetysti että paikallisesti.

2.2 Toiminnan haasteet ja tavoitteet

2.2.1 Keskitettyjen resurssien aidosti kansallinen käyttö

CSC on opetus- ja kulttuuriministeriön omistama yritys, ja sen tarjoamat tietokoneressit ja niiden varassa toimivat palvelut ovat ministeriön kustantamina tavallisesti maksuttomia yliopistojen ja korkeakoulujen tutkijoille ja opiskelijoille. Tarvetta raskaammalle laskentakapasiteetille ja mahdollisuudelle massiiviseen dataintensiiviseen prosessointiin on kuitenkin myös muiden ministeriöiden alaisissa tutkimuslaitoksissa, ja tällä hetkellä ministeriörajat ylittävä käyttö edellyttää erillisiä sopimuksia ja on maksullista. Tämä keinotekoinen raja haittaa tutkimusta ja yhteiskuntaa, ja siihen törmätään jatkuvasti. Samoin yliopistojen ja korkeakoulujen tutkijat joutuvat maksamaan CSC:n palveluista, jos palvelutaso ylittää perustason, mutta tämän tyyppisille operointikustannuksille on vaikea saada kilpailtua rahoitusta.

Vaihtoehtoisesti CSC voisi olla aidosti kansallinen resurssi ja kaikkien ministeriöiden alaisten laitosten käytettävissä ilman erillistä laskutusta. Luonnollisin ratkaisu olisi ministeriöiden yhteinen rahoitus tai korkealla tasolla solmitut ministeriöiden väliset puitesopimukset.

Tutkimuslaitosten tarvitseman raskaan laskentakapasiteetin keskittäminen olisi myös kustannustehokkaampaa ja säästäisi valtion rahaa.

2.2.2 CSC:n in house -asema

CSC:n asema kilpailutusten suhteen on nykymuotoiselle korkeakoululaitokselle ainakin jossain määrin epäselvä. Olisi selkeä etu, jos yliopistot ja ammattikorkeakoulut voisivat ostaa ilman kilpailutusta CSC:ltä myös niitä tieteellisen laskennan ja tutkimuksen tarvitsemia palveluita, jotka eivät sisälly OKM:n vuosisopimukseen. Muut ministeriöt ja niiden alaiset tutkimuslaitokset (valtion in-house-asema) voivat hankkia CSC:n palveluita kilpailuttamatta.

2.2.3 Teknologian kehityksen aiheuttamat haasteet

Tietokoneiden tehon kasvu niin sanotun Mooren lain mukaisesti perustui pitkään ainakin osittain yksittäisten prosessorien tai prosessoriytimien nopeutumiseen ja kellotaajuuden kasvamiseen. Tämän hyödyntäminen oli käyttäjille helppoa; ohjelmat nopeutuivat lähes automaattisesti seuraavaan prosessorisukupolveen siirryttäessä. Muutoksia koodiin ei siis

juurikaan tarvinnut tehdä. Samaan aikaan koneiden teho kasvoi myös rinnakkaisuutta eli prosessorien ja ydinten määrää lisäämällä.

Viime aikoina tehon kasvu on kuitenkin perustunut yksinomaan rinnakkaisuuden lisäämiseen, ja kellotaajuudet ovat jopa hieman laskeneet. Tähän on osasyynä sähkönkulutuksen voimakas riippuvuus kellotaajuudesta. Rinnakkaisuus lisääntyy kaikilla loogisen hierarkian tasoilla: koneissa on enemmän laskentasoelmia (joiden välinen kommunikaatio pitää ohjelmoida eksplisiittisesti), solmujen prosessoreissa on enemmän ytimiä (joiden välillä voidaan käyttää yhteisen muistin rinnakaistusta) ja ytimet suorittavat yhdellä käskyllä enemmän laskutoimituksia (vektorointi).

Lisääntyneen rinnakkaisuuden tehokas hyödyntäminen edellyttää ohjelmien optimointia aikaisempaa enemmän ja laitearkkitehtuurin syvällisempää ymmärtämistä. Tämä korostaa tieteellisen laskennan koulutuksen ja käyttäjätuen merkitystä ja edellyttää niiden sekä itse tieteellisen laskennan ohjelmien jatkuvaa kehittämistä.

Saatavilla olevien tieteellisten ohjelmistojen lukumäärä jatkaa kasvuaan. Erityisesti aloilla, joille laskennallisen tiede tekee tuloaan, uusia ohjelmistoja syntyy kiihtyvällä tahdilla tutkimusohjelmien tuloksena. Esimerkiksi biologiassa ja ympäristön tutkimuksessa tarvitaan jatkuvasti uusia algoritmeja johtuen tutkimusinstrumentaation nopeasta kehityksestä. Jopa vuosittain uusiutuvat laboratoriolaitteet pystyvät tuottamaan joko laadullisesti tai määrällisesti kertaluokkaa paremmin dataa, joka täytyy prosessoida ennen kuin se voidaan tulkita. Tämän seurauksena laskentapalveluiden tarjoajien pitää pystyä tarjoamaan ratkaisu, jossa kunkin tieteenalan tai sen osa-alueen tai poikkitieteellisen yhteistyön data-analyysin vaadittavat ohjelmistot ovat erittäin nopealla aikataululla saatavilla. Tieteen kehityksen seuraamisen tekee mahdolliseksi muun muassa CSC:n laskentaympäristön virtualisointi (esim. Pouta-pilvipalvelu), joka mahdollistaa tietoverkon välityksellä toimitettavat tieteellisten ohjelmistojen tasolla räätälöivät laskentapalvelimet.

Yksi tieteellisen laskennan suurimmista haasteista on koneiden tehon mukana kasvava sähkönkulutus; sähkölaskun osuus laskennan kokonaiskustannuksista alkaa lähestyä kipurajaa. Tätäkin ongelmaa pyritään ratkaisemaan rinnakkaisuutta lisäämällä: sekä kiihdyttimissä (esimerkiksi Nvidian GPGPU:t) ja apuprosessoreissa (esimerkiksi Intelin MIC/Xeon Phi) on suuri määrä suhteellisen hitaita prosessointiyksiköitä. Nämä erikoisprosessorit eivät sovellu kaikkiin tehtäviin, mutta parhaimmillaan niistä saatava teho on moninkertainen tavallisiin prosessoreihin verrattuna ja niiden energiatehokkuus on erinomainen. Valittavasti erikoisprosessorien tehokas ohjelmointi on ainakin toistaiseksi hyvin hankalaa ja toistaiseksi sovellusten tuki erikoisprosessoreille on vielä harvinaista. Siitä huolimatta vaikuttaa todennäköiseltä, että CSC:n on seuraavassa konehankinnassaan panostettava myös näihin teknologioihin suomalaisten tutkijoiden kilpailukyvyyn säilyttämiseksi.

Selkeä haaste liittyy myös valintaan näiden erikoisteknologioiden välillä seuraavassa hankinnassa. Tällä hetkellä ei ole selvää tulevatko molemmat säilymään vai kuoleeko toinen. Virheellinen päätös olisi erittäin kallias sekä taloudellisesti että palvelun tason kannalta. Riskiä voidaan vähentää huolellisella teknologiaseurannalla ja omalla testaamisella, mutta tällainen vaatii resursseja.

2.2.4 Laitteistohankinnat ja niiden rahoitus

Tekniikan kehityksen ja erityisesti jatkuvasti parantuvan suorituskyvyn ja sähkön kulutuksen suhteen vuoksi suurteholaskennan laitteistojen kustannustehokas käyttöikä on maksimissaan

viisi vuotta. Myös kapasiteetin tarve kasvaa ja elinkaaren loppupäässä laitteistojen käyttöaste on tyypillisesti erittäin korkea ja jonotusajat pitenevät.

CSC on perinteisesti joutunut hakemaan rahoituksen uusiin laitteistoihin erikseen joka kerralla. Epävarmuus rahoituksen tasosta ja aikataulusta vaikeuttaa pitkäjännteistä suunnittelua. Rahoituksen tasaisempi jakautuminen estäisi säännöllisin väliajoin toistuvan resurssipulan koneiden elinkaaren loppuvaiheessa.

Rahoituksen luonteen vuoksi CSC joutuu uusimaan kaikki suurimmat laitteistonsa yhtä aikaa, mikä tekee hankintaprojekteista vaativampia ja työlämpiä. Kapasiteetin tarjonnan kannalta voisi esimerkiksi olla parempi uusia superkone ja klusteri eri tahdissa, jolloin asiakkailta olisi useammin uutta teknologiaa käytettävissä ilman että kokonaisinvestointi kasvaisi.

Hankintoja ohjaa lainsäädäntö, joka pyrkii turvaamaan tasapuolisuuden toimittajien kannalta, mutta samalla kankeudellaan pakottaa hankkijan määrittelemään arviointiperusteensa erittäin perusteellisesti jo aikaisessa vaiheessa. Tämä lisää hankinnan epäonnistumisen riskiä jos tarjottava teknologia poikkeaa ennakoitusta. Huolellisella valmistautumisella, neuvotteluilla ja tietoa keräämällä riskiä voidaan pienentää, mutta on osoittautunut, että valmistajat ovat hyvin vastahakoisia antamaan tarkkoja tietoja hinnoista ja jopa konfiguraatioista ennen varsinaista tarjousta.

Esimerkkinä lainsäädännön ja käytännön laiteinvestointien rahoituksen haasteista on yhteistyö Suomen tutkimusinfrastruktuurien tiekartan (FIRI) kanssa. Suomen Akatemia on erikseen kieltänyt CSC:n hankkeilta edulliseksi havaitun leasing rahoitusmallin. CSC:n hallintaan päätyvien laitteiden pääomakustannus on pitänyt tehdä kertaostoina ja kertapoistoina yhden budjettikauden sisällä. Tämän seurauksena FIRI hankkeissa kehitettävien laskentapalveluiden vuosikustannuksiin päätyy suuriakin vaihteluita. Parempi malli olisi, että palveluiden tuottamisen tietotekniset edellytysten eli laitekannan kustannukset voitaisiin jakaa tasaisesti koko luvatuille palvelun elinkaarelle, ja tarvittaessa kasvattaa laitekantaa käyttöasteen kasvun myötä.

Hankkijan liikkumavaran vähentyminen on johtanut arviointikriteeristön monimutkaistumiseen, mikä puolestaan on osoittautunut ongelmaksi toimittajille ja johtanut epätäydellisiin tarjouksiin, jotka on jouduttu hylkäämään.

2.2.5 Kansainvälisen käytön ja tukipalveluiden rahoitus

Kuten edellä todetaan, suurteholaskenta-arkkitehtuurien kasvavan rinnakkaisuuden tehokas hyödyntäminen vaatii asiantuntemusta ja paljon työtä. Tällainen työ on välttämättömyydestään huolimatta väliinpuotoajan asemassa rahoitusta haettaessa: ohjelmankehitys ja suorituskyvyn optimointi ei varsinaisesti ole tiedettä, joten tutkimusrahan saaminen sitä varten on vaikeaa.

Samaten on tyypillistä, että laitteistohankintojen rahoituksen yhteydessä operointiin ja tukeen ei myönnetä resursseja. Nämä ovat kuitenkin oleellisia tekijöitä laitteistojen hyödyntämisen kannalta, ja asiaa tulisi käsitellä kokonaisuutena. Erityisesti tuen merkitys kasvaa jatkuvasti, kuten edellä todetaan.

CSC tuottaa palvelut OKM:lle omakustannusperusteisesti mikä on kustannustehokasta, mutta samalla aiheuttaa kaikenlaisen puskurin puuttumisen tukipalveluiden rahoituksesta. OKM pystyy rahoittamaan vain osan tukipalveluiden henkilöstöstä ja lopuille on haettava rahoitus ulkopuolelta joko osittain rahoitettuna projekteina tai täysin rahoitettuna alihankintatyönä. Kun

sitoumukset työmääristä on täytettävä kummassakin kategoriassa, lopputulos on erittäin kankea, ja uusiin hankkeisiin on vaikeaa löytää vapaita resursseja. Ministeriön rahoitukseen pitäisi sisältyä joustovaraa palveluiden kehitykseen. Kehitystyötä voidaan priorisoida esim. kytkemällä ministeriön rahoitus kilpailtuun ulkoiseen kehitysrahoitukseen (esim. FIRI, H2020).

Kansainvälisten e-Infrastruktuurien operointi ja käyttäjätuki rahoitetaan tyypillisesti osin EU-rahoituksella, joka vaatii käytännössä aina lisäksi kansallisen rahoitusosuuden. Hankkeet on usein monivuotisia projekteja vaikka niiden kautta tarjottavat palvelut mielletään pysyviksi ja on usein tunnistettu strategisesti välttämättömiksi Euroopan kilpailukyvyille ja tutkimukselle (mm. ESFRIt). Näin iso osa tieteellisen laskennan tarvitsemasta tuesta, erityisesti kansainvälisten resurssien ja palveluiden osalta toteutetaan ankarasti kilpaillun rahoituksen turvin. Tähän toiminta malliin liittyy merkittävä riski palvelujen epäjatkuvuudelle. Toisaalta EU-rahoitteiset kehitys- ja tutkimusinfrastruktuurihankkeet mahdollistavat paremman asiakaspalvelun samalla panostuksella tai saman palvelutason kansantaloudellisesti edullisemmin. Tällöin kuitenkin haasteeksi muodostuu tarvittavien omarahoitusosuuksien hankkiminen kun lukuisiin tutkimusinfrastruktuureihin pitäisi pystyä sitoutumaan samanaikaisesti useiksi vuosiksi kerrallaan.

2.2.6 Kokonaisoptimointi ja kustannustietoisuus

Idealisessa tilanteessa Suomen tieteellisen laskennan resurssit järjestettäisiin kokonaisuutena mahdollisimman kustannustehokkaasti ottaen huomioon kaikkien osapuolien tarpeet. Suurin este tällaisen kokonaisoptimoinnin tiellä on se, että toimijat ovat käytännössä itsenäisiä yksiköitä, jotka tekevät päätöksiä omista lähtökohdistaan.

Kokonaisvaltaisen kustannustehokkuuden kannalta ongelmaksi on havaittu, että eri toimijat arvioivat laskentaresurssien kustannukset eri tavalla. Esimerkiksi ylläpitokustannuksia tai sähkölaskua ei aina kohdisteta laskennan kustannuksiin.

Minimitavoite on kustannusten laskeminen totuuden mukaisesti samoja sääntöjä käyttämällä. Tämä lisää kustannustietoisuutta ja ohjaa rahan tehokkaaseen käyttämiseen.

2.2.7 Tieteellinen laskenta teollisuudessa

Vaikka laskennallinen tiede on Suomessa korkeatasoista, yrityksissä matemaattista mallinnusta ja tietokonesimulaatioita on hyödynnetty enimmäkseen melko huonosti, mikä on hukattu tilaisuus kilpailukyvyyn kannalta. Myös numeerisen matematiikan tutkimus ja opetus yliopistoissa on nykyään vähäisempää kuin aikaisemmin.

Tilanne on ehkä muuttumassa ns. Big Datan myötä, kun yrityksissä on havaittu massatiedon mahdollisuudet esimerkiksi tuotteiden kohdistamisessa oikealle kuluttajaryhmälle tai erilaisissa optimointiongelmissa. Toinen kasvava liiketoiminta-alue on nk. yksilölliset hoitomuodot ja erilaiset terveydenhoidon sovellukset, jotka pohjautuvat genomitason tiedon, instrumenttidatan ja rekisteritietojen yhdistämiseen. Suomessa on poikkeuksellisen paljon systemaattisesti kerättyjä näytekokoelmia ja rekisteriaineistoja, joita on kerätty 1900-luvun puolivälistä saakka.

Suomeen on viimeisten kymmenen vuoden aikana syntynyt tutkimusryhmiä ja niistä ponnistavia pieniä yrityksiä, jotka ovat erikoistuneet eri tieteenalojen datan prosessointiin ja laskentaan tarvittavien referenssitietokantojen tulkintaan. Startup-yrityksillä ei kuitenkaan ole pääomia tai intressiä luoda tarvittavaa tietoteknistä laskenta-alustaa. CSC on ELIXIR (FIRI) -hankkeen myötä yhteistyössä kolmen bioalan pienen tai keskisuuren yrityksen kanssa

laskentapalveluiden tarjoajana löytääkseen rahoituksellisesti ja sopimuksellisesti kestävä mallin infrastruktuuritason yhteistoimintaan.

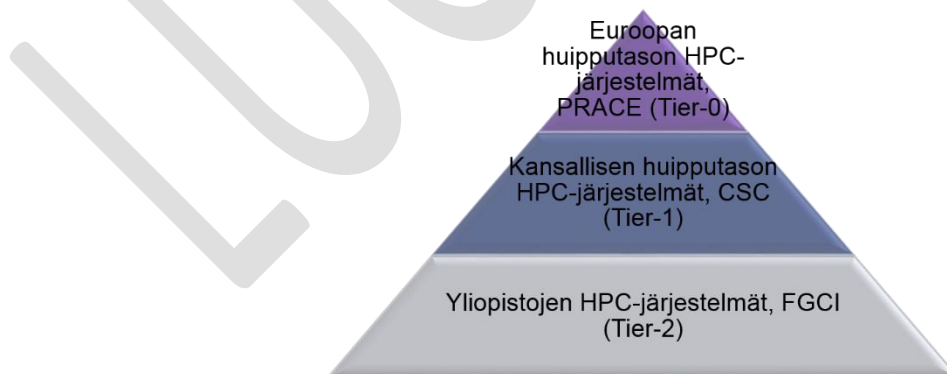
2.3 Palvelut

Palvelut	CSC	Yliopistot	Ammattikorkeakoulut	Sektoritutkimuslaitokset
Superkoneresurssit (Tier-1)	T	K		T,K
Klusteriresurssit (Tier-2)	T	T,K	K	T,K
Grid-palvelut	T	T,K	K	
Pilvipalvelut	T	T,K	T,K	K
Ohjelmistopalvelut	T	T,K	K	
Tietokantapalvelut	T	K	K	
Käyttäjätuki	T	T,K	K	
Koulutus	T	T,K	K	K

Kuva 13. Keskeisimpien palveluiden tuottajat ja käyttäjät: T = tuottaja, K = käyttäjä. T, K = tuottaja ja käyttäjä.

2.3.1 Laskentaresurssit

Tärkein tieteellisen laskennan palveluista on laskentaresurssien tarjoaminen. Suurteholaskentaresurssit luokitellaan usein kolmeen kategoriaan laskevan tehon mukaan: Tier-0, Tier-1 ja Tier-2. Laajin käyttäjäkunta on Tier-2-koneilla, ja vain äärimmäistä tehoa tarvitsevia tutkimusongelmia kannattaa ratkoa Tier-0-laitteilla.



Kuva 14. Laskentaresurssien eri tasot.

Tier-0:lla tarkoitetaan maailman tehokkaimpia tietokoneita, eli käytännössä Top 500 -listan 10–20 tehokkaimman koneen joukkoon kuuluvia laitteistoja. Tier-0-koneissa on tyypillisesti satoja tuhansia laskentatuntia sekä nopea erikoiskytkentäverkko, ja niiden tehokas hyödyntäminen edellyttää erittäin hyvin rinnakkaistuvaa ohjelmaa: tyypilliset ajot käyttävät kymmeniä tuhansia

ytimiä. Tällainen kone edellyttää suurta ja edistynyttä käyttäjäkuntaa, ja esimerkiksi Suomen kaltaisessa maassa Tier-0-koneen hankkiminen ei olisi kustannustehokasta. Tier-0:aa karakterisoi usein myös kansainvälinen käyttö ja vaativa vertaisarviointiprosessi; suomalaiset tutkijat pääsevät käyttämään tällaisia koneita PRACEn kautta.

Tier-1-koneita voidaan pitää kansallisen tason resursseina, ja esimerkiksi CSC:n nykyinen superkone Sisu 40 000 ytimellään kuuluu tähän kategoriaan. Tier-1-koneet on tarkoitettu isoille rinnakkaisajajille, jotka pystyvät hyödyntämään tuhansia ytimiä.

Tier-2 koostuu paikallisista yliopisto- tai laitostason laitteista, jotka ovat tyypillisesti nopealla InfiniBand-kytkentäverkolla varustettuja klustereita. Näitä koneita käytetään sekä peräkkäisajoihin että pieniin muutaman sadan ytimen rinnakkaisajoihin.

Hierarkian kaikki tasot ovat välttämättömiä ja tukevat toisiaan; tie raskaimpaan Tier-0-laskentaan vaatii kokemusta ja kehitystyötä ja kulkee Tier-1- ja Tier-2-tasojen kautta. Lisäksi Tier-2-resurssit voidaan paremmin räätälöidä paikallisiin tarpeisiin, joiden tyydyttäminen keskitetysti on joskus vaikeaa tai epätarkoituksenmukaista.

Edellä kuvattua luokittelua voidaan pitää suurteholaskennan ekosysteeminä. Kaikkia tasoja tarvitaan ja ne ovat riippuvaisia toisistaan. Laskentatehon kasvattaminen vaatii kalliita erikoisratkaisuja, joista ei ole hyötyä pienemmissä koneissa. Tier-0-koneiden hyödyntäminen edellyttää myös alemmilla tasoilla tehtyä kehitystyötä, hyppy suoraan huipputasolle on käytännössä mahdoton myös siksi, että resurssihakemuksessa on pystyttävä näyttämään ongelman sopivuus ja ohjelmiston suorituskyky isoilla koneilla.

Monissa supertietokoneissa on nykyisin erikoisprosessoreita (GPU tai MIC), ja on todennäköistä että tämä pätee CSC:n seuraaviin superkoneisiin. CSC:llä on muutaman vuoden ajan ollut erikoisprosessoreita pienessä mittakaavassa ja lähinnä testikäytössä. Ensimmäinen GPU/MIC-kone yleiseen asiakaskäyttöön avattiin tämä syksynä. Vaikka erikoisproessorit eivät sovi kaikkien ongelmien ratkaisemiseen, niille voidaan kustannustehokkaasti siirtää osa tavallisten prosessorien kuormasta. Erikoisproessorit ja niiden ohjelmointimallit ovat edelleen kehitysvaiheessa. Ainakin toistaiseksi erikoisprossorien tehokas hyödyntäminen edellyttää erittäin korkeatasoista teknistä osaamista ja paljon ohjelmointityötä; erikoisprosessoreita ei siis tässäkään mielessä voi pitää yleiskäyttöisinä.

Eri paikoissa olevia laskentaresursseja voidaan yhdistää suuremmaksi kokonaisuudeksi hajautetun laskentamallin (grid-laskenta) avulla. Siinä mallissa käyttäjät lähettävät työnsä ajoon erillisen väliohjelman avulla. Työt ohjautuvat automaattisesti niille laskentaresursseille joissa on vapaata tilaa. Näin saadaan hajautetut resurssit tehokkaaseen käyttöön. Suomessa hajautettua laskentaa on tehty mm. M-Grid ja sitä seuranneen FGI/FGCI-konsortion puitteissa. Molemmista tapauksissa Tier-2-tason koneita on ollut useassa yliopistossa ja kullakin laitteistolla omat ylläpitäjänsä. CSC on antanut keskitettyä ylläpitotukea.

Hajautettu järjestelmä mahdollistaa yksittäisten resurssien räätälöimisen paikallisten käyttäjien tarpeisiin, mutta kokonaisuudessaan lisää ylläpitoon tarvittavaa työtä verrattuna keskitettyyn järjestelmään.

2.3.2 Virtualisoidut laskentaresurssit

Yllä kuvatuissa järjestelmissä on käytössä etukäteen valittu käyttöjärjestelmä ja työt menevät ajoon eräajojärjestelmän kautta. Perinteinen malli ei enää riitä kaikkiin tarpeisiin.

Infrastruktuuri palveluna pilvilaskentamallissa (IaaS) palveluntarjoaja ylläpitää fyysistä laskentaresurssin ja sen päällä olevaa virtualisointikerrosta. Loppukäyttäjälle annetaan virtualisoidusta laskentaympäristöstä oma virtuaalinen palvelin. Loppukäyttäjä voi valita virtuaalikoneeseen tarvitsemansa käyttöjärjestelmän ja muokata ohjelmistovalikoiman palvelimessa tutkimusongelmaan ja prosessointiin sopivaksi. Tämä antaa käyttäjälle vapauden räätälöidä ohjelmistovalikoima, mutta vaatii samalla käyttäjältä sitoutumista ohjelmistoympäristön ja palvelimen ylläpitotyöhön. CSC on pilotoinut IaaS-pilvipalvelua käyttäjien kanssa vuodesta 2011 ja tarjonnut tuotantotasoin (Pouta IaaS) palvelua huhtikuusta 2014 saakka. Pilviympäristö soveltuu hyvin myös dataintensiiviseen laskentaan ja ympäristöä ollaan kehittämässä laajemmin datankäsittelyn tarpeita huomioivaksi vuoden 2015 aikana.

2.3.3 Korotetun tietoturvatason laskenta

Virtualisointiin perustuva Pouta Infrastructure as a Service (IaaS) pilvipalvelun soveltuu myös arkaluonteisen datan prosessointiin. Palvelun arkaluonteiseen prosessointiin suunnattu ePouta on korotetun tietoturvatason ratkaisu, joka on suunnattu tutkimusryhmien ja organisaatioiden käyttöön. Palvelussa prosessoitava data voi olla esimerkiksi ihmisiin liittyviä tutkimuskäyttöön hyväksytyjä tietoaaineistoja. ePouta-palvelussa on asiakkaina eri organisaatiota, joiden eristäminen toisistaan on keskeinen ominaisuus palvelussa. CSC:n näkökulmasta luotettuja tahoja kokonaislaskentapalvelun tuotannossa ovat IT-ylläpitäjät organisaatioissa, kuten yliopistossa, instituutissa, biokeskuksessa tai yksittäisessä tutkimusryhmässä. Palvelun avulla organisaatio voi toteuttaa haluamiaan loppukäyttäjäpalveluita. Loppukäyttäjäpalvelut voidaan suunnata organisaatioiden tutkijoille tai muille henkilöille tai organisaatioille.

Asiakasorganisaation näkökulmasta palvelu on tietoverkkoyhteydeltään oman organisaation sisäpuolella, mutta palvelun fyysinen infrastruktuuri sijaitsee oman organisaation ulkopuolella. Siten arkaluonteisen datan säilytys ja käsittely tapahtuu tietoarkkitehtuurin näkökulmasta oman organisaation sisällä. Kontrolli ympäristön virtualisoiduista laitteista ja laitetilasta loppukäyttäjien säilyy oman organisaation hallinnassa.

Ottaakseen IaaS-palvelun käyttöön arkaluonteisen tietoaaineistojen analyysissä organisaation tulee voida luottaa palvelun tuottajaan, tässä tapauksessa CSC:hen. Pouta palvelu tuotetaan Suomessa sijaitsevista konesaleista ja virtualisoitujen CSC:n resurssien käyttöönotto ja integroiminen olemassa olevaan ympäristöön vaatii käyttöönottoprojektin, kuten verkkoyhteyksien toteuttamisen ja virtuaalikoneiden hallinnan organisoinnin. Palvelun käyttöönotossa CSC:n ja asiakasorganisaation välillä luodaan tietoturvasopimus pääsopimuksen liitteeksi, jossa määritellään mm. sovellettavat käytännöt, vastuut ja oikeudet, yhteystiedot kriisitilanteissa sekä kokouskäytännöt tietoturvallisuuden laadun varmistamiseksi.

ePouta-palvelu tulee tuotantoon vuoden 2015 alkupuolelta ja on mahdollinen ratkaisu mm. biopankkien data-aineistojen käsittelyyn.

Edellä on kuvattu laskentakapasiteettia lähinnä erilaisten käyttötarpeiden mukaisesti. Teknologia seurannan vuoksi on kuitenkin oleellista, että uusia laitearkkitehtuureja pystytään hankkimaan myös testikäyttöön ilman olemassa olevaa asiakastarvetta. Ainoastaan näin pystytään varmistamaan oikeat päätökset hankintoja tehtäessä.

2.3.4 Käyttäjätuki

Perustukeen kuuluvat käyttö- ja ohjelmointiympäristön tuki (eräajot, levy/tallennusjärjestelmä, kääntäjät, kirjastot, ohjelmointi- ja suorituskykytyökalut), joka hoidetaan yleensä helpdesk-luonteisesti eikä normaalisti vaadi paljon aikaa kyselyä kohden.

Vaativampaan tukeen tai asiantuntijapalveluihin kuuluvat yleisen tieteellisen laskennan menetelmätuen lisäksi ohjelmien siirto, optimointi ja rinnakkaistaminen. Varsinkin kaksi viimeksi mainittua saattavat vaatia hyvinkin paljon perehtymistä ja aikaa, mutta parhaimmillaan johtavat huomattavaan suorituskyvyn paranemiseen ja sitä kautta tieteelliseen impaktiin. Asiantuntijapalveluihin lukeutuu myös tiedetuki, joka monesti liittyy tieteellisten ohjelmistojen käytön tukeen.

Tärkeä osa käyttäjätukea on dokumentaation ja käyttäjäoppaiden tuottaminen ja ylläpito.

Osa kansainvälisiin resursseihin, kuten PRACEen, liittyvästä tuesta on toteutettu ulkoisesti rahoitetuissa projekteissa. PRACEn tapauksessa tuki on kuulunut EU:n rahoittamiin PRACE IP -projekteihin.

2.3.5 Koulutus

Perinteinen ohjenuora koulutukseen liittyen CSC:n ja yliopistojen välillä on ollut päällekkäisyyden välttäminen; CSC järjestää kurssuja tieteellisistä ohjelmistoista, Fortran-ohjelmoinnista, optimoinnista ja rinnakkaistamisesta ja yliopistot ovat huolehtineet numeerisesta matematiikasta ja laskennallisesta tieteestä. Edellä kuvatun teknologian kehityksen myötä vaikeutuneen koneiden tehon hyödyntämisen myötä koulutuksen merkitys lisääntyy jatkuvasti.

Tavoitteena on yhdistää CSC:n tarjoama tieteellisen laskennan koulutus tiiviimmin yliopistojen opetukseen erityisesti jatko-opinnoissa ja tutkijakouluissa.

Osa kansainvälisiin resursseihin, kuten PRACEen, liittyvästä koulutuksesta on toteutettu ulkoisesti rahoitetuissa projekteissa. PRACEn tapauksessa koulutus on kuulunut EU:n rahoittamiin PRACE IP -projekteihin. NordForskin NeGI-tutkimusohjelmassa (Nordic e-Science Globalisation Initiative) suunnitellaan parhaillaan tukea yhteispohjoismaiselle eScience-koulutukselle⁵.

2.3.6 Ohjelmisto- ja tietokantapalvelut

Ohjelmistopalveluihin kuuluu valmisohjelmistojen tarjoamisen lisäksi omien tieteellisen laskennan ohjelmistojen kehitys. Esimerkiksi CSC:llä kehitetään Elmeriä, GPAWia ja Chipsteriä, jotka kaikki perustuvat avoimeen lähdekoodiin.

CSC tarjoaa korkeakoulujen tutkijoiden käyttöön laajaa asiakasjoukkoa hyödyttäviä ja CSC:n laskentakapasiteetin tehokkaasti käyttäviä tieteellisiä ohjelmistoja ja tietokantoja (sekä avoimen lähdekoodin ohjelmistoja että kaupallisia tuotteita). Kaupalliset tuotteet pyritään hankkimaan keskitetysti kansallisille lisenssikonsortioille, jolloin useat käyttäjäryhmät saavat ohjelmiston/tietokannan käyttöönsä yhteisen hankinnan kautta. Tällaiset ohjelmistot on usein mahdollista ajaa myös paikallisilla työasemilla. CSC tarjoaa lisenssipalvelua kansallisten ohjelmistolisenssien jakoon. Ohjelmistovalikoimaa kehitetään yhteistyössä mm. korkeakoulujen Lisenssi SIG -työryhmän kanssa.

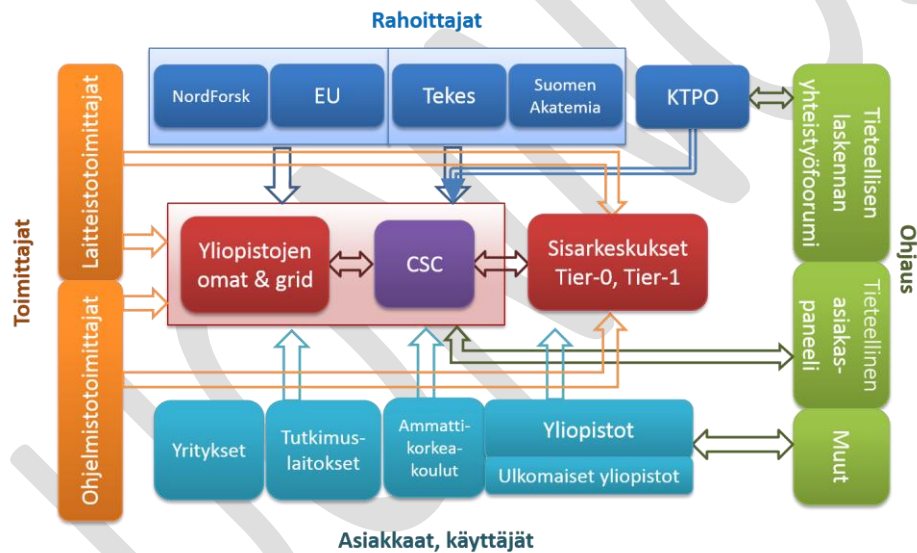
⁵ <http://www.nordforsk.org/en/programmes/programmer/escience/escience-globaliseringsinitiativ>

2.3.7 Kriittiset palvelut muualta

- Tutkimuksen tietoverkko FUNET ja sen lisäpalvelut
 - Eduroam
- Tallennus ja arkistointipalvelut
 - Työtilat datan reaaliaikaiseen käsittelyyn
 - Lyhytaikainen ja keskipitkä varastointi
 - Varmennuspalvelut (Back-up)

2.4 Toimijat ja sidosryhmät

Seuraavassa tieteellisen laskennan toimijat ja sidosryhmät on luokiteltu palveluiden tuottajiin, loppukäyttäjiin, rahoittajiin ja ohjaaviin elimiin (kuva 9).



Kuva 9. Tieteellisen laskennan toimijat ja sidosryhmät.

2.4.1 Palveluiden tuottajat

- CSC
 - Laskenta- ja tallennusresurssit
 - Ohjelmistot ja tietokannat
 - Laskentaympäristön käytön tuki
 - Tiede- ja menetelmätuki
 - Koulutus
 - Konesalipalvelut
- Yliopistojen ja ammattikorkeakoulujen ATK-keskukset/laitokset/laboratoriot
 - Esim. Helsingin yliopisto, Aalto, TTY, jne.
 - Laskentaresurssit
 - Laskentaympäristön käytön tuki

- Tutkimuslaitosten ATK-keskukset
 - Esim. Ilmatieteen laitos, VTT, Luke jne.
 - Laskentaresurssit
 - Laskentaympäristön käytön tuki
- Eurooppalaiset laskennalliset tutkimusinfrastruktuurit
 - Laskentaresurssit
 - Laskentaympäristön käytön tuki
 - Koulutus
 - PRACE
 - EGI.eu hajautetut laskentaresurssit (grid, pilvi)
- Muut kv. (alakohtaiset) resurssit ja ESFRI
 - NeIC
 - ECMWF
 - ELIXIR
 - Pohjoismainen sääkonehanke
 - Kv. laskentakeskukset
 - Hajautettu laskentainfrastruktuuri (Grid, pilvi IaaS)
 - Kaupalliset pilvipalvelut ja -brokerit
 - Jne.

2.4.2 Loppukäyttäjät

- Yliopistojen tutkijat ja opiskelijat
- Ammattikorkeakoulujen tutkijat ja opiskelijat
- Tutkimuslaitokset
- Yliopistolliset sairaalat
- Yritykset.

2.4.3 Rahoittajat

- Opetusministeriö
 - CSC
 - Yliopistot ja ammattikorkeakoulut
- Muut ministeriöt
 - Tutkimuslaitokset
 - Yliopistolliset sairaalat, biopankit jne.
- Suomen Akatemia
 - Yliopistot
 - Tutkimuslaitokset
 - Ammattikorkeakoulut
 - CSC
- Tekes
 - Yliopistot
 - Ammattikorkeakoulut
 - Tutkimuslaitokset
- EU
 - CSC
 - Yliopistot
 - Ammattikorkeakoulut
 - Tutkimuslaitokset

2.4.4 Ohjaus

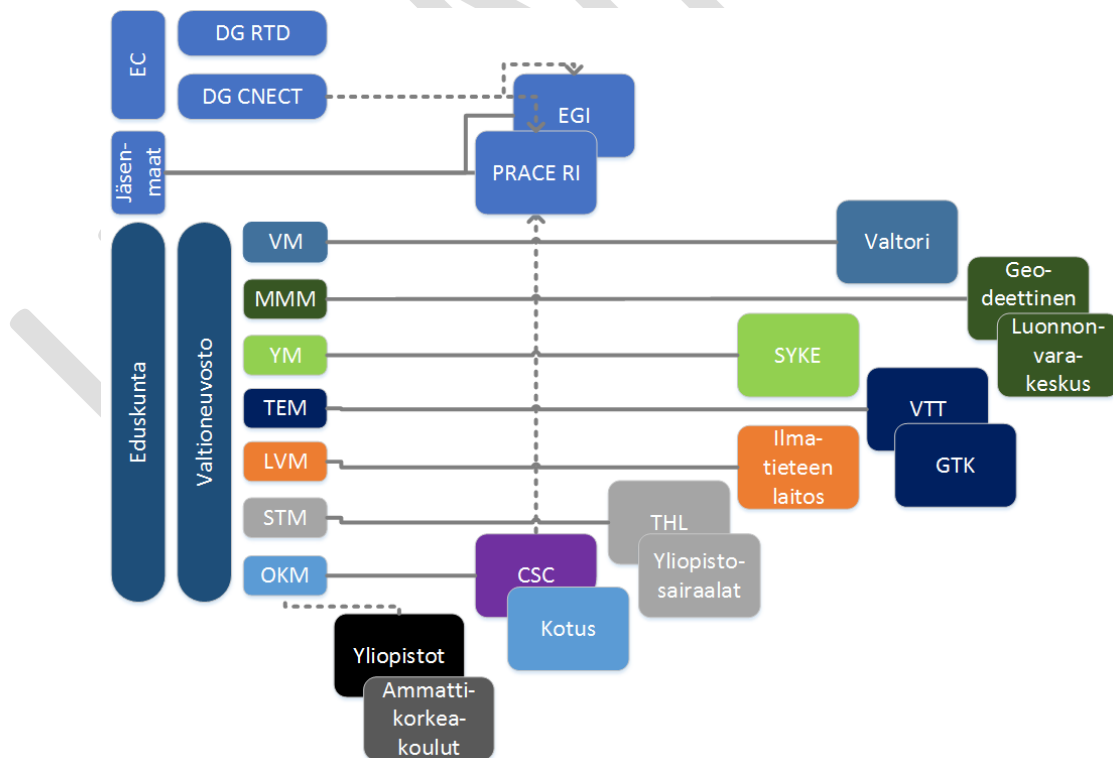
- Opetusministeriö
 - Yliopistot, ammattikorkeakoulut
 - CSC
- Tieteellisen laskennan yhteistyöfoorumi
 - CSC
- Tieteellisen laskennan asiakaspaneeli
 - CSC

2.5 Organisaatio

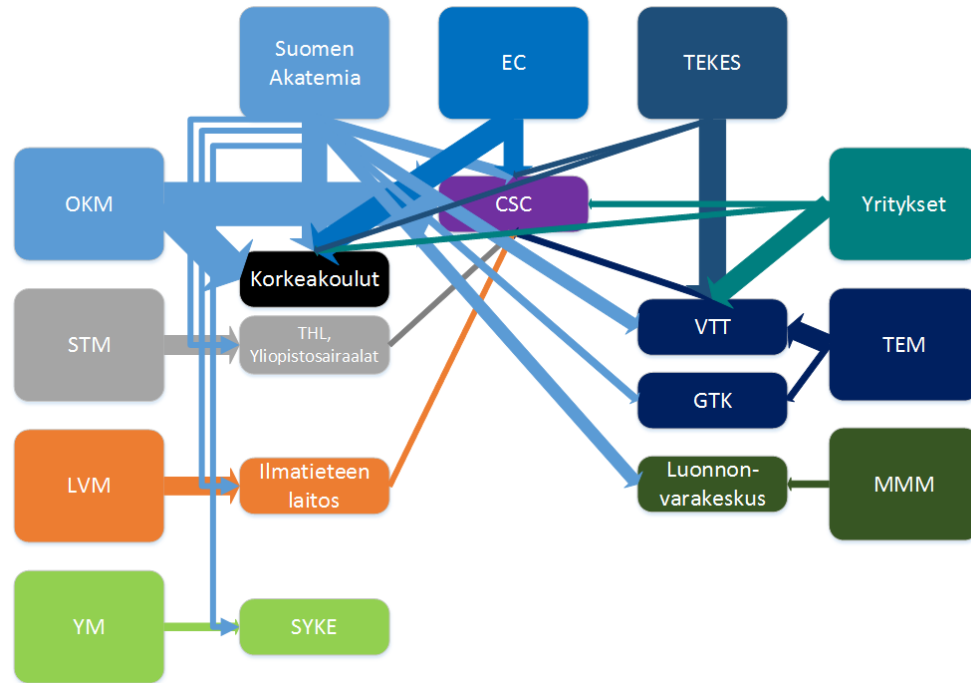
Tieteellisellä laskennalla ei Suomessa ole yhteistä organisaatiota, vaan toimijat ovat pitkälti itsenäisiä. Tutkimuslaitokset toimivat omien ministeriöidensä alla, yliopistot ja ammattikorkeakoulut ovat itsenäisiä oikeushenkilöitä ja CSC on opetusministeriön omistama osakeyhtiö.

Kuva 10 esittää tieteellisen laskennan yleistä organisaatiota Suomessa. Kuva 11 havainnollistaa tieteellisen laskennan rahoitusta ja kuvassa 12 on esitetty tieteellisen laskennan ohjausrakenteet CSC:hen liittyen.

Tavoitteena on yhteinen, laajapohjainen yli sektorirajojen ulottuva ohjaus/yhteistyöelin, joka koordinoi tieteellisen laskennan kokonaisuutta.



Kuva 10. Tieteellisen laskennan organisaatio Suomessa.



Kuva 11. Tieteellisen laskennan rahoitus.

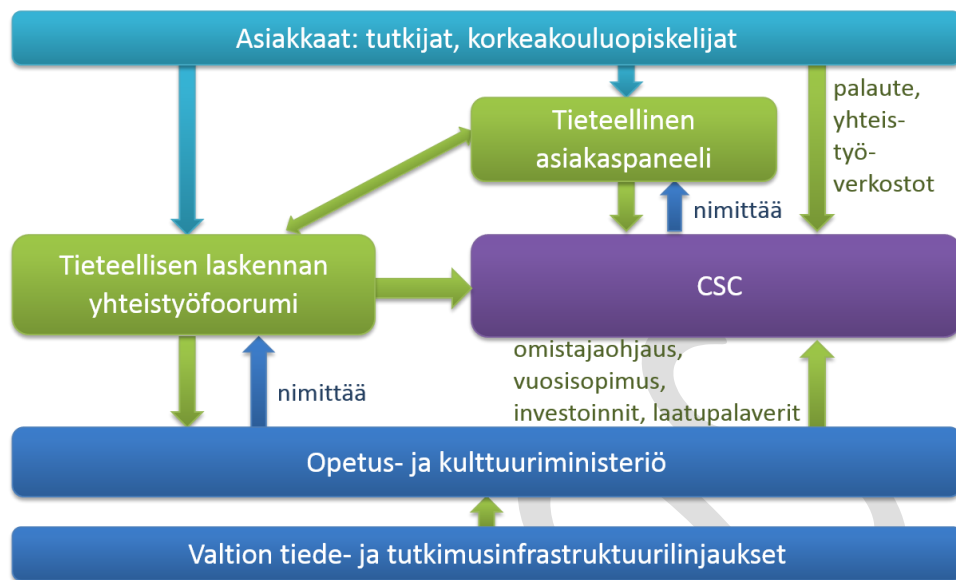
2.5.1 CSC

CSC:ssä tieteellinen laskenta ja siihen liittyvät palvelut kuuluvat tieteen palvelu -yksikköön (SR), ja laitteistojen ylläpidon ja operoinnin tallennuspalvelut mukaan lukien SR:lle tuottaa teknologia-alustat -yksikkö (ICTP). Opetus- ja kulttuuriministeriön kulttuuri- ja tiedepolitiikan osasto (KTPO) ostaa CSC:ltä korkeakouluille suunnatun tieteellisen laskennan palvelun vuosittain tarkistettavan sopimuksen mukaisesti.

Laskentapalvelinten hankinta rahoitetaan erikseen ministeriön budjetista, ja laitteistot ovat valtion omaisuutta. Tukipalvelut kuuluvat ministeriön sopimuksen piiriin, mutta niiden kehitystyötä tehdään paljon myös erilaisissa osittain ulkoisesti rahoitetuissa projekteissa.

KTPO on nimittänyt tieteellisen laskennan yhteistyöfoorumin ohjaamaan ja arvioimaan CSC:ltä ostettavia palveluita. Yhteistyöfoorumi edustaa tieteellisen laskennan palveluiden käyttäjiä ja rahoittajia, ja siinä on edustajia yliopistoista, ammattikorkeakouluista, tutkimuslaitoksista, Suomen Akatemiasta, TEKESistä ja opetus- ja kulttuuriministeriöstä. Palautetta ja ohjausta CSC saa myös tieteelliseltä asiakaspaneelilta, jonka jäsenet ovat laskennallisen tieteen professoreja.

Normaalit laskentaresurssihakemukset käsittelee CSC:n resurssienjakoryhmä (RJR), joka arvioi töiden sopivuuden CSC:n koneille ja myöntää laskentakiintiöt. RJR kokoontuu kolmen viikon välein. Suurempia resursseja tarvitsevia hankkeita varten järjestetään vuosittain tyypillisesti kaksi Grand Challenge -hakua (GC). GC- ja PRACE DECI -hakemusten tieteellisestä arvioinnista vastaa Tieteellinen asiakaspaneeli.



Kuva 12. Tieteellisen laskennan ohjausrakenteet CSC:n kannalta.

Yliopistot

Yliopistoissa tieteellinen laskenta hoidetaan kunkin yliopiston tietotekniikkaosaston ja alan tutkimusta tekevien tiedekuntien ja laitosten yhteistyönä. Yliopistot voivat tehdä myös suoraostoja CSC:ltä niistä palveluista, jotka eivät sisälly OKM:n vuosisopimukseen.

Yliopistojen eri hallintomallien takia yhteistyön formaali muoto vaihtelee, mutta pääasiallisena periaatteena on, että tietotekniikkaosastot hoitavat tietoliikenneyhteydet yliopiston sisällä ja FUNETin kautta CSC:lle ja maailmalle ja pohjainfrastruktuurin kuten konehuoneet, kun taas tutkimuksesta vastaavat tiedekunnat ja laitokset hoitavat klusterien ja muun tieteellisen laskennan spesifisen infrastruktuurin hankinnat ja hallinnoinnin. Tämä on tarkoituksenmukaista koska sillä tavalla tutkimusryhmät voivat räätälöidä paikallisen kapasiteetin optimaalisesti tutkimusta palvelemaan.

Esimerkiksi Helsingin yliopistossa tieteellistä laskentaa koordinoi matemaattis-luonnontieteellisen tiedekunnan Kumpulan kampuksen tietotekniikkatoimikunta, joissa on edustajat tietotekniikkaosastosta ja kaikista kampuksen laskentaa tekevästä laitoksista.

2.5.2 Ammattikorkeakoulut

...

2.5.3 Tutkimuslaitokset

...

2.6 Prosessit

2.6.1 Resurssien jako

Resurssien jakoprosessit toimivat eri resursseille eri periaattein ja ne noudattavat siten erityyppisiä vaikka osin samankaltaisia menettelyjä.

CSC:n hallinnoimien resurssien jako jakautuu eri tyypeihin myönnettävien resurssien suuruuden ja luonteen mukaan. Perusresurssit korkeakoulututkija tai -opiskelija saa käyttöönsä Haka-rekisteröinnillä. Keskikokoiset laskentaresurssit eli laskentaprojekti haetaan projektihakemuksella. Nämä resurssit käsittelee CSC:n Resurssienjakoryhmä. Suurimmat laskentaresurssit jaetaan ns. *Grand Challenge* -hakujen kautta. Hakuja järjestetään keskimäärin 1–2 kertaa vuodessa. Projektihakemukset arvioi CSC:n Tieteellinen asiakaspaneeli. Muiden kuin korkeakoulututkijoiden tai -opiskelijoiden laskentaresurssit sovitaan maksullisilla sopimuksilla.

PRACE Tier-0 -resurssien jako perustuu keskitettyyn malliin, jossa ns. Project access -hakuja järjestetään kaksi vuodessa ja saapuneet hakemukset arvioi yhteinen eurooppalainen tieteellinen paneeli. Lopulliset myönnot päättää ns. *Access Committee* (AC), jonka valitsee PRACEn *Scientific Steering Committee* (SSC). Prosessi on viisivaiheinen ja se sisältää lisäksi valitusmahdollisuuden:

- 1) Perustaminen ja määrittely
- 2) Haun valmistelu ja hakemusten teko
- 3) Hakemusten evaluointi ja prosessointi
- 4) Resurssien jako
- 5) Projektien käynnistäminen ja arviointi
- 6) Valitusmenettely

Lisäksi on ollut mahdollisuus hakea resursseja projektin valmistelua varten eli ns. Preparatory access -hakuja, jotka ovat periaatteessa jatkuvasti auki ja hakemukset on arvioitu tietyin väliajoin. Sen lisäksi on myös ollut mahdollisuus ns. Multiyear access -projekteihin.

PRACE DECI eli Tier-1 -resursseja on jaettu ns. *juste retour* -periaatteella, jossa kukin maa saa saman verran resursseja kuin se hakuun tarjoaa. Projektihakemusten valinta perustuu osin hajautettuun ja osin keskitettyyn malliin. Ne maat, joilla on olemassa sopiva tieteellinen elin valintaan, käyttävät omia kansallisia elimiään arvioimaan omasta maasta tulevat hakemukset. Näin on menetellyt mm. Suomi, jossa evaluointipaneelina on toiminut CSC:n Tieteellinen asiakaspaneeli. Hakemuksille, jotka tulevat maista, joissa ei tällaista elintä ole, tai jotka tulevat DECI-ryhmän ulkopuolisista maista, on käytetty yhteistä tieteellisiin perusteisiin valittua paneelia, *Non-contributing Partner Committee for Tier-1 Peer Review*. Hakuja on järjestetty 0,5–1 vuoden välein. Projektien laskenta-alustat määrää ja lopulliset resurssimyönnot tekee ns. *DECI Access and Allocations Committee* (DAAC), joka koostuu DECI-osapuolien edustajista.

2.6.2 Käyttäjähallinta

Oikeuksien hallinta

Käyttöoikeudet perustuvat käyttäjän hyväksyntään käyttöehtoihin. CSC:n yleiset käyttöehdot on sovittu OKM:n kanssa ja julkaistu [www-sivulla https://www.csc.fi/yleiset-kayttoehdot](https://www.csc.fi/yleiset-kayttoehdot).

CSC:n laskentaympäristöt integroidaan 2015 aikana CSC:n uuteen käyttäjähallintojärjestelmään. Paperisten lomakkeiden täytöstä pyritään eroon jolloin projektien vastuuhenkilöt voivat itse hallinnoida projektinsa tunnuksia sähköisesti.

Pilvipalveluissa käyttöoikeudet koskevat pilviväliohjelmistoa. Käyttäjät hallinnoivat itse virtuaalikoneidensa käyttäjätunnuksia käyttöehtojen mukaan.

HAKA-autentikointi

Käyttäjien rekisteröinti CSC:n käyttäjäksi tapahtuu suomalaisten korkeakouluasiakkaiden osalta Haka-autentikoinnilla. Muiden käyttäjien rekisteröinti tehdään käsin erikseen.

Haka on korkeakoulujen ja tutkimuslaitosten käyttäjätunnistusjärjestelmä, joka perustuu luottamusverkostoon. Haka-verkoston jäsenet, tutkijat, opiskelijat ja henkilöstö, voivat käyttää kotiorganisaationsa käyttäjätunnuksia kirjautuessaan. Myös käyttäjien henkilötietoja voidaan siirtää turvallisesti palveluihin kirjautumisen yhteydessä. Haka on yhteensopiva muiden pohjoismaiden korkeakoulujen luottamusverkostojen kanssa, joten käytettävissäsi ovat myös pohjoismaiset palvelut.

Kotiorganisaation tietohallinto vastaa käyttäjiensä käyttäjätiedoista ja henkilöllisyyden todentamisesta. Hakassa olevien palvelujen käyttäjätiedot saadaan suoraan käyttäjän kotiorganisaatiosta.

Ulkomaalaiset käyttäjät

CSC:n hallinnoimien laskentaresurssien kansainvälisen käytön toimintaperiaatteet on sovittu OKM:n kanssa vuonna 2009⁶. Näitä periaatteita noudatetaan edelleen. Pääperiaatteena on, että resursseja voidaan myöntää sellaisille ulkomailla toimiville korkeakoulututkijoille, jotka ovat suomalaisen tutkijaryhmän kanssa yhteistyössä suomalaisessa korkeakoulussa työskentelevän ryhmänvetäjän ohjauksessa, jos työn tulokset (tieto, kokemus ja julkaisut) hyödyttävät suomalaisia tutkimusryhmiä. Päätöksen tekee tältä pohjalta CSC:n johtoryhmän valtuuttama henkilö tai ryhmä. Sellaisena toimii CSC:n resurssienjakoryhmä.

Sektoritutkimuslaitosten ym. käyttöoikeudet

Sektoritutkimuslaitokset käyttävät tieteelliseen laskentaan pääsääntöisesti omia laskentaresurssejaan.

CSC:n hallinnoimien resurssien käyttö on tällä hetkellä mahdollista vain maksullisella sopimuksella.

Kansainväliset yhteistyöprojektit

CSC:n kansainvälisiin yhteistyöprojekteihin sovelletaan samaa sopimusta kuin ulkomaalaiskäyttöönkin. Pääperiaatteena on resurssien vaihto ulkomaisten keskusten kanssa, jolloin suomalaisille käyttäjille annetaan ulkomaille projektiin myönnettyä kotimaista resurssia vastaavat resurssit ulkomailta. Vaihdon ei välttämättä tarvitse tapahtua samalla kaudella vaan resurssien kysynnän ja tarjonnan kannalta sopivaan aikaan. Päätös CSC:n sitoutumisesta

⁶ Policy for international use of CSC's computing resources, 2009.

projekteihin tehdään yhteistyössä OKM:n kanssa, mutta resurssipäätökset tehdään yleensä projektien sisällä.

Käyttöprojektien elinkaaren hallinta

CSC:n koneilla käyttöprojekteissa on seuraavat periaatteet:

- Käyttäjä on ainakin yhden projektin jäsen.
- Kullakin projektilla on yksi käyttäjä, joka on projektin vastuhenkilö. Oletusarvoisesti hän on ensimmäinen käyttäjä.
- Projektin vastuhenkilö voi pyytää CSC:tä lisäämään käyttäjiä projektiinsa.
- Jäsenen lähtiessä projektista projektin vastuhenkilön on ilmoitettava tästä CSC:lle.
- CSC:n pyynnöstä projektin vastuhenkilön on raportoitava projektin etenemisestä ja palveluiden käytöstä.
- Käyttäjätunnus on voimassa enintään kaksi vuotta kerrallaan myöntöpäivästä alkaen, ellei toisin ole sovittu. CSC lähettää käyttäjälle muistutuksen hyvissä ajoin ennen tunnuksen viimeistä voimassaolopäivää.
- Palveluiden käyttöoikeus päättyy, kun käyttötarkoitus ei ole enää voimassa tai kun käyttäjä ei enää ole tutkimusprojektissa tai opiskelu- tai työsuhteessa yliopistoon tai ammattikorkeakouluun.
- Päättyvää käyttäjätunnusta ja siihen liittyvä sisältöä käsitellessään CSC noudattaa palvelukuvausta. CSC ei automaattisesti tuhoa sisältöä yrittämättä ottaa yhteyttä käyttäjään.

Näistä ehdoista sovitaan asiakkaan kanssa jo käyttäjäksi tullessa Laskentapalvelujen yleisissä käyttöehdoissa.

Käytön raportointi

CSC raportoi yleistiedot laskentaresurssien käytöstä ja projektista. Näihin tietoihin kuuluu projektin nimi, vastuhenkilö, organisaatio, kuvaus ja resurssien käyttö.

2.6.3 Resurssien ylläpito ja uudistaminen

- Hankintaprosessi
- Laitteistojen ylläpitoprosessi
- Laitteistojen elinkaaren hallintaprosessi
- Ohjelmistojen hankinta ja elinkaaren hallinta
- Tietoturva
- Palvelukehitys ja teknologiaseuranta
- Palveluiden monitorointi ja raportointi
- Palvelujen myynti
- Viestintä ja tiedotus

3 Tieteellisen laskennan tietoarkkitehtuuri

Kartturi-mallin mukaisesti tietoarkkitehtuuri kuvaa käsitteellisellä tasolla käsitteistön ja roolit ja loogisella tasolla tietomallit, loogiset tietovarannot, tietovirrat, järjestelmät ja tietovarannot

sekä prosessitiedot. Tässä vaiheessa tieteellisen laskennan kokonaisarkkitehtuuri sisältää kuitenkin vain käsitteistön ja muita arkkitehtuurin osia lisätään työn seuraavassa vaiheessa.

3.1 Käsitteitä

Laskennallinen tiede ja tieteellinen laskenta

Laskennallinen tiede on tietokoneiden kehityksen myötä vakiintunut kolmanneksi paradigmaksi perinteisten kokeellisten ja teoreettisten menetelmien rinnalle. Laskennallisen tieteen prosessi voidaan karkeasti jakaa neljään vaiheeseen:

- 1) Tutkittavasta ilmiöstä tehdään matemaattinen malli.
- 2) Matemaattiselle mallille etsitään numeerinen ratkaisumenetelmä, jota kutsutaan usein algoritmiksi.
- 3) Numeerinen algoritmi ohjelmoidaan ja ajetaan tietokoneella. Tätä kutsutaan usein simuloimiseksi.
- 4) Tulosten tulkinta ja havainnollistaminen, mikä tyypillisesti vaatii lisää laskentaa tietokoneella.

Erityisesti kohdat 2) ja 3) karakterisoivat laskennallista tiedettä. Laskennallisen tieteen ja tietokonesimulaatioiden avulla voidaan tarkastella ilmiöitä, joiden tutkiminen kokeellisesti on liian hidasta, kallista, vaarallista tai yksinkertaisesti mahdotonta. Puhtaasti teoreettinen lähestymistapa muuttuu laskennalliseksi tieteeksi, kun käytettävät matemaattiset mallit tulevat niin monimutkaisiksi, että niitä voidaan analysoida vain numeerisesti.

Tieteellinen laskenta käsittää pääasiassa edellä olevan listan kohdan 3), mutta käytettävissä olevan tietokoneen tekniset ominaisuudet voivat vaikuttaa kohtaan 2) eli algoritmin valintaan. Jatkoissa tarkastellaan tieteellisen laskennan kokonaisarkkitehtuuria tai -kuvaa, jonka keskeisimmät osakokonaisuudet ovat:

- 1) Laskentakapasiteetti käyttöympäristöineen.
- 2) Oheispalvelut kuten tallennuskapasiteetti, ohjelmistot, aineistot ja tietokannat jne.
- 3) Laskentakapasiteetin käyttöön ja ohjelmointiin liittyvä käyttäjätuki.
- 4) Edellisiin liittyvä koulutus.

Pilvilaskenta

Tieteelliseen laskentaan tarkoitettu pilvipalvelu käsittää datan prosessoinnin, tallennuksen, verkkoratkaisun ja muut olennaiset laskennallisen tieteen tarvitsemat resurssit. Palvelun käyttäjä ei operoi tai hallinnoi pilven fyysistä infrastruktuuria, mutta voivat valita haluamansa käyttöjärjestelmät, levyjärjestelmät, sekä käytettävät sovellusohjelmistot ja mahdollisesti myös verkkoliikenteen komponentit (esim. palomuri, dynaamiset valopolut). Pilvipalvelukonsepti pitää sisällään myös resurssien hallinnan ja käyttöympäristön ylläpidon.

Pilvipalveluiden kolme keskeisintä palvelumallia ovat:

- Infrastrukturi palveluna (IaaS; Infrastructure as a Service), jossa palveluntarjoaja rakentaa virtualisointikerroksen laskentaresurssien päälle. Asiakas voi rakentaa oman ympäristönsä tämän kerroksen päällä ja hallinnoida sitä. Tämä palvelumalli sopii

parhaiten käyttäjille, jotka tarvitsevat räätälöidyn ympäristön ja osaavat itse rakentaa sen.

- Sovelluslusta (PaaS; Platform as a Service). Tässä palveluntarjoaja tuottaa ja ylläpitää ympäristön, johon asiakas voi asentaa valitsemiaan sovelluksia tai palveluita.
- Sovellus (SaaS; Software as a Service). Tässä palveluntarjoaja tuottaa ja ylläpitää sovelluksen, jota asiakas käyttää.

Grid

Hajautettu eli grid-laskenta muistuttaa paljolti normaalia klusterilaskentaa. Merkittävin ero tavalliseen laskentaklusterin käyttöön verrattuna on se, että grid-laskennassa käyttäjä ei ota suoraan yhteyttä laskentapalvelimeen. Sen sijaan yhteydenoton ja tiedostojen siirron suorittaa automaattisesti niin sanottu *väliohjelmisto*. Grid-ympäristöön on tyypillisesti liitetty useita laskentaklustereita, joista väliohjelmisto valitsee tehtävään sopivat klusterit, joille työt ohjataan.

Suurteho- ja rinnakkaislaskenta

Suurteholaskenta on käsitteenä hieman epämääräinen, mutta yleensä se tarkoittaa laskentaa, jossa tarvitaan huomattavasti tavallista pöytäkonetta (tai vastaavaa) enemmän tehoa. Käytännössä teho saavutetaan rinnakaistamalla eli hyödyntämällä useita laskentaytimiä saman tehtävän ratkaisemiseen. Rinnakkaislaskenta edellyttää lähes aina ohjelman kirjoittamista tai muokkaamista siten, että se voi hyödyntää käytettävissä olevan konearkkitehtuurin rinnakkaisuutta.

Rinnakkaislaskennan peruskäsitteisiin kuuluu (vahva) skaalautuvuus, joka kuvaa ohjelman suoritusajan lyhentymistä laskentaytimien lukumäärää lisättäessä. Ihanteellisessa tapauksessa suoritus aika puoliintuu aina kun ytimien määrä kaksinkertaistuu, mutta tyypillisesti nopeutuminen pysähtyy tietyn rajan jälkeen, ja ytimien määrän kasvattaminen ei enää kannata. Hyvin kirjoitetut rinnakkaisohjelmat toimivat tehokkaasti kymmenillä ja jopa sadoilla tuhansilla ytimillä.

Tavallisimmat rinnakkaisohjelmointimallit ovat viestinvälitys, joka toteutetaan kirjastokutsuilla, ja yhteisen muistin rinnakaistus, joka toteutetaan kääntäjän direktiiveillä tai pragmoilla. Myös kiihdyttimien ja apuprosessorien teho perustuu rinnakkaisuuteen. Tehokkaan rinnakkaisohjelman kirjoittaminen vaatii yleensä hyvää asiantuntemusta ja paljon työtä.

Tier-0,1,2

Supertietokoneet ja suurteholaskentaresurssit voidaan tehon perusteella jakaa kolmeen ryhmään. Tier-0 käsittää maailman tehokkaimmat supertietokoneet, jotka ovat Top500-listalla tyypillisesti 20 ensimmäisen joukossa. Näissä koneissa on tavallisesti satoja tuhansia laskentaytimiä ja tehokas kytkentäverkko. Tier-1:een kuuluvat kansallisen tason koneet, kuten CSC:n Sisu. Tier-1-koneissa on tyypillisesti kymmeniä tuhansia laskentaytimiä. Tier-2 muodostuu paikallisen tason, kuten yliopistojen, koneista, jotka ovat tavallisesti klustereita.

Hierarkian kaikki tasot ovat välttämättömiä ja tukevat toisiaan; tie raskaimpaan Tier-0-laskentaan vaatii kokemusta ja kehitystyötä ja kulkee Tier-1- ja Tier-2-tasojen kautta. Lisäksi Tier-2-resurssit voidaan paremmin räätälöidä paikallisiin tarpeisiin, joiden tyydyttäminen keskitetysti on joskus vaikeaa tai epätarkoituksenmukaista.

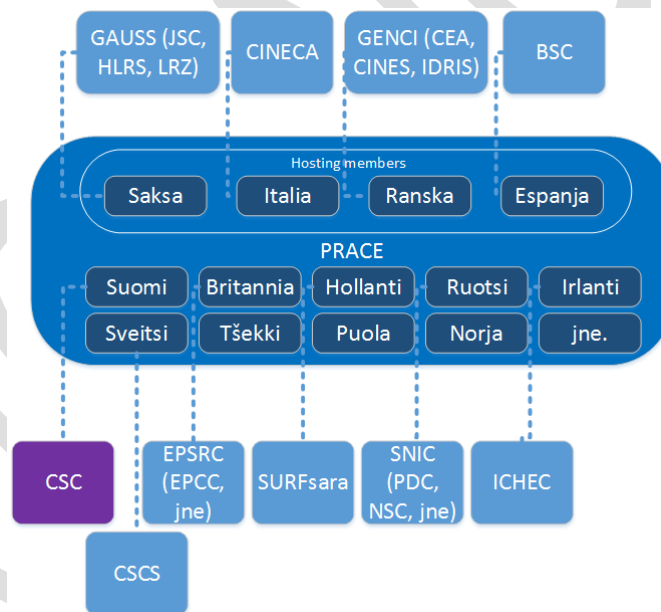
PRACE

PRACE on vuonna 2010 perustettu eurooppalainen suurteholaskennan infrastruktuuri, jonka tarkoitus on taata laskennallisen tieteen tutkijoille kansainvälisesti huippuluokan resurssit. PRACEssa on 25 jäsenmaata, ja Suomi on edustettuna CSC:n kautta (kuva 16). PRACE on hajautettu infrastruktuuri, johon tällä hetkellä kuuluu kuusi Tier-0-luokan konetta Saksassa, Ranskassa, Italiassa ja Espanjassa. Laskenta-aikaa on tarjottu kahdesti vuodessa järjestettävissä hauissa, ja ehdotukset käyvät läpi tiukan tieteellisen ja teknisen arvioinnin.

Viite: www.prace-ri.eu

PRACEn laskentaresurssit ovat toistaiseksi olleet jäsenmaille ilmaisia, mutta tilanne tulee muuttumaan infrastruktuurin toisessa vaiheessa, joka alkaa vuoden 2015 keväällä. Tarkka malli kustannusten kattamisesta on vielä auki, mutta se tulee perustumaan jäsenten suhteellisiin BKT-osuuksiin ja käyttöön ensimmäisen vaiheen aikana. Suomessa PRACE on hyväksytty Akatemian FIRI-tiekartalle.

PRACEn tulevaisuus, jatkuvuus ja toimintamalli ovat neuvotteluvaiheessa ja ne täydennetään kun kokonaisuus on selvillä.



Kuva 16. PRACEn jäsenet.

Kiihdyttimet ja apuprosessorit

Kiihdyttimillä tarkoitetaan yleiskäyttöisiä grafiikkaprosessoreita (GPGPU tai yksinkertaisesti GPU), joiden merkittävien valmistaja on Nvidia. Intelin kilpaileva arkkitehtuuri on MIC (many integrated cores), jota kutsutaan apuprosessoriksi (co-processor). Molempien perusajatus on sama: tehoa ja energiatehokkuutta on haettu rinnakkaisuudella ja usealla alhaisen kellotaajuuden fyysisellä tai loogisella laskentayksiköllä. Parhaimmillaan erikoisprosessoreilla, ja varsinkin GPU:illa, saavutetaan moninkertainen suorituskyky tavallisiin prosessoreihin verrattuna ja laskentatehon ja suorituskyvyn suhde on myöskin selkeästi parempi. Erikoisprosessorit kehittyvät jatkuvasti, ja niiden ominaisuudet sopivat yhä paremmin

laskentaan, mutta siitä huolimatta niiden ohjelmointi on edelleen hankalampaa kuin tavallisten prosessorien. Erityisesti hyvän suorituskyvyn saavuttaminen vaatii toistaiseksi hyvää asiantuntemusta ja paljon työtä.

Tällä hetkellä (vuoden 2014 lopussa) Nvidian Tesla on Intelin Knights Corner -MICin edellä käytettävyydellä ja sovellusten määrällä mitattuna, mutta Intelin seuraava versio, Knights Landing, tulee suurella todennäköisyydellä tasoittamaan kilpailutilannetta.

Mooren laki

Mooren laki on historiaan perustuva havainto, jonka alkuperäisen muodon mukaan yhden integroidun piirin tai prosessorin transistorien määrä kaksinkertaistuu kahden vuoden välein. Tämä itseään toteuttava havainto on pitänyt hyvin paikkansa 1960-luvulta lähtien nykypäiviin asti. Myös pitkään jatkuneeseen prosessorien tehon kaksinkertaistumiseen 18 kuukauden välein viitataan usein Mooren lakina. Vaikuttaa kuitenkin vahvasti siltä, että tällä hetkellä yleisesti käytössä olevalla teknologialla kasvu ei enää voi jatkua 2020-luvulla. Vaihtoehtoisia ratkaisuja kuitenkin tutkitaan intensiivisesti.

Tallennuskapasiteetti

Tässä yhteydessä tallennuskapasiteetilla tarkoitetaan nopeita levyjärjestelmiä, joilla olevaa dataa voidaan käyttää laskennan aikana. Datan prosessointiin ja hallintaan tarvittavan tallennuskapasiteetin nopeus on nykyisin useissa dataintensiivisissä hankkeissa tieteellisen laskennan pullonkaula: prosessoreille ei pystytä syöttämään dataa suurista satunnaisesti luettavista tiedostoista niin nopeasti kuin prosessorit pystyisivät sitä käsittelemään. Tietenkin jos data on kooltaan niin suurta, että sitä ei pystytä tuomaan laskentaympäristöön, laskentaa ei voida edes käynnistää. Tämän vuoksi CSC on aktiivisesti kasvattanut laskentaympäristön kapasiteettia tieteellisen datan prosessoinnille.

4 Arkkitehtuuriperiaatteet

4.1 Sidosarkkitehtuurit

Tieteellisen laskennan keskeiset sidosarkkitehtuurit ovat tutkimukseen, tietoverkkoon ja tallennukseen liittyvät sekä opetus- ja kulttuuriministeriön Opetus, tiede ja kulttuuri -kohdealueen arkkitehtuurit.

- Tallennusarkkitehtuuri
- Tutkimuksen tietoaineistot -kokonaisarkkitehtuuri
 - TTA-hanke
 - luonnos <https://www.tdata.fi/documents/10180/44526/TTA-kokonaisarkkitehtuuri/7df43aee-1e0c-4c41-b6cd-a64aaced15a4>
- Verkkoarkkitehtuurit
- Funet-kokonaisarkkitehtuuri
- Tutkimuksen kokonaisarkkitehtuuri
 - Avoimen tieteen ja tutkimuksen asiantuntijaryhmä
 - tekeillä
- Tutkimuksen tuen ja hallinnon viitearkkitehtuuri
 - Tutkimuksen tuen ja hallinnon verkosto

- tekeillä
- Korkeakoulujen opiskelun ja opetuksen viitearkkitehtuuri
 - Digitalisoituvan opiskelun ja opetuksen yhteistyöryhmä
 - tekeillä
- Korkeakoulujen ICT-palveluiden nykytila ja tavoitearkkitehtuuri
 - Korkeakoulujen tietohallinto- ja ICT-ohjausryhmä
 - tekeillä
- Yhteentoimivuuden tietoarkkitehtuuri ja yhteentoimivuusmalli
 - Korkeakoulujen tietohallinto- ja ICT-ohjausryhmä
 - tekeillä
- Kansallinen opintohallinnon viitearkkitehtuuri
 - [http://www.minedu.fi/OPM/Koulutus/artikkelit/oppijanpalvelut/liitteet/Opintohal
linnonkansallinenxviitearkkitehtuuri xv1_0.pdf](http://www.minedu.fi/OPM/Koulutus/artikkelit/oppijanpalvelut/liitteet/Opintohal
linnonkansallinenxviitearkkitehtuuri xv1_0.pdf)
- Opetus, tiede ja kulttuuri -kohdealueen kokonaisarkkitehtuurit
 - [https://www.yhteentoimivuu.s/fi/aihealue/Kohdealueet/Koulutus, tiede ja kulttu
uri](https://www.yhteentoimivuu.s/fi/aihealue/Kohdealueet/Koulutus,_tiede_ ja_kulttu
uri)
- Kansallisen Digitaalisen Kirjaston kokonaisarkkitehtuuri
 - julkaistu:
[https://www.yhteentoimivuu.s/fi/semic/view/Asset/Asset.SingleView.xhtml?id=6
0214](https://www.yhteentoimivuu.s/fi/semic/view/Asset/Asset.SingleView.xhtml?id=6
0214)
- Korkeakoulujen muut kokonaisarkkitehtuurit
- Valtionhallinnon yhteinen kokonaisarkkitehtuuri
 - [https://www.yhteentoimivuu.s/fi/aihealue/Valtionhallinnon_yhteinen kokonaisark
kitehtuuri](https://www.yhteentoimivuu.s/fi/aihealue/Valtionhallinnon_yhteinen_kokonaisark
kitehtuuri)



Kuva 15. Tieteellisen laskennan läheisimmät kokonaisarkkitehtuurit

4.2 Yleiset arkkitehtuuriperiaatteet

1. Arkkitehtuuri seuraa yleisiä ohjaavia periaatteita
2. Arkkitehtuuri on strategialähtöistä
3. Arkkitehtuuri on asiakas- ja toimintalähtöistä

4. Kokonaisarkkitehtuuri muodostaa yhtenäisen toimintaa ohjaavan kokonaisuuden ottaen huomioon läheiset toiminnot ja arkkitehtuurit
5. Arkkitehtuuri sisältää koko OKM:n tieteellisen laskennan kohdealueen
6. Arkkitehtuurin kehittäminen on avointa, ketterää ja mukautuvaa

4.3 Toiminnalliset arkkitehtuuriperiaatteet

2. Infrastruktuuri palvelee tutkimusta ja yhteiskuntaa
3. Infrastruktuuri mukautuu eri tieteenalojen tarpeisiin
4. Infrastruktuuri koostuu paikallisista ja keskitetyistä palveluista
5. Järjestelmien ja niiden välinen toiminta tukeutuu rajalliseen määrään erilaisia tekniikoita
6. Palvelut ovat tarkoituksenmukaiset, tehokkaat, luotettavat ja hyvin resursoidut

4.4 Tietoturva- ja tietosuojaperiaatteet

CSC:n laskentapalveluissa ja niiden ylläpitokäytännöissä noudatetaan CSC:n sisäistä tietoturvapoliittikkaa sekä yleisiä hyviä tietoturvakäytäntöjä. CSC:n järjestelmät on suunniteltu toteuttamaan valtionhallinnon tietoturvallisuuden (Vahti 2/2010) perustason vaatimukset.

Lisäksi laskenta-alustat sekä CSC:n datakeskukset on auditoitu tietoturvan hallintaan liittyvän standardin ISO/IEC 27001:2013 mukaisesti ja myös pilvipalvelut on tarkoitus auditoida vuoden 2015 aikana.

5 Liitteet

5.1 Lait ja säädökset

- o Laki julkisista hankinnoista (348/2007), EU:n direktiivi julkisista hankinnoista 2014 (2014/24/EU)
- o Laki julkisen hallinnon tietohallinnon ohjauksesta (634/2011)
- o Sähköisen viestinnän tietosuojalaki (516/2004), Henkilötietolaki (523/1999)
- o Laki biopankeista (688/2012)
- o Laki (409/2014) ja asetus (701/2014) biologisen tiedon ja bioinformatiikan (ELIXIR) tutkimusinfrastruktuurista
- o Tekijänoikeuslaki (404/1961), lisenssisopimukset
- o Yliopistolaki (558/2009), Ammattikorkeakoululaki (351/2003), tutkimuslaitoksia koskeva lainsäädäntö
- o EU:n, Suomen ja muiden maiden vientirajoitukset

5.2 CSC:n tämän hetkisten koneiden tekniset yksityiskohdat

Sisussa (Cray XC40) on kesällä 2014 tehdyn päivityksen jälkeen noin 40 000 laskentaydintä, ja sen teoreettinen kaksoistarkkuuden huipputeho on noin 1,7 Pflop/s ja Linpack-teho noin 1,25 Pflop/s. Kaikissa laskentanoodeissa on 64 GB muistia 24 ydintä kohden. Uusimmalla, vuoden 2014 marraskuussa julkaistulla Top 500 -listalla Sisun päivitetty konfiguraatio on sijalla 37.

Sisussa on isoille rinnakkaisajoille optimoitu erittäin nopea Cray Aries -kytkentäverkko, mikä on sen merkittävin tekninen ero klusterikoneisiin verrattuna. Tehokkaan osittain nesteeseen perustuvan jäähdytyksen vuoksi laitteisto on myös tiivis ja energiatehokas.

Taito-klusterissa (Hewlett-Packard SL) on syksyn 2014 laajennuksen jälkeen noin 19 000 laskentaydintä ja siinä on suhteellisen tehokas InfiniBand-kytkentäverkko. Laajennetun Taidon teoreettinen kaksoistarkkuuden huipputeho on noin 580 Tflop/s. Taidon muistikonfiguraatio on heterogeeninen: muistin määrä noodia kohden on 64, 128 tai 256 GB tai jopa 1,5 TB.

Sisun ja Taidon tallennusjärjestelmän (DDN- ja Lustre-tiedostojärjestelmän) kapasiteetti on 4 PB.

Espoossa sijaitsevassa pilvipalvelussa on n. 2900 laskentaydintä (Hyper-Threading) ja yhteensä 8900 GB keskusmuistia. Palvelimia on tarkoitus uusia 2015 aikana. Verkkona on 10 Gb Ethernet. Projektikohtaisen NFS-tila tulee NetApp-laitteistosta (kapasiteettia yhteensä noin 1 PB). Lisäksi 2015 aikana otetaan käyttöön CEPH-teknologiaan pohjautuva 1 PB -tallennusjärjestelmä.