

Tieteellisten laskentaresurssien ja niiden oheispalvelujen kokonaisarkkitehtuuri



Opetus- ja
kulttuuri-
ministeriö

Tämän määrittelyn on laatinut opetus- ja kulttuuriministeriön asettama
tieteellisen laskennan yhteistyöfoorumi (asettamispäätös OKM/20/040/2013)

Tieteellisten laskentaresurssien ja niiden oheispalvelujen kokonaisarkkitehtuuri

1	Yleistä	5
1.1	Laskennallinen tiede ja tieteellinen laskenta.....	5
1.2	Rajaukset ja reunaehdot	6
1.3	Nykytila.....	6
1.3.1	Laskennallinen tiede Suomessa	6
1.3.2	Tieteellisen laskennan käyttäjät	8
1.3.3	Tutkimuslaitosten näkökulmia	11
1.3.4	Resurssien ja käyttäjien (hallintoalueiden) väliset raja-aidat Suomessa	12
1.3.5	Laskentaresurssit	14
2	Tieteellisen laskennan toiminta-arkkitehtuuri	19
2.1	Toiminnan haasteet ja tavoitteet	19
2.1.1	Keskitettyjen resurssien aidosti kansallinen käyttö.....	19
2.1.2	Teknologian kehityksen aiheuttamat haasteet.....	19
2.1.3	Laitteistojen elinkaari ja rahoitusyksiö	20
2.1.4	Laskentapalveluiden kehittämisen haasteita	21
2.1.5	Kokonaisoptimointi ja kustannustietoisuus.....	22
2.1.6	Tieteellinen laskenta elinkeinoelämässä	22
2.2	Palvelut	23
2.2.1	Laskentaresurssit	23
2.2.2	Ohjelmisto- ja tietokantapalvelut	25
2.2.3	Tukipalvelut	25
2.2.4	Kriittiset palvelut muualta.....	26
2.3	Toimijat ja sidosryhmät	27
2.3.1	Palveluiden tuottajat.....	27
2.3.2	Loppukäyttäjät	28
2.3.3	Rahoittajat	28
2.3.4	Ohjaus	28
2.3.5	Asiantuntijaneuvonanto CSC:lle.....	29
2.4	Tieteellisen laskennan organisointi Suomessa	29
2.4.1	CSC – Tieteen tietotekniikan keskus Oy	29
2.4.2	Yliopistot.....	30
2.4.3	Ammattikorkeakoulut	30
2.5	Prosessit	31
2.5.1	Resurssien jako	31
2.5.2	Käyttäjähallinta	32
2.5.3	Resurssien ylläpito ja uudistaminen.....	34
2.5.4	Muut prosessit	34
3	Ohjaavat strategiat ja tavoitetila.....	34
3.1	Olemassa olevat strategiat.....	34
3.2	Ohjeellinen tieteellisen laskennan strategia	35
3.3	Tavoitetila	36
3.3.1	Tieteellisen laskennan kokonaisuus tukee kilpailukykyistä ja kansainvälisesti houkuttelevaa tutkimusjärjestelmää ja mahdollistaa uusia tieteellisiä avauksia.	37
3.3.2	Tieteellisen laskennan kokonaisuus organisoidaan mahdollisimman kustannustehokkaasti yhdistelmänä keskitettyä ja hajautettua ratkaisua	38

4	Taustalla vaikuttavat arkkitehtuurit ja periaatteet	39
4.1	Sidosarkkitehtuurit	39
4.2	Yleiset arkkitehtuuriperiaatteet	40
4.3	Toiminnalliset arkkitehtuuriperiaatteet	40
4.4	Tietoturva- ja tietosuojaperiaatteet.....	41
5	Liitteet	41
5.1	Lait ja säädökset.....	41
5.2	CSC:n tämänhetkisten koneiden tekniset yksityiskohdat	41
5.3	Käsitteitä/Sanasto.....	42

Tiivistelmä

Tässä dokumentissa kuvataan ensimmäistä kertaa tieteellisten laskentaresurssien ja niiden oheispalvelujen kokonaisarkkitehtuuri. Arkkitehtuurikuvaus keskittyy nyky- ja tavoitetilan kuvaukseen sekä tieteellisen laskennan kehittämisen strategiaan. Arkkitehtuurikuvauksessa keskeisimmäksi asiaksi nousee eri sektorirajojen luomat raja-aidat, jotka haittaavat kansallisesti tehokasta tieteellisen laskennan toimintaa. Tulevaisuuden tavoitetilana nähdäänkin paikallisen ja keskitetyn toiminnan laajapohjainen koordinointi kansallisella tasolla. Etenkin dataintensiivisen laskennan menetelmiä otetaan käyttöön yhä useammilla tieteenaloilla ja jatkossa tieteellisen laskennan merkitys korostuu entisestään.

Johdanto

Laskennallinen tiede 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 Suomen 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 – Tieteen tietotekniikan keskus Oy:n (CSC) tarjoama kansainvälisestikin erittäin korkea määrä laskentakapasiteettia ryhmää kohden.

Lisääntyvän digitalisaation myötä tieteellinen laskenta laajenee perinteisten laskennallisten tieteen alalta kovaa vauhtia muihin tieteesiin ja esimerkiksi dataintensiivisen laskennan menetelmiä sovelletaan jo rutiininomaisesti mm. kielitieteesiin ja biologiaan.

Laskennalliseen tieteeseen ja tieteellisen laskentaan panostaminen on kansallisesti erittäin tärkeää, koska

- 1) Laskennallinen tiede tukee vahvasti poikki- ja monitieteellisyyttä, jotka tuottavat tulevia innovaatioita ja liiketoimintaa.
- 2) Suomen Akatemian valitsemista tutkimuksen huippuyksiköistä iso osa tarvitsee tutkimuksessaan laskennallisen tieteen osaamista ja välineitä.
- 3) Laskennallinen tiede on monesti kustannustehokasta suhteessa kokeelliseen tutkimukseen, ja sen avulla voidaan selvittää asioita, joita ei voi aina tutkia kokeellisin tai analyyttisin menetelmin.
- 4) Samoja fasiliteetteja ja osaamista voi käyttää monilla tieteenaloilla.
- 5) Laskennallinen tiede on jatkuvasti kasvava trendi, kun digitalisaatio lisääntyy (internet of things, teollinen internet).
- 6) Lähes kaikille suurille kansainvälisille tutkimushankkeille ja infrastruktuureille tieteellisen laskennan resurssit ja osaaminen ovat pakollinen edellytys.
- 7) Hyvät tieteellisen laskennan fasiliteetit luovat Suomeen kansainvälisesti kilpailukykyisen ja kansainvälistä yhteistyötä houkuttelevaa tutkimusympäristön.

Tieteellisten laskentaresurssien ja niiden oheispalveluiden kokonaisarkkitehtuuri kuvaa suomalaisten laskennallisen tieteen ja tutkimuksen käytössä olevaa laskenta- ja tallennuskapasiteettista, 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.

Tieteellisten laskentaresurssien ja niiden oheispalveluiden kokonaisarkkitehtuurin tarkoitus on kuvata suomalaista tieteellistä laskentaa, edistää eri toimijoiden välistä yhteistyötä ja ohjata toimintaa ja sen suunnittelua siten, että kokonaisuus palvelee mahdollisimman hyvin laskennallisen tieteen yhteisöä ja koko yhteiskuntaa. Kokonaisarkkitehtuuri on ennen kaikkea johtamisen ja toiminnan kehittämisen tukimenetelmä.

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

Kokonaisarkkitehtuurin keskeisimmät osakokonaisuudet ovat nyky- ja tavoitetilän kuvaukset sekä tieteellisen laskennan strategia. Tämä käsillä oleva arkkitehtuurikuvaus vastaa tietohallintolain (10.6.2011/634) opetus- ja kulttuuriministeriölle (OKM) antamaan velvoitteeseen edistää oman toimialan tietoteknistä yhteentoimivuutta soveltamalla kokonaisarkkitehtuurimenetelmää niin kansallisesti keskitetyllä kuin paikallisten palvelujen tasolla. Tämä on ensimmäinen kerta, kun tieteellinen laskenta kuvataan kokonaisarkkitehtuurin avulla ja siten tämä luo pohjan tulevien vuosien päivityksille. Tällä tieteellisen laskennan kuvauksella on oma tärkeä arvonsa niin sanotun TULA-periaatepäätöksen (5.9.2013) pohjalta käynnistetyssä korkeakoulujen ja tutkimuslaitosten yhteistyön syventämisen prosessissa.

Käsillä oleva kuvaus on syntynyt tieteellisen laskennan yhteistyöfoorumin nimittämän ryhmän toimesta. Ryhmään kuuluivat Paula Eerola (Helsingin yliopisto), Juha Herrala (Tampereen teknillinen yliopisto), Jyri Naarmala (Yrkeshögskolan Novia), Minna Palmroth (Ilmatieteen laitos), Juha Fagerholm (CSC), Sami Niinimäki (OKM), Juha Haataja (OKM) ja ryhmän puheenjohtajan roolissa Jussi Heikonen (CSC).

Arvokkaita kommentteja eri versioihin ovat antaneet seuraavat tahot joille suuri kiitos: Heikki Räisänen (työ- ja elinkeinoministeriö), Ilpo Vattulainen (Tampereen teknillinen yliopisto), Ari Kuusio (Hämeen ammattikorkeakoulu), Antti Poso (Itä-Suomen yliopisto), Kai Nordlund (Helsingin yliopisto), Kai Puolamäki (Työterveyslaitos), Kaisa Miettinen (Jyväskylän yliopisto), Tero Tuononen (CSC), Janne Ignatius (CSC), Eija Korpelainen (CSC), Per Öster (CSC), Pekka Lehtovuori (CSC), Klaus Lindberg (CSC), Suvi Remes (CSC), Jura Tarus (CSC), Aleks Kallio (CSC), Ville Savolainen (CSC), Tommi Nyrönen (CSC), Jarno Laitinen (CSC), Peter Råback (CSC), yliopistojen IT-johtajien verkosto FUCIO, Helsingin yliopisto, Oulun yliopisto, Aalto-yliopisto, Terveyden ja hyvinvoinnin laitos, Tampereen teknillinen yliopisto, Luonnonvarakeskus ja Teknologian tutkimuskeskus VTT Oy.

1 Yleistä

1.1 Laskennallinen tiede ja tieteellinen laskenta

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

- 1) Tutkittavasta ilmiöstä tehdään matemaattinen malli tai teoria.
- 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.
- 5) Tulosten verifiointi ja/tai kalibrointi mittausten avulla => palaa kohtaan 1, kunnes tulokset ovat riittäviä soveltuvilla kriteereillä mitaten.

Monilla sovellusaloilla pelkkä simulointi ei kuitenkaan riitä. Mallin ja simuloinnin edut saadaan täysimääräisesti hyödynnettyä, kun kohtiin 1-3 kytetään mukaan optimointi, jonka avulla valittujen kriteerien mukaan etsitään parasta ratkaisua (esim. kestävin, edullisin, ympäristöystävällisin tai tasapaino näiden väliltä).

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.

Tieteellistä laskentaa hyödyntää kuitenkin myös edellä esitetyn laskennallisen tieteen mallin rinnalle 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 ja vastaavia tieteitä datatieteiksi.

Klassinen esimerkki jälkimmäisistä on bioinformatiikka, jossa tietojenkäsittelytiede ja biologia/genetiikka ovat yhdistyneet aivan uudeksi tieteenalaksi tietojenkäsittelyn ja biologian tutkijoiden tiiviinä yhteistyönä. Seurauksena on syntynyt aivan uusia tieteellisiä käsitteitä: esimerkiksi geenien vuorovaikutusverkosta on tullut biologian käsite, mikä ei olisi ollut mahdollista pelkästään biologien voimin. Osa datatieteestä vaatii suurta laskentakapasiteettia ja muuta infrastruktuuria, mutta merkittävä osa työstä voidaan tehdä vaatimattomallakin infrastruktuurilla, jopa "kuluttajataso" laitetehoilla. Oleellista on osata käsitellä dataa uusilla tavoilla, ja infrastruktuuri on työkalu tähän.

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 sekä
- 4) edellisiin liittyvä koulutus.

Tässä arkkitehtuurissa tallennuksen kohdalla rajoitetaan käsittelemään vain välittömästi laskentaan tai analysointiin liittyvää nopeaa levykapasiteettia tai vastaavaa. Suurten datamäärien hallintaa, jakamista ja pitkäaikaista tallentamista ei kuitenkaan tarkastella, koska ne on kuvattu omassa arkkitehtuurissaan (ks. 4.1).

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 kustannustehokas käyttöikä on noin neljä vuotta, Suomen tämän hetkinen tutkimusinfrastruktuurien tiekartta¹ on voimassa vuosina 2014–2020, ja EU-infrastruktuurien rahoitusinstrumenttien näkymä H2020-ohjelman kautta on noin viiden vuoden päässä, joten tarkastelujaksossa on pyritty keskittymään seuraaviin viiteen 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 parhailaan kaikille – esimerkiksi kännyköihin, paikantamiseen, terveydenhoitoon ja energiantuotantoon, ja sen ohessa mitä luontevimmin uuden tieteellisen tiedon tuottamiseen.

Suomessa laskennallinen tiede kattaa oleellisesti kaikki tieteenalat (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 kasvat-

¹ Suomen tutkimusinfrastruktuurien strategia ja tiekartta 2014–2020, FIRI-asiantuntijaryhmä, Suomen Akatemia ja Opetus- ja kulttuuriministeriö 28.2.2014.

taen. Samassa hengessä laskennallinen tiede on Suomessa saanut merkittävän aseman eri tietealojen 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 (Espoo, Helsinki, Kuopio, Tampere).

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 huipputason 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 lukuisissa² on merkittävä laskennallisen tieteen osuus, joka tukee muuta, useimmiten kokeelliseen tutkimukseen pohjautuvaa tutkimusta. Mainittakoon lisäksi, että Suomen Akatemian vuonna 2012 tekemässä fysiikan tieteenala-arvioinnissa todetaan laskennallisen tieteen olevan Suomessa erittäin korkeatasoista³.

Kattavaa arviota laskennallisen tieteen osuudesta suhteessa kaikkeen suomalaiseen tutkimukseen ei liene vielä tehty, mutta suuntaa antavana arviona voidaan pitää yksittäisten yliopistojen tilanteita. Esimerkiksi Tampereen teknillisessä yliopistossa laskennallisen tieteen osuus sen kaikista 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

² 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.

³ "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

laskennalliselle tieteelle ja sen kehitykselle antanut. Jotta laskennallinen tiede yhtenä suomalaisen tieteen keihäänkärkialana ja korkean vaikutuksen tuottajana kehittyisi myös jatkossa, on CSC:n tieteellisen laskennan toiminnan tukeminen ja edelleen vahvistaminen välttämätöntä.

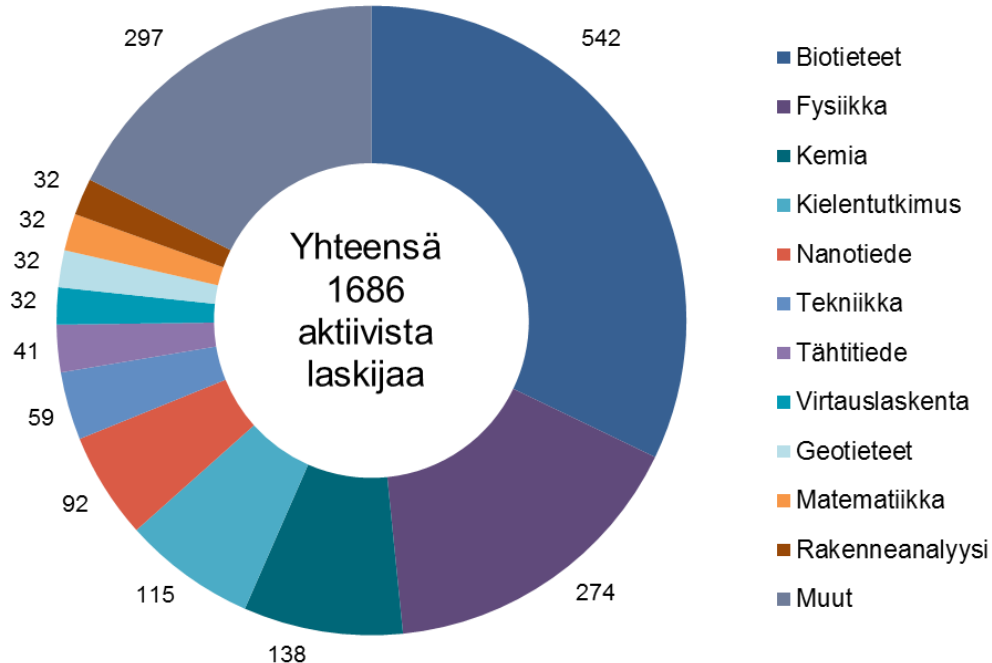
Dataintensiivisen tieteen ja laskennan yleistymisen tukee omalta osaltaan laskennallisen tieteen kasvua.

1.3.2 Tieteellisen laskennan käyttäjät

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

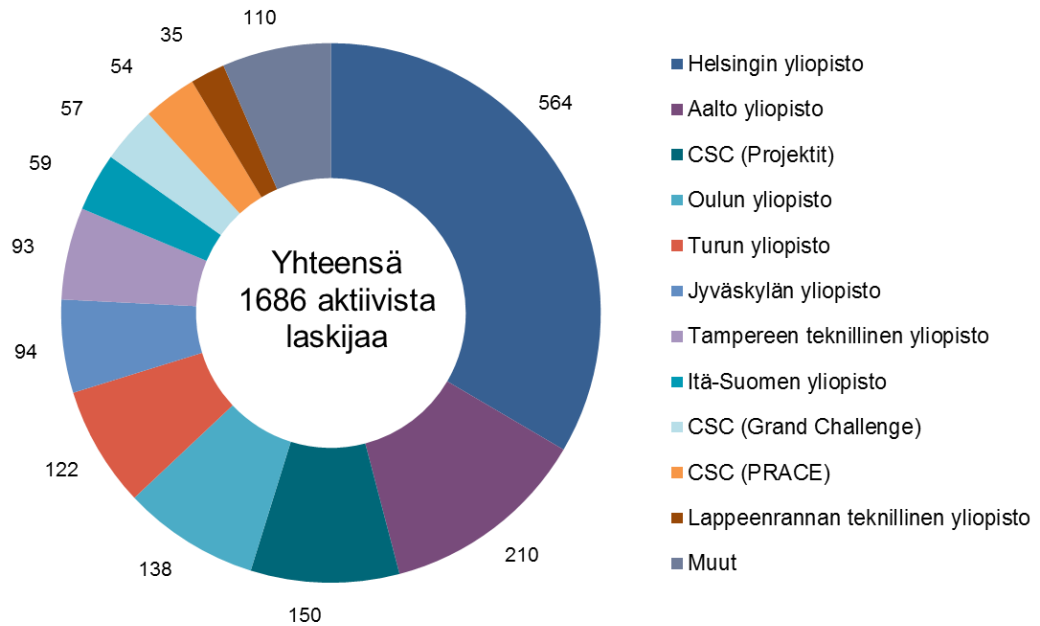
CSC:n keskitettyjä tieteellisen laskennan resursseja käytetään pääasiassa yliopistoissa, ja pienessä määrin myös ammattikorkeakouluissa ja tutkimuslaitoksissa. Käyttö jakaantuu monen tieteenalan kesken. Lukumääräisesti suurin käyttäjäryhmä on biotieteen tutkijat ennen fyysikoita, kemistejä ja kielentutkimusta (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



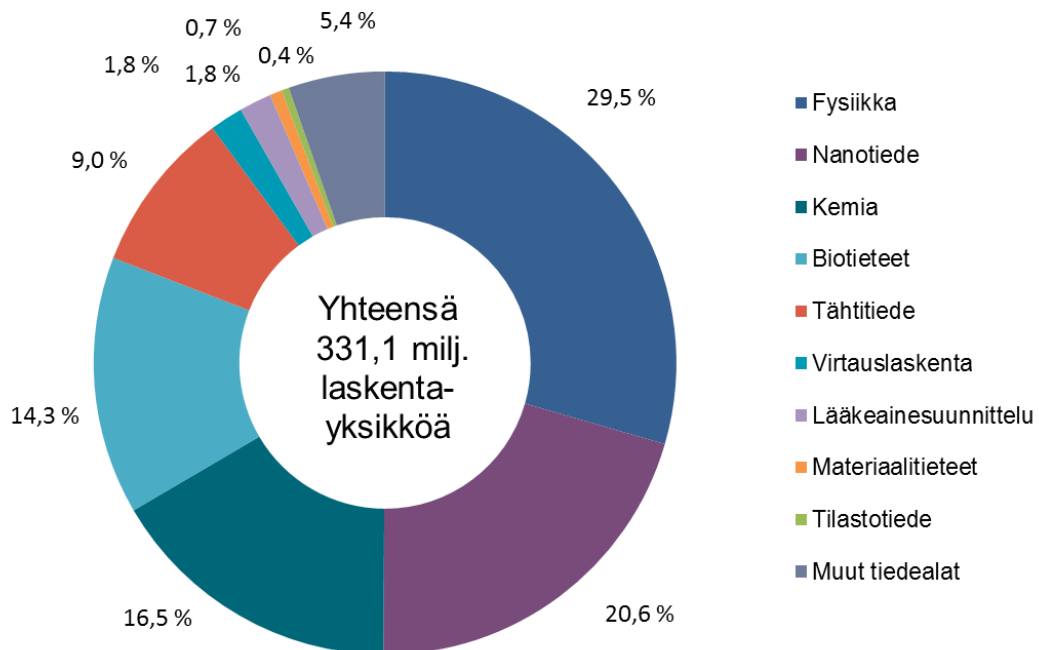
Kuva 1. CSC:n laskenta-asiakkaat tiedealoittain 2014.

Tietokoneressurssien käyttäjät organisaatioittain 2014



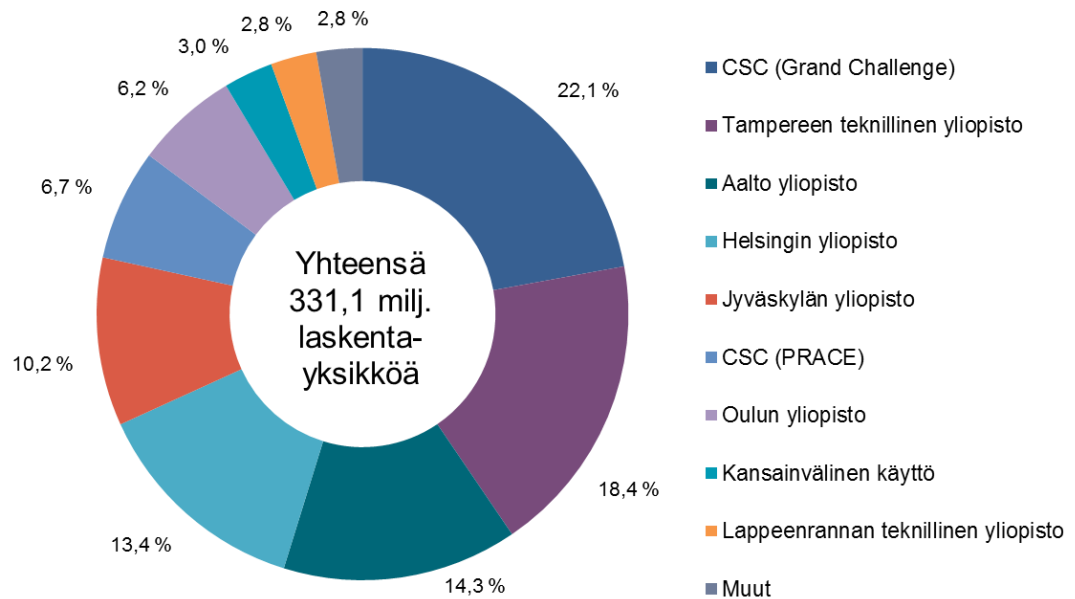
Kuva 2. CSC:n laskenta-asiakkaat organisaatioittain 2014.

Tietokoneressurssien käyttö tiedealoittain 2014



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

Tietokoneresurssien 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 säännestämisen ja tutkimuksen kesken. Ilmatieteen laitos on myös käyttänyt huomattavan määrän PRACE Tier-0 -kapasiteettia avaruussään tutkimukseen liittyen. PRACE (*Partnership for Advanced Computing in Europe*) on eurooppalainen suurteholaskennan ekosysteemi, jolla pyritään takaamaan suurteholaskentaresurssien laaja saatavuus ja tasapuoliset käyttömahdollisuudet EU:ssa.

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

On kuitenkin huomattava, että eri sovellukset tarvitsevat erityyppisiä resursseja: optimaaliset resurssit ovat keskenään hyvin erilaisia laskenta- 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-pilvipalvelu (*Infrastructure as a Service*) toi uuden tavan käyttää CSC:n laskenta- resursseja. Käyttäjät voivat valita haluamansa käyttöjärjestelmän ja asentaa ohjelmistoja täysin pääkäyttäjäoikeuksin. 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 laskentaklusteriympäristöllä ei ole pystytty vastaamaan.

Kattava kuvaus laskennallisen tieteen suurteholaskentatarpeista Euroopassa on esitetty raportissa PRACE Scientific Case (www.prace-ri.eu/prace-the-scientific-case-for-hpc).

1.3.3 Tutkimuslaitosten näkökulmia

Simulaatioperustaisen raskasta numeronmurskauskapasiteettia vaativan tieteellisen laskennan ohella tutkimuslaitoksissa on paljon mielenkiintoa dataintensiivistä laskentaa ja sensitiivisen datan käsittelyä tukevia ratkaisuja kohtaan. Seuraavassa on muutamia esimerkkejä.

Työterveyslaitos

Työterveyslaitos (TTL) on voimakkaasti siirtämässä toimintaansa laskennallisen tieteen ja erityisesti datatieteiden suuntaan. Tarkoituksena on pystyä hyödyntämään TTL:n omia ja muiden tietolähteitä mahdollisimman tehokkaasti eri toiminta-alueilla ja integroimaan datatieteet toimintoihin. Tätä varten mietitään, miten infrastruktuuri kannattaa järjestää, mutta vähintään yhtä tärkeä kysymys on, miten datatiede saadaan integroiduksi mm. tutkimustoimintaan ja miten saadaan paras hyöty tietojenkäsittelytieteen ja substanssialueen (TTL:ssä esim. lääketiede, psykologia, kognitiotiede jne.) osaamisesta.

Keskeistä on datan ymmärtäminen ja uusien laskennallisten käsitteiden ja menetelmien tuominen laitoksen toiminta-alueille. Esimerkiksi datan hallintaratkaisut ovat tärkeitä.

Pelkkä tietokonetehon tai muun kapasiteetin määrä ei anna kattavaa kuvaa. Pelkästään infrastruktuuritason ratkaisut eivät yksin riitä, vaan tarvitaan myös laskennallisen tieteen ja datatieteen huippuosaamista.

Luonnonvarakeskus

Luonnonvarakeskuksessa (Luke) tieteelliselle laskennalle on kasvava tarve. Lukessa tieteellinen laskenta on tärkeä osa eläin-, kasvi- ja mikrobigenomiikan, biometrisen genetiikan ja bioinformatiikan tutkimusta sekä monia diversiteetin projekteja. Laskentakapasiteettia tarvitaan kaikkeen genomisen datan analysointiin, homologia-hauista geeniekspressioanalyysiin, mukaan lukien RNAseq- ja metagenomiikka-analyysit. Teknologisen kehityksen myötä tutkimuksessa käsiteltävän genomidatan määrä ja sitä myöten laskentakapasiteetin tarve tulee kasvamaan. Lisäksi suurehkoa laskentakapasiteettia tarvitaan mm. kaukokartoituspohjaisessa mallinnuksessa.

Luken mukaan eri tutkimustoimijoilla olisi oltava yhtäläiset mahdollisuudet saada tieteellistä laskentakapasiteettia käyttöönsä joustavasti ja kohtuullisilla kustannuksilla. Tällä hetkellä LUKE käyttää osin omaa kapasiteettia, osin ostaa CSC:n palveluita ja osin ostopalveluita kaukallisilta palveluntarjoajilta.

Lukessa joillakin genetiikan tutkimusaloilla käytettävissä olevan laskentatehon puute aiheuttaa jo ongelmia, mm. hidastaa tutkimusta ja aiheuttaa lisätyötä. Toisissa, keskisuuria aineistoja käsittelevissä hankkeissa oma laskentakapasiteetti riittää toistaiseksi. Erätyömuotoinen toimintatapa on koettu osin kankeaksi jonotusaikoinen. Lisäksi osin pyritään käyttämään kapasiteettia, jota tutkimuksen kehittämien sovellusten käyttäjät (elinkeinot) voivat realistisesti saada käyttöönsä esimerkiksi viiden vuoden sisällä.

Laskentakapasiteetin lisäksi on Luken mukaan erittäin tärkeää tavoitella hyvää tutkimusyhteistyötä laskentamenetelmien ja laskennan asiakasratkaisujen kehityksessä. Tässä yhteydessä on mietittävä toimintamallit, joilla CSC:n ja tutkimuslaitosten yhteisen kehitystyön tulokset ovat osapuolten käytettävissä kohtuullisilla ehdoilla myös hankkeen päätyttyä. Riittävä laskentakapasiteetti varmistaisi ja voimistaisi suomalaisten tutkimusryhmien asemaa kansainvälisissä tutkimuskonsortioissa.

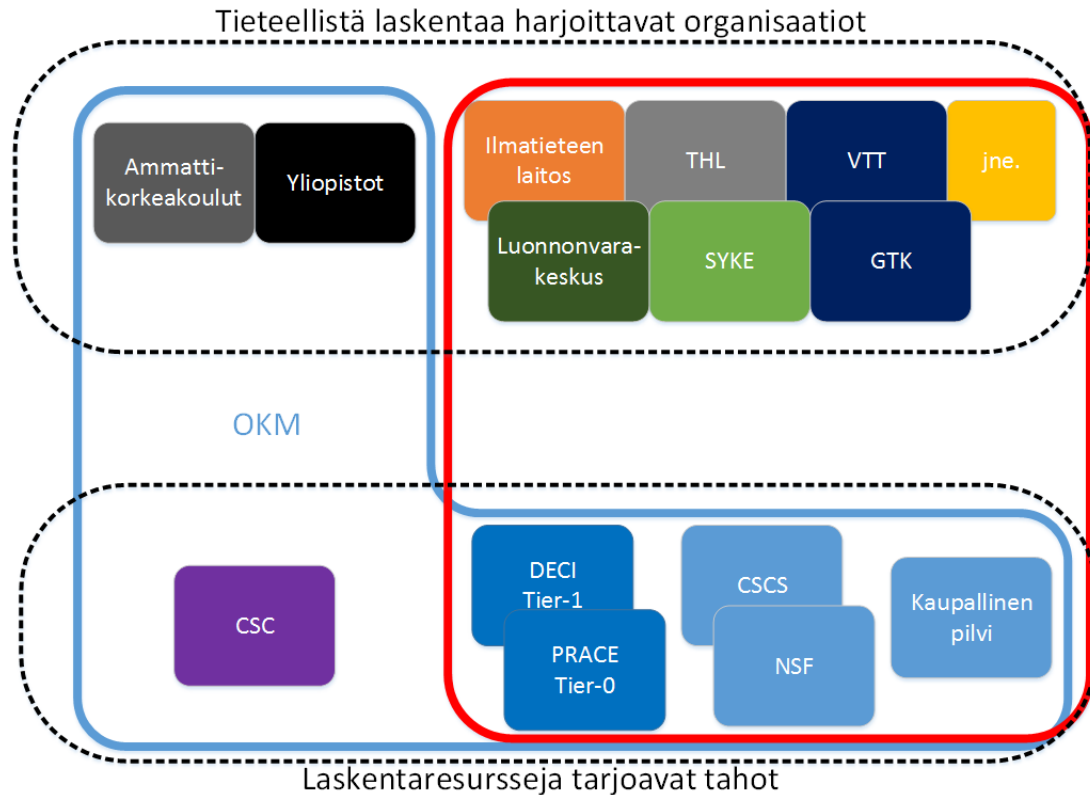
Teknologian tutkimuskeskus VTT Oy

VTT:n mukaan tieteellisen laskennan tarve on kiistatta kasvava ja suomalaisilla toimijoilla on jatkossa kasvava tarve käyttää laskentakapasiteettia monimuotoisesti. CSC:n lisäksi VTT on käyttänyt muiden toimijoiden tarjoamaa kapasiteettia.

VTT:n toiminta vaatii joustavien liiketoimintamahdollisuuksien hyödyntämisen ja mahdollisuuden oppia uusia teknologioita eturintamassa. VTT:n toiminnasta suuri osa on kansainväliseen asiakasliiketoimintaan tai ei-julkiseen puoleen perustuvaa projektiluonteista innovaatio-toimintaa, joka edellyttää joustavuutta niin teknologioiden, tietoturva vaatimusten ja erityisesti oikea-aikaisen resurssisaatavuuden suhteen. VTT:n liiketoimintaan liittyvä tarve ja kysyntä vaihtelevat jatkuvasti, joten tarkka pitkänajan sitoutuminen tiettyihin teknologioihin ei ole aina tarkoituksenmukaista.

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

CSC:n keskitettyjen laskentaresurssien veloitukseton käyttö on mahdollista vain opetus- ja kulttuuriministeriön toimialalle eli yliopistoille ja ammattikorkeakouluille, koska toiminta rahoitetaan niin sanotun korkeakoulujen kehittämismomentin kautta. 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. Tier-0 ja Tier-1 tason tieteellinen laskenta Suomessa. THL = Terveystieteiden tutkimuslaitos, VTT = Teknologian tutkimuskeskus VTT Oy, SYKE = Suomen ympäristökeskus, GTK = Geologian tutkimuskeskus, DECI = Distributed European Computing Initiative (PRACE), CSCS = Sveitsin kansallinen superkonekeskus, NSF = National Science Foundation (USA).

Kansainvälinen käyttö (suomalaisten tutkijoiden kansainvälinen käyttö ja päinvastoin) 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 (eurooppalainen datainfrastruktuuri, www.eudat.eu) ja EGI (European Grid Infrastructure, www.egi.eu) myötä.

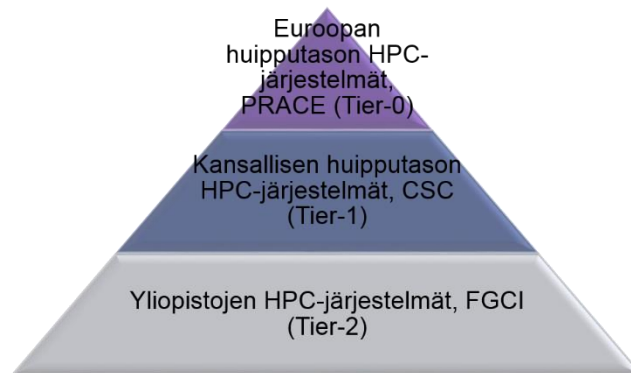
On myös mahdollista ostaa pilvilaskentakapasiteettia kaupallisilta tarjoajilta kuten Amazonilta. Tällaiset palvelut sopivat kuitenkin parhaiten toisistaan riippumattomien töiden suorittamiseen ja mittakaavaltaan vaatimattomaan rinnakkaislaskentaan, eivätkä pysty korvaamaan tehokkailla kytkentäverkoilla varustettuja supertietokoneita. Myöskään varsinaista käyttäjätukea ei ole saatavilla.

Lisäksi esimerkiksi korkeakoulut ja Ilmatieteen laitos ylläpitävät laskentakapasiteettia omaan käyttöönsä.

1.3.5 Laskentaresurssit

Laskentaresurssien hierarkia

Tieteellisen laskennan kapasiteetin voidaan ajatella muodostavan hierarkkisen kokonaisuuden, jossa suurteholaskentaresurssit luokitellaan usein kolmeen kategoriaan laskevan tehon mukaan: Tier-0, Tier-1 ja Tier-2. Tarjolla olevan laskentatehon ja käyttäjien lukumäärän suhdetta havainnollistaa seuraava pyramidi:



Laajin käyttäjäkunta on Tier-2-koneilla, ja vain äärimmäistä tehoa tarvitsevia tutkimusongelmia kannattaa ratkoa Tier-0-laitteilla. Tier-1 on teholtaan ja käyttäjämäärältään näiden kahden välissä.

Pyramidin voi ajatella jatkuvan alaspäin kerroksella, johon kuuluvat mm. tutkijoiden omat pöytäkoneet ja muut vastaavan tason resurssit. Edellisessä luvussa kuvattu pilvikapasiteetti sijoittuu suorituskyvyltään Tier-2 -tasoon.

Tier-0:lla tarkoitetaan maailman tehokkaimpia tietokoneita, eli käytännössä Top500-listan 10–20 tehokkaimman koneen joukkoon kuuluvia laitteistoja. Tier-0-koneissa on tyypillisesti satoja tuhansia laskentaytimiä 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 (ks. luku 1.3.5) kautta.

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

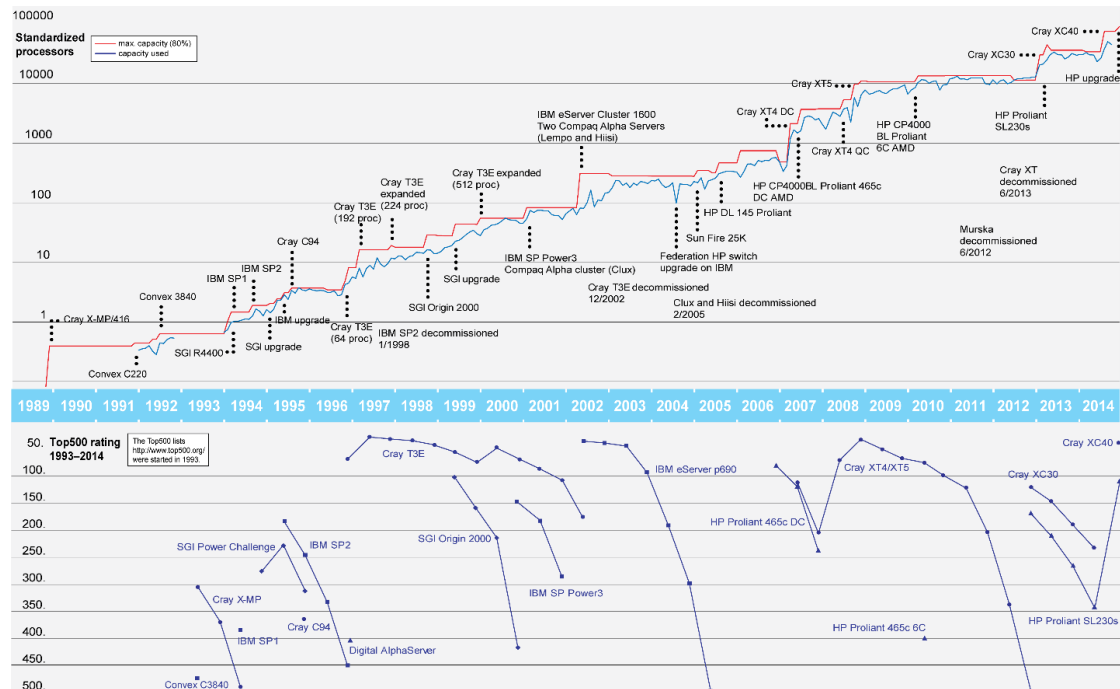
Laskentatehon kasvattaminen Tier-0 ja -1-tasolle vaatii kalliita erikoisratkaisuja, joista ei ole hyötyä pienemmissä koneissa. Tier-2 koostuu paikallisista yliopisto- tai laitostason laitteista, jotka ovat tyypillisesti nopealla InfiniBand-kytkentäverkolla ja tehokkailla prosessoreilla varustettuja klustereita. Näitä koneita käytetään sekä peräkkäisajoihin että pieniin muutaman sadan ytimen rinnakkaisajoihin.

Edellä kuvattua luokittelua voidaan pitää suurteholaskennan ekosysteeminä: hierarkian kaikki tasot ovat välttämättömiä ja tukevat toisiaan. Tie raskaimpaan Tier-0-laskentaan vaatii koke-musta ja kehitystyötä ja kulkee Tier-1- ja Tier-2-tasojen kautta. Hyppy suoraan huipputasolle on käytännössä mahdoton myös siksi, että resurssihakemuksessa on tyypillisesti pystyttävä näyttämään ongelman sopivuus ja ohjelmiston suorituskyky isoilla koneilla.

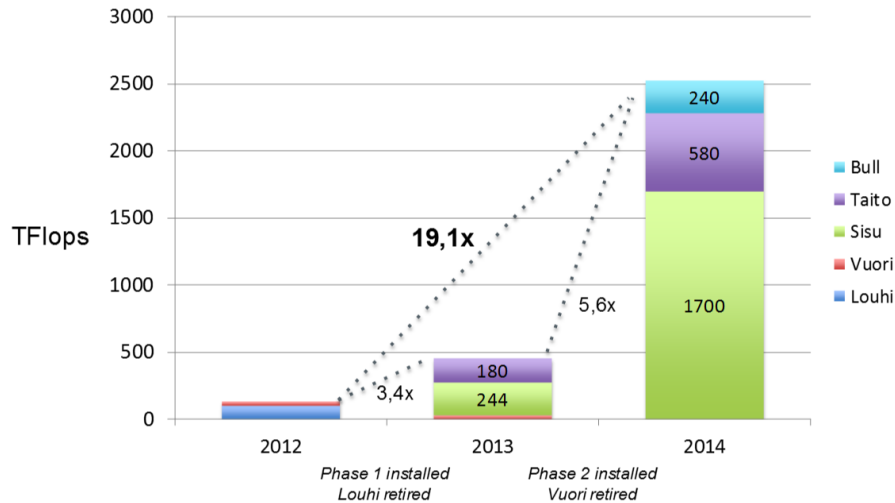
Suomessa jako CSC:n tarjoaman keskitettyjen Tier-1- ja yliopistojen hajautettujen Tier-2-re-surssien välillä on tarkoituksenmukainen ja toimii hyvin. Raskaan kapasiteetin keskittäminen on kustannustehokasta ja toisaalta hajauttamalla saavutetaan toimintavarmuutta ja joustavuutta. Tier-2-resurssit voidaan paremmin räätälöidä paikallisiin tarpeisiin, joiden tyydyttäminen kes-kitetysti on joskus vaikeaa tai epätarkoituksenmukaista. Lisäksi hajauttaminen tukee osaamisen säilymistä ja kehittämistä.

CSC:n laskentaresurssit

CSC:n laskentaresurssien kasvu vuodesta 1989 lähtien on esitetty kuvassa 6. Vuonna 2014 suo-ritettu 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 hyödyntä-mään tarjotut resurssit tehokkaasti, ja koneiden käyttöaste on tyypillisesti ollut erittäin korkea ja elinkaaren loppupäässä jonotusajat pitkiä. Koneiden elinkaaren alkupäässä niiden suoritus-kyky on ollut kansainvälisellä superkoneiden Top500-listalla viidenkymmenen paremmalla puolella, ja yleensä Euroopan ja Pohjoismaiden nopeimpien järjestelmien joukossa.

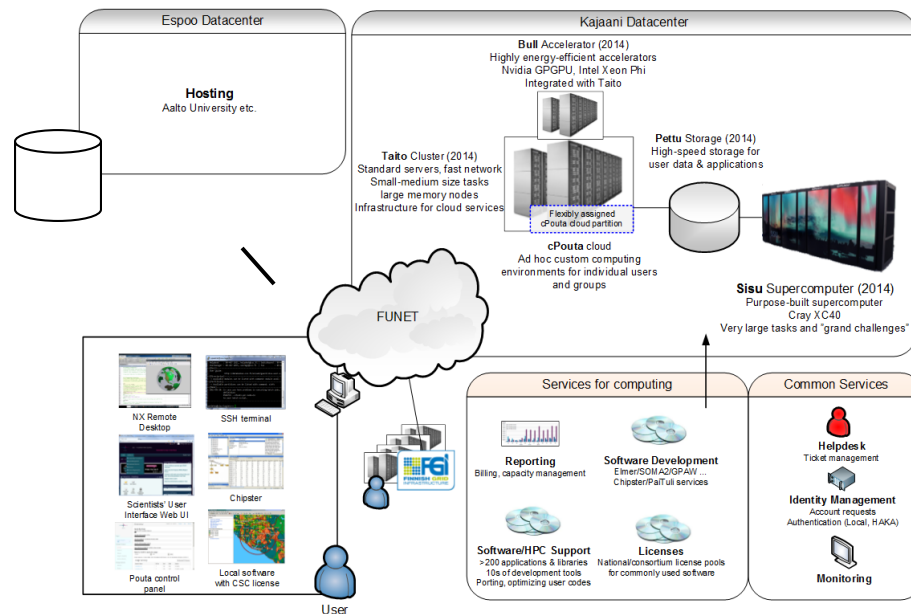


Kuva 6. CSC:n laskentakapasiteetin ja käytön kehitys 1989–2014.



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

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



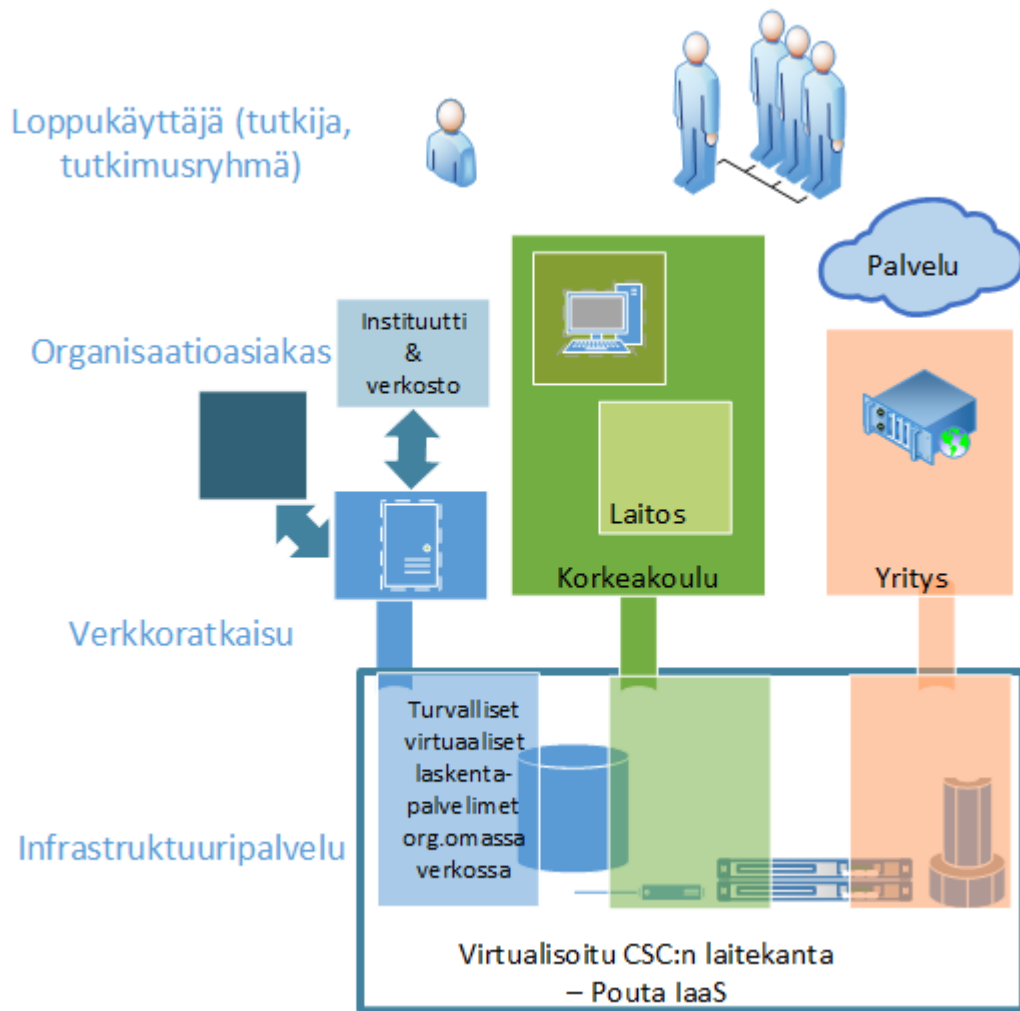
Kuva 8. CSC:n laskentaympäristö 2015 (suunnitelma).

CSC on profiloitunut Sisun isoja rinnakkaisajoja varten: normaalin käytön puitteissa suurin sallittu työn koko on 9600 ydintä eli noin 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 kehittämiä 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 noin 10 miljoonaa core-tuntia. (Sisussa vuodessa teorettinen maksimi on 350 M coreh 100 % käytöllä.)

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.



Kuva 9. IaaS-pilvipalvelumalli organisaatioille.

Laskentaresurssit yliopistoissa, ammattikorkeakouluissa ja tutkimuslaitoksissa

Yliopistot ovat tuottaneet omaa laskentakapasiteettia 1960-luvulta lähtien, jolloin Fysiikan laskentatoimisto perustettiin Helsingin yliopistossa. 1990-luvun alkuun 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, integrointi paikalliseen arkkitehtuuriin, lisenssikysymykset ja äärimmäisen korkea luotettavuus.

Kun grid-teknologia, jolla laskentaresursseja 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 sitä seuranneissa 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 pilviteknologioiden kehitys- ja testausalustana.

Yhteispohjoismaiset resurssit: NeIC

Nordic e-Infrastructure Collaboration (NeIC) on pohjoismainen hajautettu e-infrastruktuuri⁴. Sen tehtävänä on kehittää ja tuottaa korkeatasoisia laskenta- ja tallennuspalveluja pohjoismaisille 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 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 PRACEen 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 (2002–2011).

⁴ <http://neic.nordforsk.org/>

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.

2 Tieteellisen laskennan toiminta-arkkitehtuuri

2.1 Toiminnan haasteet ja tavoitteet

2.1.1 Keskitettyjen resurssien aidosti kansallinen käyttö

CSC on opetus- ja kulttuuriministeriön omistama yritys, ja sen tarjoamat tietokoneressit ja niiden varassa toimivat palvelut ovat ministeriön korkeakoulu- ja tiedepolitiikan osaston (KTPO) kustantamina pääsääntöisesti maksuttomia yliopistojen ja korkeakoulujen tutkijoille ja opiskelijoille. KTPO-CSC-vuosisopimus rahoitetaan niin sanotulta korkeakoulujen kehittämisen momentilta. Tarvetta raskaammalle laskentakapasiteetille ja mahdollisuudelle massiivisen 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 julkisen sektorin tutkijoiden käytettävissä samoilla käyttöehdoilla. Tutkimuslaitosten tarvitseman raskaan laskentakapasiteetin keskittäminen olisi myös kustannustehokkaampaa ja säästäisi valtion rahaa.

2.1.2 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: ko-

neissa on enemmän laskentasoimuja (joiden välinen kommunikaatio pitää ohjelmoida eksplisiittisesti), solmujen prosessoreissa on enemmän ytimiä (joiden välillä voidaan käyttää yhteisen muistin rinnakkaisuutta) ja ytimit 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 räätälöitävät 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 erikoisprosessorille 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 kallis sekä taloudellisesti että palvelun tason kannalta.

Haasteisiin voidaan vastata huolellisella teknologiaseurannalla ja -testaamisella sekä korkeatasoisen asiantuntemuksen ylläpidolla ja jatkuvalla kehittämisellä.

2.1.3 Laitteistojen elinkaari ja rahoitussykli

Tekniikan kehityksen ja erityisesti jatkuvasti parantuvan suorituskyvyn ja sähkön kulutuksen suhteen vuoksi suurteholaskennan laitteistojen 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 pitkäjänteistä 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 asiakkailla 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 ennakoidusta. 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.

CSC:n hallintaan päätyvien laitteiden pääomakustannuksia on usein jouduttu käsittelemään kertaostoina ja kertapoistoina yhden budjettikauden sisällä (esim. laitteiden leasing ei ole aina sallittua). Tämän seurauksena kehitettävien laskentapalveluiden vuosikustannuksiin syntyy suuriakin vaihteluita. Parempi malli olisi, että palveluiden tuottamisen tietotekniset edellytysten eli laitekannan kustannukset voitaisiin jakaa tasaisesti koko luvutulle 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.

2.1.4 Laskentapalveluiden kehittämisen haasteita

Kuten edellä todetaan, suurteholaskenta-arkkitehtuurien kasvavan rinnakkaisuuden tehokas hyödyntäminen vaatii asiantuntemusta ja paljon työtä. Tällainen työ on välttämättömyydestään huolimatta väliinputoajan 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 henkilötyöstä 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 ovat usein monivuotisia projekteja, vaikka niiden kautta tarjottavat palvelut mielletään pysyviksi ja ne 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 toimintamalliin 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.

2.1.5 Kokonaisoptimointi ja kustannustietoisuus

Ideaalitilanteessa 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 totuudenmukaisesti samoja sääntöjä käyttämällä. Tämä lisää kustannustietoisuutta ja ohjaa rahan tehokkaaseen käyttämiseen. Käytettävien periaatteiden tulisi olla myös kansainvälisesti yhteensopivia (esim. EUNIS-organisaation Bencheit-työtä voidaan hyödyntää).

2.1.6 Tieteellinen laskenta elinkeinoelämässä

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.

Toisaalta niin sanotun big datan osalta tilanne on ehkä muuttumassa, kun yrityksissä on havaittu massatiedon mahdollisuudet esimerkiksi tuotteiden kohdistamisessa oikealle kuluttajaryhmälle tai erilaisissa optimointiongelmassa. 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 näytekoelmia 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 (www.elixir-europe.org, FIRI) -hankkeen myötä yhteistyössä kolmen bioalan pienen tai keskisuuren yrityksen kanssa laskentapalveluiden tarjoajana löytääkseen rahoituksellisesti ja sopimuksellisesti kestävä mallin infrastruktuuritason yhteistoimintaan.

2.2 Palvelut

Tuottajat ja käyttäjät 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	T,K	K	
Käyttäjätuki	T	T,K	K	
Koulutus	T	T,K	K	K

Kuva 10. Keskeisimpien palveluiden tuottajat ja käyttäjät: T = tuottaja, K = käyttäjä. T, K = tuottaja ja käyttäjä.

2.2.1 Laskentaresurssit

Superkoneet, klusterit ja erikoisprosessorit

Tärkein tieteellisen laskennan palveluista on laskentaresurssien tarjoaminen. Laskentaresurssien hierarkia ja ekosysteemi on kuvattu luvussa 1.3.5.

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 tänä syksynä. Vaikka erikoisprosessorit eivät sovi kaikkien ongelmien ratkaisemiseen, niille voidaan kustannustehokkaasti siirtää osa tavallisten prosessorien kuormasta. Erikoisprosessorit ja niiden ohjelmointimallit ovat edelleen kehitysvaiheessa. Ainakin toistaiseksi erikoisprosessorien tehokas hyödyntäminen edellyttää erittäin korkeatasoista teknistä osaamista ja paljon ohjelmointityötä; erikoisprosessoreita ei siis tässä mielessä voi pitää yleiskäyttöisinä.

Grid-laskenta

Eri paikoissa olevia laskentaresursseja voidaan myös yhdistää suuremmaksi kokonaisuudeksi hajautetun laskentamallin (grid-laskenta) avulla. Tässä 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. MGrid ja sitä seuranneen FGI/FGCI-konsortion puitteissa.

Molemmissa 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.

Virtualisoidut laskentaresurssit ja pilvi

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 laskentaresurssia 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 tuotantotasosta (Pouta IaaS) palvelua huhtikuusta 2014 saakka. Pilviympäristö soveltuu hyvin myös dataintensiiviseen laskentaan ja ympäristöä ollaan kehittämässä laajemmin datankäsittelyn tarpeita huomioivaksi vuoden 2015 aikana.

Pilvipalveluiden mahdollistama käyttöympäristön räätälöinti on ja tulee jatkossakin olemaan osalle käyttäjistä ratkaisevan tärkeää. Ainakaan tällä hetkellä pilviteknologia ei kuitenkaan mahdollista vaativimpien rinnakkaisajojen tehokasta ajamista, mutta on odotettavissa, että suorituskyky paranee tulevaisuudessa. On silti epätodennäköistä, että kaupalliset pilvipalvelut lähituisivat kilpailemaan kansallisesti strategisesti tärkeiden superlaskentakeskusten asiakkaista, jotka saavat laskenta-aikansa käytännössä maksutta. Lisäksi tieteellinen ja suurteholaskenta ovat mittakaavaltaan vaatimattomia muuhun tietokoneiden hyödyntämiseen verrattuna. Pilvipalvelut tulevat siis olemaan osa kokonaisuutta perinteisen superkonekapasiteetin rinnalla, mutta eivät korvaa sitä.

Korotetun tietoturvatason laskenta

Virtualisointiin perustuva Pouta Infrastructure as a Service (IaaS) -pilvipalvelu 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 tietoaineistoja. 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 järjestelmä- ja teknologia-arkkitehtuurin näkökulmasta oman organisaation sisällä. Kontrolli ympäristön virtualisoiduista laitteista ja laitetilasta loppukäyttäjiin säilyy oman organisaation hallinnassa.

Ottaakseen IaaS-palvelun käyttöön arkaluonteisen tietoaineistojen 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 pääsopimuksen liitteeksi tietoturvasopimus, 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. Teknolgiaseurannan vuoksi on kuitenkin oleellista, että uusia laitearkkitehtuureja pystytään hankkimaan myös testikäyttöön ilman olemassa olevaa asiakastarvetta. Ainoastaan näin pystytään varmistamaan oikeat päätökset hankintoja tehtäessä.

2.2.2 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.

2.2.3 Tukipalvelut

Käyttäjätuki

Perustukeen kuuluu 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 menettämätuen 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. Asiantuntija-palveluihin 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.

Kuten luvuissa 2.1.2 ja 2.1.4 todetaan, korkeatasoiset tukipalvelut ovat avainasemassa laitteistojen tehokkaan hyödyntämisen kannalta ja niiden riittävä resursointi on erittäin tärkeää. Osamisen jatkuvaa kehittämistä voidaan edistää muun muassa toimijoiden välisellä yhteistyöllä ja tiedon vaihdolla.

Koulutus

Perinteinen ohjenuora koulutukseen liittyen CSC:n ja yliopistojen välillä on ollut päällekkäisyyden välttäminen; CSC järjestää kurseja tieteellisistä ohjelmistoista, eri ohjelmointikielistä, optimoinnista ja rinnakkaistamisesta ja yliopistot ovat huolehtineet numeerisesta matematiikasta ja laskennallisesta tieteestä. Tieteelliseen laskentaan liittyvän koulutuksen merkitys kasvaa jatkuvasti uusien käyttäjäryhmien ja luvussa 2.1.2 kuvatun teknologian kehityksen myötä.

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. CSC tekee aktiivista kansainvälistä koulutusyhteistyötä myös nopeasti kehittyvän bioinformatiikan alalla. Rahoitusta korkealaatuisiin koulutustilaisuuksiin on saatu useista EU-projekteista, ja eurooppalaisen ELIXIR-hankkeen ja maailmanlaajuisen GOBLET-organisaation jäsenenä CSC saa mm. koulutusmateriaaleja, ulkomaalaisia luennoitsijoita sekä parhaiden koulutustapojen neuvontaa.

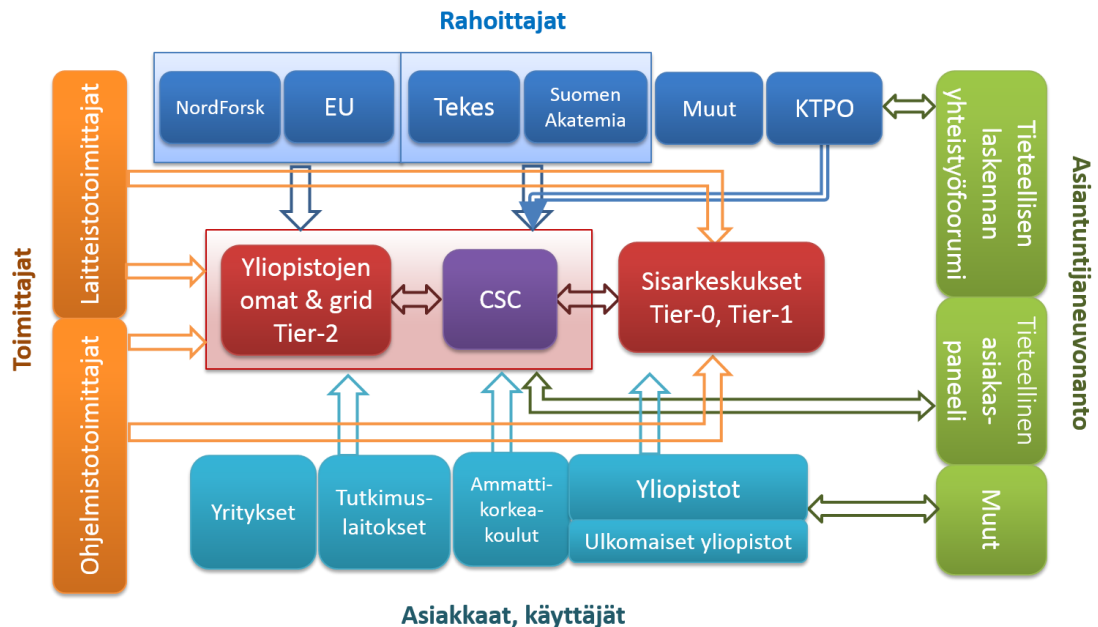
2.2.4 Kriittiset palvelut muualta

- Tutkimuksen tietoverkko Funet ja sen lisäpalvelut
 - eduroam
- Tallennus ja arkistointipalvelut (ks. 4.1)
 - Työtilat datan reaaliaikaiseen käsittelyyn
 - Lyhytaikainen ja keskipitkä varastointi
 - Varmennuspalvelut (back-up)

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

2.3 Toimijat ja sidosryhmät

Seuraavassa tieteellisen laskennan toimijat ja sidosryhmät on luokiteltu palveluiden tuottajiin, loppukäyttäjiin, rahoittajiin ja ohjaaviin elimiin (kuva 11).



Kuva 11. Tieteellisen laskennan toimijat ja sidosryhmät.

2.3.1 Palveluiden tuottajat

- CSC – Tieteen tietotekniikan keskus Oy
 - Laskenta- ja tallennusresurssit
 - Ohjelmistot ja tietokannat
 - Laskentaympäristön käytön tuki
 - Tiede- ja menetelmätuki
 - Koulutus
 - Konesalipalvelut
- Yliopistojen ja ammattikorkeakoulujen tietotekniikkakeskukset/laitokset/laboratoriot
 - Esim. Helsingin yliopisto, Aalto, TTY, jne.
 - Laskentaresurssit
 - Laskentaympäristön käytön tuki
- Tutkimuslaitosten tietotekniikkakeskukset
 - Esim. Ilmatieteen laitos, VTT, Luke jne.
 - Laskentaresurssit
 - Tiede- ja menetelmätuki
 - 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
 - European Center for Medium-Range Weather Forecasts (ECMWF)
 - ELIXIR
 - Pohjoismainen sääkonehanke
 - Kv. laskentakeskukset
 - Hajautettu laskentainfrastruktuuri (Grid, pilvi IaaS)
 - Kaupalliset pilvipalvelut ja -brokerit
 - Jne.

2.3.2 Loppukäyttäjät

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

2.3.3 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

2.3.4 Ohjaus

- Opetusministeriö
 - Yliopistot, ammattikorkeakoulut
 - CSC

2.3.5 Asiantuntijaneuvonanto CSC:lle

- Tieteellisen laskennan yhteistyöfoorumi
- Tieteellisen laskennan asiakaspaneeli

2.4 Tieteellisen laskennan organisointi Suomessa

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.

Kuvassa 12 on esitetty tieteellisen laskennan ohjausrakenteet CSC:hen liittyen.

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

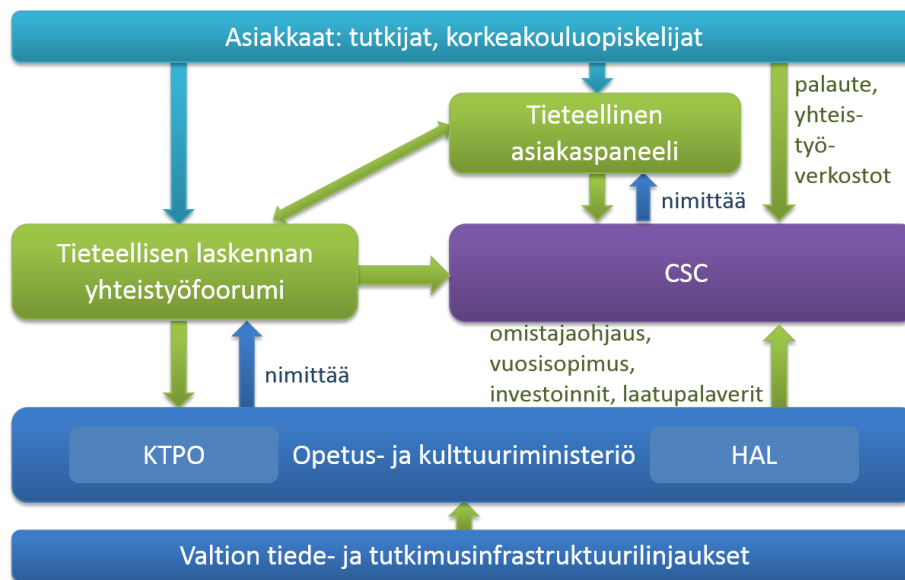
2.4.1 CSC – Tieteen tietotekniikan keskus Oy

CSC:ssä tieteellinen laskenta ja siihen liittyvät palvelut kuuluvat tieteen palvelut -yksikköön (Services for Research, SR), ja laitteistojen ylläpidon ja operoinnin tallennuspalvelut mukaan lukien SR:lle tuottaa teknologia-alustat -yksikkö (ICT Platforms, 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 sekä 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.

2.4.2 Yliopistot

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

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

Esimerkiksi Helsingin yliopistossa tietotekniikkakeskus hoitaa keskitetysti yliopiston IT-palveluita ja -hankkeita sekä mm. auttaa tarvittaessa tutkijoita CSC:n palveluiden hyödyntämisessä. Tieteellisen laskennan infrastruktuurihankkeet, esim. FGCI, koordinoidaan laitos- tai tutkimusryhmävetoisesti.

2.4.3 Ammattikorkeakoulut

Ammattikorkeakouluissa tieteellisen laskennan rooli on ollut perinteisesti yliopistosektoria vähäisempi. Syynä ovat lähtökohtaisesti olleet ammattikorkeakoulujen käytännönläheisen opetuksen ja TKI-toiminnan painotukset. Uusi ammattikorkeakoululaki (14.11.2014/932) tuo tähän todennäköisesti muutospaineita, sillä ammattikorkeakouluja veloitetaan panostamaan tulevaisuudessa enemmän myös tutkimustoimintaan. Tieteellinen laskenta sisältyy joissakin koulutusohjelmissa opetukseen, sekä erilaiset yhteistyöprojektit ja TKI-toiminta osaltaan mahdollistavat tieteellisen laskennan palveluiden hyödyntämisen ammattikorkeakoulujen projekteissa.

Tieteellisen laskennan rooli tulee jatkossa dataintensiivisen laskennan ja suurten tietomassojen vuoksi olemaan merkittävämpi myös ammattikorkeakoulusektorilla.

Tieteellisen laskennan rahoituksen osalta CSC:n tarjoaminen palveluiden tehokkaampi hyödyntäminen sekä vaikuttavuus ammattikorkeakoulukentässä ovat olennaisen tärkeitä tekijöitä kansallista kilpailukykyä ajatellen. Korkeakoulusektorin rahoituksessa tapahtuneet viime vuosien muutokset ovat pakottaneet korkeakoulut uudelleen miettimään rahoitusta ja painopistealueitaan. Kalliiden ohjelmistojen ja järjestelmien tehokas yhteiskäyttö tarjoaa nykyisessä taloustilanteessa ammattikorkeakouluille mahdollisuuden laadukkaan ja korkeatasoisen tutkimus- ja kehitystyön tekemisen jatkossakin.

2.5 Prosessit

2.5.1 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 Hakarekisteröinnillä. Keskikokoiset laskentaresurssit eli laskentaprojektit haetaan projektihakemuksella, jonka käsittelee CSC:n resurssienjakoryhmä. Hakemukset arvioidaan tieteellisen vaikuttavuuden ja laadun, kuten esimerkiksi tieteellisten julkaisujen perusteella. Resurssienjakoryhmä seuraa kansallista tiedepolitiikkaa ja noudattaa Suomen tiedehallinnon määrittelemiä prioriteetteja. Laskentaresurssit kohdennetaan ensisijaisesti kansalliseen tutkimukseen. Myönnettyistä resursseista raportoidaan kerran vuodessa CSC:n hallitukselle sekä opetus- ja kulttuuriministeriölle.

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 myönnöt 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 elinvalintaan, käyttävät omia kansallisia elimiään arvioimaan omasta maasta tulevat hakemukset. Näin on menetelty 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 perustein 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.

2.5.2 Käyttäjähallinta

Oikeuksien hallinta

Käyttöoikeudet perustuvat käyttäjän hyväksymiin käyttöehtoihin. CSC:n yleiset käyttöehdot on sovittu OKM:n kanssa ja julkaistu www-sivulla <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

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

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 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 vastuuhenkilö. Oletusarvoisesti hän on ensimmäinen käyttäjä.
- Projektin vastuuhenkilö voi pyytää CSC:tä lisäämään käyttäjiä projektiinsa.
- Jäsenen lähtiessä projektista projektin vastuuhenkilön on ilmoitettava tästä CSC:lle.
- CSC:n pyynnöstä projektin vastuuhenkilön 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, vastuuhenkilö, organisaatio, kuvaus ja resurssien käyttö.

2.5.3 Resurssien ylläpito ja uudistaminen

- Teknologiaiseuranta
- Hankintaprosessi
- Laitteistojen ylläpitoprosessi
- Laitteistojen elinkaaren hallintaprosessi
- Ohjelmistojen hankinta ja elinkaaren hallinta

2.5.4 Muut prosessit

- Tietoturva
- Palvelukehitys
- Palveluiden monitorointi ja raportointi
- Palveluiden myynti
- Viestintä ja tiedotus

3 Ohjaavat strategiat ja tavoitetila

3.1 Olemassa olevat strategiat

Suomen Akatemian vuonna 2014 julkistama "Suomen tutkimusinfrastruktuurien strategia ja tiekartta 2014–2012"⁷ listaa e-tutkimusinfrastruktuurikategoriassa kolme kansallista tutkimusinfrastruktuuria: CSC RI (CSC tutkimusinfrastruktuurina), PRACE Suomi (Suomen osuus PRACE:sta) ja FGCI, joka on merkitty potentiaaliseksi hankkeeksi. ESFRI-tutkimusinfrastruktuurikumppanuuksissa e-tutkimusinfrastruktuureista on tiekartalle päässyt mukaan PRACE. Jäsenyyksiä kansainvälisissä e-tutkimusinfrastruktuureissa on yksi, NeIC.

Tiekartan johdannon mukaan "Tutkimusinfrastruktuureilla tarkoitetaan tutkimusvälineitä, laitteistoja, aineistoja ja palveluita, jotka mahdollistavat innovaatiotoiminnan eri vaiheissa tapahtuvan tutkimus- ja kehitystyön ja tukevat järjestäytynyttä tutkimustyötä, tutkijankoulutusta ja opetusta sekä kehittävät tutkimus- ja innovaatiokapasiteettia. Ajanmukainen tutkimusinfrastruktuurien ekosysteemi on edellytys kärkeen pyrkivälle tutkimukselle ja laadukkaalle tutkijankoulutukselle sekä kansainväliselle yhteistyölle ja maan tutkijakunnan kansainvälistymiselle. Innovaatiotoiminta ja teollisuuden uudistuminen sekä valtion tutkimuslaitosten palvelutuotanto edellyttävät laadukasta infrastruktuuria samoin kuin tutkimuslaitosten ja korkeakoulujen tutkimusyhteistyö." Tutkimusinfrastruktuuritiekartta tulee voimakkaasti ohjaamaan tutkimusinfrastruktuurien resursointia tulevina vuosina.

On tärkeää, että tieteellisen laskennan kokonaissuunnittelu tehdään tiiviissä yhteistyössä tutkimusinfrastruktuuritiekartan implementoinnin kanssa, ja että tiekartan päivityksessä e-tutkimusinfrastruktuureihin kiinnitetään erityistä huomiota.

⁷ http://www.aka.fi/Tiedostot/Tiekartta/Tutkimusinfrastruktuurien_strategia_ja_tiekartta_2014_20.pdf

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 tiekartan 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.

3.2 Ohjeellinen tieteellisen laskennan strategia

Tieteellisen laskennan hyödyntäminen lisääntyy kaikilla tieteenaloilla ja tukee suomalaisen yhteiskunnan kehittämistä. Tieteellisen laskennan yhteiskunnallinen merkittävyys tunnustetaan 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 yhteiskuntatieteistä.

Tehokkuus ja tarkoituksenmukaisuus

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

Kustannustehokkuus

Tarjoamalla raskain Tier-1 -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 jatkuvan 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.

3.3 Tavoitetila

Kuten luvussa 1.3.1 todettiin, laskennallinen tiede on Suomessa korkeatasoista, mihin on osaltaan vaikuttanut hyvä tieteellisen laskennan kokonaisuus. Yleisenä tavoitteena on tämän hyvän tilan säilyttäminen ja kehittäminen siten, että

- **Tieteellisen laskennan kokonaisuus tukee kilpailukykyistä ja kansainvälisesti houkuttelevaa tutkimusjärjestelmää ja mahdollistaa uusia tieteellisiä avauksia.**
- **Tieteellisen laskennan kokonaisuus organisoidaan mahdollisimman kustannustehokkaasti yhdistelmänä keskitettyä ja hajautettua ratkaisua.**

Seuraavassa kuvataan yksityiskohtaisemmin näiden tavoitteiden edellytyksiä ja osatekijöitä.

3.3.1 Tieteellisen laskennan kokonaisuus tukee kilpailukykyistä ja kansainvälisesti houkuttelevaa tutkimusjärjestelmää ja mahdollistaa uusia tieteellisiä avauksia.

Menestyäkseen Suomen laskennallisen tieteen on pystyttävä houkuttelemaan kansainvälistä yhteistyötä, ja saatavilla olevan keskitetyn ja hajautetun kapasiteetin on toimittava yliopistojen kilpailuetuna kansainvälisiä huippututkijoita Suomeen houkuteltaessa. Tieteellisen laskennan kokonaisuuden on vastattava tutkijoiden tarpeita ja resurssien oltava ajanmukaisia sekä riittävän tehokkaita, jotta ne mahdollistavat laskennallisen tieteen kehittymisen. Nousevat dataintensiivisen tieteen tarpeet otetaan huomioon hyödyntäen synergioita olemassa oleviin tieteellisen laskennan palveluihin ja infrastruktuureihin.

Tämä edellyttää seuraavaa:

Toiminta on kansainvälistä ja verkostoitunutta

- Tiedonvaihto ja yhteistyö kansainvälisten toimijoiden kanssa.
- Kansainvälisten ja kansallisten data-aineistojen saatavuus laskentaympäristössä. Mahdollisimman saumaton liittymä kansainväliseen Tier-0 - ja grid-kapasiteettiin.

Laitteistot ja palvelut ovat tarkoituksenmukaisia

- Tehokkaalla kytkentäverkolla varustettu superkone raskaaseen rinnakkaislaskentaan.
- Klusterit peräkkäis- ja pienille rinnakkaistöille sekä dataintensiiviseen laskentaan.
- Tehokkaat levyjärjestelmät sekä perinteisen laskennan että dataintensiivisen tieteen tarpeisiin.
- Suurimuistisia solmuja.
- Virtualisoinnin laajentaminen tiedealoille räätälöityjen sovellusympäristöjen tukemiseksi.
- Erikoisproessoreita, niitä hyödyntämään pystyviä sovelluksia ja äärimmäistä tehoa vaativaa laskentaa varten.
- Tietoverkkojen laatu ja ajanmukaisuus.
- Sensitiivisten tutkimusaineiston laskentaan soveltuvat, korotetun tietoturvatason järjestelmät.
- Interaktiivinen käyttö.
- 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).

Laskentakapasiteettia on riittävästi, jotta suuret rinnakkaisajot ovat mahdollisia, jonotusajat pysyvät kohtuullisina ja kokonaiskapasiteetti (throughput, työt/aikayksikkö) vastaa tarpeita

- Keskitetysti operoitavan kansallisen superkoneen tulisi parhaimmillaan olla sijoilla 30–40 Top500-listalla. CSC:n tarjoamat superkoneet ovat perinteisesti toteuttaneet tämän vaatimuksen, ja taso tulisi säilyttää.
- Keskitetysti operoitavan klusterin tulisi kooltaan (laskentaytimien lukumäärällä mitattuna) olla vähintään puolet superkoneesta.
- Laitteistot uusitaan riittävän usein eli käytännössä noin neljän vuoden välein.
- Mahdollisuus sujuvaan siirtymiseen Tier-0-tasolle.

Riittävä, tehokas ja tarkoituksenmukainen tallennuskapasiteetti

- Nopeaa koneiden ”lähellä” olevaa tallennuskapasiteettia (computable storage): mahdollisuus suurien data-aineistojen (yli 100 TB) tehokkaalle käsittelylle osana laskentaa.
- Helppokäyttöinen arkistointi, pitkäaikaissäilytys ja jakaminen.
- Tallennetun tiedon nopea palautusaika tutkijan käyttöön etenkin pysyvän (hitaan) tallennuksen osalta.

Kapasiteetin saavutettavuus ja luotettavuus on mahdollisimman korkea (tarkoituksenmukaisuuden rajoissa)

- Ylläpito resursoidaan riittävästi kaikkien toimijoiden osalta.
- Ylläpidossa hyödynnetään verkostoja ja yhteistyötä muiden toimijoiden kanssa.
- Tarkoituksenmukainen tietoturva.

Korkeatasoisilla laskentaympäristöön, menetelmiin, ohjelmointiin, ohjelmistoihin ja laskennalliseen tieteeseen liittyvillä tukipalveluilla ja koulutuksella tuetaan suomalaisen tieteellisen laskennan ohjelmoinnin ja laskennallisen tieteen sekä yhteiskunnan kansainvälistä kilpailukykyä ja varmistetaan laitteistojen tehokas hyödyntäminen (ks. 2.1.2 ja 2.1.4)

- Tuki ja koulutus resursoidaan riittävästi ja niitä kehitetään jatkuvasti.
- Hyödynnetään verkostoja ja yhteistyötä muiden toimijoiden kanssa.
- Lisätään mallinnuksen ja laskennallisten menetelmien opetusta kaikilla tasoilla (ks. 2.1.6).

3.3.2 Tieteellisen laskennan kokonaisuus organisoidaan mahdollisimman kustannustehokkaasti yhdistelmänä keskitettyä ja hajautettua ratkaisua

Laskennallisen tieteen kilpailukyky ja sen tarvitsemat palvelut pyritään turvaamaan niukassa resurssitilanteessa kokonaisuuden huolellisella suunnittelulla. Hyväksi osoittautunutta luvussa 1.3.5 kuvattua keskitetyn ja hajautetun ratkaisun yhdistelmää sovelletaan harkitusti ja kokonaisuus huomioon ottaen. Palvelujen järjestämisestä aiheutuneet kustannukset lasketaan yhteisesti sovittujen periaatteiden mukaisesti, hyödyntäen linkaarimallia mikä edistää kustannustietoisuutta ja kokonaistaloudellisuutta.

Päällekkäisyyden välttämiseksi ja kustannustehokkuuden edistämiseksi keskitettyjen palveluiden pitäisi olla kaikkien julkisen sektorin tutkijoiden käytettävissä samoilla ehdoilla.

Tämä edellyttää seuraavaa:

Tarkoituksenmukainen hierarkia ja yhteistyö paikallisten ja keskitettyjen resurssien välillä

- Vältetään tarpeettomia päällekkäisyyksiä.
- Keskitetyt ja paikalliset resurssit täydentävät toisiaan.
- Suurteholaskennan palvelut toteutetaan keskitetysti.
- Rajoitetun käyttäjäkunnan (erikois)palvelut, kuten esimerkiksi sensitiivisen datan palvelut ja erityisen IO-intensiivinen laskenta, toteutetaan koordinoitusti.
- Tiedon jakaminen ja yhteistyö uusia teknologioita kokeiltaessa.

Infrastrukturi on toteutettu kustannustehokkaasti

- Kokonaiskustannukset lasketaan yhteisesti sovittujen periaatteiden mukaisesti.
- Pyritään kokonaistaloudellisesti edullisimpaan ratkaisuun.
- Yhteistyö hankinnoissa (ks. 2.1.3).

- Turvattu rahoitus ja rahoitusmalli (jaksotus jne.), joka mahdollistaa pitkän tähtäimen suunnittelun ja tukee kustannustehokkuutta. Luopumista CSC:n superkoneen ja superklusterin hankkimisesta samanaikaisesti tulisi harkita (ks. 2.1.3).

Keskittetyt resurssit ovat aidosti kansalliset, ei keinotekoisia raja-aitoja (ks. 2.1.1)

- CSC:n palvelut ovat samoilla ehdoilla koko julkisen sektorin tiedeyhteisön käytettävissä.

Yhteinen ohjaus

- Tieteellisen laskennan kokonaisuus noudattaa kokonaisarkkitehtuuria.
- Kehitetään ohjauksen yhteistä laajapohjaista koordinoitua.
- Kaikki käyttäjät ja rahoittajat selkeän sitoutuneita yhteiseen toimintaan.

4 Taustalla vaikuttavat arkkitehtuurit ja periaatteet

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>
- Verkoarkkitehtuurit
- Funet-kokonaisarkkitehtuuri
- Avoimen tieteen ja tutkimuksen (ATT) kokonaisarkkitehtuuri
 - Avoimen tieteen ja tutkimuksen asiantuntijaryhmä ohjaa laatimista
 - Tekeillä
- Tutkimuksen tuen ja hallinnon viite- ja muut arkkitehtuurikuvaukset
 - Tutkimuksen tuen ja hallinnon (TUHA) 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/Opintohallinnonkansallinenviitearkkitehtuurixv1_0.pdf
- Opetus, tiede ja kulttuuri -kohdealueen kokonaisarkkitehtuurit
 - https://www.yhteentoimivuuus.fi/aihealue/Kohdealueet/Koulutus,_tiede_ja_kulttuuri

- Kansallisen Digitaalisen Kirjaston kokonaisarkkitehtuuri
 - <https://www.yhteentoimivuus.fi/semic/view/Asset/Asset.SingleView.xhtml?id=60214>
- Korkeakoulujen muut kokonaisarkkitehtuurit ja omat kohdearkkitehtuurit
- Valtionhallinnon yhteinen kokonaisarkkitehtuuri
 - https://www.yhteentoimivuus.fi/aihealue/Valtionhallinnon_yhteinen_kokonaisarkkitehtuuri



Kuva 14. 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

1. Palveluiden tuotanto ja kehittäminen on käyttäjälähtöistä
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 resursoituidut

4.4 Tietoturva- ja tietosuojaperiaatteet

CSC:n laskentapalveluissa ja niiden ylläpitokäytännöissä noudatetaan CSC:n sisäistä tietoturvapolitiikkaa 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.

Tietoturva on CSC:n palveluiden merkittävä etu ja hyöty, ja tästä näkökulmasta on tulevaisuudessaakin tärkeää pitää kiinni. On kansallisesti merkittävää, että on olemassa palveluita, joiden hallinta pysyy täydellisesti kotimaassa.

5 Liitteet

5.1 Lait ja säädökset

- Laki julkisista hankinnoista (348/2007), EU:n direktiivi julkisista hankinnoista 2014 (2014/24/EU)
- Laki julkisen hallinnon tietohallinnon ohjauksesta (634/2011)
- Sähköisen viestinnän tietosuojalaki (516/2004), Henkilötietolaki (523/1999)
- Laki biopankeista (688/2012)
- Laki (409/2014) ja asetus (701/2014) biologisen tiedon ja bioinformatiikan (ELIXIR) tutkimusinfrastruktuurista
- Tekijänoikeuslaki (404/1961), lisenssisopimukset
- Yliopistolaki (558/2009), Ammattikorkeakoululaki (14.11.2014/932), tutkimuslaitoksia koskeva lainsäädäntö
- EU:n, Suomen ja muiden maiden vientirajoitukset
- Tietoyhteiskuntakaari (2014/917)

5.2 CSC:n tämänhetkisten 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 Top500 -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.

5.3 Käsitteitä/Sanasto

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.

Kiihdyttimet ja apuprosessorit (GPGPU/GPU ja MIC)

Kiihdyttimillä tarkoitetaan yleiskäyttöisiä grafiikkaprosessoreita (GPGPU tai yksinkertaisesti GPU), joiden merkittävin 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 erikoisprosesseilla, 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.

Pilvilaskenta

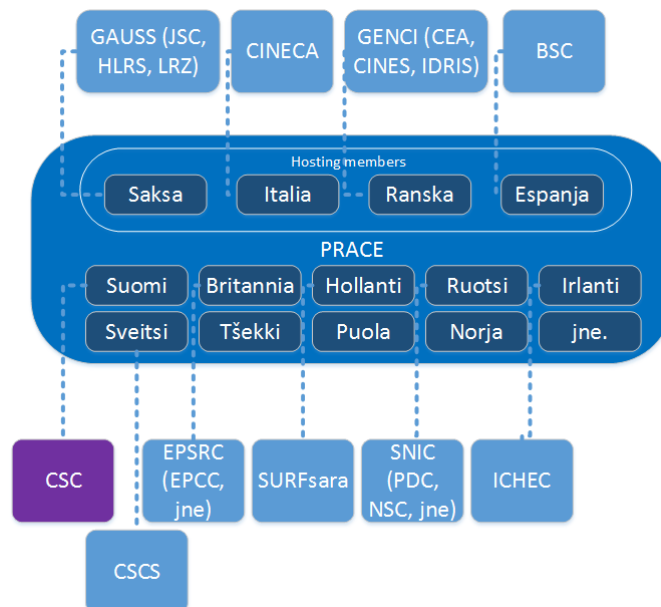
Tieteelliseen laskentaan tarkoitettu pilvipalvelu käsittää datan prosessoinnin, tallennuksen, verkkoratkaisun ja muut olennaiset laskennallisen tieteen tarvitsemat resurssit. Palvelun käyttäjä ei operoi tai hallinnoi pilven fyysistä infrastruktuuria, mutta voivat valita haluamansa käyttöjärjestelmät, 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.
- Sovellusalusta (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.

PRACE (www.prace-ri.eu)

PRACE on vuonna 2010 perustettu eurooppalainen suurteholaskennan infrastrukturi, jonka tarkoitus on taata laskennallisen tieteen tutkijoille kansainvälisesti huippuluokan resurssit. PRACEssa on 25 jäsenmaata, ja Suomi on edustettuna CSC:n kautta (kuva 13). PRACE on hajautettu infrastrukturi, 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.



Kuva 13. PRACEn jäsenet.

PRACEn laskentaresurssit ovat toistaiseksi olleet jäsenmaille ilmaisia, mutta tilanne tulee muuttamaan 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.

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 rinnakkaistamalla 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 rinnakkaistus, joka toteutetaan kääntäjän direktiiveillä tai pragmoilla. Myös kiihdyttimien ja apuprosessorien teho perustuu rinnakkaisuuteen. Tehokkaan rinnakkaisohjelman kirjoittaminen vaatii yleensä hyvää asiantuntemusta ja paljon työtä.

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.

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. Tier-3 tarkoittaa paikallisia muutaman koneen verkkoja ja Tier-4 yksittäistä pöytäkonetta.