

Grid on tulevaisuuden tehoverkko

Internetin www-samoilua seuraava verkkoaskel on grid. Se tuo internetin laskenta-resurssit yhteiseen käyttöön.

TEKSTI: ERKKI NIIRANEN
KUVA: ERIC LERAILLEZ

Grid on internetin tietokoneet yhdistävä järjestelmä, jossa internetiä käytetään jakamaan maantieteellisesti eri puolilla sijaitsevien koneiden laskenta- ja tallennuskapasiteettia. Gridiä (suomennokseksi on ehdotettu hilaa tai ritilää) testataan jo suurta tiedonkäsittelykykyä vaativissa tehtävissä.

Gridillä tarkoitetaan nykyisin kahta vähän eri asiaa. On "globaali-grid", joka on syntynyt tiedeorganisaatioissa, kuten Euroopan hiukkastutkimuksen keskuksessa Cernissä; Cern muistetaan myös webin syntytöinä.

Globaali-gridin lisäksi on palomuurin sisäpuolelle jäävä "yritys-grid". Se on muun muassa Oraclen ja Sunin myyntijohtajien mieluinen esitelmäaihe. Kyseessä on yrityksen sisäisen tietokonekapasiteetin hajautettu hyödyntäminen.

Vaikka yritys-gridin toivotulla toteutumisella on nopeampia taloudellisia seurauksia kuin globaali-gridillä, sitä on vaikeampi erottaa yritysten muusta tietoteknisestä verkottumisesta. Siksi tässä käsitellään lähinnä pitemmälle tulevaisuuteen vaikuttavaa globaali-gridiä.

YKSINKERTAINEN ON VOIMAKASTA

Gridin ajatus on yksinkertainen. Tehdään monista pienistä tietokoneista yksi iso tietokone. Toteutuksen tekniikka vain on monimutkainen.

Ajatuksena on, että käyttäjä liittyy omin tunnuksin grid-verkkoon ja antaa sovel-
lusohjelmalle raskaan laskentatyön tai



vaativan tiedonhakutehtävän. Gridiin liittyneet koneet tekevät annetun tehtävän vapaa-aikoinaan, yhteisten hajautusohjelmistojen ohjaamina. Kun yhden koneen kapasiteetti ei riitä tai se on uppoutunut haltijansa ensisijaiseen työhön, tehtävää hajautetaan muille.

Nopean tietoliikenneverkon kautta saavutettu yhteinen käsittelykyky nousee superlaskennan tasolle. Se ylittää yksittäisen supertietokoneen tai jopa niiden ryppäiden tehot.

Englanniksi sana grid tarkoittaa myös sähköverkkoa. Kun napsauttaa lamppuun valon, ei napsauttajan tarvitse tietää, missä voimassa tai edes missä maassa sähkö on tuotettu. Yhtä vähän gridiä käyttävä tutkija tai tulevaisuudessa perus-kotikäyttäjä tietää tai on edes kiinnostunut tietämään, missä maassa hänen antamansa tehtävä suoritetaan.

Suorittava grid-ohjelmisto voi nojata linuxiin, windowsiin tai Maciin, alustalla ei ole merkitystä. Nykyisin gridissä on supertietokoneita, tulevaisuudessa myös yrityksen tai kodin mikrotietokone.

Kun organisaatioiden ensisijainen teh-

tävä on palvella omia asiakkaitaan, osallistuminen gridiin ei saa aiheuttaa merkittäviä muutoksia laitteisiin, ohjelmistoihin tai tallennusjärjestelmiin.

KAIKKI OVAT MUKANA

Yritys-grid hankitaan tuottamaan kustannussäästöjä ja helpottamaan ylläpitoa. Sen sijaan, että yritysverkkoon hankitaan yhä jyrkempiä palvelimia tai niiden ryppäitä, sovellusohjelmien tehtävät hajautetaan pienemmille, yrityksessä jo oleville koneille.

Gridiä ohjaamaan on tehty kaupallisen lisenssin alaisia ohjelmistoja, mutta pääasiassa se on avoin moniorganisaatiohankke. Sitä kehittävät akateemiset ja valtiolliset tietotekniikan tutkimuslaitokset, kuten Suomessa Tieteen tietotekniikan keskus CSC, sekä yksityiset kaupalliset osallistujat, kuten Cisco, HP, IBM, Oracle ja Sun.

Molemmilla kehittäjäpuolilla on oma lehmä ojassa. Akateeminen maailma tarvitsee tutkimukseen suurta laskentakapasiteettia ja mahdollisuutta etäkäyttää kal-

liita tutkimuslaitteita, ja yritykset tarvitsevat leviämisalueita tietoliikenne- ja hallintajärjestelmilleen.

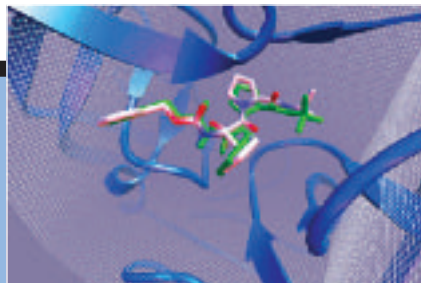
SETI ON GRIDIN SUKULAINEN

Mikä sitten erottaa grid-laskennan muusta tietoverkkoja käyttävästä laskennasta? Tarkasti ottaen ei mikään. Grid lainaa piirteitä tutummista käsittelytavoista, ja muutoin ei voisi ollakaan. "Ex nihilo nihil", mikään ei synny tyhjästä ilman jo olevaa tekniistä tietämystä.

Tietokoneiden yhteiskäyttöä on harjoitettu kauan. Tämänkin lukemiseen kuluvaan minuutin aikana sadattuhannet ja vuorokauden aikana miljoonat mikrotietokoneet kodeissa ja yrityksissä heräävät loppoaikoinaan osallistumaan seti@home-projektiin. Siinä lasketaan tietosiivuja radioteleskoopin kuuntelemasta materiaalista, ja tavoitteena on löytää avaruuden älyllistä elämää.

Vastaavia projekteja on muitakin: alkulukujen laskentaa, salausalgoritmien murtamista, dna:n ja proteiinien tutkimusta,

Tieteen tehosyöpöt



Laskentakapasiteetin tarjoaminen on tärkein gridin palveluista. Monimutkaisten proteiinien rakenteiden laskeminen on esimerkki vaativasta laskentatehtävästä. Tämän päivän tehokkaalta mikrotietokoneelta laskenta saattaisi viedä seitsemänkymmentä vuotta.

Kymmenien supertietokoneyksiköiden ryppäältä laskenta kestäisi vuoden, jos yhteen projektiin olisi mahdollista antaa niin paljon laskuaita. Jos kymmenen superkoneen välillä olisi grid-verkko, laskenta kestäisi päivän.

Kun joskus tulevaisuudessa rakennetaan grid-verkko, joka sisältää kaikkien koneiden tehot miljoonista kotimikroista kymmeneen supertietokoneisiin, proteiinirakenteiden selvittämiseen riittää tunti.

Tulevaisuudessa tämä miljardien eurojen arvoinen laskentakapasiteetti on periaatteessa jokaisen tutkijan käytettävissä.

Suuria laskentatarpeita löytyy monelta aluelta:

- Aivotoiminnan tutkimuksesta: yksi korkearesoluutioinen aivoskannaus tuottaa kolme tera-

tavua dataa, teratavu = 1000 Gt.

- Lääketieteestä: Yhdysvaltain kansallisen mamografiatietokannan arvioidaan vaativan 10 petatavua.

- Molekyylibiologiasta: eräs hajautetusti suoritettu koe tutki 35 miljoonaa molekyyliä, jotka voivat olla 750 muodossa eli yhdistelmiä laskettiin 26 miljardia; näistä jatkotutkimuksiin valittiin 44 yhdistelmää.

- Tähtitaivaan tutkimuksesta: vuosittaiseksi datamääräksi nousee 10 petatavua.

Vastaavia tietomääriä ennustetaan kerättävän biofysiikassa, genetiikassa, ilmaston muutosten ennustamisessa, kemiassa, suurenergiafysiikassa, maailmanlaajuisessa sään ennustamisessa, tieteellisten tutkimuslaitteiden suunnittelussa, talousanalyseissa, vaativassa konesuunnittelussa ja alkeishiukkassimuloinneissa.

Toteutunut laskentamahdollisuus poikii uusia laskentatarpeita myös yritys-gridiin.

SUOMI ON MUKANA

Euroopan yliopistojen yhteisverkkoon ja grid-projekteihin on EU:n puiteohjelmissa varattu useiden vuosien ajaksi 300 miljoonaa euroa. Grid-hankkeista suurin on Cernin hiukkastutkimuskeskuksen Datagrid-projekti, jossa suuren LHC-hiukkaskiihdyttimen (Large Hadron Collider) tuottamaa dataa tutkitaan EU-maiden laskentakeskuksissa. Yksi suurimmista tutkimuskohteista on maailmankaikkeuden alkuhetki, Big Bang.

LHC-hiukkaskiihdyttimessä protonit törmäävät 40 miljoonaa kertaa sekunnissa ja syntyvän tiedon (arvion mukaan 13 petatavua vuodessa, petatavu = 1 000 000 Gt) käsittelemiseen tarvittaisiin 100 000 nykytietokoneen kapasiteettiä.

Suomikin osallistuu Datagridiin pohjoismaiden joukossa Tieteen tietotekniikan keskuksen CSC:n ja Fysiikan tutkimuslaitoksen (HIP) välityksellä. Grid-järjestelmä valmistuu vuonna 2006 tai 2007.

syövän, isorokon ja aidsin tutkimuksen laskentaa. Näiden projektien, joille on ehdotettu nimeä "internet-laskenta", ja grid-verkon välillä on silti eroja, joita grid-tutkijat mielellään korostavat.

Seti-laskenta ei ole tosiaikaista. Seti-paketit käydään hakemassa Berkeleyn yliopiston keskuksesta, käsitellään mikroissa erikseen ja aikanaan valmiit paketit lähetään takaisin Berkeleyyn. On-linea Seti ei myöskään ole, sillä käsittelyvaiheessa ei olla yhteydessä Seti-keskukseen.

ENTÄ METAKONEET

Grid-verkko on läheistä sukua myös metakoneille. Näillä tarkoitetaan kokonaisuutta, joka muodostuu fyysisesti eri paikoissa sijaitsevista yhteen kytketyistä (super)tietokoneista, jotka käyttäjä näkee yhtenä kokonaisuutena. Mutta grid ei ole metakone. Metakoneen osat on yhdistetty joko lähiverkossa tai maantieteellisesti läheisesti hajautettuna, gridissä resursseja hajautetaan laajemmin joko kansallisesti tai kansainvälisesti, yhteisen grid-ohjelmiston avulla.

Suomessa CSC on rakentanut omaan verkkoonsa metakoneen, mutta se on myös tehnyt molekyylibiologian geenien laskennan grid-koeajoa Ruotsin vastaavan PDC- ja Norjan Parallah-keskuksien kanssa.

Grid sisältää piirteitä myös vertaiskäsitelystä (peer to peer) entisen Napsterin tapaan, kun siinä käytetään tasavertaisina käsittely- ja tallennusyksikköinä ympäri maailmaa sijaitsevia tietokoneita.

MUUTAKIN KUIN LASKENTAA

Laskennan lisäksi tärkeä palvelualue on tietokantapalvelut. Tietokannat ja tiedonhaku niistä hajautetaan maantieteellisesti. Kun käyttäjä tallentaa tietomassaa, hänen kannaltaan tieto tallentuu yhteen paikkaan ja haetaan yhdestä paikasta, vaikka tieto fyysisesti tallentuisi Filippiineille ja Meksikoon.

Kolmas palvelutapa on etäohjelmien käyttö. Kun nykyäänkin lähiverkossa tai osittain internetissä ohjelman voi käynnistää välittämättä siitä, missä ohjelmopalvelin fyysisesti sijaitsee, grid tarjoaa lisäksi maantieteellisen hajautuksen.

Teknisten ongelmien ratkettua ei ole estettä kotikoneiden osallistumiselle grid-laskentaan. Vastavuoroisesti kotikoneet saisivat käyttöönsä gridin tehoja, jolloin kotonakin olisi käytössä supertietokone –

Gridejä kaupasta

Useat tietotekniikkayhtiöt kehittävät kaupallisia grid-ratkaisuja. Tavoitteena on tasata suuren yrityksen palvelimien kuormaa sekä parantaa järjestelmien hallittavuutta ja vikasietoisuutta.

IBM kehittää ratkaisuja yhteistoimin kumppanien kanssa erityisesti pankki- ja rahoitusallalle. Palvelinlaitteistojen välisellä resurssienjaolla pyritään lyhentämään talusanalyysien käsittelyaikoja.

Oraclen 10g-tuoteperhe käsittelee grid-perustaisen tietokannan, sovelluspalvelimen ja hallintavälineet.

Sunin grid-strategia kohdistuu resurssien tehokkaaseen jakoon, tiedon hallintaan, pienten palvelimien ryvästämiseen ja graafiseen työhön.

Adobe aikoo sisällyttää After Effects Professional -ohjelmaansa grid-lisäkkeen, joka hajauttaa tehosteiden laskennan usealle koneelle verkossa.

HP ja **Unesco** toimittavat grid-ratkaisun eräille Balkanin maiden yliopistoille. Se parantaa tutkijoiden mahdollisuutta osallistua kansainvälisiin tutkimushankkeisiin. ■

MIRJA PAATERO

ehkä ei suoraan, mutta yksityisten ja julkisten palvelujen välityksellä, kenties ensin elokuvien ja multimedian käsittelyn yhteydessä.

VIELÄ ON ONGELMIA

Muotitermi tuo mukanaan muotiodotuksia, vaikka ratkaisematta on teknisiä ongelmia. Kun käsittely on jopa globaalia, tietoturvapohjan tulee myös olla jaettu. Tarvitaan yhteinen autentikointi, salaus ja muu turvallisuusstandardointi.

Vaikeimmat ongelmat liittyvät ohjelmistojen kehittämiseen. Ohjelmistojen pitää olla yhteisiä ja toistensa kanssa keskustelomaan kykeneviä. Niiden pitää olla tehokkaita hajautettujen resurssien varaajia. Käyttöliittymienkin pitäisi olla samanlaisia.

Grid jakautuu kolmeen ohjelmistokerrokseen. Alinna ovat käyttöjärjestelmät, tietoliikenneohjelmat ja muu perusarkkitehtuuri. Ylin kerros ovat käyttäjälle näkyvät ohjelmistot ja niiden käyttöliittymät.

Välikerros on gridin ydin. Sen pitää yhdistää muut kerrokset niin, että alin kerros kätkeytyy käyttäjältä ja ylimmältä kerrokselta, vaikka se olisi maantieteellisesti laajasti hajautettu ja tekniikaltaan heterogeeninen.

TULEVAISUUS MULLISTUU

Internet lewahti lentoon kun saatiin helpokäyttöiset selaimet. Webin arkipäiväistyminen kesti silti vuosia. Gridin arvioidaan leviävän nopeammin, koska se vaatii vähemmän rakentamista valmiin internetin johdosta. Silti kaikesta gridin potentiaalista on toistaiseksi toteutettu tuskin muutama prosentti.

Tietotekniikan keksintöjen suhteen on monesti oltu holtittoman optimistisia (ja myös päinvastoin: IBM:n suuruuteen kasvattanut Thomas J. Watson arvioi 1940-luvulla, että koko maailmaan tarvitaan vain viisi tietokonetta). Eräs IBM:n myyntijohtaja visioi syksyllä 2003, että grid-laskenta tarkoittaa samanlaista hyppäystä kehityksessä kuin oli nisäkkäiden kehittyminen ohi dinosaurusten tai hermosoluverkoston kehitysluokan esikambriin kaudella. Niinpä niin.

Gridin vaikutus on rinnastettu sähkön yleistymiseen edellisen vuosisadan vaihteissa. Vahvimmin gridiin uskovat ennustavat, että viimeinkin syntyy tietämyksen yhteiskunta pelkän tietoyhteiskunnan sijaan.

Olkoit ennustajat vaikka oikeassa, silti gridin osaa laskea vain sen, mikä on laskehtavissa. Kaaoksia ei sekään osaa laskea: esimerkiksi säätä yli viikon päästä tai ihmisten käyttäytymistä. ■

▶ OSOITTEITA

<http://gridcafe.web.cern.ch> "The place for everybody to learn about the grid."

www.gridcomputingplanet.com
Grid-tietoa.

www.grid.org Grid-projekti, johon kuka vain voi liittyä.

<http://setiathome.ssl.berkeley.edu>
Seti@home-projekti.

www.csc.fi Tieteen tietotekniikan keskus.