



# **Grid: Käsitteet, teknologiat, sovellukset sekä vaikutus CSC:läisten työhön**

Arto Teräs

[arto.teras@csc.fi](mailto:arto.teras@csc.fi)

CSC:n Grid-seminaari 13.10.2004



## Sisällys

- Mitä grid tarkoittaa?
- Mitä grid muuttaa?
- Yleisiä harhaluuloja
- Sovellusten siirtäminen grid-ympäristöön
- Grid-väliohjelmiston rooli
- Grid-tyypit: laskenta- ja datagrid
- Tärkeimmät väliohjelmistot
- Tietoturva
- Varmenteet ja käyttäjien hallinta
- Grid eri CSC:n pääyksiköissä



## Mitä grid tarkoittaa?

- Muotisana: monta käyttäjää, monta merkitystä
- Grid-tekniikoilla pyritään tarjoamaan helppo ja tietoturvallinen pääsy hajallaan sijaitseviin tieto- ja laskentaresursseihin
- Grid ylittää organisaatorajat!
- Grid-yhteistyö on paljon muutakin kuin pelkkää tekniikkaa



## Mitä muuttuu kun grid tulee?

- Koneiden ja datan fyysisen sijainnin merkitys pienenee
- Käyttäjätunnuksen tilalle tulee varmenne, salasanojen määrä vähenee
- Eri laitearkkitehtuureihin päästään käsiksi saman liittymän kautta, mutta kirjastoriippuvuuksien ja muiden ongelmien kanssa painitaan entiseen tapaan
  - Virtuaalikoneiden käyttö voi tuoda helpotusta tulevaisuudessa
- Halutun koneen valintamahdollisuus säilyy, mutta joustavasti eri ympäristöissä toimiville ohjelmille tarjolla lisäresursseja
  - Virheiden raportoinnissa ja niistä toipumisessa puutteita, kehittynee väliohjelmien kypsymisen myötä



## Yleisiä harhaluuloja

- Grid moninkertaistaa resurssit
  - Yleinen vertaus Internetiin ja www-palvelimiin ontuu:
    - Yksi www-palvelin ja 1000 käyttäjää => hinta 1/1000 x
    - Yksi grid-käyttäjä haluaa 1000 konetta => hinta 1000 x...
  - Kuorman tasaus ja optimointi mahdollista, mutta olennaisempaa on uudet palvelut ja helppo pääsy niihin
- Grid on musta laatikko, jonka sisällä kaikki on automaattista
  - Vision mukaan laskentateho tulee kuin sähkö töpselistä, mutta käytännössä tästä ollaan vielä kaukana
    - Mahdollisuus seurata työn etenemistä tärkeää — jos yritetään tehdä musta laatikko niin syntyy helposti vain musta aukko

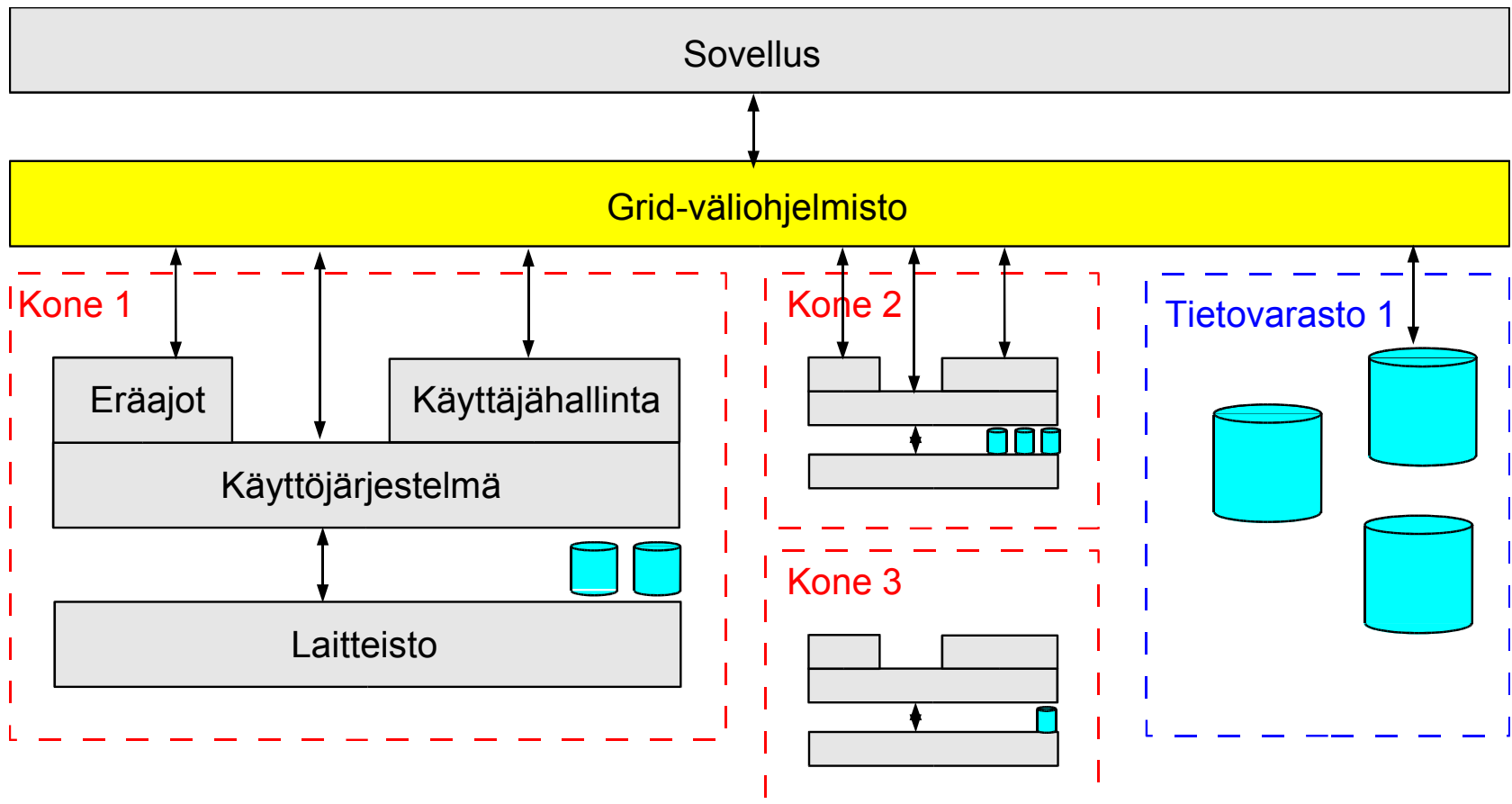


## Tarvitseeko sovelluksia muuttaa?

- Kolme erilaista lähestymistapaa:
  - 1) Käytetään sovellusta sellaisenaan, grid-väliohjelmisto vain siirtää sen ja datat haluttuun paikkaan suoritettavaksi
    - Kääntäminen staattisesti linkitettyinä helpottaa kirjasto-ongelmissa
  - 2) Asennetaan sovellus kohdekoneelle ja käytetään sitä grid-liittymän kautta
    - Eräajotyypiseen sovellukseen ei yleensä tarvita mitään muutoksia, mutta graafinen liittymä voi vaatia paljonkin työtä
  - 3) Tehdään aidosti hajautettua ympäristöä hyödyntävä sovellus tai muokataan vanha sellaiseksi
    - Laaja maantieteellinen hajautus ei ole järkevää ellei sovellus osaa itse pilkkoa laskentatehtävää toisistaan riippumattomiin palasiin



## Väliohjelmiston rooli





## Laskenta- ja datagrid

- Laskentagrid kaikkein tunnetuin grid-tyyppi: laskentatehtävät suoritetaan hajautetussa ympäristössä
  - ”Hajautettu eräajojärjestelmä”
  - Syötetiedostot siirretään kohteeseen yhdessä sovelluksen kanssa, tulostiedostot haetaan työn loputtua
  - Latenssiongelma: valon nopeutta ei ole onnistuttu ylittämään...  
=> Yhtä laskentatehtävää ei yleensä hajauteta maantieteellisesti
- Datagridissä pyritään tuomaan tieto helposti ja tehokkaasti saataville
  - Yhtenevä liittymä eri puolilla sijaitsevaan dataan
  - Suurten datamäärien hajautus ja replikointi





## Grid-väliohjelmistoja

### Globus ja johdannaiset



### Muut





## Globus Toolkit

- Tunnetuin Grid-väliohjelmisto, toteutettu pääosin C-kielellä
- Tukevasti hypen eturintamassa: uusien tuulien myötä aina epäyhteensopivia muutoksia ja epävakaita versioita
- Nimensä mukaisesti työkalupakki => ei käyttökelpoinen sellaisenaan
  - Mm. töiden aikataulutus (scheduler) puuttuu kokonaan
  - Ei aina pidetty, mutta silti paljon käytetty, vaihtoehtoja niukasti
- Open source: <http://www.globus.org>
- Synnyttänyt suuren joukon projekteja, jotka käyttävät Globusin tietoturvainfrastruktuuria ja valikoiden muita komponentteja sekä lisäävät omia laajennuksia



## UNICORE

- Pääosin Saksassa kehitetty ja erityisesti siellä suosiota saanut väliohjelmisto
- Toteutettu Javalla, voi toki ajaa myös muita ohjelmia
- Huomattavasti tiukemmin integroitu paketti kuin Globus
  - Laskentatehtävälle tehdään ajomääritys graafisen käyttöliittymän kautta, sen jälkeen ajo myös komentoriviltä
  - Käyttäjä valitsee kohdekoneen, ei automaattista valintaa
  - Paljon käytetyille sovelluksille syytä tehdä UNICORE-plugin, jolloin käyttö kätevämpää
- Ei vielä kokemusta CSC:llä
- Open source: <http://unicore.sourceforge.net>



## Avaki

- Datapohjainen lähestymistapa, ”laajennettu NFS”
  - Täysin läpinäkyvä käyttäjän kannalta
- Suljettu kaupallinen tuote, perustuu Virginian yliopistossa kehitettyyn Legion-järjestelmään (<http://legion.virginia.edu/>)
  - Ohjelman kaikki oikeudet nykyisin Platform Computingilla, joka myy Avaki-tuotetta
- Ei kokemusta CSC:llä
- <http://www.avaki.com>



## Tietoturva

- ”Vähän sama kuin silloin kun PC tuli...” (Urpo Kaila, CSC)
  - Grid-tunnus on lupa käyttää koneita yli organisaatorajojen!
    - Tyypillisesti lupa ajaa kohdekoneessa mielivaltaista ohjelmaa, vastaa käytännössä tavallista unix-käyttäjätunnusta
  - Uusia riskejä, tietohallinnot ymmällään
  - Paljon tehoa => paljon tuhoja
- Toisaalta: Grid-kehitys alkoi vasta kun virukset ja madot jo mellastivat — tietoturvaan on kiinnitetty huomiota
  - Vahva tunnistus ja salaus: ei selväkielisiä salasanoja
  - Pääsy sidottu varmenteisiin: mitätöinti (revocation) sulkee pääsyn kaikkiin koneisiin

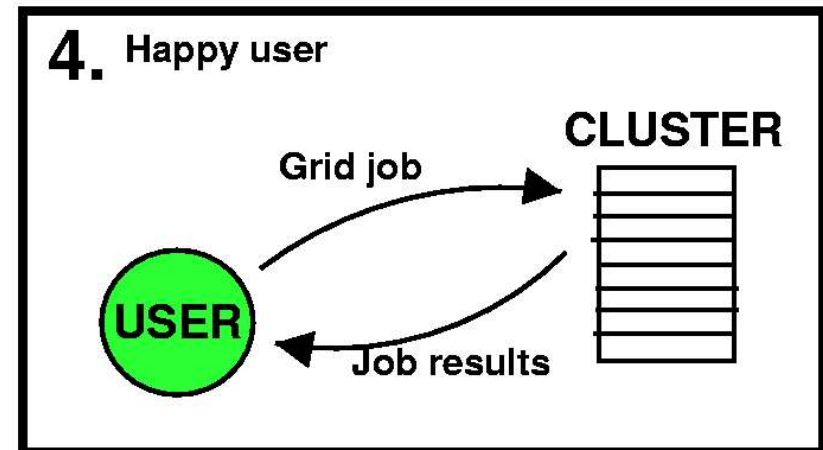
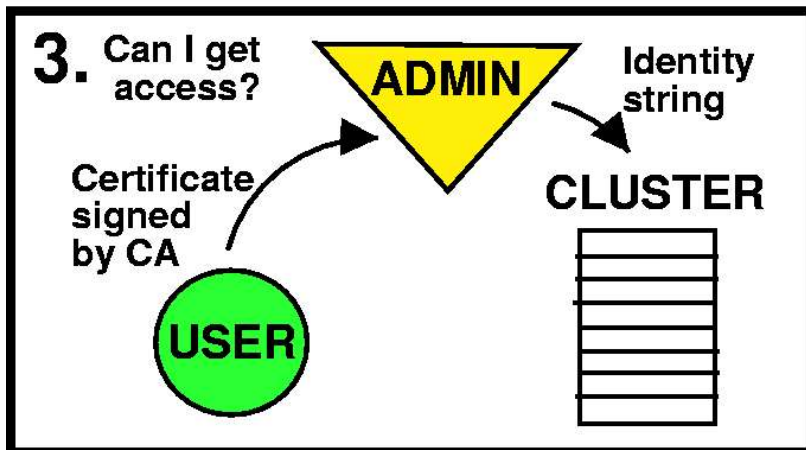
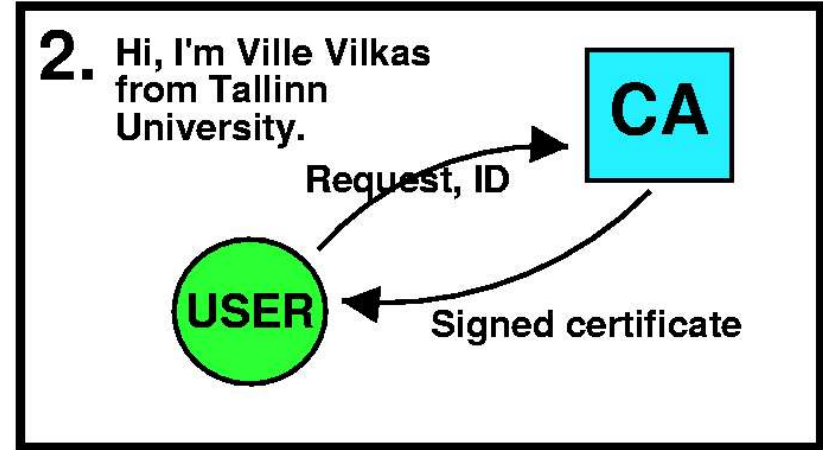
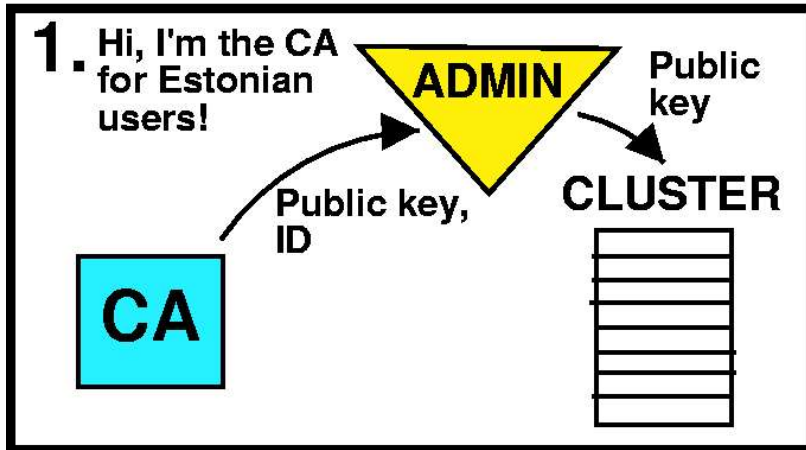


## Varmenteet

- Lähes kaikissa grid-väliohjelmistoissa käyttäjien tunnistus tapahtuu X.509-varmenteilla
- Varmenne on käyttäjän identiteetti gridissä
  - Joka käyttäjällä oma henkilökohtainen varmenne
  - Perustuu julkisen avaimen salaukseen, vakiintunut teknologia
  - Ei automaattisesti takaa pääsyä resursseihin, käyttäjän nimi pitää löytyä kohdekoneen pääsyylistasta
- Tarvitaan luotettu kolmas osapuoli
  - Vahvistaa käyttäjän henkilöllisyyden allekirjoittamalla varmenteen
  - Ylläpitää mitätöityjen varmenteiden listaa



## Varmenteet: luottamusketju





## Virtuaaliorganisaatiot

- Virtuaaliorganisaatiot (Virtual Organization, VO) ovat listoja käyttäjien identiteeteistä
  - Mahdollistavat käyttäjien hallinnan ryhminä
- Tyypillisesti pääsy koneeseen avataan ja oikeudet määritellään kokonaiselle VO:lle kerrallaan
  - Vrt. ryhmäviisumi
  - Luotettava kyseisen virtuaaliorganisaation ylläpitäjään
- Toteutus yksinkertaisina tekstitiedostoina, LDAP-palvelimilla tai virtuaaliorganisaatioiden hallintaan tarkoitettujen tietokantojen avulla





## Global Grid Forum

- Gridin edistämisen- sekä standardointiorganisaatio, mukana sekä akateemista maailmaa että yrityksiä
- Luo teknisiä määrittämiä, ”best practices” -käytäntöjä sekä kartoittaa käyttäjien ja kehittäjien tarpeita
- Perustettu vuonna 2000, ryhmät avoimia kaikille, iso kokoontuminen kolme kertaa vuodessa
  - Mukaan ja vaikuttamaan pääsee helposti, jos aikaa ja intoa on
- Työ vielä alussa, grideissä käytetään toki osin vakiintuneita teknologioita, mutta varsinaiset grid-standardit vielä melko varhaisessa vaiheessa
- <http://www.gridforum.org>



## Grid CSC:llä: Teknologiapalvelut

- Uusia lähes käyttöjärjestelmätason ohjelmia, uusia tiedonsiirtoprotokollia
- Verkon merkitys korostuu entisestään, kapasiteetin tarve kasvaa
- Koneiden huoltokatkot helpottuvat
  - ei enää yhtä suurta merkitystä sillä että yksi kone pois pelistä
- Haasteita kokonaisuuden hallinnassa ja valvonnassa



## Grid CSC:llä: Ohjelmistopalvelut

- Päävastuu grid-ympäristöjen kehittämisestä CSC:llä
- Käyttäjien koulutus
- Mahdollisuus tarjota käyttäjille entistä yhtenäisempi, joustavampi ja monipuolisempi ympäristö
- Integroinnissa monenlaisia haasteita
- Tähän mennessä tehty lähinnä laskentagridiä, tutustuttava myös datapuoleen



## Grid CSC:llä: Tiedepalvelut

- Sovellusten sovittaminen grid-ympäristöihin
- Käyttäjien neuvonta yhdessä ohjelmistopalveluiden kanssa
- Kysyntä älykkäisiin käyttöliittymiin kasvanee, grid-teknologia tarjoaa uusia mahdollisuuksia
- Löydettävä gridin mahdollisuudet kunkin oman tieteenalan sisällä, sekä tunnistettava myös mihin grid ei sovi
- Verkostojen luominen: grid-hankkeet yleensä edistävät ryhmien ja organisaatioiden välistä yhteistyötä



## Grid CSC:llä: Hallinnon palvelut

- Grid tunkeutuu syvälle organisaation sisään, osa aiemmin talon sisäisistä toiminnoista tuleekin nyt sovittaa yhteen kumppanien ratkaisujen kanssa
  - Käyttäjähallinta
  - Käytön seuranta, mittarien kehitys, resurssien jako
- Varmenteiden myöntäminen tai sopiminen käytännöistä miten käyttäjät saavat varmenteensa muuta kautta
- Kansainväliset kontaktit lisääntyvät



## Kiitos mielenkiinnosta!

- Grid-sivut CSC:llä: <http://www.csc.fi/grid/index.phtml.fi>
- CSC:n kirjastossa saatavilla ainakin seuraavat grid-aiheiset kirjat:
  - Foster, Kesselman: The Grid: Blueprint for a New Computing Infrastructure, 2nd Edition, 2003
  - Berman, Fox, Hey: Grid Computing: Making The Global Infrastructure a Reality, 2003
- Kysyttävää?