Korkean käytettävyyden klusteriteknologioiden vertailu

Pro gradu -tutkielma

Juha Petteri Salonvaara

Helsinki 28.10.2007
HELSINGIN YLIOPISTO
Tietojenkäsittelytieteen laitos

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET – UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty/Section	Laitos – Institution – Department	
Matemaattis-luonnontieteellinen tiedekunta	Tietojenkäsittelytieteen laitos	
Tekijä – Författare – Author		
Juha Petteri Salonvaara		
Työn nimi – Arbetets titel – Title		
Korkean käytettävyyden klusteriteknologioiden vertailu		
Oppiaine – Läroämne – Subject		
Tietojenkäsittelytiede		
Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Pro gradu -tutkielma	28.10.2007	68 sivua

Tiivistelmä – Referat – Abstract

Tämä tutkielma käsittelee korkean käytettävyyden klustereita. Tavoitteena on vertailla eri klusteriteknologioiden eroja ja arvioida tulosten perusteella sopivia käyttökohteita eri klusterituotteille. Samoja funktioita tarjoavat eri klusterituotteet asetetaan paremmuusjärjestykseen. Arviointi perustuu erilaisiin määrällisiin arvoihin, kuten solmulaitteiden maksimilukumäärä, sekä laadullisiin arvoihin, kuten käyttöönoton ja hallinnoinnin helppous. Erityisesti tavoitteena on tuoda esille eri tuotteiden vahvuuksia ja heikkouksia. Vertailtavia tuotteita ovat Microsoft Cluster Service (MSCS), TruCluster Server for Tru64 UNIX, Steeleye Lifekeeper for Windows ja Sun Cluster.

ACM Computing Classification System (CSS):
D.4 OPERATING SYSTEMS (C)
D.4.5 Reliability

Avainsanat – Nyckelord – Keywords

klusteri, korkea käytettävyys, Microsoft Cluster Service (MSCS), TruCluster Server for Tru64 UNIX, Steeleye Lifekeeper for Windows, Sun Cluster

Säilytyspaikka – Förvaringställe – Where deposited

Kumpulan tiedekirjasto, sarjanumero C-2007-

Muita tietoja – Övriga uppgifter – Additional information

Sisältö

1 Johdanto

Organisaation toiminnalle elintärkeät tietojärjestelmät vaativat palvelinalustaltaan lähes katkotonta käytettävyyttä. Tavallisten palvelimien käytettävyyttä parannetaan yleisesti kahdentamalla komponentteja, mutta kahdentamatta jäävien komponenttien rikkoutumisista aiheutuvat katkot estävät korkean käytettävyyden saavuttamisen. Klusteri on hyvän kustannus/laatusuhteensa ja tehokkuutensa ansiosta suosiotaan kasvattava teknologia saavuttaa korkea käytettävyys kahdentamalla kokonaisia palvelimia. Useat varusohjelmistovalmistajat ovat kehittäneet omat kilpailevat klusterituotteensa ja tuetut alustat kattavat jo kaikki yleisimmät palvelinkäyttöjärjestelmät.

Klusterit jaotellaan kahteen päätyyppiin: korkean käytettävyyden klusterit (high-availability-cluster, HA-cluster) ja korkean suorituskyvyn klusterit (high-performance-cluster, HPC-cluster). Tämä tutkielma käsittelee korkean käytettävyyden klustereita, joista on tehty lukumääräisesti hyvin vähän eri tuotteiden ominaisuuksia kokonaisvaltaisesti vertailevia tutkimuksia. Jeffrey Absher DePaul-yliopistosta vertaili vuonna 2003 Microsoftin klusteripalvelun, AIX HA:n ja Linux HA:n HTTP-palvelun käytettävyyttä tavalliseen klusteroimattomaan HTTP-palveluun verrattuna [Abs03]. Muita korkean käytettävyyden klusterituotteiden vertailuja on esiintynyt lähinnä tietotekniikka-alan ammattilehdistössä, kuten Tietokone-lehden numerossa 8/2002 [Häm02].

Tämän tutkielman tavoitteena on vertailla eri klusteriteknologioiden eroja ja arvioida tulosten perusteella sopivia käyttökohteita eri klusterituotteille. Samoja toimintoja tarjoavat eri klusterituotteet asetetaan paremmuusjärjestykseen. Arviointi perustuu erilaisiin määrällisiin arvoihin, kuten solmulaitteiden maksimilukumäärä, sekä laadullisiin arvoihin, kuten käyttöönoton ja hallinnoinnin helppous. Erityisesti tavoitteena on tuoda esille eri tuotteiden vahvuuksia ja heikkouksia. Vertailtavia tuotteita ovat Microsoft Cluster Service (MSCS), TruCluster Server for Tru64 UNIX, Steeleye Lifekeeper for Windows ja Sun Cluster.

Tutkielma rakentuu seuraavista luvuista. Luvussa kaksi käsitellään klustereiden yleistä terminologiaa ja historiaa. Luvuissa kolmesta kuuteen käydään läpi tutkittavat klusterituotteet kukin omassa luvussaan. Luvussa seitsemän esitellään vertailun tulokset ja luvussa kahdeksan on yhteenveto tutkielmasta.

2 Klusteriarkkitehtuuri

Tämän tutkielman aluksi on hyödyllistä määritellä klustereiden yhteydessä käytettyä terminologiaa.

2.1 Korkea käytettävyys ja vikaantuminen

ISO 9241-11 -standardi määrittelee käytettävyyden (availability) seuraavasti: "käytettävyys mittaa sitä, missä määrin tietyt käyttäjät tietyssä tilanteessa voivat käyttää tuotetta tiettyyn tarkoitukseen, kriteereinä käytön tehokkuus ja vaikuttavuus sekä käyttäjän subjektiivinen tyytyväisyys" [Ran99]. Tietojärjestelmien palveluiden käytettävyys lasketaan jakamalla järjestelmän palveluiden käytettävissä ollut aika kuluneella aikajaksolla. Käytettävyys ilmaistaan tyypillisesti prosentteina tietyllä aikajaksolla, esimerkiksi 99,9 % vuodessa, jota kutsutaan myös numeron yhdeksän yhteenlasketun lukumäärän mukaan ”kolmen yhdeksikön käytettävyydeksi”.

Eri järjestelmien asiakkailla on erilaisia vaatimuksia käytettävyyden suhteen. Vähemmän tärkeä järjestelmä voi olla alhaalla vuorokausia, kun taas kriittisessä järjestelmässä lyhytkin katko voi aiheuttaa suuria menetyksiä. Korkea käytettävyys (high-availability) kuvaa palvelutasoa, jossa suunnitellut ja suunnittelemattomat katkot ovat mahdollisimman lyhyitä ja erityisesti eivät ylitä määritettyä raja-arvoa. Korkea käytettävyys on siten subjektiivinen käsite. HP:n johto määritteli vuonna 1998 korkean käytettävyyden rajaksi 99,999 %, joka tarkoittaa, että kyseisellä järjestelmällä saisi olla maksimissaan vuodessa yhteenlaskettuna vain viiden minuutin katkoaika [Wey01, s. 7]. Taulukossa 1 on kuvattu maksimikatkoaikoja vuoden tarkasteluaikajaksolla tyypillisille palvelusopimuksissa esiintyville käytettävyyden prosenttilukemille.

käytettävyys	99 %	99,5 %	99,95 %	99,999 %	100 %
maksimi katkoaika	88 h	44 h	5 h	5 min.	0

Taulukko 1. Maksimi katkoaika vuoden aikajaksossa eri käytettävyyksille [Wey01, s. 15].

Vaikka käyttäjät voivat toivoa kaikille järjestelmilleen korkeata käytettävyyttä, käytännössä näin kovat vaatimukset tuottavat järjestelmille niin suuret kustannukset, etteivät käyttäjät suostuisi palvelua maksamaan. Näin ollen korkean käytettävyyden saavuttavaa teknologiaa käytetään yleensä vain liiketoiminnan kannalta kriittisiin erikoistarkoituksiin. HP:n klusteriasiantuntija Peter Weygant esittää seitsemän menetelmää korkea käytettävyyden saavuttamiseksi, jotka ovat [Wey01, s. 37]:

1. komponenttien monentaminen (redundancy)
2. ohjelmiston ja laitteiston vaihtomenetelmät, esimerkiksi hot-swap
3. ennakoitujen katkojen huolellinen suunnittelu
4. käyttäjien ja järjestelmän välisen interaktion eliminointi operointivirheiden estämiseksi
5. automaattiset reagoinnit vikatilanteisiin
6. kokonaisvaltainen hyväksymiskoe (acceptance test)
7. ylläpitäjien ohjeistaminen sellaisiin vikatilanteisiin, joihin ei ole automaattista reagointia

Järjestelmiä suunniteltaessa ja palvelun laatua mitattaessa tavoiteltavan käytettävyyden saavuttaminen riippuu järjestelmän vikasietoisuudesta. Yksittäisen palvelimen laitetason vikasietoisuutta voidaan parantaa kahdentamalla komponentteja esimerkiksi RAID-levyjärjestelmällä (redundant array of inexpensive disks) sekä kahdennetuilla virtalähteillä, lähiverkkokytkennöillä ja tuulettimilla. Jos koko järjestelmä on yhden palvelimen varassa, aiheuttaa esimerkiksi emolevyllä olevan komponentin hajoaminen palvelimen ja koko järjestelmän kaatumisen. Huollon saapuminen paikan päälle vaihtamaan uutta emolevyä kestää helposti useita tunteja parhaimmallakin huoltosopimuksella, jolloin kriittisen järjestelmän käytettävyyden tavoite voi jäädä saavuttamatta.

Komponenttia, jonka hajoaminen aiheuttaa koko järjestelmän kaatumisen, kutsutaan yksittäiseksi vikaantumispisteeksi (SPOF, single point of failure) [Wey01, s. 42]. Tätä ongelmaa voidaan helpottaa kahdentamalla kokonaisia palvelimia, jolloin järjestelmä esimerkiksi siirtää kaatuneelta palvelimelta palveluja toimiville palvelimille. Eräs tällaisen palvelun tarjoava korkean käytettävyyden teknologia on klusteri (cluster). Muita korkean käytettävyyden ratkaisuja palvelinlaitteilla ovat esimerkiksi ylimääräiset prosessorikortit (SPU, system processor unit), jotka muodostavat partitioita (hardware partition), mutta järjestelmä voi jakaa yhteisiä komponentteja luoden yksittäisiä vikaantumispisteitä [Wey01, s. 56].

2.2 Klusteri ja failover-operaatio

Klusteri on joukko erillisiä yhteen kytkettyjä tietokoneita, jotka yhdessä takaavat jatkuvan palvelun, vaikka jokin yksittäinen tietokone vikaantuisi [Wey01, s. 41]. Gregory Pfisterin määritelmän mukaan klusteri on: ”rinnakkainen tai hajautettu kokoelma yhteen liitettyjä kokonaisia tietokoneita, jotka toimivat yhtenä yhdistyneenä laskentaresurssina” [MST06]. Klusterin jäseninä olevia tietokoneita kutsutaan solmulaitteiksi (node).

Korkean käytettävyyden klustereilla pyritään lähes katkottomaan palveluun. Korkean käytettävyyden klusterissa vikasietoisuutta parannetaan kahdentamalla kokonaisia palvelimia, jolloin yksittäisen solmulaitteen hajoaminen ei aiheuta koko järjestelmän ja palvelun kaatumista. Klusterin näkökulmasta kadonneen solmulaitteen palvelut siirretään toiminnassa oleville klusterin muille jäsenille. Palvelujen siirtoa vikatilanteessa klusterin solmulaitteelta toiselle kutsutaan failover-operaatioksi.

Failover-operaation voi aloittaa myös toiminnassa oleva solmulaite, jos sen hallinnassa oleva klusteriresurssi (cluster resource) ei ole saatavilla, eikä kyseinen solmulaite saa käynnistettyä resurssia takaisin toimintaan. Palvelujen käyttäjälle failover-operaatio ei näy muuten kuin itse failover-operaation aikana lyhyenä katkona palvelussa. Kun palvelut ovat failover-operaation jälkeen siirretty toiselle solmulaitteelle, jatkuu palvelu entiseen tapaan [Wey01, s. 59].

Failover-operaation kuluttamaa aikaa käytetään klusterin suorituskykymittarina ja kutsutaan failover-ajaksi. Failover-aika riippuu luonnollisesti käytetystä klusteriteknologiasta, laitteiston suorituskyvystä, klusteroitujen resurssien lukumäärästä ja yksittäisten klusteriresurssien failover-operaation kuluttamasta ajasta. Eri klusteriresursseilla kuluu eri aika prosessien sammuttamiseen toiselta solmulaitteelta ja käynnistämiseen toiselle solmulaitteelle. Esimerkiksi klusteroidun tietokantaprosessin failover-operaatio on tyypillisesti pidempi kuin yksinkertaisen klusteroidun tiedostojaon (file share). Yksittäisen tiedostojaon failover-operaatio suoriutuu sekunnin osissa, kun tietokantaprosessin failover-operaatioon voi kulua kymmeniä sekunteja.

2.3 Klusterikonfiguraatiot

Klusterin solmulaitteet voivat toimia aktiivi- tai passiivitilassa. Aktiivitilassa oleva solmulaite suorittaa omia tehtäviään ja saman klusterin toisen jäsenen kaaduttua, sille tai jollekin toiselle jäljelle jääneelle solmulaitteelle siirretään kaatuneen solmulaitteen mahdolliset tehtävät. Passiivitilassa oleva solmulaite on vain toimettomana varalaitteena, kunnes failover-operaatio määrää sille tehtäviä ja resursseja, jolloin kyseinen passiivitilassa ollut solmulaite siirtyy aktiivitilaan [Wey01, s. 57].

Klusterissa olevien aktiivisten ja passiivisten jäsenten lukumäärällä voidaan jakaa klusterin konfiguraatiot kahteen luokkaan, jotka ovat aktiivi/passiivi-konfiguraatio (active/standby configuration) tai aktiivi/aktiivi-konfiguraatio (active/active configuration). Aktiivi/passiivi-konfiguraatiossa on mukana passiivisia jäseniä. Aktiivi/passiivi-konfiguraation etuna on, että failover-operaation tapahtuessa dedikoidulla varalaitteella on yleensä riittävästi suorituskykyä, jolloin palvelun laatu ei heikkene. Haittapuolena on luonnollisesti ylimääräisten tyhjäkäynnillä pyörivien laitteiden hankinta- ja ylläpitokulut. Aktiivi/aktiivi-konfiguraatiossa kaikilla klusterin jäsenillä ajetaan palveluja, jolloin failover-operaation tapahtuessa yhdelle solmulaitteelle tulee omien tehtäviensä lisäksi ylimääräisiä töitä ja suorituskyky heikkenee tyypillisesti jonkin verran [Wey01, s. 92].

Levy- ja tiedostopalvelujen yhteiskäytön osalta klusterien toimintatapa jaetaan kahteen päätyyppiin: ”shared resource” ja ” shared nothing”. Shared resource -mallissa klusterin kaikki jäsenet pääsevät käsiksi kaikkiin levyresursseihin samanaikaisesti. Tällöin useat solmulaitteet voivat yrittää avata samaa tiedostoa, jolloin sovelluksen ja käyttöjärjestelmän huolenaiheeksi jää eheyden säilyttäminen. Koska lukkotaulu (lock table) on palvelinkohtainen, tarvitaan shared resource -mallin klustereissa hajautettua lukkojenhallintaohjelmistoa (distributed lock manager) välittämään tiedot lukoista kaikille solmulaitteille. Hajautetun lukkojenhallintaohjelmiston toiminta aiheuttaa itsessään järjestelmään kuormitusta, kun lukkojen tiedot täytyy välittää kaikille prosessoreille. Hajautetun lukkojenhallintaohjelmiston monimutkaisuutta lisää tarve huolehtia välimuisteista sekä erikoistilanteiden hallinta, esimerkiksi kaatuneen solmulaitteen lukkojen siirto klusterin muille jäsenille [Lib00, s. 9 - 13].

Shared nothing -mallissa ainoastaan yksi solmulaite kerrallaan voi käyttää tiettyä levyresurssia. Muut solmulaitteet eivät pysty käyttämään kyseiseltä levyresurssilta mitään tiedostoja. Tällöin erillistä hajautettua lukkojenhallintaohjelmistoa ei tarvita, järjestelmä on yksinkertaisempi ja prosessorit kuormittuvat vähemmän. Shared nothing -mallin haitta on, että kaatuneen solmulaitteen levyresurssit täytyy siirtää failover-operaatiolla klusterin muille jäsenille. Lisäksi levyresursseja täytyy olla perustettuna mahdollisesti useita, koska jokainen aktiivitilassa oleva solmulaite tarvitsee vähintään yhden levyresurssin [Lib00, s. 13 - 14].

2.4 Klustereiden maantieteellinen jaottelu

Korkean käytettävyyden klusterit jaotellaan eri tyyppeihin maantieteellisen etäisyyden mukaan. Jos klusterin kaikki solmulaitteet sijaitsevat yhdessä konesalissa, muodostavat ne paikallisen klusterin (local cluster). Suurin osa klustereista on paikallisia klustereita, joten lähtökohtaisesti korkean käytettävyyden klustereista puhuttaessa tarkoitetaan paikallisia klustereita. Muilla maantieteellisen jaottelun klusterityypeillä haetaan pidempää etäisyyttä solmulaitteiden välille, jotta järjestelmä sietäisi paikalliseen infrastruktuuriin kohdistuvia katastrofeja. Kampusklusterin (campus cluster) solmulaitteet sijaitsevat kahdessa tai useammassa konesalissa, joilla voi olla jonkin verran

etäisyyttä keskenään. Kampusklusterin konesalit ovat tyypillisesti vierekkäisiä rakennuksia samassa korttelissa. Maantieteellisesti kampusklusterin osalta odotetaan, että maa-alue rakennuksineen on klusterin ylläpitäjäorganisaation hallinnoima, eikä siten tarvita ulkopuoliselta verkko-operaattorilta vuokrattua linjaa (leased line) konesalien väliseen kommunikointiin [Wey01, s. 183].

Kampusklusteria suurempaa etäisyyttä varten ovat tarjolla kaupunkialueklusteri (metropolitan cluster) ja mantereinen klusteri (continental cluster). Kaupunkialueklusterin konesalien maksimietäisyys toisistaan on määritelty 43 kilometriksi, joka johtuu tiettyjen levyjärjestelmien spesifikaatioiden asettamasta ylärajasta. Kyseinen 43 kilometrin raja on ollut esimerkiksi järjestelmillä HP Surestore Disk Array ja EMC Symmetrix Array [Wey01, s. 189]. Suurempi etäisyys aiheuttaisi liian pitkän siirtoviiveen [HP03, s. 4-196]. Mantereinen klusteri toteutetaan peilaamalla konesalit kokonaisuudessaan, jolloin samasta klusterista on kaksi kopiota [Wey01, s. 75]. Mantereisen klusterin konesalit voivat sijaita nimensä mukaisesti eri maanosissa, eikä mantereiselle klusterille ole asetettu etäisyysrajoituksia.

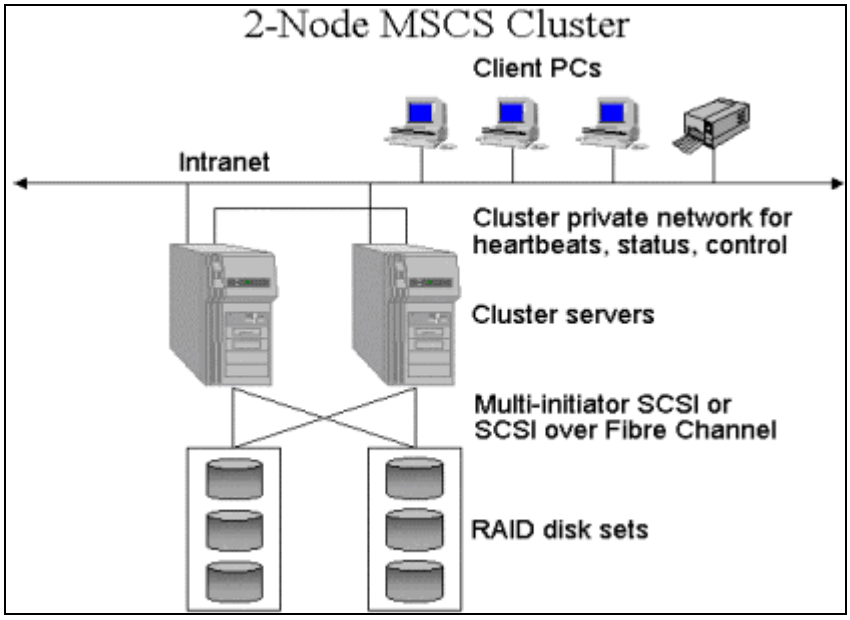
2.5 Solmulaitteiden kommunikointi

Korkean käytettävyyden klusterit käyttävät lähiverkkoa asiakkaiden palvelun lisäksi klusterin solmulaitteiden väliseen kommunikointiin. Klusterin solmulaitteiden toisilleen välittämiä statusviestejä kutsutaan heartbeat-signaaleiksi ja tähän viestintään tarkoitettua verkkoa heartbeat-aliverkoksi (heartbeat subnet). Heartbeat-signaalien avulla klusteri pystyy päättelemään onko jokin sen jäsensolmuista mahdollisesti kaatunut. Mikäli heartbeat-signaalien välittämisessä tapahtuu katko, muodostetaan klusteri uudestaan jäsenien enemmistöstä, jonka solmulaitteiden välinen kommunikointi toimii vielä. Tätä enemmistöä kutsutaan quorum:iksi. Vähemmistöön jääneet sekä heartbeat-kommunikoinnin tavoittamattomissa olevat jäsenet poistavat itsensä klusterista fail-operaatiolla [Wey01, s. 63].

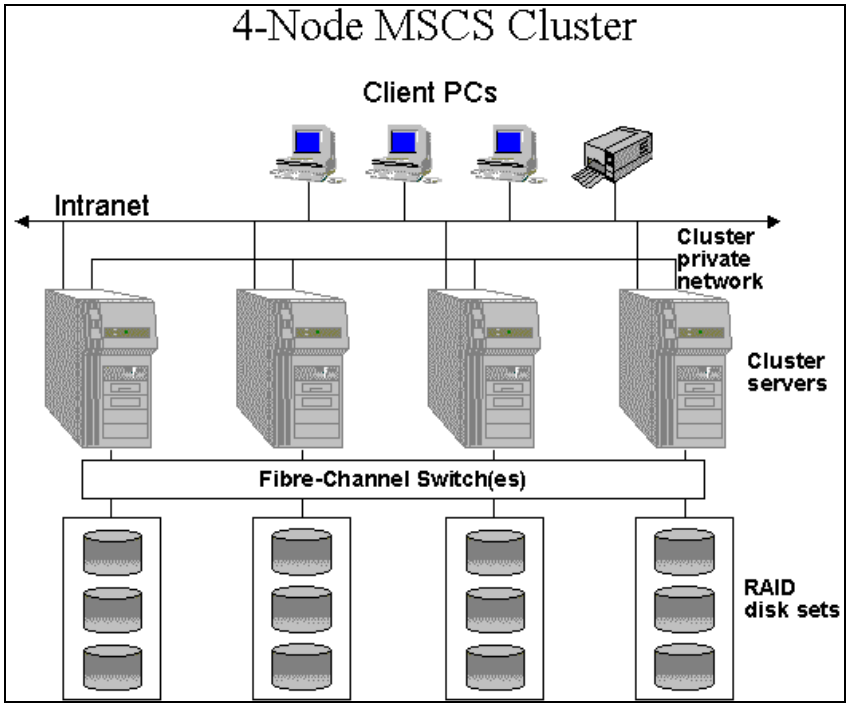
Mikäli klusterin uudelleenmuodostamistilanteessa ei saada solmujen joukosta aikaiseksi enemmistöä, vaan jako on tasan 50 %, joudutaan split-brain-tilanteeseen. Tällöin

kumpikin 50 % joukko yrittää saada quorum:in itselleen ja muodostaa klusterin samanaikaisesti, mutta vain toinen joukoista voi onnistua tässä. Split-brain-tilanteesta selvitään erityisellä klusterilukkoresurssilla (cluster lock), joka on tyypillisesti jokin kaikille klusterin jäsenille näkyvä looginen levy, levyalue tai tiedosto. Se kummassa joukossa oleva jäsen ehtii varata lukon itselleen, pääsee muodostamaan klusterin. Hävinnyt osapuoli pysäyttää klusteriprosessiensa toiminnan havaittuaan lukon olevan lukittuna [Wey01, s. 86].

Klusterin resursseina olevat levyt jaetaan fyysisesti yhteisesti kaikkien jäsenten kesken, joten kaikki jäsenet pääsevät kiinni jaettuihin levyihin. Klusterin shared nothing malli tai hajautettu lukkojenhallintaohjelmisto pitää kuitenkin huolen, että tietty resurssi on luovutettu kerrallaan vain yhden jäsenen käyttöön. Fyysisesti levyjen näyttäminen kaikille jäsenille hoituu SAN:in (Storage Area Network) avulla tai pienissä kahden solmulaitteen SCSI-levyratkaisussa (Small Computer System Interface) Y-kaapelilla. Näistä SAN kuitukeskittimineen (fibre channel hub) tai kuitukytkimineen (fibre channel switch) on huomattavasti edistyneempi vaihtoehto, eikä Y-kaapeliratkaisua käytetä vanhentuneena juuri kuin opetus- ja testauskäyttöön. Kuvissa 1 ja 2 on kuvattu esimerkkeinä SCSI-levy- ja SAN-ratkaisut [MST06]:



Kuva 1. Kahden solmulaitteen Microsoftin klusteri SCSI-Y-kaapelilla [MST06].



Kuva 2. Neljän solmulaitteen Microsoftin klusteri SAN-järjestelmällä [MST06].

Quorum-enemmistö voidaan muodostaa eräissä klusteritoteutuksissa jaetun levyn sijasta erillisellä quorum-palvelimella (quorum node, arbitrator node). Quorum-palvelin ei kuulu kyseiseen klusteriin varsinaisten solmulaitteiden tapaan, vaan quorum-palvelimen tehtävänä on ratkaista split-brain-tilanteet. Quorum-palvelin voi kylläkin olla jonkin toisen klusterin tavallisena jäsenenä ilman quorum-palvelin-statusta, esimerkiksi Linux-klustereissa sekä HP-UX-klusterissa. Quorum-palvelinta käytetään yleensä kaupunkialueklusterissa, jolloin kahden konesalin väliin perustetaan usein kolmas konesali, jossa sijaitsevat klusterin quorum-palvelimet [Wey01, s. 188].

2.6 Klusteriresurssit

Klusteriresurssit tarjoavat asiakkaille klusterin palveluja. Resurssit voivat olla fyysisiä, kuten kiintolevy, tai loogisia, kuten IP-osoite. Klusteriresurssi on klusterin palvelujen pienin hallinnollinen yksikkö, eli niitä voidaan perustaa, muokata ja poistaa yksitellen. Klusteriresurssit kootaan erityisiin resurssiryhmiin (resource group), jotka muodostavat fail-over-operaatiossa yhtenäisen atomisen kokonaisuuden. Yhden resurssiryhmän kaikki resurssit siirretään siis yhdessä samalle solmupalvelimelle. Resurssiryhmille voidaan määritellä ensisijainen solmupalvelin (primary node, preferred

owner), jolle resurssiryhmä pyritään siirtämään. Jos resurssiryhmän ensisijainen solmupalvelin ei ole käytettävissä, siirretään resurssiryhmä toiselle solmupalvelimelle, joka kuuluu ennalta määriteltyjen mahdollisten solmupalvelinten joukkoon (adoptive node, possible owner) [Wey01, s. 96].

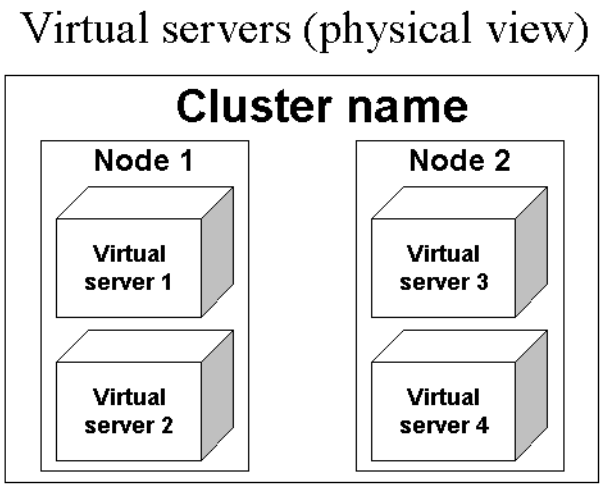
Kun kaatunut ensisijainen solmupalvelin jossain vaiheessa saadaan takaisin toimintaan, voidaan sille automaattisesti siirtää takaisin toisille jäsenille siirretyt resurssiryhmät. Tämä tapahtuu määrittelemällä ennakkoon resurssiryhmälle failback-politiikka (failback-policy), jolloin kyseinen resurssiryhmä siirretään failback-operaatiolla ensisijaiselle solmupalvelimelle heti kun solmupalvelin on liittynyt takaisin klusteriin [Wey01, s. 100].

Klusteriresursseja voidaan sammuttaa ja käynnistää yksitellen, mutta kuten aiemmin todettiin, klusteriresursseja voidaan siirtää solmulta toiselle vain kokonaisina resurssiryhminä. Jos yksittäinen klusteriresurssi kaatuu, yrittää klusteri käynnistää resurssin uudestaan samalla solmulaitteella, mutta jos uudelleenkäynnistys ei onnistu, suoritetaan koko resurssiryhmälle failover-operaatio toiselle solmulaitteelle [VDB98].

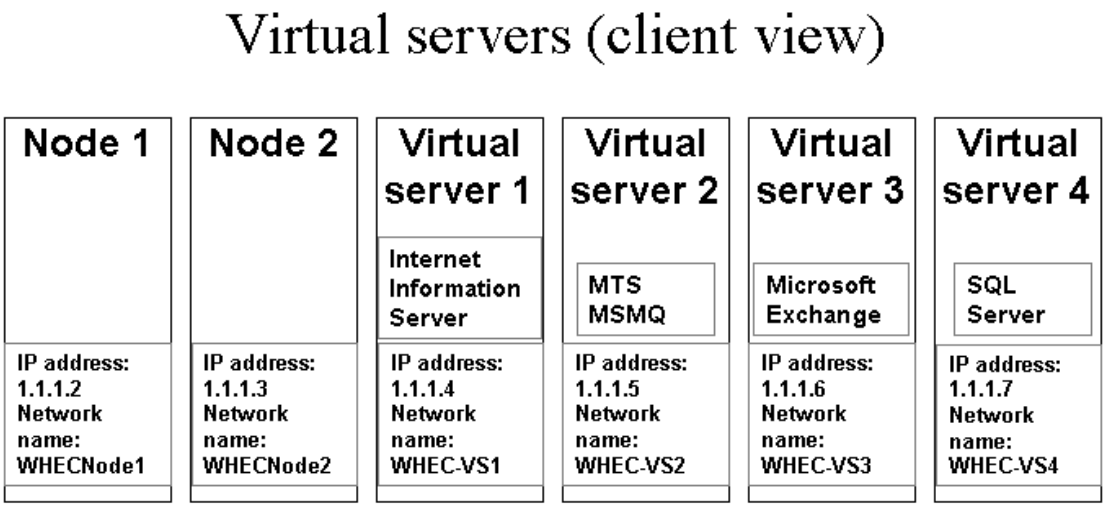
Klusteriresurssit muodostavat toisiinsa nähden riippuvuuksia (dependency). Esimerkiksi tietokantaresurssi riippuu levyresurssista, jossa tietokanta sijaitsee. Riippuvuuksista muodostuu riippuvuuspuita, jotka määräävät missä järjestyksessä klusterin täytyy käynnistää ja sammuttaa klusteriresursseja. Riippuvuudet kohdistuvat vain saman resurssiryhmän resursseihin, eli riippuvuudet eivät voi ylittää resurssiryhmien rajoja [VDB98].

Resurssit esitetään asiakkaille virtuaalipalvelimen (virtual server) muodossa. Yhdessä klusterissa voi olla useampia virtuaalipalvelimia, jotka jakautuvat resurssiryhmien mukaan. Virtuaalipalvelimilla on omat nimet ja IP-osoitteet, jolloin asiakkaat voivat käyttää niiden palveluja tarvitsematta tietää mikä on solmulaitteen nimi ja IP-osoite [MSP01]. Esimerkki solmulaitteiden ja virtuaalipalvelinten jaosta on kuvassa 3 sekä IP-osoitteiden ja palveluiden jaosta kuvassa 4 [MST06]. Molemmat kuvat esittävät samaa klusteria eri näkökulmista. Kuva 3 näyttää virtuaalipalvelinten 1 ja 2 olevan solmulaitteella 1 ja virtuaalipalvelinten 3 ja 4 solmulaitteella 2. Kuvasta 4 selviää solmulaitteiden ja virtuaalipalvelinten nimet, IP-osoitteet ja palvelut. Solmulaitteella 2

on siis käynnissä virtuaalipalvelimet 3 ja 4, jotka tarjoavat palveluina Microsoftin Exchange-sähköpostipalvelun ja SQL-tietokantapalvelun.



Kuva 3. Esimerkki solmujen ja virtuaalipalvelinten jaosta [MST06].



Kuva 4. Esimerkki IP-osoitteiden ja palveluiden jaosta [MST06].

2.7 Sovellusten klusteritietoisuus ja klusterituki

Mitä tahansa sovellusta ei voida suorittaa klusteriresurssina. Klusterin tarjoamilta palveluilta vaaditaan, että kyseiset palvelut ovat klusteritietoisia (cluster-aware) kyseisen klusterituotteen suhteen. Muuten klusteri ei pysty päättelemään toimivatko nämä palveluprosessit vai onko syytä tehdä failover-operaatio. Sovelluksen, joka ei ole klusteritietoinen, failover-operaatio voi aiheuttaa ongelmia datan eheyteen ja sovelluksen vakauteen. Erityisesti varusohjelmistojen, esimerkiksi nauhavarmistusohjelmistojen ja

virustorjuntatuotteiden, osalta edellytetään, että ne ovat klusterituettuja (cluster supported) kyseisen klusterituotteen suhteen. Tällöin varusohjelman resursseja ei klusteroida, mutta kyseinen varusohjelma osaa toimia klusterin yhteydessä haittaamatta klusterin toimintaa.

Ohjelmistotoimittajat ilmoittavat ohjelmistojensa spesifikaatioissa ovatko kyseiset tuotteet klusteritietoisia ja tuettuina kunkin klusterituotteen osalta. Jos klusterituesta tai -tietoisuudesta ei erityisesti mainita mitään, ei sovellus silloin yleensä ole klusterituettu eikä -tietoinen. Luonnollisesti sovellusvalmistajan teknisestä tuesta voidaan varmistaa sovelluksen klusterituki ja -tietoisuus. Kun sovellusta lähdetään suunnittelemaan klusteriympäristöä varten, on siihen Weygantin määritelmään mukaan sisällytettävä seuraavat ominaisuudet [Wey01, s. 77 - 78]:

1. sovelluspalvelun kyky tehdä failover-operaatio
2. sovelluspalvelun kyky käynnistyä uudelleen ilman ylläpitäjän toimenpiteitä
3. sovelluspalvelun kyky toimia klusterin millä tahansa solmulaitteella
4. sovelluspalvelun kyky monitoroida omaa toimintaansa ja erityisesti onko palvelu toiminnassa
5. sovelluspalvelun hyvin määritellyt käynnistys- ja sammutusproseduurit
6. sovelluspalvelun hyvin määritellyt varmistus-, palautus- ja päivitysproseduurit (backup, restore and upgrade procedures)

Korkean käytettävyyden klusteriratkaisun suunnittelu, toteutus ja toiminta koostuvat usean eri tason tekijän yhteistoiminnasta. Tasot on lueteltu taulukossa 2 [Wey01, s. 81 - 82].

systeemitaso	kuinka korkea käytettävyys saavutetaan
laitteistokomponenttitaso	komponenttien redundanssilla tai vaihtomenetelmillä
mikrokooditaso (firmware level)	virheenkorjaavalla mikrokoodilla
palvelintaso	redundanssilla ja usealla datapolulla
käyttöjärjestelmätaso	käyttöjärjestelmän peilauksella
järjestelmän ja verkonhallintaso	hajautetulla hallinnalla ja monitorointityökaluilla
klusteritaso	Datan pitää olla suojattu. Solmupalvelinten välinen kommunikointi pitää olla korkean käytettävyyden tasolla. Solmulaitteita täytyy olla useita.
tietokantataso	Tietokannan täytyy kyetä käynnistymään toisella solmulaitteella tai toimia usealla solmulaitteella samanaikaisesti.
transaktioiden käsittelyn taso	transaktioiden valvontatyökalujen (monitor) korkea käytettävyys
sovellustaso	Sovellusten täytyy olla vikasietoisia ja kyetä toipumaan vikatilanteista. Sovellusten on kyettävä vaihtamaan solmulaitetta ilman ylläpitäjän toimenpiteitä.

Taulukko 2. Korkean käytettävyyden tasot [Wey01, s. 81 - 82].

2.8 Korkean käytettävyyden klusterituotteiden historiaa

Kunniaa korkean käytettävyyden klusterin käsitteen luomisesta ja ensimmäisestä klusterituotteesta ei ole yksimielisesti myönnetty kenellekään, mutta Digitalin 1980-luvulla kehittämää [Lib00, s. 7] ja toukokuussa 1983 esiteltyä VAXclusteria VAX/VMS-käyttöjärjestelmälle [DEC97, s. 46] pidetään yleisesti ensimmäisenä kaupallisesti menestyneenä korkean käytettävyyden klusterituotteena [Abs03]. VAXclus-

ter on yhä tänäkin päivänä markkinoilla VMScluster-nimisenä Hewlett-Packardin kehittämänä.

1990-luvulla akateeminen yhteisö ja Digitalin lisäksi muut varusohjelmistoja kehittävät yritykset kiinnostuivat klustereista. Nykyään erilaisia korkean käytettävyyden klusterituotteita on jo kymmeniä. Wikipedia määrittelee korkean käytettävyyden klustereiksi seuraavat 20 tuotetta, jotka on kerätty taulukkoon 3 [Wik07]:

tuote	järjestelmä
GoAhead SelfReliant	Linux, Windows, VxWorks ja Solaris
HA/FST	Solaris
HA-OSCAR	avoimen lähdekoodin järjestelmät
HP ServiceGuard	HP-UX ja Linux
IBM HACMP	AIX
iCluster	iSeries
Linux-HA	Linux
Microsoft Cluster Server (MSCS)	Windows
Novell Cluster Services	Novell NetWare [Nov07]
OpenClovis ASP	avoimen lähdekoodin järjestelmät
OpenSSI	Linux
OpenVMS	VMS
Parallel Sysplex	IBM mainframe
PRIMECLUSTER	Solaris ja Linux
Red Hat Cluster Suite	Linux
RSF-1	AIX, HP-UX, Linux, Solaris ja OS X
SteelEye LifeKeeper	Linux ja Windows
Sun Cluster	Solaris
TruCluster	Tru64 Unix
Veritas Cluster Server	AIX, HP-UX, Linux, Solaris ja Windows [Ran02, s. 403]

Taulukko 3. Korkean käytettävyyden klustereita [Wik07].

Valitsin tähän tutkielmaan vertailtavaksi kaupallisia klusterituotteita, jotka edustavat eri käyttöjärjestelmiä sekä tulevat omien kokemuksieni mukaan useimmiten vastaan palvelininfrastruktuuriprojekteissa valittaessa klusteriratkaisuja. Tuotteet ovat Microsoft Cluster Service (MSCS), TruCluster Server for Tru64 UNIX, Steeleye Lifekeeper for Windows ja Sun Cluster. Seuraavissa luvuissa tutkitaan näiden tuotteiden ominaisuuksia.

3 Microsoft Cluster Service (MSCS)

3.1 Versiot

Microsoftin klusteripalvelin MSCS (Microsoft Clustering Server) julkaistiin ensimmäisen kerran Windows NT 4.0 server enterprise edition -palvelimelle [MSP01]. Sen jälkeen klusteri on ollut tarjolla käyttöjärjestelmissä Windows 2000 advanced server, Windows server 2003 enterprise edition sekä edellisten datacenter-versioissa, jolloin akronyymin MSCS merkitys vaihdettiin tarkoittamaan klusteripalvelua (Microsoft Cluster Service). Klusteripalvelu kuuluu näiden käyttöjärjestelmäversioiden lisenssin hintaan. Sen sijaan edullisempiin palvelinversioihin Windows NT 4.0 server, Windows 2000 server ja Windows server 2003 standard edition klusteripalvelua ei saa lainkaan. Windows server 2003 enterprise edition ja datacenter edition tukevat molemmat maksimissaan kahdeksaa solmulaitetta klusterissa [MS03a, s. 6]. Käytännössä klusteripalvelu asennetaan Windows server 2003 enterprise edition -alustalle. 64-bittinen Windows server 2003 enterprise edition 25 CAL-lisenssillä maksoi maaliskuussa 2007 Verkkokauppa.com:ssa 2848,90 € [Ver07].

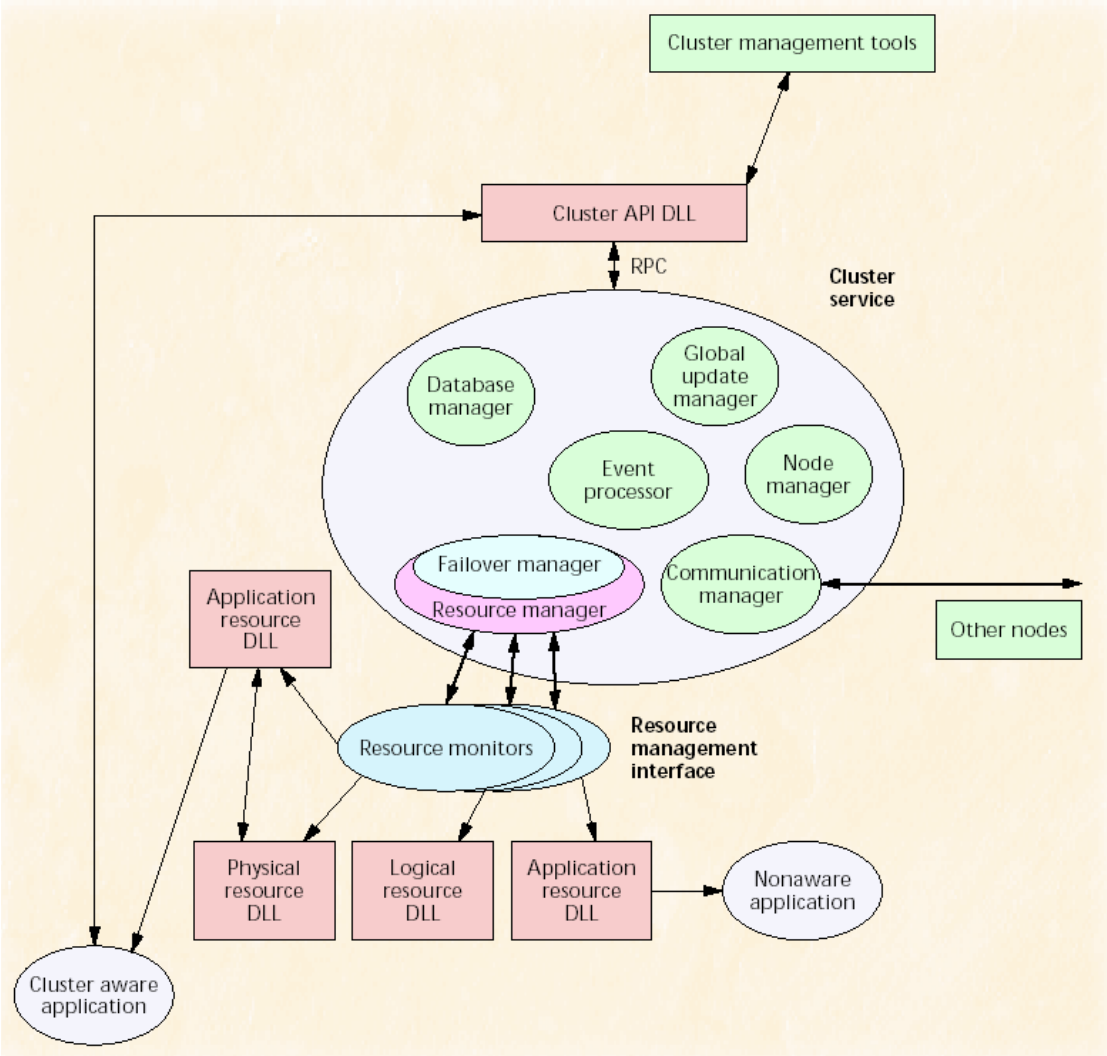
3.2 Arkkitehtuuri

Klusteripalvelussa on 14 komponenttia [MSP01, s. 9]:

1. Checkpoint manager
2. Communications manager
3. Configuration database manager
4. Event processor
5. Event log manager
6. Failover manager
7. Global update manager
8. Log manager
9. Membership manager

10. Node manager
11. Object manager
12. Resource manager
13. Resource monitors
14. Resource DLLs

Klusteripalvelu näkyy yhtenä käyttöjärjestelmän taustalla suoritettavana palveluna, mutta koostuu useasta manager-nimisestä komponentista, jotka suorittavat erillisiä ja hyvin erilaisia tehtäviä [Lib00, s. 15]. Komponenttien välisen yhteistoiminnan havainnollistaa kuva 5 [GSM98].



Kuva 5. Microsoft NT 4.0 Clustering Server -arkkitehtuuri [GSM98].

Checkpoint manager varmistaa, että sovellusten data saadaan siirrettyä failover-operaatiossa toiselle solmulaitteelle tarkistuspistetiedon (checkpoint data) avulla. Checkpoint manager kirjoittaa myös quorumin lokiin tiedot klusteriresurssien käynnistämisistä ja sammuttamisista. Communications manager huolehtii solmulaitteiden välisestä klusteripalvelun viestinnästä heartbeat-signaalin avulla sekä tiedottaa klusterin solmulaitteille klusteriresurssien käynnistämisistä ja sammuttamisista. Configuration database manager tallentaa klusterin asetukset konfiguraatiotietokantaan solmulaitteiden rekistereihin sekä valvoo konfiguraatiotietokannan eheyttä. Event processor alustaa klusteripalvelun ja välittää tapahtumasignaalit (event signals) solmulaitteille. Tapahtumasignaalit kertovat klusterin tapahtumista, esimerkiksi statustietojen muutoksista ja klusteriresurssien käynnistymisistä [MSP01, s. 10].

Event log manager replikoi tapahtumalokin (event log) tapahtumat solmulaitteiden välillä. Failover manager vastaa klusteriresurssien failover-operaatiosta ja määrää mille solmulaitteille klusteriresurssit siirretään. Jos solmulaitteiden failover managerit eivät pysty kommunikoimaan toistensa kanssa, quorum-resurssin omistaja kaappaa päätösvallan. Global update manager päivittää klusterin tilatiedot solmulaitteiden välillä. Log manager kirjoittaa klusterin transaktiolokia quorum.log-tiedostoon. Membership manager ylläpitää listaa klusterin jäsenistä [MSP01, s. 11].

Node manager jakaa resurssiryhmien hallinnan solmulaitteille. Object manager pitää keskusmuistissa tietokantaa klusterin objekteista. Resource manager vastaa resurssiryhmistä ja klusteriresursseista. Resurssimonitorit toimivat kommunikointikanavana klusteripalvelun ja sovelluskohtaisten resurssi-DLL-kirjastojen välillä. Käyttöjärjestelmän mukana tulevat resurssi-DLL-kirjastot Microsoftin omille klusteriresursseille. Kun kolmannet osapuolet haluavat tehdä klusteritietoisia sovelluksia toimimaan klusteripalvelun päällä, joutuvat kyseiset tahot kehittämään omat resurssimonitorit ja resurssi-DLL-kirjastot [MSP01, s. 12 - 13].

Klusteripalvelu tutkii LooksAlive-tarkistuksella (LooksAlive check) näyttääkö resurssi olevan toimintakunnossa. LooksAlive-tarkistus on nopea ja yksinkertainen operaatio, jolla resurssimonitori arvioi resurssin toimintaa. Jos LooksAlive-tarkistus epäonnistuu, tehdään seuraavaksi sitä perusteellisempi IsAlive-tarkistus (IsAlive check). Jos IsAlive-tarkistuskin epäonnistuu, sammutetaan kyseinen resurssiryhmä ja sille teh-

dään failover-operaatio. LooksAlive- ja IsAlive-tarkistusten aikavälit ovat säädettävissä resurssikohtaisesti [MSP01, s. 109]. Klusteripalvelun failover-operaation nopeus on pienellä resurssiryhmällä alle sekunnin luokkaa, kun taas suuren tai monimutkaisia klusteriresursseja sisältävän resurssiryhmän failover-operaatio voi kestää minuutteja.

Klusteripalvelu käyttää sisäiseen kommunikointiinsa TCP/IP:n päällä kahta menetelmää: etäproseduurikutsua (RPC, remote procedure call) ja heartbeat-signaalia. Etäproseduurikutsua käytetään solmulaitteiden ohjaamiseen. Klusterin ensimmäisenä käynnistyneen solmulaitteen Node manager valvoo muiden solmulaitteiden päällä oloa heartbeat-signaalilla, johon muiden solmulaitteiden on vastattava. Heartbeat-signaalia lähetetään 0,5 sekunnin välein [MSP01, s. 13 - 14].

3.3 Laite- ja infrastruktuurivaatimukset

Microsoft edellyttää, että klusteripalvelua käyttävä laitteisto löytyy Microsoftin omalla laitteiston yhteensopivuuslistalta (HCL, Hardware Compatibility List) [Chr03, s. 7]. Microsoftin klusteripalvelulla on lisäksi seuraavia rajoituksia laitteiston suhteen. Jaetulle levyjärjestelmälle täytyy olla solmupalvelimessa oma levyohjain, joka on eri kuin käyttöjärjestelmän käynnistävän levyn levyohjain [MSP01, s. 23]. IDE-levyjä ei voi käyttää, koska IDE-teknologia ei tue levyjen jakamista eri palvelinten kesken [MSP01, s. 21]. Täten Microsoft on sulkenut pois IDE-levyjen käytön klusteripalvelussa. SCSI-Y-kaapeliratkaisussa voidaan käyttää maksimissaan kahta solmulaitetta [MSP01, s. 22]. Useamman solmulaitteen ympäristössä tarvitaan kallis kuitupohjainen SAN-levyjärjestelmä. Quorum suositellaan perustettavaksi omalle levylle, jolloin kyseisellä levyllä ei olisi muita klusteriresursseja [MSP01, s. 48]. Dynaamiset levyt (dynamic disk) eivät ole tuettuja, jolloin myöskään Microsoftin ohjelmistopohjaista RAID-järjestelmää ei voi käyttää. Osioiden kryptaaminen EFS:llä (encryption file system) ei ole tuettu Windows 2000:ssa [MSP01, s. 45], mutta on tuettu Windows 2003:ssa [MS03a, s. 15].

Verkkokortteja tarvitaan kaksi per solmulaite [MSP01, s. 23]. Klusteripalvelun saa asennettua yhdelläkin verkkokortilla, mutta Microsoft ei tue yhdellä verkkokortilla

toimivia klustereita [MSP01, s. 24]. Microsoft neuvoo käyttämään toista verkkokorttia vain klusterin omaan sisäiseen kommunikointiin. Toinen verkkokortti on asiakkaiden palvelua varten sekä varalla klusterin sisäistä kommunikointia varten.

Klusteripalvelu vaatii toimiakseen Microsoftin Active Directoryn [MSP01, s. 24] ja kaikkien solmulaitteiden on kuuluttava samaan toimialueeseen (domain) [MSP01, s. 7]. Active Directoryn tilalla voidaan käyttää Active Directorystä muokattua kevyempää domainlet-järjestelmää [MSP01, s. 24]. Domainlet toteutetaan asentamalla solmulaitteet toimialueen ohjainpalvelimiksi (domain controller) ja ottamalla pois käytöstä klusterin kannalta ylimääräiset palvelut, kuten global catalog -palvelu. Microsoftin Exchange-postipalvelun klusterointi ei ole tuettu domainlet-järjestelmässä [MSKB07], koska Exchange tarvitsee global catalog -palvelua. Käytännössä klusteripalvelut toteutetaan Active Directoryyn tukeutuen, eikä domainlet-vaihtoehtoa juuri käytetä. Active Directoryn käyttö nostaa kustannuksia, kun redundanssitarpeen vuoksi toimialueen ohjauskoneita tarvitaan vähintään kaksi kappaletta. Käytännössä dedikoituja toimialueen ohjauskoneita on kolme tai enemmän per toimialue.

Klusteripalvelun maantieteellisesti hajautettu konfiguraatio, Majority Node Set (MSN), ei tarvitse jaettua levyä lainkaan, vaan quorumista on kopio kaikkien solmulaitteiden paikallisilla levyillä [MS03a, s. 7]. Siten Majority Node Set -klusterin päätösvaltaa ei voida ratkaista tavallisen klusteripalvelun tapaan lukitsemalla quorumin levy, vaan päätösvalta muodostuu solmulaitteiden enemmistöstä. Majority Node Set klusterin klusteripalvelu käynnistyy vain, kun enemmistö solmulaitteista näkee toisensa. Vähemmistössä olevat solmulaitteet sammuttavat automaattisesti klusteripalvelunsa [MS04, s. 14]. Majority Node Set -ratkaisun haittapuolena klusteriresurssien datoja ei kopioida klusteripalvelun puolesta, vaan järjestelmän rakentajan on itse ratkaistava datan replikointi solmulaitteiden välillä [MS03a, s. 7].

3.4 Klusteriresurssit, resurssiryhmät ja resurssien parametrit

Microsoft ei vaivaudu mainostamaan mitkä kaikki muut sovellustoimittajat ovat tehneet sovelluksistaan klusteritietoisia Microsoftin klusteripalvelun suhteen. Käytännös-

sä Microsoftin klusteripalvelulle sovellusten määrä on erittäin laaja. Jos Windows-alustalle tehdään klusteritietoinen sovellus, on Microsoftin klusteripalvelu yleensä mukana.

Klusteripalvelun resurssit organisoidaan resurssiryhmiin. Klusteriresurssityypit asettavat rajat mitä sovelluksia ja palveluja klusteri voi tarjota. Jokainen resurssiryhmä muodostaa oman virtuaalipalvelimen. Kullakin resurssiryhmällä on klusteriresursseina ainakin IP-osoite, verkkonimi (network name) ja levy (physical disk) [MSP01, s. 32]. Muut tarjolla olevat resurssityypit ovat Windows 2000:ssa DHCP-palvelin, WINS-palvelin, tulostusjono (print spooler), tiedostojako (file share), yleinen sovellus (generic application), yleinen palvelu (generic service), Internet-palvelimen instanssi (IIS server instance), viestin välitys (message queuing) ja hajautettu transaktioiden käsittelijä (DTC, distributed transaction coordinator) [MSP01, s. 98 - 99]. Tiedostojaoille on valittavissa klusteripalvelun toimesta erikoisominaisuutena alihakemistojen automaattinen jako. Tämän avulla esimerkiksi käyttäjien kotilevyt voidaan jakaa omina erillisinä tiedostojakoina käyttäen kuitenkin vain yhtä klusteroitua tiedostojakoa [MSP01, s. 98]. Windows 2003:n mukana tuli uusia resurssityyppejä: yleinen skripti (generic script), Majority Node Set ja ”volume shadow copy service task”.

Yleinen sovellus, yleinen palvelu ja yleinen skripti ovat klusteriresursseja, joiden avulla voidaan klusteroida sovelluksia, jotka eivät ole klusteritietoisia [MSP01, s. 97]. Yleinen sovellus, yleinen palvelu ja yleinen skripti eivät tarvitse sovelluskohtaisia resurssimonitoreita ja resurssi-DLL-kirjastoja, vaan ne käyttävät klusteripalvelun oletusresurssimonitoria ja oletusresurssi-DLL-kirjastoa [MSP01, s. 13]. Sovellusten, jotka eivät ole klusteritietoisia, klusteroiminen ja käyttö klusterissa voi aiheuttaa vakausongelmia sekä kyseisessä sovelluksessa että koko klusterissa.

Kokoamalla erityyppisiä klusteriresursseja voidaan klusteroida laajempia järjestelmiä. Esimerkiksi Microsoftin sähköpostipalvelin Exchange ja tietokantapalvelin SQL server klusteroidaan joukolla klusteriresursseja. Exchangen ja SQL serverin asennusohjelmat ymmärtävät asennuksen tapahtuvan klusterin solmulaitteella, jolloin asennus suoritetaan klusteriasennuksena ja palvelut asentuvat klusteriresursseiksi automaattisesti.

Exchange 2000:n klusteriasennuksessa on muutamia rajoitteita, joita tavallisessa Windows-asennuksessa ei ole. NNTP-protokolla, avaintenhallintapalvelu (key management service) ja pikaviestipalvelu (instant messaging service) eivät ole tuettuja [MSP01, s. 248]. Klusteriasennus tukee maksimissaan neljää tietokantaryhmää (storage group) per solmulaite [MSP01, s. 249]. SQL server 2000:n klusteriasennuksen rajoitteena on 16 instanssin maksimimäärä [MSP01, s. 272]. Käytännössä SQL serveriin asennetaan vain oletusinstanssi tai korkeintaan muutama instanssi, joten 16 instanssin raja tulee harvoin vastaan.

Resurssiryhmille on muutamia attribuutteja, joilla voidaan säädellä ryhmän käyttäytymistä tai jotka kertovat resurssiryhmän tilasta. Kaikille resurssiryhmille yleiset resurssiryhmän attribuutit ovat taulukossa 4 [MSP01, s. 370]. Lisäksi cluster-resurssiryhmällä on omia attribuutteja, jotka kohdistuvat koko klusteriin.

attribuutti	kuvaus
Description	Kuvaus resurssiryhmästä.
PersistentState	Kertoo resurssiryhmän tilan (tosi = online, epätosi = offline).
FailoverThreshold	Kertoo kuinka monta kertaa klusteripalvelu yrittää tehdä failover-operaation, ennen kuin luovuttaa jättäen ryhmän resursseja offline-tilaan.
FailoverPeriod	Määrittelee failover-operaatioiden välisen ajan.
AutoFailbackType	Sallii tai kieltää failback-operaation (0 = kielletty, 1 = sallittu).
FailbackWindowStart	Asettaa failback-operaation aloittamisen kellonajan.
FailbackWindowEnd	Asettaa failback-operaation lopettamisen kellonajan.

Taulukko 4. Yleiset resurssiryhmän attribuutit [MSP01, s. 370].

Myös klusteriresursseilla on attribuutteja, joilla voidaan säätää resurssin käyttäytymistä klusterissa. Yleiset klusteriresurssin attribuutit ovat taulukossa 5 [MSP01, s. 372]. Osalla klusteriresursseista on lisäksi omia resurssityyppikohtaisia attribuutteja.

attribuutti	kuvaus
Description	Kuvaus resurssista.
DebugPrefix	Määrittelee resurssille määrätyn debuggerin.
SeparateMonitor	Kuvaa jakaako resurssi resurssimonitorin jonkin toisen resurssin kanssa (tosi tai epätosi).
PersistentState	Kertoo resurssin tilan (tosi = online, epätosi = offline).
LooksAlivePollInterval	Määrittelee aikavälin, jolla tutkitaan näyttääkö resurssi toimivan.
IsAlivePollInterval	Määrittelee aikavälin, jolla tutkitaan toimiiko resurssi.
RestartAction	Kuvaa klusteripalvelun toiminnan, jos resurssi todetaan toimimattomaksi. Vaihtoehdot ovat: ClusterResourceDontRestart (0) ClusterResourceRestartNoNotify (1) ClusterResourceRestartNotify (2)
RestartThreshold	Määrittelee kuinka monta kertaa klusteripalvelu yrittää käynnistää resurssin, ennen kuin klusteripalvelu aloittaa failover-operaation.
RestartPeriod	Määrittelee ajan, jonka kuluessa klusteripalvelu saa yrittää käynnistää resurssin, ennen kuin klusteripalvelu aloittaa failover-operaation.
PendingTimeout	Määrittelee ajan, jonka kuluessa klusteripalvelu päättää resurssin olevan offline- tai failed-tilassa.

Taulukko 5. Yleiset klusteriresurssin attribuutit [MSP01, s. 372].

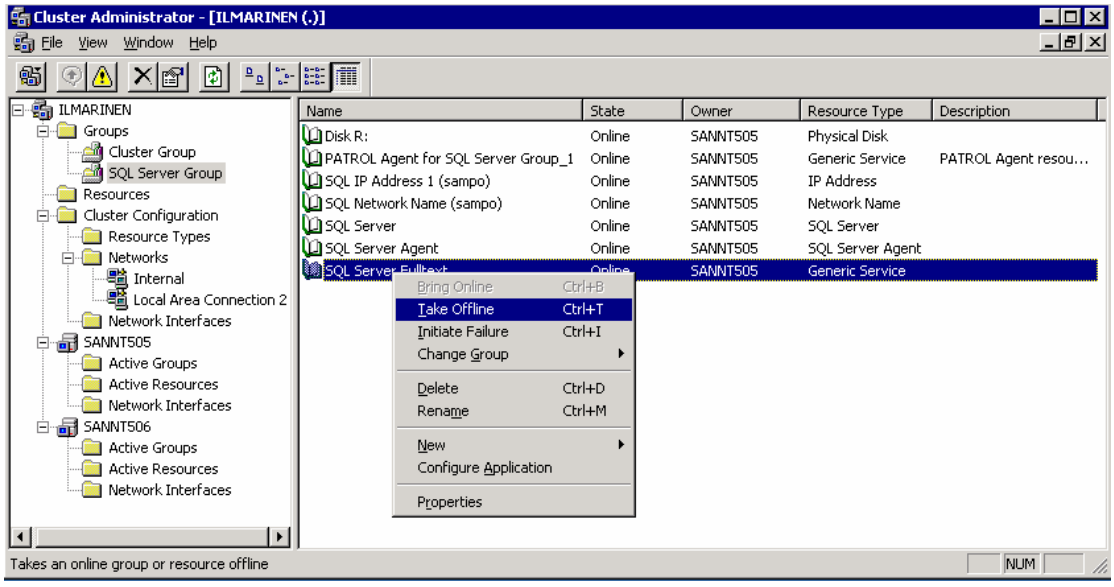
3.5 Asennus- ja hallintatyökalut

Klusteripalvelun asennukseen ja ylläpitoon Microsoftilla on sekä komentorivipohjaisia että graafisen käyttöliittymän omaavia työkaluja. Klusteripalvelun graafisia työkaluja pidetään helppokäyttöisinä ja havainnollisina [VDB98]. Komentorivipohjaisilla klusteripalvelun työkaluilla voi tehdä samat operaatiot kuin graafisilla työkaluilla muutamaa poikkeusta lukuun ottamatta, joten on lähinnä makuasia kumpia työkaluja käyttää. Komentorivipohjaisten työkalujen etuna on mahdollisuus automatisoida operaatioita tekemällä skriptejä. Esimerkiksi sadan klusteroidun levynjaon perustaminen skriptillä on tehokkaampaa kuin graafisen käyttöliittymän kautta monivaiheisesti suoritettuna puhumattakaan jälkimmäisen tavan puuduttavasta vaikutuksesta ylläpitäjään.

Klusteripalvelun voi asentaa komentoriviltä cluscfg.exe-työkalulla. Cluscfg.exe tukee vastaustiedoston käyttöä asennuksessa, jossa vastaustiedostoon muokataan valmiiksi klusterin asennusparametrit [MSP01, s. 57]. Vastaustiedostoa hyödynnetään lähinnä

massa-asennuksessa, kun asennetaan useita samanlaisia klustereita. Jos cluscfg.exe käynnistetään ilman parametreja, tapahtuu klusteripalvelun asennus graafisen velhon (wizard) avustuksella [MSP01, s. 54]. Saman velhon voi käynnistää myös ohjauspaneelin lisää/poista sovellus -kohdasta Windowsin komponentit.

Klusteripalvelun ylläpitäjän tärkein graafinen työkalu on cluster administrator (cluadmin.exe) [MSP01, s. 86], jolla hallinnoidaan klusteria. Työkalu näyttää havainnollisesti solmulaitteiden tilat ja klusteriresurssit, jolloin ylläpitäjä hahmottaa nopeasti symboleista mikä on klusterin tila. Erikoisuutena cluster administratorilla voidaan simuloida resurssien vikatilanteita, joihin reagoimista klusterin käyttäytymistä voidaan tutkia [MSP01, s. 89]. Cluster administratorin käyttöliittymästä on kuva 6. Kyseissä esimerkkiklusterissa on klusteroitu Microsoft SQL server 2005 Enterprise edition sp1 ja käyttöjärjestelmänä on Windows server 2003 R2 Enterprise edition sp1. Esimerkkiklusterin nimi on Ilmarinen ja klusteriin kuuluvat solmulaitteet Sannt505 ja Sannt506.



Kuva 6. Cluster administratorin käyttöliittymä Windows server 2003 R2 Enterprise edition sp1 -palvelimella.

Komentorivipohjaisia työkaluja suosiville cluster administratorin vastaavia toimintoja tarjoaa cluster.exe [MSP01, s. 91]. Cluster.exe ilman parametreja näyttää klusterin statuksen, mutta operointi ja erityisesti konfigurointimuutokset vaativat lukuisia parametreja, jolloin komennon pituus ylittää helposti komentokehotteen yhden rivin 80

merkkiä. Cluster.exe:n käyttö vaatiikin opasteiden tai manuaalin runsasta selaamista. Esimerkiksi seuraava komento tekee valitusta levystä klusteriresurssin [MSP01, s. 139]:

```
CLUSTER mycluster RESOURCE mydisk /Create /Group:mygroup
/Type:"Physical Disk"
```

Vaikka klusteripalvelun asennukseen ja operointiin on helppokäyttöiset graafiset työkalut, vikatilanteista toipuminen vaatii pahimmillaan monivaiheista operointia eri työkaluilla. Klusterin asetukset voidaan palauttaa ntbackup.exe:n ”system state backup”-palautuksella, jolloin HKEY_LOCAL_MACHINE\Cluster-haaran tiedot palautetaan rekisteriin [MSP01, s. 202]. Klusteripalvelua voi yrittää käynnistää erilaisilla korjausparametreilla, joita ovat esimerkiksi fixquorum [MSP01, s. 203] ja noquorumlogging [MSP01, s. 238]. Quorumin datan palautusta varten on oma työkalu clusrest.exe [MSP01, s. 240]. Diagnoosia varten ylläpitäjä käy läpi käyttöjärjestelmän tapahtumalokia apunaan Microsoftin Knowledge Base -tietokannan artikkelit. Tapahtumalokia tarkemmin klusteripalvelu kerää lokia cluster.log-tiedostoon [MSP01, s. 240], jonka sisältö on vaikeasti tulkittavaa. Mahdollisten ongelmien varalta ylläpitäjän kannattaa tehdä tukisopimus Microsoftin Premier Support -palveluun, jolloin asiantuntija-apua on saatavissa tarvittaessa nopeasti klusteriongelmiin. Ilman tukisopimusta ylläpitäjä voi jäädä yksin ongelmiensa kanssa.

4 TruCluster Server for Tru64 UNIX

4.1 Versiot

HP TruCluster Server on Tru64 UNIX -käyttöjärjestelmän päälle asennettava korkean käytettävyyden klusteri. TruCluster ja Tru64 UNIX vaativat molemmat omat lisenssit. TruClusterin versionumero seuraa UNIX:in versionumeroa. Siis HP Tru64 UNIX Version 5.1B-3 päälle asennetaan HP TruCluster Server Version 5.1B-3 [HP05]. Viimeisin TruCluster Server Version 5.1B tukee yhdestä kahdeksaan solmulaitetta

[Tru02a]. TruClusterista ovat ilmestyneet taulukossa 6 luetellut versiot [HP06 ja CGS96].

tuote	julkaisuajankohta
TruCluster Server Version 5.1B	marraskuu 2002
TruCluster Server Version 5.1A	syyskuu 2001
TruCluster Server Version 5.1	lokakuu 2000
TruCluster Server Version 5.0A	huhtikuu 2000
TruCluster Server Version 5.0	heinäkuu 1999
TruCluster Software Products Version 1.6	huhtikuu 1999
TruCluster Software Products Version 1.5	tammikuu 1998
TruCluster version 1.0	1996

Taulukko 6. TruCluster Server versiot [HP06 ja CGS96].

TruClusterin ja Tru64 UNIX -käyttöjärjestelmän kehitysinto tuntuu laantuneen Compaqin ja HP:n vuonna 2001 tapahtuneen fuusion myötä. HP:n epäiltiin lopettavan Tru64 UNIX -käyttöjärjestelmän kehitystyön [Wri02] ja kehittävän jatkossa vain HP-UX:ia, mutta HP on kuitenkin ilmoittanut jatkavansa kehitystyötä sekä tukea ainakin vuoteen 2012 saakka [Tru07]. Uusia versioita TruClusterista ei ole kuitenkaan ilmestynyt viiteen vuoteen. Markkinoinnissaan HP kuitenkin suositteleekin Tru64 UNIX:n vaihtamista jo HP-UX:iin, johon HP kehuu siirtäneensä Tru64 UNIX:n parhaat ominaisuudet [HP07a].

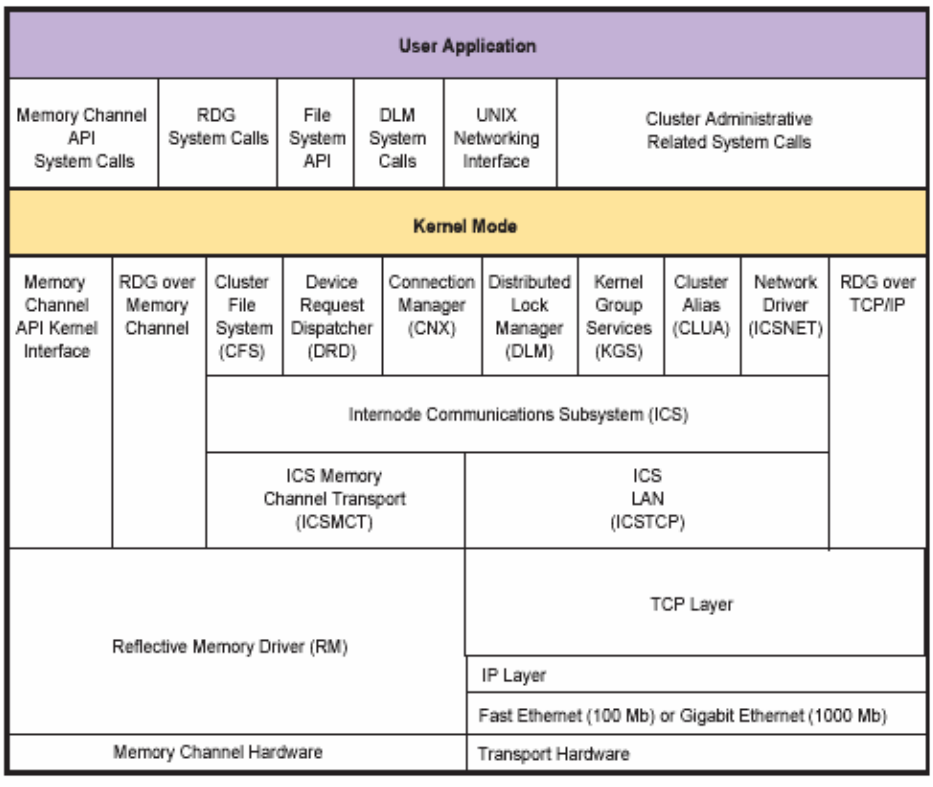
Vuonna 2002 TruCluster Server Version 5.1A -lisenssin hintahaarukka oli 3000 - 48000 $ per solmulaite riippuen palvelinmallista [Wri02].

4.2 Arkkitehtuuri

TruClusterin päätoiminnot tarjoava Cluster Executive koostuu kolmesta komponentista, jotka ovat Connection Manager (CNX), hajautettu lukkojenhallintaohjelmisto (Distributed Lock Manager, DLM) ja Kernel Group Services (KGS) [HP04, s. 4]. Connection Manager pitää kirjaa klusterin jäsenistä ja valvoo, että solmulaitteet kykenevät

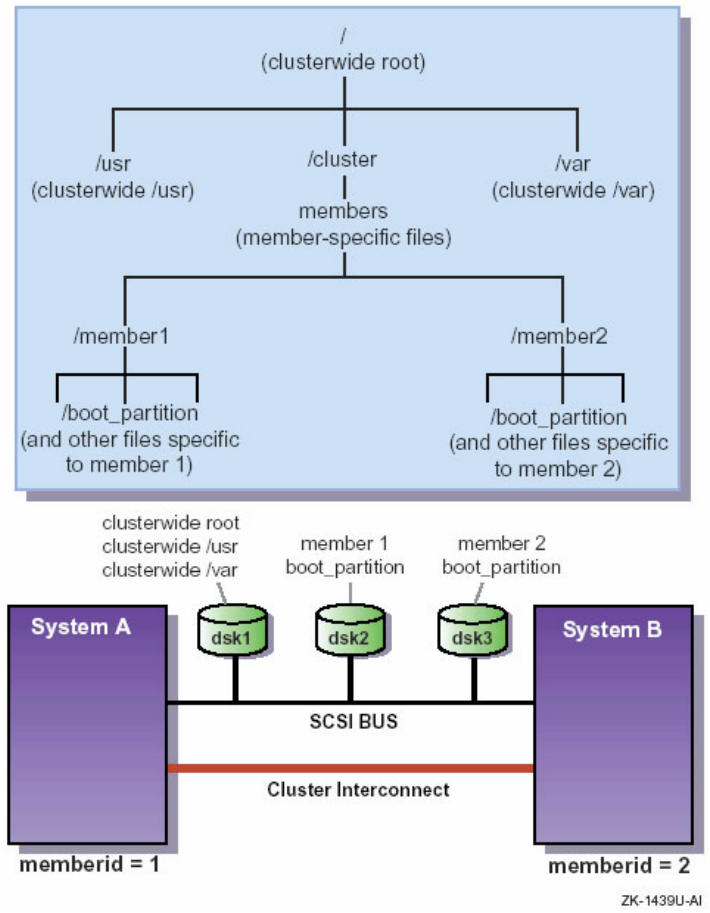
kommunikoimaan keskenään. Connection Manager muodostaa klusterin, lisää jäseniä, poistaa jäseniä sekä kirjoittaa Event Managerin välityksellä lokiin tietoja klusterin statusmuutoksista. Klusteria muodostaessaan Connection Manager laskee yhteen niiden solmulaitteiden äänet (vote), joihin se on yhteydessä, sekä näistä solmulaitteista jonkin mahdollisesti lukitseman quorum-levyn äänen. Kullakin solmulaitteella sekä quorum-levyllä on oletusarvoisesti yksi ääni. Connection Manager muodostaa klusterin vain solmulaitteiden omistaessa äänten enemmistön [Tru02a, s. 3-1 ja 3-2].

Palvelujen virtuaalisoinnissa auttaa Cluster Alias -alijärjestelmä (CLUA), jonka avulla klusteri näyttäytyy asiakkaille yhdellä IP-osoitteella per klusterialias (cluster alias) [HP04, s. 5]. Klusterialias vastaa siis muiden klusterituotteiden virtuaalipalvelinta. Ensimmäisen solmulaitteen asennuksen myötä TruCluster muodostaa oletusklusterialiaksen (default cluster alias), mutta klusterialiaksia voi tehdä lisää tarpeen mukaan [Tru02a, s. 6-5]. Klusterialiasten maksimilukumäärä on oletuksena kahdeksan, mutta max_aliasid-attribuuttia muuttamalla maksimilukumäärä voidaan korottaa arvoon 102 400 [Tru02a, s. 6-6]. Käytännössä klusterialiaksia tarvitaan vain muutama. TruClusterin arkkitehtuuria esittää kuva 7 [HP04, s. 4].



Kuva 7. TruCluster Server Version 5.1A -arkkitehtuuri [HP04, s. 4].

Klusterin tiedostojärjestelmää pitää yllä cluster file system (CFS), jonka ansiosta tiedostojärjestelmä näyttäytyy asiakkaille yhtenä kokonaisuutena [HP04, s. 5]. Cluster file system -järjestelmää esittää kuva 8 [Tru02a, s. 2-7]. Kuvasta näkee klusterin juuressa olevat klusterin yhteiset hakemistot, /cluster-hakemiston alla olevat solmulaitteiden omat hakemistorakenteet sekä hakemistojen sijoittumisen fyysisille levyille. Kuvassa 8 näkyy myös cluster interconnect -väylä, jota käytetään klusterin sisäiseen viestintään. Cluster interconnect -väylästä kerrotaan lisää luvussa ”Laite- ja infrastruktuurivaatimukset”.



Kuva 8. TruCluster Server Version 5.1B cluster file system [Tru02a, s. 2-7].

TruClusterissa klusteroitavat sovellukset jaetaan kolmeen ryhmään: yhden instanssin sovellus (single-instance application), monen instanssin sovellus (multi-instance application) ja hajautettu sovellus (distributed application). Yhden instanssin sovellusta ajetaan kerrallaan vain yhdessä solmulaitteessa. Yhden instanssin sovelluksen tilaa klusterissa valvoo Cluster Application Availability -alijärjestelmä (CAA) [Tru02a, s.

4-1]. Yhden instanssin sovellus ei ole klusteritietoinen. Se, että yksittäistä yhden instanssin sovellusta voi ajaa vain yhtä koko klusterissa, on selkeä puute, sillä kuormantasauksen kannalta olisi hyödyllistä kyetä ajamaan sovellusta samanaikaisesti usealla solmulaitteella.

Monen instanssin sovellusta voidaan ajaa samanaikaisesti useammalla solmulaitteella. Monen instanssin sovellus on klusteritietoinen ja itsessään vikasietoinen, koska yhden instanssin vikatilanne ei vaikuta muihin instansseihin [Tru02a, s. 4-1]. Monen instanssin sovellus tarvitsee klusterilta vain tiedonsiirron reitityspalvelut Cluster Alias alijärjestelmän kautta [Tru02e, s. 2-2].

Hajautettu sovellus on alusta lähtien suunniteltu toimimaan TruClusterissa ja käyttää Memory Channel -pohjaista cluster interconnect -väylää, hajautettua lukkojenhallintaohjelmistoa ja Cluster Alias -alijärjestelmää [Tru02a, s. 4-1]. Memory Channel pohjaisesta cluster interconnect -väylästä kerrotaan lisää seuraavassa luvussa.

4.3 Laite- ja infrastruktuurivaatimukset

TruCluster vaatii alleen HP:n valmistaman Alphaserver-palvelinlaitteiston ja SAN-levyjärjestelmän. HP:n spesifikaatiot [HP05] määrittelevät tarkalleen mitkä versiot Alphaserver-palvelimista ja yksittäisistä komponenteista kelpaavat TruClusterille. TruCluster vaatii vähintään 64 megatavua keskusmuistia ja lisäksi käyttöjärjestelmä oman minimimääränsä muistia [Tru02b, s. 1-7].

Kaikki solmulaitteet yhdistäväksi cluster interconnect -väyläksi voi valita HP:n oman Memory Channel -pohjaisen tai edullisemman LAN-pohjaisen. Memory Channel kytkeytyy suoraan solmulaitteiden keskusmuistin väylälle (memory bus). TruClusterin ensimmäisissä versioissa Memory Channel oli ainoa vaihtoehto ja LAN-vaihtoehto tuli tarjolle myöhemmin. Memory Channel rakentuu solmulaitteiden PCI-väylälle kytkettävistä Memory Channel -korteista sekä kolmen tai useamman solmulaitteen klusterissa solmulaitteiden väliin tulevasta Memory Channel -keskittimestä. HP on asetta-

nut rajoitteen, että Memory Channel korttien maksimietäisyys toisistaan saa olla enintään 6 km [HP04, s. 11].

LAN-pohjainen cluster interconnect -väylä toteutetaan verkkokorteilla ja kahden solmulaitteen klusterissa ristiinkytketyllä parikaapelilla tai useamman solmulaitteen klusterissa tavallisella lähiverkon kytkimellä. Hyppyjen määrä on rajoitettu kolmeen, joten solmulaitteiden välillä voi olla maksimissaan kaksi kytkintä, mikä rajoittaa solmulaitteiden etäisyyttä toisistaan [HP04, s. 11]. Molempien cluster interconnect väylien etäisyysrajoitukset tarkoittavat, ettei TruCluster sovellu kaupunkialueklusteriksi eikä mantereiseksi klusteriksi.

Ethernet-teknologian kehittyminen on kaventanut Memory Channel -toteutuksen ylivoimaa suorituskyvyn suhteen. HP on omissa tutkimuksissaan osoittanut, että 1 gigabitin Ethernet-verkkokortilla päästään jo samaan suorituskykyyn kuin Memory Channel -toteutuksella. Kahdeksan solmulaitteen maksimilukumäärärajoitus koskee sekä Memory Channel että LAN-vaihtoehtoa [HP04, s. 11]. Memory Channel -toteutuksen etuna on klusterin nopea reagointi odottamattomiin vikatilanteisiin, kun solmulaitteen katoaminen havaitaan lähes välittömästi. LAN-pohjainen cluster interconnect puolestaan kärsii TCP/IP-protokollan keep alive timer -aikakatkaisumekanismista, joten solmulaitteen katoamisen havaitseminen on hitaampaa [HP04, s. 15].

Kiintolevyjä tarvitaan kahden solmulaitteen klusterissa vähintään neljä kappaletta. Yksi levy tarvitaan dedikoiduksi käyttöjärjestelmälevyksi (Tru64 UNIX disk) [Tru02d, s. 1-3]. Toinen levy tarvitaan klusterin jaetuksi levyksi (clusterwide disk), johon sijoittuvat juuri-, /usr- ja /var-hakemistot. Lisäksi kullekin solmulaitteelle tulevat omat käynnistyslevyt (member boot disk) [Tru02d, s. 1-4]. Minimikokoonpanon lisäksi HP suosittelee quorum-levyn käyttöä, jolloin levyjen lukumäärä on vähintään viisi kappaletta [Tru02b, s. 2–15].

4.4 Klusteriresurssit

Mikä tahansa Tru64 UNIX -käyttöjärjestelmässä toimiva sovellus voidaan klusteroida yhden instanssin sovellukseksi. Yhden instanssin sovelluksen toimintaa klusterissa valvoo ja hallinnoi Cluster Application Availability -alijärjestelmä, eikä klusteroitavaan sovellukseen tarvitse tehdä muutoksia. Tällaiselle sovellukselle on joko valmiina tai tehdään itse resurssiprofiili (resource profile), jossa kerrotaan sovelluksen resurssivaatimukset sekä Action-skriptit sovelluksen käynnistämistä, sammuttamista, valvontaa ja toiselle solmulaitteelle siirtämistä varten [Tru02a, s. 5-1]. Resurssiprofiileja tehdään caa_profile-työkalulla ja niitä voi muokata myös tekstieditorilla /var/cluster/caa/profile -hakemistossa. Käsin muokatut resurssiprofiilit on syytä tarkistuttaa caa_profile-työkalun ”-validate”-vivulla syntaksin oikeellisuuden varmistamiseksi [Tru02a, s. 5-4]. Esimerkkinä resurssiprofiilista on BIND-nimipalvelimen named.cap-niminen resurssiprofiili kuvassa 9 [Tru02a, s. 5-8]. Resurssiprofiilin kaikki attribuutit kuvataan taulukossa 7 [Tru02e, s. 2-4 - 2-6].

```
TYPE = application
NAME = named
DESCRIPTION = BIND Server
CHECK_INTERVAL =
FAILURE_THRESHOLD = 0
FAILURE_INTERVAL = 0
REQUIRED_RESOURCES =
OPTIONAL_RESOURCES =
HOSTING_MEMBERS =
PLACEMENT = balanced
RESTART_ATTEMPTS =
FAILOVER_DELAY =
AUTO_START =
ACTION_SCRIPT = named.scr
```

Kuva 9. Nimipalvelimen named.cap-resurssiprofiili [Tru02a, s. 5-8].

attribuutti	kuvaus
TYPE	Resurssin tyyppi. Vaihtoehdot ovat: application, network, tape tai changer.
NAME	Resurssin nimi.
DESCRIPTION	Resurssin kuvaus.
FAILURE_THRESHOLD	Resurssin epäonnistuneiden käynnistysyritysten lukumäärä, jonka jälkeen CAA-alijärjestelmä merkitsee, ettei resurssi ole saatavilla (unavailable), jolloin resurssi jätetään sammutetuksi (offline).
FAILURE_INTERVAL	Resurssin epäonnistuneiden käynnistysyritysten tarkistuksen aikaväli sekunneissa.
REQUIRED_RESOURCES	Lista resursseista, joista tämä resurssi on riippuvainen.
OPTIONAL_RESOURCES	Lista valinnaisista resursseista. Klusteri käyttää listaa arvioidessaan missä solmulaitteessa resurssi käynnistetään.
PLACEMENT	Määrittelee resurssin sijoituspolitiikan (placement policy), vaihtoehdot ovat: balanced = Käynnistetään sovellus solmulaitteella, jolla on vähiten kuormaa. favored = Valitaan solmulaite HOSTING_MEMBERS-attribuutin listasta ja valintakriteerinä on OPTIONAL_RESOURCES-attribuutti. Jos HOSTING_MEMBERS-listan solmulaitteita ei ole käytettävissä, kelpaa mikä tahansa solmulaite. restricted = Muuten sama kuin ”favored”, mutta jos HOSTING_MEMBERS-listan solmulaitteita ei ole käytettävissä, resurssia ei käynnistetä lainkaan.
HOSTING_MEMBERS	PLACEMENT-attribuutin käyttämä lista solmulaitteista.
RESTART_ATTEMPTS	Resurssin epäonnistuneiden käynnistysyritysten lukumäärä, jonka jälkeen resurssi sijoitetaan toiselle solmulaitteelle.
FAILOVER_DELAY	Käynnistysyritysten aikaväli sekunneissa.
AUTO_START	Käynnistetäänkö resurssi klusterin käynnistyessä. 1= käynnistetään aina, 0 = käynnistetään, jos resurssi oli käynnissä klusteria ajettaessa alas.
ACTION_SCRIPT	Skripti resurssin käynnistykseen, sammutukseen sekä toiminnan tarkistamiseen.
ACTIVE_PLACEMENT	Lippu, jolla määrätään CAA-alijärjestelmä arvioimaan resurssin sijoitus klusterissa.
SCRIPT_TIMEOUT	ACTION_SCRIPT-attribuutin määrittelemän skriptin suorituksen aikakatkaisu sekunneissa.
CHECK_INTERVAL	Action-skriptin toistuvien suoritusten aikaväli sekunneissa.
REBALANCE	Aika, jonka jälkeen klusteri pohtii resurssille optimaalisen sijoituspaikan, siis sopivimman solmulaitteen. Kentän muoto on: ”t:day:hour:min”.

Taulukko 7. Resurssiprofiilin attribuutit [Tru02e, s. 2-4 - 2-6].

Yhden instanssin sovellusten Action-skriptien rakentaminen tyhjästä voi olla työlästä. Action-skriptin tekemisessä kannattaa käyttää pohjana /var/cluster/caa/script hakemiston valmiita skriptejä, joita on valmiina muutamalle sovellukselle [Tru02e, s. 2-16]. HP:n tuotteistamia yhden instanssin sovelluksia ovat DHCP-palvelin [Tru02c, s.7-1] ja BIND-nimipalvelin [Tru02c, s.7-5], joiden Action-skriptit ovat dhcp.scr ja named.scr. Internet Express for Tru64 UNIX -paketin mukana tulevan yhden instanssin OpenLDAP (Lightweight Directory Access Protocol) -hakemistopalvelun Action-skriptin openldap.scr sisältö on esimerkkinä kuvassa 10 [Tru02e, s. 2-34].

```
#!/sbin/sh
#
# Start/stop the OpenLDAP Directory Server.
#
OLPIDFILE=/data/openldap/var/openldap_slapd.pid
OPENLDAP_CAA=1
export OPENLDAP_CAA
case "$1" in
’start’)
/sbin/init.d/openldap start
;;
’stop’)
/sbin/init.d/openldap stop
;;
’check’)
# return non-zero if the service is stopped
  if [ -f "$OLPIDFILE" ]
  then
    MYPID=‘cat $OLPIDFILE‘
    RUNNING=‘/usr/bin/ps -e -p $MYPID -o command | grep slapd‘
  fi
  if [ -z "$RUNNING" ]
  then
  exit 1
  else
  exit 0
  fi
  ;;
*)
echo "usage: $0 {start|stop|check}"
;;
esac
```

Kuva 10. OpenLDAP-hakemistopalvelun Action-skripti openldap.scr [Tru02e, s. 2-34].

TruClusterin sähköposti sendmail on vain SMTP-protokollan osalta klusteritietoinen käyttäen siis cluster alias -alijärjestelmää. Tarvittaessa muita sähköpostiprotokollia (DECnet, MTS, UUCP tai X.25) sähköposti pystytetään yhden instanssin sovelluk-seksi [Tru02c, s. 7-24].

TruCluster tarvitsee toimintaansa ajan synkronointia, jonka vuoksi klusterin jokaiseen solmulaitteeseen asentuu automaattisesti NTP-aikapalvelu [Tru02c, s.7-6]. Muita TruClusterin tukemia monen instanssin klusteritietoisia sovelluksia ovat: NFS-tiedostopalvelu (Network File System) [Tru02c, s.7-7], Internet Server Daemon

(inetd) [Tru02c, s.7-23] ja RIS-asennuspalvelu (Remote Installation Services) [Tru02c, s.7-28]. Tulostus (printer daemon, lpd) konfiguroidaan TruClusterissa klusteroimattoman Tru64 UNIX-palvelimen tapaisesti [Tru02c, s.7-4], joten tulostuspalvelu ei ole klusteroitu.

TruClusterille on saatavilla hajautettuina sovelluksina Oraclen tietokantajärjestelmät Oracle Parallel Server (OPS) ja Oracle 9i Real Application Cluster (RAC) [Tru02e, s. 1-2]. Oraclea voi ajaa myös yhden instanssin sovelluksena [Tru02e, s. 2-37]. Muita sovellustaloja, jotka ovat tehneet sovelluksistaan klusteritietoisia TruClusterin osalta, ovat esimerkiksi Informix, Sybase, BEA/Tuxedo, Baan ja SAP [Wri02].

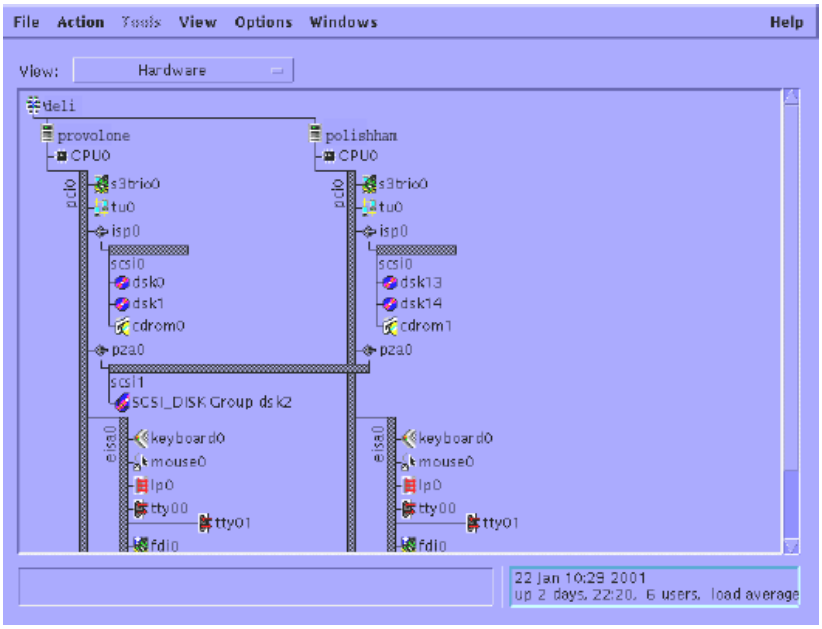
4.5 Asennus- ja hallintatyökalut

HP suosittelee, että solmulaitteille asennetaan kaikki fyysiset komponentit ennen Tru64 UNIX -käyttöjärjestelmän asennusta. Jos komponentti, esimerkiksi verkkokortti, lisätään jälkikäteen, joudutaan ydin kääntämään uudestaan [Tru02b, s. 2-8]. Vertailuksi Microsoftin Windowsissa komponenttien lisääminen on helpompaa, kun käyttöjärjestelmä tunnistaa lisätyt komponentit yleensä automaattisesti.

TruClusterin asennus sisältää muutaman välivaiheen. Asennus aloitetaan kopioimalla Tru64 UNIX -järjestelmän levylle TruCluster-ohjelmisto ja -lisenssi. Klusteri perustetaan clu_create-komennolla, joka luo ensimmäiselle solmulaitteelle käynnistysosion sekä perustaa klusterin jaetut hakemistot (clusterwide root (/), /usr ja /var). Tämän jälkeen käynnistetään ensimmäinen solmulaite käyttäen clu_create-komennon luomaa käynnistysosiota. Solmulaite käynnistää klusterin ollen sen ainoa jäsen ja jakaa klusterin jaetut hakemistot. Klusterin solmulaitteella voidaan lisätä klusteriin lisää jäseniä clu_add_member-komennolla [Tru02a, s. 9-1 ja 9-2].

TruClusterin graafinen hallintatyökalu on SysMan, joka kokoaa yhteen joukon erillisiä työkaluja [Tru02c, s. 2-1]. SysManilla voidaan hallita yksittäisiä solmulaitteita tai koko klusteria [Tru02a, s. 2-5]. Kaikkia klusterin komentoja ei ole liitetty SysManiin, vaan hieman yli puolet kahdestakymmenestä komennosta on suoritettavissa ainoas-

taan komentoriviltä [Tru02c, s. 2-4]. SysManin käyttöliittymää esittää kuva 11 [Tru02a, s. 2-5]. Kuvassa on klusteri nimeltään Deli, jolla on kaksi solmulaitetta: Provolone ja Polishham.



Kuva 11. SysMan-työkalun käyttöliittymä [Tru02a, s. 2-5].

Merkkipohjaisista työkaluista on esimerkkinä kuvassa 12 CAA-alijärjestelmän caa_stat, joka näyttää yhden instanssin sovellusten tilat [Tru02c, s. 8-8]. Kuvasta näkee, että cluster_lockd-sovellus on päällä solmulaitteella Provolone. Sen sijaan DHCP-palvelin ja nimipalvelin ovat offline-tilassa. Solmulaitteella Polishham on ongelmia, kun laite on päällä, mutta verkkoliityntä on offline-tilassa.

```
# caa_stat -v -t
Name Type R/RA F/FT Target State Host Rebalance
-------------------------------------------------------------------
----------------
cluster_lockd application 0/30 0/0 ONLINE ONLINE provolone
dhcp application 0/1 0/0 OFFLINE OFFLINE
named application 0/1 0/0 OFFLINE OFFLINE
ln0 network 0/5 ONLINE ONLINE provolone
ln0 network 1/5 ONLINE OFFLINE polishham
```

Kuva 12. caa_stat-työkalun näkymä [Tru02c, s. 8-8]

TruClusterin solmulaitteen pysäytysoperaatio eroaa tavallisen Tru64 UNIX palvelimen tavasta. Jos ylläpitäjä pysäyttää (halt) solmulaitteen, jolla on ratkaiseva ääni, kaatuu koko klusteri [Tru02c, s. 5-6]. Ennen solmulaitteen pysäytystä klusterin ylläpitäjän täytyy selvittää clu_quorum-komennolla onko solmulaitteella ratkaiseva ääni ja onko solmulaitteella mahdollisesti käynnissä klusteriresursseja, joiden suorittaminen on rajoitettu vain kyseiselle solmulaitteelle. Jos solmulaitteella on ratkaiseva ääni, täytyy ennen pysäytysoperaatiota saada tilanne muutettua esimerkiksi käynnistämällä mahdollinen alhaalla oleva solmulaite, poistamalla ääni solmulaitteelta tai lisäämällä quorum-levylle ääni [Tru02c, s. 5-7].

TruClusterin vikatilanteissa diagnoosi tehdään analysoimalla lokeja. Event manager kirjoittaa lokiinsa tapahtumia koko klusterin näkökulmasta. Klusterin lokitiedostoon /var/cluster/members/{solmulaitteen nimi}/adm/syslog.dated/{päiväys}/daemon.log kirjoitetaan tapahtumat yksittäisen solmulaitteen näkökulmasta. Molempia lokeja on syytä tutkia, kun alijärjestelmät, esimerkiksi CAA-alijärjestelmä (Cluster Application Availability subsystem) kirjoittaa näihin kahteen lokiin osittain erityyppisiä tietoja [Tru02c, s. 8-28]. Vikatilanteissa auttaa HP:n huoltohenkilökunta ja voimassa oleva huoltosopimus mahdollistaa nopean reagoinnin tukipyyntöihin.

5 Steeleye Lifekeeper for Windows

5.1 Versiot

Lifekeeperin alkuperäinen kehittäjä on AT&T, joka toi tuotteen UNIX-järjestelmille vuonna 1992. Steeleye Technology Inc. osti Lifekeeper-tuotteen oikeudet vuonna 1999 [BoC01] ja jatkoi sovelluksen kehittämistä Linux-, Solaris- ja Windows-alustoilla [Ste01]. Steeleye kuitenkin luopui Solaris-version kehittämisestä ja tänä päivänä tarjolla olevat alustat ovat Linux ja Windows [Car05].

LifeKeeperin uusin Windows-versio on LifeKeeper for Windows v6. Windows-alustoista tuetut versiot ovat Windows 2000 server, Windows 2000 advanced server,

Windows 2000 server data center edition sekä Windows server 2003 web edition, Windows server 2003 standard edition, Windows server 2003 enterprise edition, Windows server 2003 data Center edition 32- ja 64-bittisinä [Ste07a, s. 6]. Microsoftin klusteripalveluun verrattuna Lifekeeperille kelpaavat siis huomattavasti edullisemmat web- ja standard-versiot.

Lifekeeper tukee maksimissaan 32 solmulaitetta [Ste04, s. 5]. Lifekeeper for Windows -lisenssi maksaa 3.000 $ per solmulaite [Mic07]. Lifekeeperin failover-operaation nopeus vastaa palveluiden sammutus- ja käynnistysnopeuksia ollen tyypillisesti sekunteja [Car05].

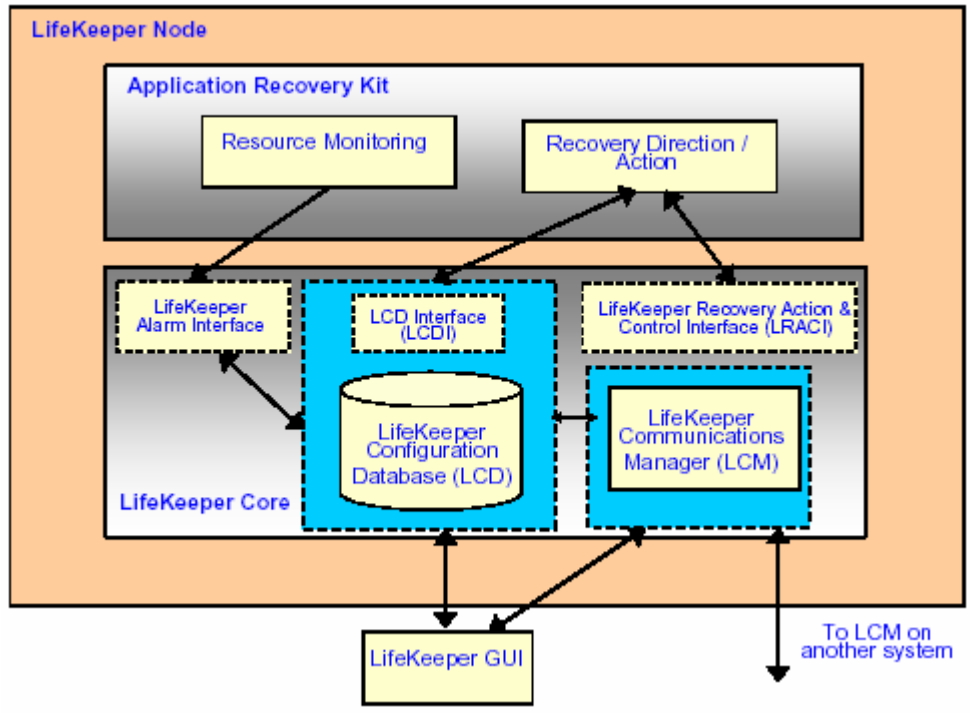
5.2 Arkkitehtuuri

Lifekeeper koostuu kolmesta komponentista, jotka ovat Lifekeeper Core, Application Recovery Kit ja LifeKeeper GUI. Lifekeeper Core on nimensä mukaisesti tuotteen ydin ja tarjoaa klusterin toiminnot. Lifekeeper Core jakaantuu seuraaviin osiin: LifeKeeper Configuration Database (LCD), LCD Interface (LCDI), LifeKeeper Communications Manager (LCM), LifeKeeper Alarm Interface ja LifeKeeper Recovery Action and Control Interface (LRACI) [Ste03, s. 7].

LCD tallentaa staattiset tiedot klusterin asetuksista ja resursseista sekä dynaamiset tiedot klusterin tilasta [Ste07b]. Muista vertailun klusterituotteista poiketen Lifekeeper ei käytä quorum-levyä lainkaan. Quorumin sijaan klusterin tiedot tallennetaan hajautetusti jokaiselle solmulaitteelle erikseen [Ste04, s. 8].

LCDI:tä käytetään rajapintana LCD:n muokkaamisen ja kyselyjen tekemiseen LCD:lle sekä klusteriresurssien tilatietojen tarkistamiseen käyttäen apunaan LifeKeeper Recovery Kit -paketteja [Ste07b]. LifeKeeper Recovery Kit -paketit ovat Lifekeeperin sovellusaluekohtaisia laajennuksia, joiden avulla Lifekeeper kykenee valvomaan ja operoimaan sovellusta klusterissa. Recovery Kit -paketeista kerrotaan lisää kohdassa 5.4 Klusteriresurssit ja Recovery Kit -paketit.

LCM koordinoi heartbeat-liikennettä ja pitää kirjaa klusterin toiminnassa olevista solmulaitteista. LifeKeeper Alarm Interface tarjoaa rajapinnan ja funktiot vikatilanteista toipumista varten. LifeKeeper Recovery Action and Control Interface sisältää liittymän klusteriresurssien käsittelyyn komentoskriptien avulla [Ste07b]. LifeKeeper GUI on client/server-sovellus Lifekeeperin asetusten muokkaamiseen ja hallinnointiin. Työkalusta kerrotaan lisää kohdassa 5.5 Asennus- ja hallintatyökalut. Lifekeeperin arkkitehtuuria esittää kuva 13 [Ste03, s. 7].



Kuva 13. Lifekeeperin arkkitehtuuri [Ste03, s. 7].

5.3 Laite- ja infrastruktuurivaatimukset

Lifekeeperin laitevaatimukset poikkeavat huomattavasti kilpailevista tuotteista, kun Lifekeeper ei vaadi mitään erityisiä laitekomponentteja [Ste04]. Siten Lifekeeper-klusterin voi rakentaa edullisista komponenteista. Steeleye edellyttää vain SCSI-ohjainten ja FCA-korttien (Fibre Channel Adapter) osalta käytettävän Microsoftin laitteiston yhteensopivuuslistalla (HCL) olevia laitteita [Ste06a, s. 11].

Lifekeeper tarjoaa mahdollisuuden käyttää tai olla käyttämättä jaettua levyä. Kilpailevilla järjestelmillä pakollinen SAN-verkko ja -levyjärjestelmä nostavat kustannuksia. Jos jaettua levyä ei käytetä, Lifekeeper replikoi datat solmulaitteiden välillä erillisen ja erikseen ostettavan Lifekeeper Data Replication (LKDR) for Windows -tuotteen avulla. LKDR replikoi verkon yli tiedot lohkotason kopioinnilla, joka ohittaa tiedostojärjestelmän. Steeleye kehuu lohkotason kopioinnin olevan käyttöjärjestelmän suorittamaa kopiointia nopeampaa ja varmempaa, koska tiedostotason lukitusmekanismista ei välitetä, eikä avaus- ja sulkemisoperaatiota tarvitse siis käyttää [Ste04, s. 6].

Heartbeat-kommunikointia varten Lifekeeper tarjoaa kolmea eri siirtotietä, joita voi käyttää useampia yksittäisen vikaantumispisteen poistamiseksi. Suositeltavin tapa välittää heartbeat-signaalia on verkon yli TCP/IP-protokollalla. Heartbeat-signaalia voidaan välittää myös jaetun levyn sekä sarjaportin (RS-232) kautta, mutta näillä molemmilla menetelmillä solmulaitteiden maksimi lukumäärä rajoittuu kahteen [Ste06a , s. 9].

Levyjakoklusteriresurssissa käytettävä LAN Manager recovery kit vaatii NetBIOS-protokollan ajamista [Ste07a, s. 6] ja broadcast-liikenteen vähentämiseksi suosituksena on käyttää nimenselvityksessä WINS-palvelinta (Windows Internet Name Service) [Ste07a, s. 13]. Microsoftin strategiana on päinvastoin päästä kokonaan eroon vanhentuneesta NetBIOS-protokollasta, jolloin nimenselvityksessä käytetään ainoastaan DNS-palvelimia. Steeleyen kannattaisi hankkiutua eroon riippuvuudesta NetBIOS-protokollaan, kun riskinä on, että Microsoft voi luopua koko protokollasta tulevissa Windows-versioissa.

Lifekeeper GUI server edellyttää, että palvelimelle on asennettu Java kehitysympäristö (SDK) tai ajonaikainen käyttöympäristö (JRE). Lifekeeperin asennuksen myötä asennetaan JRE 1.5.0_06. Muiden ja erityisesti uudempien Java-versioiden asentaminen saattaa aiheuttaa ongelmia klusterin hallinnoinnissa, mikäli Lifekeeper GUI server lopettaa toiminnan [Ste07a, s. 8].

Lifekeeperillä on seuraavia rajoituksia. Lifekeeper ei tue ohjelmistopohjaisia RAID-levyjärjestelmiä [Ste07a, s. 12]. Klusteroituja tiedostojakoja voi olla maksimissaan 9999 kappaletta [Ste07a, s. 13]. Käytännössä 9999 jaon raja ei tule vastaan. Steeleye

edellyttää keskusmuistia olevan riittävästi. Jos keskusmuisti ja virtuaalimuisti ovat loppumassa, saattaa se lamauttaa Lifekeeperin toiminnan ja aiheuttaa vääriä statustietoja solmulaitteen tilasta, joka voi johtaa failover-operaatioon [Ste07a, s. 13].

5.4 Klusteriresurssit ja Recovery Kit -paketit

Lifekeeperin jaottelee klusteriresurssit Recovery Kit -paketteihin. Klusteroitavaan sovelluksiin ei tehdä muutoksia, vaan Lifekeeper valvoo ja operoi tuotekohtaisen Recovery Kit -paketin avulla klusteriresursseja [Ste04]. Tuotteen mukana vakiona tuleva Core Recovery Kit sisältää taulukon 8 sisältämät Recovery Kit -paketit [Ste07b].

Recovery Kit	funktio
Volume Recovery Kit	Levyjen ja tiedostojakojen klusterointi.
LAN Manager Recovery Kit	LAN Manager ja NetBIOS -palveluiden klusterointi.
IP Recovery Kit	IP osoitteen klusterointi.
DNS Recovery Kit	DNS-nimipalvelimen host-tietueiden (A) ja pointer-tietueiden (PTR) automaattinen päivitys.
Microsoft IIS Recovery Kit	Internet Information Services (IIS) klusterointi.
Generic Application Recovery Kit	Yleisen sovelluksen klusterointi, eli sovelluksen jolla ei ole valmista Recovery Kit -pakettia.

Taulukko 8. Core Recovery Kit -paketit [Ste07b].

Core Recovery Kit -paketin lisäksi Lifekeeperille on erillisiä ja erikseen palvelinkohtaisesti ostettavia valinnaisia Recovery Kit -paketteja. Valinnaiset Recovery Kit paketit [Ste07a, s. 8] on lueteltu hintoineen [Mic07] taulukossa 9.

Recovery Kit	tuetut sovellusversiot	hinta per solmulaite
Microsoft Exchange Server Recovery Kit v5.3	Exchange 2000 Server ja Exchange Server 2003	3250 $
Microsoft SQL Server Recovery Kit v5.2	SQL server 2000 ja SQL server 2005	960 $
Microsoft Exchange 5.5 Recovery Kit v5.0	Exchange Server 5.5	?
Oracle Recovery Kit v5.3	Oracle 9i R2 ja Oracle 10g R2	900 $
SAP Recovery Kit v5.3	SAP Web Application Server v. 6.40	5400 $
IBM Director Recovery Kit v6.0	IBM Director v. 5.10 ja 5.20	ilmainen [Ste06b]

Taulukko 9. Valinnaiset Recovery Kit -paketit [Ste07a, s. 8] [Mic07] [Ste06b].

Steeleye tarjoaa omille sovelluksille Recovery Kit -pakettien tekemistä varten LifeKeeper Extender -kehitystyökalun [Ste07a, s. 17]. LifeKeeper Extender sisältää dokumentaation sekä malliskriptejä [Ste07c].

Lifekeeperin klusteriresursseilla on neljä kaikille klusteriresurssityypeille yhteistä attribuuttia, jotka ovat Resource Tag, Quickcheck Interval, Deepcheck Interval ja Local Recovery. Attribuutit on kuvattu taulukossa 10 [Ste07b].

attribuutti	kuvaus
Resource Tag	Klusteriresurssin nimi muodossa App.x, jossa x on juokseva numero.
Quickcheck Interval	Klusteriresurssin tilan vakiotarkistuksen aikaväli minuuteissa.
Deepcheck Interval	Klusteriresurssin tilan perinpohjaisen tarkistuksen aikaväli minuuteissa.
Local Recovery	Yritetäänkö klusteriresurssia käynnistää vikatilanteessa samalla solmulaitteella uudestaan.

Taulukko 10. Klusteriresursseille yhteiset attribuutit [Ste07b].

Loput attribuutit täydentävät klusteriresurssin ominaisuuksia ja valikoimat eroavat Recovery Kit -pakettikohtaisesti huomattavasti toisistaan. Esimerkiksi levyjakoklusteriresurssilla on edellisten attribuuttien lisäksi vain ”File Share and Path Name”-attribuutti, jossa on jaon nimi ja hakemistopolku. Sen sijaan yleinen sovellus klusteriresurssilla on yhteisten attribuuttien lisäksi seitsemän muuta attribuuttia, jotka on kuvattu taulukossa 11 [Ste07b].

attribuutti	kuvaus
Restore Script	Klusteriresurssin käynnistävän skriptin hakemistopolku ja skriptin nimi.
Remove Script	Klusteriresurssin sammuttavan skriptin hakemistopolku ja skriptin nimi.
Quick Check Script	Klusteriresurssin tilan vakiotarkistuksen tekevän skriptin hakemistopolku ja skriptin nimi.
Deep Check Script	Klusteriresurssin tilan perinpohjaisen tarkistuksen tekevän skriptin hakemistopolku ja skriptin nimi.
Local Recovery Script	Vikaantuneen klusteriresurssin uudestaan käynnistävän skriptin hakemistopolku ja skriptin nimi.
Application Information	Vapaasti käytettävä attribuutti sovellusta varten.
Bring Resource In Service	Käynnistetäänkö klusteriresurssi. Attribuuttia käytetään lähinnä riippuvuuksien hallintaan.

Taulukko 11. Yleinen sovellus -klusteriresurssin laajennetut attribuutit [Ste07b].

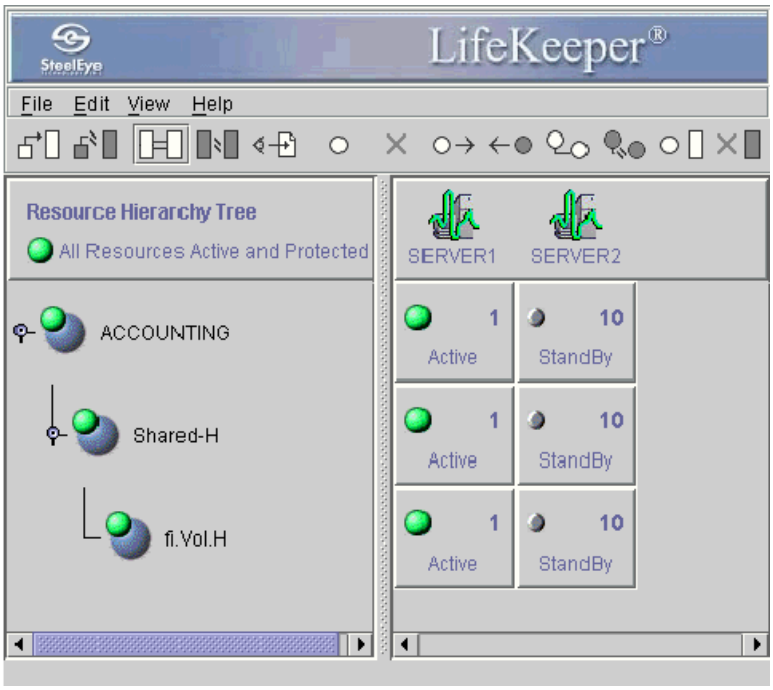
5.5 Asennus- ja hallintatyökalut

Lifekeeper asennetaan InstallShieldillä tehdyn asennusvelhon avulla. Lifekeeper asentuu oletuksen C:\LK-hakemistoon, jonka voi asennuksen yhteydessä muuttaa, mutta vaihdettavan hakemiston nimi ei saa sisältää välilyöntejä ja maksimi pituus on 8 merkkiä. Asennusvelhon asennus on lyhyt. Velho kysyy hakemiston lisäksi asennustavan, jonka vaihtoehtoja ovat ”Typical”, ”Compact” ja ”Custom”, jotka eroavat toisistaan asennettavien komponenttien joukolla. Seuraavaksi kysytään halutaanko kytkeä pois verkkokortin ”media sense”-ominaisuus, joka suositellaan pois kytkettäväksi,

jotta Lifekeeper voi käyttää IP local recovery -toimintoaan. Custom-asennuksessa kysytään vielä käynnistetäänkö Lifekeeperin palvelut asennuksen lopuksi [Ste06a, s. 19].

Asennukseen kuuluu lisenssien hallinnointi. Jokainen solmulaite tarvitsee oman lisenssiavaintiedoston, joka saadaan Steeleyeltä asennuksen luomaa Host ID -tunnusta ja asennusmedian mukana tulevaa valtuutuskoodia (authorization code) vastaan [Ste06a, s. 21]. Host ID on sidottu ensisijaiseen verkkokorttiin. Jos verkkokortti vaihdetaan, täytyy hankkia uusi lisenssiavaintiedosto tai Lifekeeper ei lähde enää käyntiin [Ste06a, s. 24]. Valinnaiset Recovery Kit -paketit tarvitsevat nekin omat solmulaitekohtaiset lisenssiavaintiedostonsa [Ste06a, s. 25].

Lifekeeperin ylläpitoa varten on graafinen työkalu Lifekeeper GUI server, jota voi käyttää etäisesti selaimen kautta [Ste06a, s. 30]. Komennot ovat suoritettavissa myös komentokehotteesta. Osaa Lifekeeperin komennoista ei voida ajaa lainkaan tavallisen terminal server -istunnon kautta, vaan ylläpitäjällä täytyy olla konsoli-istunto [Ste07a, s. 11]. Windows 2003:ssa konsoli-istunto saadaan terminal server -istunnolle ”/console”-parametrilla, mutta Windows 2000:ssa ominaisuuden puuttuessa on käytettävä jotain muuta etähallintatuotetta. Lifekeeper GUI server -työkalun käyttöliittymästä on kuva 14 [Ste06a, s. 31]. Kuvassa näkyvät aktiivitilassa oleva solmulaite Server1 ja passiivitilassa oleva solmulaite Server2. Klusteri jakaa fi.vol.H-levyresurssia, joka on järjestetty klusterissa Accounting- ja Shared-H-hierarkioiden alaisuuteen.



Kuva 14. Lifekeeper GUI server -työkalun käyttöliittymä [Ste06a, s. 31].

Lifekeeperin asennusoperaatiota ja ylläpitoa pidetään helppoina, kun tuotetta oppii ensin käyttämään [Car05]. Testiryhmä ei tosin saanut Lifekeeperin failover-toimintoa toimimaan lainkaan Tietokone-lehdessä 8/2002 [Häm02]. Lifekeeperin tukea tarjoaa Suomessa paikallinen maahantuoja Nordicmind. Puhelintukea on saatavissa arkisin klo 9 - 17. Sähköpostitse ja webbilomakkeella jätetyille tukipyynnöille Nordicmind lupaa kahden tunnin vasteajan 24/365 [Nor07].

6 Sun Cluster

6.1 Versiot

Sun Microsystems toi ensimmäiset korkean käytettävyyden klusterituotteensa markkinoille vuonna 1995 kahden erillisen tuotteen voimin. Ensin valmistui SPARCcluster PDB (parallel database) 1.0 tarjoamaan tietokantapalveluja [BDV01]. SPARCcluster PDB 1.0 toimi maksimissaan kahden solmulaitteen klusterina ja laitteistovaihtoehtoina oli tarjolla kahta tuotetta: SPARCserver 1000E ja SPARCcenter 2000E sekä käyttöjärjestelmänä Solaris 2.4 [ElR01].

SPARCcluster PDB:n jälkeen valmistui SPARCcluster HA 1.0 tukemaan klusteroitavaksi muita palveluja, esimerkiksi NFS- ja DNS-palveluja. Molemmasta SPARCcluster-tuotteesta julkistettiin muutama paranneltu versio, kunnes tuotteet yhdistettiin Sun Cluster 2.0:n myötä yhdeksi tuotteeksi [BDV01]. Sun kehitti Sun Clusterista uusia versioita noin vuoden välein, mutta viime vuosina julkistustahti on hidastunut. Tammikuussa 2007 julkistettiin Sun Clusterin uusin versio: Sun Cluster 3.2 [SUN07b]. Sun Clusterin versiot on lueteltu taulukossa 12 [BDV01 ja SUN07g].

tuote	julkaisuajankohta
SPARCcluster PDB 1.0	kesä 1995
SPARCcluster HA 1.0	talvi 1995
SPARCcluster PDB 1.1	kevät 1996
SPARCcluster HA 1.1	kesä 1996
SPARCcluster PDB 1.2	syksy 1996
SPARCcluster HA 1.2	syksy 1996
SPARCcluster HA 1.3	syksy 1997
Sun Cluster 2.0	kevät 1997
Sun Cluster 2.1	kevät 1998
Sun Cluster 2.2	maaliskuu 1999
Sun Cluster 3.0	marraskuu 2000
Sun Cluster 3.1	toukokuu 2003
Sun Cluster 3.2	tammikuu 2007

Taulukko 12. Sun Clusterin versiot [BDV01 ja SUN07g].

Sun Cluster 3.2 kuuluu Sun Java Availability -tuotepakettiin (suite), jonka muut jäsenet ovat Sun Cluster Geographic Edition, klusteriagentit (Sun Cluster agents) ja kehitystyökalut. Sun Cluster 3.2:ta ajetaan Solaris-käyttöjärjestelmän versioiden 8, 9 ja 10 päällä [SUN07c]. Tuetut käskykannat ovat SPARC, x64 ja x86 sekä prosessorityypit UltraSPARC, SPARC64 ja AMD64 [SUN06a, s.18]. x64-käskykantaa tukee vain Solaris 10. SPARC- ja x86-tuet löytyvät Solaris-versioista 8, 9 ja 10 [SUN07c].

Sun Cluster 3.2 tukee maksimissaan 16 solmulaitetta SPARC-käskykannalla ja neljää solmulaitetta x86-käskykannalla [SUN06b, s. 18]. Edellinen versio Sun Clusterista,

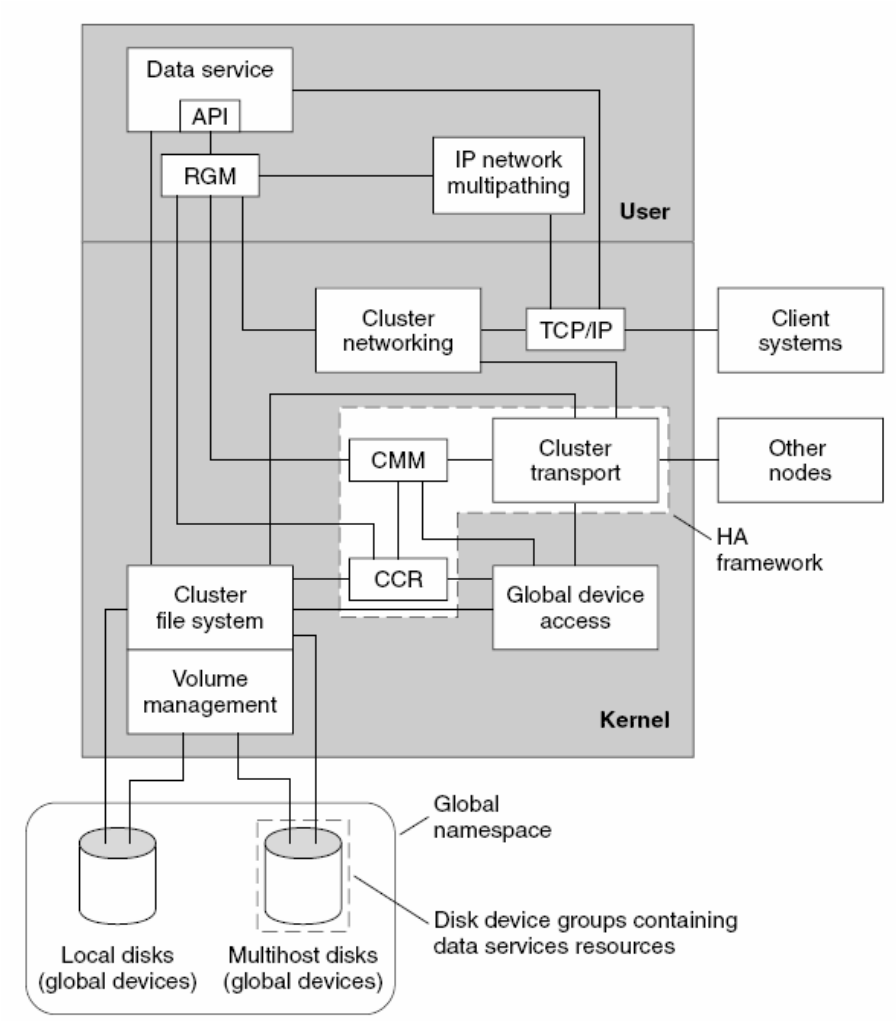
3.1, tukee maksimissaan kahdeksaa solmulaitetta [SUN03, s. 21]. Sun Cluster 3.2 tarjoaa kaikki maantieteellisen jaottelun mukaiset konfiguraatiotyypit: paikallinen klusteri, kampusklusteri, kaupunkialueklusteri ja mantereinen klusteri, joista mantereinen klusteri on omana tuotteena: Sun Cluster Geographic Edition [SUN07a].

Sun Java Availability Suiten lisenssi on hinnoiteltu käyttäjämäärän mukaan: 50 dollaria vuodessa per käyttäjä [SUN07a]. Myös muita lisensiointimalleja on tarjolla [SUN07d].

6.2 Arkkitehtuuri

Sun Clusterin pääkomponentit ovat ytimessä ajettavat Cluster Membership Monitor (CMM) ja Cluster Configuration Repository (CCR). CMM pitää kirjaa klusterin jäsenistä ja huolehtii heartbeat-signaalin käsittelystä. CMM käynnistää tarvittaessa failover-operaation ja varmistaa vikaantuneen solmulaitteen poistumisen klusterista [SUN06a, s. 18]. CCR on kaikille solmulaitteille hajautettu tietovarasto, johon tallennetaan klusterin konfiguraatiotiedot sekä tilatiedot [SUN06a, s. 19].

Sun kutsuu klusterin tarjoamia palveluita termillä data-palvelut (data services), jotka koostuvat resurssiryhmistä, joissa ovat klusteriresurssit [SUN06b, s. 65]. Resource Group Manager (RGM) käynnistää ja sammuttaa klusteriresurssit [SUN06a, s. 24]. Kullekin klusteriresurssityypille on vikamonitori (fault monitor), jonka avulla RGM tutkii klusteriresurssin tilaa. Jos klusteriresurssi ei ole toimintakuntoinen, RGM voi käynnistää klusteriresurssin uudestaan tai suorittaa failover-operaation [SUN06a, s. 19]. Sun Clusterin arkkitehtuuria esittää kuva 15 [SUN06a, s. 31].



Kuva 15. Sun Cluster 3.2 -arkkitehtuuri [SUN06a, s. 31].

Sun Cluster tarjoaa jaettujen levyjen tiedostopalvelut CFS-komponentin kautta (Cluster File System). CFS käyttää ”Shared rescouce”-mallia, eli kaikki solmulaitteet pääsevät samanaikaisesti kaikille levyalueille levyresurssien näkyessä klusterille globaaleina laitteina (global devices). Tiedostojen lukitukseen käytetään viitelukkoja (advisory file locking) [SUN06a, s. 32]. Myös solmulaitteiden paikallisia levyjä (local disks) voidaan mountata klusteriin globaaleiksi laitteiksi, mutta näitä ei voida siirtää failover-operaatiolla muille solmulaitteille, vaan omistavan solmulaitteen vikaantuessa kyseinen levyresurssi häviää näkyvistä [SUN06b, s. 21].

Sun Cluster tarvitsee levyjärjestelmän hallintaan erillisen ja erikseen ostettavan volume manager -tuotteen [SUN06a, s. 30]. Tuettuja volume manager -tuotteita ovat Solaris Volume Manager, Solaris Volume Manager for Sun Cluster ja VERITAS Volume

Manager [SUN06a, s. 12]. Tuetut tiedostojärjestelmät ovat UNIX file system (UFS), Sun StorEdge QFS file system ja VERITAS file system (VxFS) [SUN06a, s. 13].

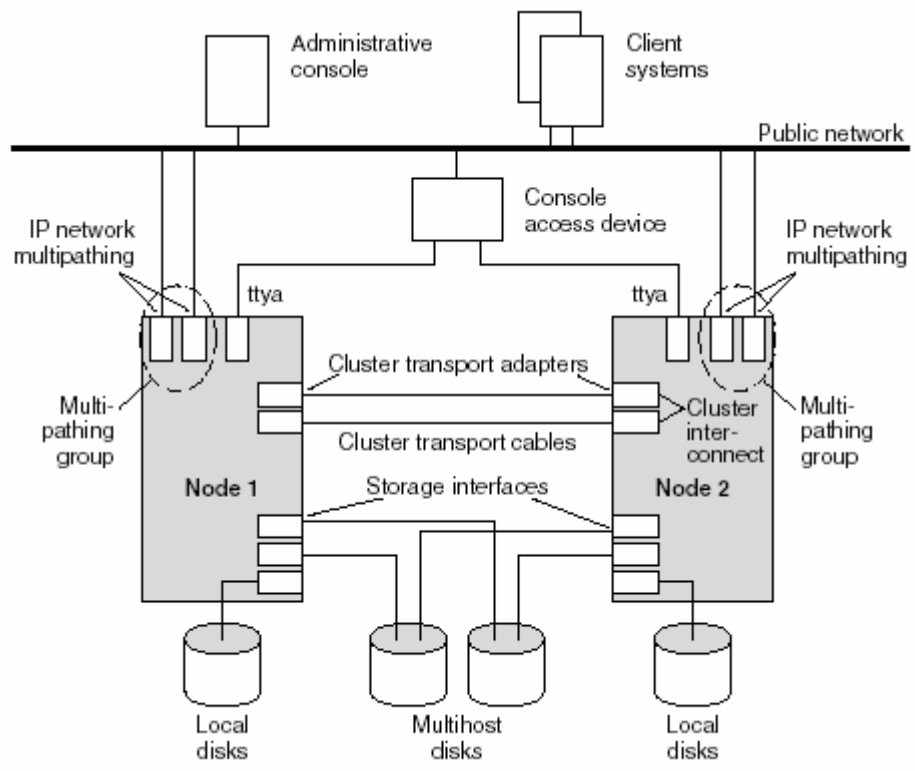
Quorum toteutetaan Sun Clusterissa joko jaetulla quorum-levyllä tai erillisellä quorum-palvelimella. Kullakin solmulaitteella sekä quorum-levyllä on oletuksena yksi ääni quorumia muodostettaessa. Yksittäisen solmulaitteen ääni voidaan poistaa asettamalla kyseinen solmulaite huoltotilaan (maintenance state) [SUN06a, s. 20]. Klusteri muodostetaan, kun jäsenillä on enemmistö äänistä [SUN06a, s. 21].

Sun Clusterin failover-operaation nopeus on yksinkertaisella resurssilla alle sekunnin luokkaa. Suuren tai monimutkaisen klusteriresurssin failover-operaatio voi kestää kymmenistä sekunneista minuutteihin [Nys07].

6.3 Laite- ja infrastruktuurivaatimukset

Sun tukee Sun Clusteria vain Sunin omalla laitteistolla. Sunin spesifikaatiot määrittelevät tarkalleen mitkä versiot Sunin palvelimista ja yksittäisistä komponenteista kelpaavat Sun Clusterille. Tuettujen palvelinten joukkoon kuuluu Sun Fire, Netra, Sun Enterprise ja Sun Blade -sarjojen palvelimia. Tuettujen SAN-kytkinten joukkoon kuuluu Sunin omien lisäksi Brocaden, Ciscon ja McDatan SAN-kytkimiä [SUN07d].

Sun Clusterin vaatima infrastruktuuri vastaa muiden kaupallisten korkean käytettävyyden klustereiden infrastruktuuria. Sun Clusteriin tarvitaan solmulaitteita omilla levyillä, jaettu levyjärjestelmä, Cluster interconnect -väylä solmulaitteiden väliseen kommunikointiin sekä verkkoliitännät julkiseen verkkoon asiakkaiden palvelua varten [SUN06a, s. 29]. Sun Clusterin fyysisiä komponentteja esittää kuva 16 [SUN06a, s. 30].



Kuva 16. Sun Clusterin fyysiset komponentit [SUN06a, s. 30].

Cluster interconnect -väylän toteuttamiseen tarjolla ovat seuraavat teknologiat: Fast Ethernet, Gigabit-Ethernet, InfiniBand ja Scalable Coherent Interface (SCI, IEEE 1596-1992). Sun suosittelee Cluster interconnect -väyläksi SCI:tä pienemmän siirto-viiveen ansiosta verrattuna Ethernet-pohjaisiin ratkaisuihin [SUN06a, s. 35].

Sun määrittelee maantieteellisen jaottelun mukaisille konfiguraatiotyypeille tarkat maksimietäisyysrajat johtuen käytetystä ohjainteknologiasta. Kampusklusterilla solmulaitteiden maksimietäisyys toisistaan on 10 kilometriä, jonka asettaa LWGBIC-teknologian (long wave gigabit interface controller) rajoitus. Kaupunkialueklusterilla maksimietäisyys on 400 kilometriä ja kyseisen rajan asettaa DWDM-teknologia (dense wave division multiplexing). Mantereisella klusterilla ei ole mitään etäisyysrajoitteita [SUN07a].

Solariksen erikoisuus, DR-alijärjestelmä (Dynamic Reconfiguration subsystem), on tuettuna myös Sun Clusterissa. Sen avulla laitteistokomponentteja, esimerkiksi muistipiirejä ja prosessoreita, voidaan vaihtaa lennossa laitteen ollessa päällä. RD-alijärjestelmä tutkii komponentin vaihto-operaation aikana onko kyseinen komponentti käytössä ja sallii vaihdon vasta, kun komponentti vapautuu käytöstä [SUN06b, s.

85]. Kilpailevilla järjestelmillä emolevyllä tapahtuvat hotswap-tyyppiset komponenttien vaihdot ovat mahdollisesti tuettuna vain laitteisto- ja ajuritasolla, eikä silloin klusteri ymmärrä tapahtumaa. DR-alijärjestelmä on tuonut viime aikoina Sunille kiusallista huomiota ohjelmointivirheestään, joiden seurauksena tietoja voi korruptoitua [TJT07].

6.4 Data-palvelut ja klusteriresurssit

Sun jakaa data-palvelut kolmeen luokkaan: failover data-palvelu (failover data service), skaalattava data-palvelu (scalable data service) ja rinnakkaiset sovellukset (parallel applications). Failover data-palvelu koostuu yhdestä failover-resurssiryhmästä, jota ajetaan kerrallaan korkeintaan yhdessä solmulaitteessa ja jonka sovellus ei ole lainkaan klusteritietoinen [SUN06a, s. 25].

Skaalattava data-palvelu koostuu kahdesta resurssiryhmästä, jotka ovat skaalattava resurssiryhmä ja failover-resurssiryhmä. Skaalattavan resurssiryhmän palveluita voidaan ajaa samanaikaisesti usealla solmulaitteella. Skaalattava resurssiryhmä on riippuvainen failover-resurssiryhmän resursseista, jotka ovat lähinnä verkkoresursseja [SUN06a, s. 25], esimerkiksi palvelun jaettu IP-osoite (shared address) [SUN06b, s. 59] tai looginen host-nimi (logical host name) [SUN06b, s. 57]. Skaalattava data-palvelu käyttää kuormantasausta, kun sovellusta ajetaan kahdella tai useammalla solmulaitteella. Kuormantasauksen säätöä varten on omat asetukset: load-balancing policies [SUN06b, s. 60].

Kolmas data-palveluluokka, rinnakkaiset sovellukset, sisältää sovelluksia, jotka ovat alusta lähtien ohjelmoitu toimimaan Sun Clusterin päällä ja joita ajetaan rinnakkain usealla solmulaitteella samanaikaisesti. Tätä data-palveluluokkaa edustaa vain Oraclen tietokantasovellus Oracle Real Application Server [SUN06a, s. 26].

Solariksen erikoisuus, vyöhykkeet (zones), ovat käytettävissä valinnaisesti myös Sun Clusterissa. Vyöhykkeillä voidaan palvelin jakaa useisiin loogisiin osiin, joissa sovellukset ajetaan eristettyinä toisten vyöhykkeiden prosesseista. Resurssiryhmät voidaan

sijoittaa eri vyöhykkeisiin ja failover-operaatio voidaan asettaa tapahtuvaksi samassa vyöhykkeessä toiselle solmulaitteelle tai samalla solmulaitteella toiselle vyöhykkeellä [SUN06c, s. 68 - 69]. Jälkimmäinen vaihtoehto ei hyödytä vikaantumistilanteessa, vaan se on tarkoitettu lähinnä testauskäyttöön [SUN06c, s. 70]. Vyöhykkeiden negatiivisena puolena niiden käyttö aiheuttaa lisää prosessointitarvetta ja failover-operaatiossa kuluu ylimääräistä aikaa vyöhykkeen siirtämiseen [SUN06c, s. 71]. Sun Clusterissa resurssiryhmät ja klusteriresurssit voidaan järjestää myös projekteihin (projects). Projekteja käytetään laskutustietojen keräämiseen sekä laitteistokapasiteetin käyttöasteen hallintaan ja valvontaan [SUN06c, s. 71].

Sovellusten klusteritietoisuus toteutetaan Sun Clusterissa klusteriagenteilla, joista osa on tehty Sunin toimesta ja osa sovellustoimittajien toimesta. Sovelluksiin ei tehdä muutoksia, vaan klusteriagentit hoitavat yhteistyön klusterin kanssa [SUN07c]. Klusteriagentit ovat muodoltaan k-sh-skriptejä, C-ohjelmia tai konekielisiä ohjelmia [SUN07e]. Kaikkiaan agentteja on noin 50 sovelluksella. Klusteriagentit on lueteltu tuoteperheittäin taulukossa 13 [SUN07c].

HA for Solaris Containers
HA Oracle 8 i, 9 i, Database 10 g, Parallel Server (OPS), 9 iRAC6, Oracle Application Server, e-Business Suite
HA DNS
HA NFS
HA Sybase, Sybase Adaptive Server Enterprise (ASE) 12.5
HA SAP liveCache database, J2EE Application Engine, Enqueue Server, Scalable SAP
HA Siebel, Siebel 7.7 – HA Java System Web, Application, Messaging, Directory, Calendar servers ja Message Queue, Java System Application Server Enterprise Edition 8.1, Java System Web Server
HA Apache Web/Proxy Server, Tomcat, scalable Apache Web/Proxy Server
HA NetBackup
HA Samba
HA DHCP
HA IBM WebSphere MQ, MQ Integrator
HA BEA WebLogic Server
HA Sun StorEdge Availability Suite
HA MySQL
HA N1™ Grid Engine, Sun N1 Service Provisioning System
IBM DB2 (EE ja EEE)
HA Informix Dynamic Server

Taulukko 13. Sun Clusterin klusteriagenttien tuoteperheet [SUN07c].

Mikäli tarkoitus on klusteroida jokin muu sovellus, johon ei ole valmista klusteriagenttia, käytetään yleistä data-palvelua (generic data service) tai tehdään oma klusteriagentti. Yleistä data-palvelua varten kehittäjän täytyy muokata omat monitorit ja skriptit käyttäen data-palvelu-API:a (data service API) ja data-palvelujen kehityskirjasto-API:a (data service development library API) [SUN06b, s. 64]. Oman sovelluksen klusteriagentin luomiseen Sunilla on kehitystyökalu Sun Cluster Agent Builder, jolla agentti tehdään käyttövalmiiksi graafisella velholla [SUN07e].

Sun tarjoaa klusteriresurssiryhmille monipuoliset säädöt laitteistokapasiteetin hallintaan. Esimerkiksi osia prosessorien kapasiteetista tai kokonaisia prosessoreita voidaan osoittaa yksittäisille resurssiryhmille. Tällöin klusterin oletusvuorottajaksi (default scheduler) pitää vaihtaa Fair-Share Scheduler (FSS) [SUN06a, s. 27], kun oletuksena on Timesharing Scheduler (TS) [SUN06d, s. 264]. Sun Clusterin resurssiryhmien attribuutit on lueteltu taulukossa 14 [SUN06c, s. 191 - 204].

attribuutti	kuvaus
Auto_start_on_new_cluster	Käynnistetäänkö resurssiryhmä automaattisesti klusteria muodostettaessa.
Desired_primaries	Resurssiryhmän solmulaitteiden suositeltu lukumäärä.
Failback	Solmulaitteen liittyessä klusteriin siirretäänkö mahdollisesti resursseja suositellulle (preferred) solmulaitteelle.
Global_resources_used	Käyttääkö resurssiryhmä globaaleja resursseja.
Implicit_network_dependencies	Käynnistetäänkö verkkoresurssit ennen muita klusteriresursseja.
Maximum_primaries	Resurssiryhmän solmulaitteiden maksimi lukumäärä.
Nodelist	Lista resurssiryhmän solmulaitteista.
Pathprefix	Hakemistopolku resurssiryhmän valinnaisiin asetustiedostoihin (administrative files).
Pingpong_interval	Resurssiryhmän heartbeat-signaalin jaksoväli sekunneissa.
Resource_list	Lista resurssiryhmän klusteriresursseita.

Taulukko 14 (osa 1/2). Sun Clusterin resurssiryhmien attribuutit [SUN06c, s. 191 - 204].

attribuutti	kuvaus
RG_affinities	Resurssiryhmän sidonta (affinity) muihin resurssiryhmiin. Arvot ovat: +++ vahva positiivinen sidonta ja failover ++ vahva positiivinen sidonta + heikko positiivinen sidonta - heikko negatiivinen sidonta -- vahva negatiivinen sidonta Esimerkiksi kentän arvo ”+RG2,--RG3” tarkoittaa, että tämä resurssiryhmä on heikosti sidottu resurssiryhmään 2 ja vahvasti negatiivisesti resurssiryhmään 3.
RG_dependencies	Resurssiryhmän riippuvuudet muihin resurssiryhmiin.
RG_description	Resurssiryhmän kuvaus.
RG_is_frozen	Ollaanko resurssiryhmän käyttämää globaalia laitetta siirtämässä.
RG_mode	Onko resurssiryhmä tyypiltään failover vai skaalattava.
RG_name	Resurssiryhmän nimi, jonka on oltava uniikki klusterissa.
RG_project_name	Resurssiryhmän projektin nimi.
RG_slm_cpu	Resurssiryhmälle osoitettujen prosessorien aikajaksojen (CPU share) lukumäärä.
RG_slm_cpu_min	Resurssiryhmän vaatima prosessorien minimi lukumäärä.
RG_slm_type	Onko resurssienhallinta päällä. Arvot ovat ”automated” ja ”manual”.
RG_slm_pset_type	Käytetäänkö automaattisessa resurssienhallinnassa dedikoitua prosessorijoukkoa (processor set).
RG_state	Resurssiryhmän tila. Vaihtoehtoiset arvot ovat: unmanaged, online, offline, pending_online, pending_offline, error_stop_failed, online_faulted ja pending_online_blocked.
Suspend_automatic_recovery	Pidättäydytäänkö resurssiryhmän automaattisesta käynnistämisestä esim. huolto-operaation vuoksi.
RG_system	Tätä attribuuttia käytetään resurssiryhmän suojaamiseen. ”True”-arvon alaisuudessa resurssiryhmään ei voi kohdistaa tiettyjä komentoja.

Taulukko 15 (osa 2/2). Sun Clusterin resurssiryhmien attribuutit [SUN06c, s. 191 - 204].

Klusteriresurssien attribuutit jaetaan Sun Clusterissa standardiominaisuuksiin (standard properties) ja laajennusominaisuuksiin (extension properties). Standardiominaisuudet ovat tarjolla kaikille resurssityypeille. Sen sijaan laajennusominaisuustyyppisiä attribuutteja on tarjolla erilaisia eri data-palvelulle [SUN06b, s. 64]. Sun Clusterin

klusteriresurssien attribuuttien standardiominaisuudet on lueteltu taulukossa 15 [SUN06c, s. 171 - 191].

attribuutti	kuvaus
Affinity_timeout	Affinity-määre sekunneissa kertoo kuinka kauan asiakkaan istunto muistetaan. Tämä affinity on siis eri asia kuin resurssiryhmillä.
Boot_timeout	Boot-metodin aikakatkaisu sekunneissa.
Cheap_probe_interval	Peräkkäisten ”quick fault probe”-operaatioiden aikaväli sekunneissa.
Failover_mode	Käytetäänkö failover-operaatiota. Vaihtoehtoiset arvot ovat: none, soft, hard, restart_only ja log_only.
Init_timeout	Init-metodin aikakatkaisu sekunneissa.
Load_balancing_policy	Skaalattavan palvelun kuormantasauspolitiikat. Vaihtoehtoiset arvot ovat: Lb_weighted, Lb_sticky ja Lb_sticky_wild.
Load_balancing_weights	Skaalattavan palvelun kuormantasauksen painotukset syntaksilla: kuorma@solmulaitteen numero. Esimerkiksi arvo ” 1@1,3@2” kertoo, että solmulaitteelle 1 tavoiteltu kuorma on 1/4 ja solmulaitteella 2 kuorma on 3/4.
Monitor_check_timeout	Monitor_check-metodin aikakatkaisu sekunneissa.
Monitor_start_timeout	Monitor_start-metodin aikakatkaisu sekunneissa.
Monitor_stop_timeout	Monitor_stop-metodin aikakatkaisu sekunneissa.
Monitored_switch	Valvotaanko resurssia vai ei.
Network_resources_used	Lista resursseista, joista tämä klusteriresurssi on riippuvainen.
Num_resource_restarts	Klusteriresurssin uudelleenkäynnistysyritysten yhteenlaskettu lukumäärä Retry_interval-attribuutin määräämällä aikajaksolla.
Num_rg_restarts	Resurssiryhmän uudelleenkäynnistysyritysten yhteenlaskettu lukumäärä Retry_interval-attribuutin määräämällä aikajaksolla.
On_off_switch	Ylläpitäjän asettama Enabled tai Disabled -tila.
Port_list	Solmulaitteen kuuntelemien TCP- ja UDP-porttien lista. Esimerkki: ”80/tcp6,40/udp6.”
Postnet_stop_timeout	Postnet_stop-metodin aikakatkaisu sekunneissa.
Prenet_start_timeout	Prenet_start-metodin aikakatkaisu sekunneissa.
Proxied_service_instances	Lista SMD-palveluista (SMF services), joita tämä klusteriresurssi käyttää proxyinä.
R_description	Kuvaus klusteriresurssista.
Resource_dependencies	Lista klusteriresursseista, joihin tällä klusteriresurssilla on vahva riippuvuus.
Resource_dependencies _offline_restart	Lista klusteriresursseista, joihin tällä klusteriresurssilla on offline-restart-riippuvuus.

Taulukko 16 (osa 1/2). Sun Clusterin klusteriresurssien attribuuttien standardiominaisuudet [SUN06c, s. 171 - 191].

attribuutti	kuvaus
Resource_dependencies _restart	Lista klusteriresursseista, joihin tällä klusteriresurssilla on restart-riippuvuus.
Resource_dependencies _weak	Lista klusteriresursseista, joihin tällä klusteriresurssilla on heikko riippuvuus.
Resource_name	Klusteriresurssin nimi.
Resource_project_name	Klusteriresurssin projektin nimi.
Resource_state	Klusteriresurssin tila. Vaihtoehdot ovat: Online, Offline, Start_failed, Stop_failed, Monitor_failed, Online_not_monitored, Starting tai Stopping.
Retry_count	Monitorin suorittamien uudelleenkäynnistysten lukumäärä, joka tullaan suorittamaan.
Retry_interval	Monitorin suorittamien uudelleenkäynnistysten välinen aikajakso.
Scalable	Onko klusteriresurssi tyypiltään skaalattava.
Start_timeout	Start-metodin aikakatkaisu sekunneissa.
Status	Monitorin asettama klusteriresurssin statustieto. Vaihtoehdot ovat: ok, degraded, faulted, unknown ja offline.
Status_msg	Monitorin Status-attribuutin asetuksen yhteydessä asettama viesti.
Stop_timeout	Stop-metodin aikakatkaisu sekunneissa.
Thorough_probe_interval	Klusteriresurssin ”high-overhead fault”-luotausten aikaväli.
Type	Klusteriresurssin tyyppi.
Type_version	Klusteriresurssin tyypin versio.
UDP_affinity	Tosi-arvolla asiakkaalle lähetetään klusteriresurssilta UDP-paketit saman solmulaitteen kautta, jolla on TCP-yhteys kyseiseen asiakkaaseen.
Update_timeout	Update-metodin aikakatkaisu sekunneissa.
Validate_timeout	Validate-metodin aikakatkaisu sekunneissa.
Weak_affinity	Kuormantasauksen heikko riippuvuus -asetus.

Taulukko 17 (osa 2/2). Sun Clusterin klusteriresurssien attribuuttien standardiominaisuudet [SUN06c, s. 171 - 191].

6.5 Asennus- ja hallintatyökalut

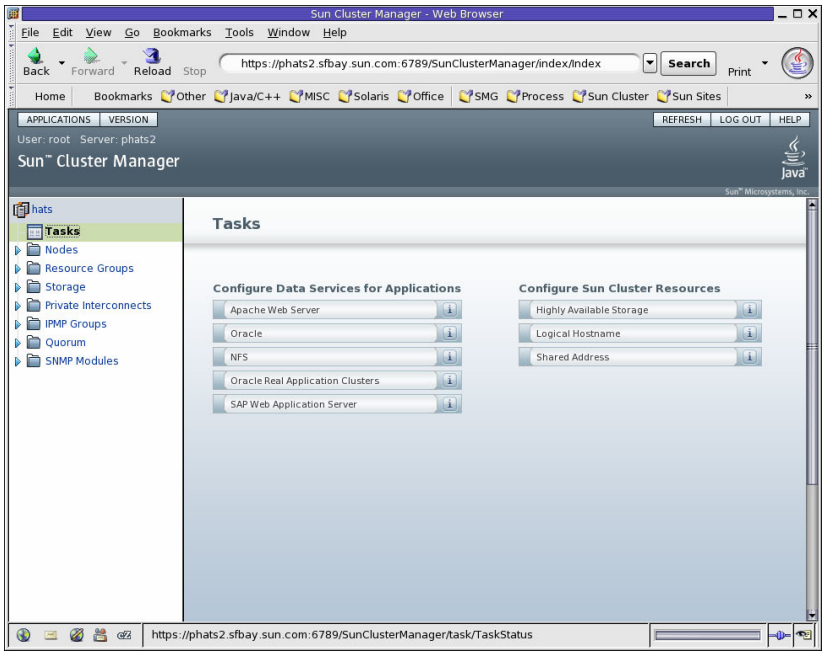
Sun Clusterin asennus sisältää useita erilaisia vaiheita. Asennusoperaation osatehtävien monimuotoisuus on omaa luokkaansa sisältäen runsaasti manuaalisia tehtäviä. Asennukseen kuuluu ascii-tiedostojen editointia, useiden asennusvelhojen ajamista sekä asetusten ja oikeuksien muutostehtäviä komentorivityökaluilla. Hyvän dokumentoinnin avulla kokenut ylläpitäjä suoriutuu asennuksesta, mutta Sun olisi voinut tehdä

Sun Clusterille viimeistellymmän asennusrutiinin. Asennuksen työvaiheista on kerätty pääkohtiin tiivistetty taulukko 16 [SUN06e, s. 17 - 41]:

nro	työvaihe
1	Cluster Control Panel -työkalun (SUNWccon) asennus hallintakonsoliin.
2	Ympäristömuuttujien ja tiedostojärjestelmän oikeuksien muokkausta.
3	Käyttöjärjestelmän asetusten muokkausta ascii-tiedostoja editoimalla.
4	Luodaan State Database Replicas (metadb).
5	Juuren tiedostojärjestelmän peilaaminen.
6	Sun Cluster -ohjelmiston ja agenttien asennus (installer).
7	Klusterin perustaminen (scinstall).
8	Klusterin levyjen ja LUN:ien luonti.
9	Klusterin tiedostojärjestelmän luonti (CFS)
10	Data-palvelujen asennus ja konfigurointi.

Taulukko 18. Sun Clusterin asennusvaiheet pääkohdittain [SUN06e, s. 17 - 41].

Sun Clusteria ylläpidetään graafisella Sun Cluster Manager -työkalulla, joka tunnettiin edellisissä Sun Clusterin versioissa nimellä SunPlex Manager. Samat toimenpiteet voidaan tehdä myös komentorivikäyttöliittymästä (command-line interface, CLI). Lisäksi osa komennoista voidaan suorittaa graafisen Management Center -ohjelmiston kautta [SUN06a, s. 13]. Sun Clusterin manuaalit käsittelevät ylläpito-operaatioita lähes ainoastaan komentorivityökalujen kautta. Oheisena on Sun Cluster Managerin käyttöliittymästä kuva 17 [SUN07f].



Kuva 17. Sun Cluster Managerin käyttöliittymä [SUN07f].

Sun Clusterin operoinnissa on omalaatuisia piirteitä, joita ei ole kilpailevissa tuotteissa. Esimerkiksi klusterin solmulaitteiden käynnistysjärjestys täytyy olla käänteinen verrattuna solmulaitteiden ja siten koko klusterin sammutukseen [SUN06b, s. 93]. Tällä estetään muistinmenetys (amnesia) eli klusterin muuttuneiden konfiguraatio- ja tilatietojen kadottaminen. Kilpailevista tuotteista esimerkiksi Microsoftin klusteripalvelussa solmulaitteiden käynnistysjärjestyksellä ei ole merkitystä, kun tilatiedot ovat jaetulla quorum-levyllä.

Ongelmatilanteista toipumisen ja tuen saamisen varmistamiseksi Sunin kanssa kannattaa tehdä tukisopimus. Sun Clusterin paikkauspaketteja (patch) toimitetaan nopeasti vain tukisopimuksen tehneille asiakkaille. Muille paikkauspaketit tulevat vapaasti jakeluun viikkoja myöhemmin [MS03b, s. 27]. Sunin maksullinen tuki koostuu neljästä vaihtoehtoisesta ja erihintaisesta tasosta: pronssi, hopea, kulta ja platina. Sopimusten hintoja ei ilmoiteta julkisesti, vaan Sun pyytää ottamaan yhteyttä myyntiin. Tukisopimusten palveluajat ovat taulukossa 17 [SUN06g].

palveluluokka	huollon vasteaika ja palveluaika	tuen palveluaika
pronssi	2 arkipäivää	klo 08 - 17 arkisin
hopea	4 h / klo 08 - 17 arkisin	klo 08 - 20 arkisin
kulta	4 h / klo 08 - 20 arkisin	24 h, 365
platina	2 h / 24 h, 365	24 h, 365

Taulukko 19. Tukisopimusten palveluajat [SUN06g].

Sun myös leasing-vuokraa palvelimiaan ”Sun System Packs”-tuotteena, jolloin vuokran hintaan kuuluu myös tuki. Leasing-sopimuksia on kolmea tasoa: hopea, kulta ja platina. Hopeatasoa on tarjolla vain edullisimpiin palvelinmalleihin ja palveluaika on 4 tunnin vasteajalla arkisin klo 8 - 17. Kulta- ja platinatasojen palveluaika on 24 h / 365 vrk ja vasteajat kulta-tasolla 4 h ja platinalla 2 h. Leasing-sopimuksen hinta riippuu laitteen kapasiteetista. Taulukossa 18 on verrattu Sun Fire v125 ja Sun Fire E6900 -palvelinten leasing-sopimusten hintoja yhden vuoden ajanjaksolla [SUN06f].

palvelin	hinta, hopeataso	hinta, kultataso	hinta, platinataso
Sun Fire v125	2324 $	2480 $	2720 $
Sun Fire E6900	*tasoa ei tarjolla*	239011 $	248743 $

Taulukko 20. Leasing-sopimusten hintoja yhden vuoden ajanjaksolla [SUN06f].

7 Ominaisuuksien vertailu

7.1 Vertailuun valitut ominaisuudet

Kaikki vertailun klusterituotteet käyttävät tyypillisiä korkean käytettävyyden klustereiden teknologioita, kuten heartbeat-signaalia ja jaettua levypintaa. Lifekeeper poikkeaa eniten muista quorumin puuttuessa kokonaan sekä jaetun levypinnan valinnaisuudella. Muuten vertailun tuotteet ovat selkeästi arkkitehtuuriltaan sukulaisia toisilleen. Pääkäyttäjän näkökulmasta yhteen tuotteeseen perehtymisen jälkeen on helppoa tutustua seuraavaan tuotteeseen.

Käytännössä monissa järjestelmähankkeissa päätetään ensin laitealusta, käyttöjärjestelmä ja sovellukset. Vasta tämän jälkeen mietitään korkean käytettävyyden vaihtoehtoja. Jos kyseiselle alustalle ei ole kuin yksi korkean käytettävyyden klusterituote, ei valinnalle ole vaihtoehtoja. Esimerkiksi, jos laitteistoksi on valittu HP:n Alphaserver-palvelin ja käyttöjärjestelmäksi Tru64 Unix, ei korkean käytettävyyden klusterituotteeksi jää muita vaihtoehtoja kuin TruCluster. Linux-, Solaris- ja Windows-alustoilla valinnanvaraa on enemmän. Seuraava rajoittava tekijä on, että onko sovellus klusteritietoinen kyseisen klusterituotteen osalta. Pienten ja vähemmän tunnettujen klusterituotevalmistajien sekä avoimen lähdekoodin klusterituotteita ei välttämättä uskalleta kuitenkaan valita, jos tuotteella ja valmistajalla ei ole riittävästi uskottavuutta. Päättäjälle tuntuu turvallisemmalta valita tutun ja suuren ohjelmistotalon klusterituote.

Mitä ominaisuuksia pitäisi sitten verrata klusterituotteen valintaratkaisua varten? Gartner esittää korkean käytettävyyden klusterituotteiden valintaan seuraavia vertailtavia ominaisuuksia [Wri02]:

1. klusteritietoisten sovellusten lukumäärä
2. solmulaitteiden maksimilukumäärä
3. klusterin palveluiden näkyvyys yhtenä objektina asiakkaille
4. solmulaitteiden välisen kommunikointiväylän nopeus
5. klusterin konfiguraatiomuutosten vaikutus ajonaikaiseen tuotantoon, siis tuleeko asiakkaille katkoja
6. kuormantasausominaisuudet

Tulen käyttämään edellisen listan ominaisuuksia vertaillessani klusterituotteita tässä dokumentissa. Edellisen listan lisäksi käytän vertailussa seuraavia valitsemiani huomionarvoisia ominaisuuksia:

7. lisenssin hinta
8. infrastruktuurin hinta
9. tuki maantieteelliselle hajauttamiselle
10. asennuksen ja ylläpidon helppous
11. hallinnan ja asetusten monipuolisuus

Muita mielenkiintoisia, mutta hankalasti vertailtavissa olevia ominaisuuksia ovat: operaatioiden suorituskyky sekä tuotteen ja valmistajan elinkaaren jatkuvuus. Operaatioiden suorituskyky on vertailun tuotteilla samaa suuruusluokkaa. Esimerkiksi failover-operaatio on kestoltaan tyypillisesti sekunteja. Erojen löytäminen vaatisi tuotteiden mittaamista identtisillä laitealustoilla, joka on vertailun tuotteiden osalta hankalaa erilaisten laitevaatimusten vuoksi.

Vertailun tuotteiden elinkaarten jatkuvuus on toistaiseksi turvattua, kun valmistajat ovat luvanneet jatkaa kehitystyötä sekä tarjota tukea. Microsoftia, Sunia ja HP:ta voidaan pitää suurempina organisaatioina vakaampina toimittajana kuin SteelEyeä. HP:n on epäilty lopettavan TruClusterin kehittämisen. Sun Cluster ja TruCluster kärsivät molemmat siitä, että kun klustereita ja Unix-järjestelmiä uusitaan, vaihtuu alusta useassa tapauksessa Linuxiin. Eräät Unix-ylläpitäjät luettelivat syitä, joiden vuoksi heidän organisaatioissaan oli siirrytty Solariksesta Redhat Linuxiin [Mat06]:

1. Linuxilla on parempi käytettävyys, laajempi tuki oheislaitteille ja valikoima sovelluksia.
2. Redhatille toimitetaan paremmin paikkauspaketteja (patch) sovelluksiin.
3. Sovellusvalmistajat ovat siirtäneet kehitystyön painopistettä Solariksesta Linuxiin ja Windowsiin, jolloin jälkimmäisiin on tarjolla enemmän sovelluksia.
4. Solariksen paikkauspakettien asennusprosessia pidetään työläänä. Ylläpitäjät kaipaavat Redhatin up2date-tyyppistä työkalua päivitysten asentamiseen.

7.2 Pistelaskujärjestelmä ja pisteytys

Tuotteiden paremmuusjärjestyksen ratkaisemiseksi asetan kunkin tuotteen paremmuusjärjestykseen edellisen luvun 11-kohtaisen listan mukaan ominaisuuskohtaisesti. Yhden ominaisuuden ”voitosta” saa 4 pistettä ja ”häviäjälle” jää 1 piste. Tasatuloksissa pisteet jaetaan, esimerkki jaettu toinen ja kolmas sija tuo 2,5 pistettä molemmille. Lopuksi pisteet lasketaan yhteen ja vertailun voittaa eniten pisteitä kerännyt.

7.2.1 Klusteritietoisten sovellusten lukumäärä

Microsoftin klusteripalvelulle on ylivoimaisesti eniten klusteritietoisia sovelluksia. Sun Cluster ja TruClusterille on karkeasti arvioiden samansuuruinen joukko klusteritietoisia sovelluksia. Lifekeeperille sovelluksia on vähiten. Kaikki vertailun tuotteet osaavat klusteroida ainakin yhden toimittajan tiedostojako-, web-, sähköposti- ja tietokantapalvelut.

tuote	Lifekeeper	MSCS	Sun Cluster	TruCluster
sovellusten lkm	10	satoja	kymmeniä	kymmeniä
Pisteet	1	4	2,5	2,5

Taulukko 21. Pisteet klusteritietoisten sovellusten lukumäärästä.

7.2.2 Solmulaitteiden maksimilukumäärä

Lifekeeperiin voi liittää eniten solmulaitteita: 32 kappaletta. Vähiten, kahdeksan kappaletta, solmulaitteita on liitettävissä Microsoftin klusteripalveluun ja TruClusteriin. Käytännössä korkean käytettävyyden klusterit koostuvat yleensä vain muutamasta solmulaitteesta, jolloin kahdeksan solmulaitteen lukumäärän yläraja tulee harvoin vastaan. Solmulaitteiden maksimilukumäärä näyttää usein kaksinkertaistuvan, kun klusterituotteesta tulee uusi versio. Näin kävi viimeksi Microsoftin klusteripalvelulla Windows 2000:sta siirryttäessä Windows 2003:een ja Sun Clusterilla 3.1:stä 3.2:een.

tuote	Lifekeeper	MSCS	Sun Cluster	TruCluster
solmulaitteita	32	8	16	8
Pisteet	4	1,5	3	1,5

Taulukko 22. Pisteet solmulaitteiden maksimilukumäärästä.

7.2.3 Klusterin palveluiden näkyvyys yhtenä objektina asiakkaille

Kaikki vertailtavat tuotteet näyttävät klusterin palvelut virtuaalisen palvelimen ja IP-osoitteen avulla, jolloin asiakkaan ei tarvitse tietää millä solmulaitteella palvelut ovat. Tämän suhteen vertailun tuotteilla ei ole eroa.

tuote	Lifekeeper	MSCS	Sun Cluster	TruCluster
Pisteet	3,3	3,3	3,3	3,3

Taulukko 23. Pisteet klusterin palveluiden näkyvyydestä yhtenä objektina asiakkaille.

7.2.4 Solmulaitteiden välisen kommunikointiväylän nopeus

Solmulaitteiden välisen kommunikointiväylän nopeus vaikuttaa kuinka nopeasti klusteri kykenee huomaamaan jäsenten tilamuutoksia, esimerkiksi solmulaitteen katoamisen. Sun Clusterin SCI on kaistanleveyden suuruuden ja siirtoviiveen pienuuden osalta ylivoimainen. TruClusterin memory channel on suorituskyvyssä lähellä SCI:tä. Microsoftin klusteripalvelulla ja Lifekeeperillä nopein kommunikointiväylä 1 Gbit Ethernet on kaventanut etumatkaa kaistanleveyden suhteen, mutta latenssin ero on yhä suuri verrattuna SCI ja memory channel -väyliin.

tuote	Lifekeeper	MSCS	Sun Cluster	TruCluster
kommunikointiväylä	1 Gbit Ethernet	1 Gbit Ethernet	SCI	memory channel
latenssi µs	< 100	< 100	1 - 2	3
kaistanleveys MB/s	< 100	< 100	< 320	> 100
Pisteet	1,5	1,5	4	3

Taulukko 24. Pisteet solmulaitteiden välisen kommunikointiväylän suorituskyvystä [HP07b ja YBP04, s. 4].

7.2.5 Konfiguraatiomuutosten vaikutus ajonaikaiseen tuotantoon

Konfiguraatiomuutokset voivat vaatia katkon klusterin toimintaan. Asiakkaiden palvelun kannalta on merkittävää voiko solmulaitteita lisätä klusteriin tai poistaa klusterista sekä tehdä muita muutoksia ajamatta koko klusteria alas. Merkittävää on myös, että voidaanko klusterin solmulaitteille tehdä klusterituotteen version päivitys ”lennossa”, eli ns. rolling upgrade, jolloin klusterin solmulaitteet päivitetään uuteen versioon solmulaite kerrallaan tuotannon jatkuessa samanaikaisesti. Kaikki vertailun tuotteet osaavat lisätä ja poistaa jäseniä ilman klusterin sammutustarvetta. Rolling upgrade -päivityksen osaavat muut, paitsi Sun Cluster [Wri02].

tuote	Lifekeeper	MSCS	Sun Cluster	TruCluster
Pisteet	3	3	1	3

Taulukko 25. Pisteet konfiguraatiomuutosten vaikutuksesta tuotantoon.

7.2.6 Kuormantasausominaisuudet

Vertailun tuotteista vain Sun Cluster osaa kuormantasauksen. Muilla vertailun tuotteilla ei ole lainkaan kuormantasausominaisuuksia, jolloin klusteriresurssit sijoitetaan solmulaitteille ennalta määriteltyjen staattisten sääntöjen mukaisesti.

tuote	Lifekeeper	MSCS	Sun Cluster	TruCluster
Pisteet	2	2	4	2

Taulukko 26. Pisteet kuormantasausominaisuuksista.

7.2.7 Lisenssin hinta

Lisenssin hintaa laskettaessa pitää ottaa huomioon käyttöjärjestelmän hinta, sekä agenttien mahdolliset erilliset hinnat. Lifekeeperin etuna sille kelpaavat edullisemmat web- ja standard-editionit Windowsista, joiden hinnat ovat noin 340 € ja 660 € [Ver07]. Microsoftin klusteripalvelu vaatii kalliimman enterprise- tai datacenter-

editionin, mutta klusterituote kuuluu lisenssin hintaan. Lifekeeper vaatii oman lisenssin ostamisen jokaiselle solmulaitteelle, sekä lisäksi agenteille mahdolliset omat lisenssit sovellus- ja solmulaitekohtaisesti. Microsoftin klusteripalvelun vaatiman Windowsin enterprise-lisenssin hinta 2849 € on kalliimpi kuin LifeKeeperin lisenssin hinta 2143 € (3000 $ / 1,4) Windows-lisenssin kera (340 € tai 660 €). Lifekeeperin maksullisia agentteja tarvittaessa Lifekeeperin lisenssistä tulee Microsoftin klusteripalvelun lisenssiä kalliimpi.

TruClusterin hinta ohittaa Windows-klustereiden hinnan, jos solmulaitteen kokoonpano on muu kuin edullisin vaihtoehto. Sun Clusterin käyttäjämäärästä riippuva lisenssin hinta nousee helposti korkeimmaksi, jos käyttäjiä on tuhansia. Solmulaitteiden lukumäärä on korkean käytettävyyden klustereissa tyypillisesti kaksi tai korkeintaan muutama, jolloin solmulaitekohtainen hinnoittelu on edullisempaa, kuin käyttäjäkohtaiset lisenssit.

tuote	Lifekeeper	MSCS	Sun Cluster	TruCluster
lisenssien hinta	3000 $ + Windowsin lisenssi (+ agentit)	3989 $ (enterprise edition)	50 $ per käyttäjä vuodessa	3000 - 48000 $
Pisteet	3,5	3,5	1	2

Taulukko 27. Pisteet lisenssin hinnasta.

7.2.8 Infrastruktuurivaatimukset

Lifekeeperillä on pienimmät vaatimukset laitteiston suhteen. Lifekeeper ei tarvitse ulkoista levyjärjestelmää, eikä muutakaan erikoista laitteistoa. Vertailun muut tuotteet tarvitsevat SCSI- tai SAN-levyjärjestelmän. Microsoftin klusteripalvelu ja Lifekeeper ovat laitevalmistajariippumattomia. Sun Cluster ja TruCluster vaativat valmistajan oman laitteiston. Microsoftin klusteripalvelu vaatii Active Directoryn, johon tarvitaan kaksi tai kolme palvelinta lisää. Active Directory tosin on monessa ympäristössä jo valmiina hoitamassa autentikointia, sekä toimialueen ohjauskoneille voi sijoittaa myös muita toimintoja.

tuote	Lifekeeper	MSCS	Sun Cluster	TruCluster
Pisteet	4	2	2	2

Taulukko 28. Pisteet infrastruktuurivaatimuksista.

7.2.9 Tuki maantieteelliselle hajauttamiselle

Vertailun tuotteista muut osaavat maantieteellisen hajauttamisen eri konesaleihin, paitsi TruCluster, joka ei sovellu kaupunkialueklusteriksi eikä mantereiseksi klusteriksi.

tuote	Lifekeeper	MSCS	Sun Cluster	TruCluster
Pisteet	3	3	3	1

Taulukko 29. Pisteet tuesta maantieteelliselle hajauttamiselle.

7.2.10 Asennuksen ja ylläpidon helppous

Kaikilla vertailun tuotteilla on graafisia työkaluja. TruClusterilla tosin suuri osa komennoista on suoritettavissa vain komentorivipohjaisina työkaluina. Lifekeeperin lisenssijärjestelmä on hankala. Sun Clusterin asennus on monivaiheisuudessaan työläs. Helpointa asennus ja ylläpito on Microsoftin graafisilla työkaluilla.

tuote	Lifekeeper	MSCS	Sun Cluster	TruCluster
Pisteet	2	4	2	2

Taulukko 30. Pisteet asennuksen ja ylläpidon helppoudesta.

7.2.11 Hallinnan ja asetusten monipuolisuus

Microsoftin klusteripalvelussa on kehittyneitä ominaisuuksia, kuten failback, autoshare subdirectories, vikatilanteen simulointi sekä solmulaitteiden massa-asennukset.

Sun Clusterin etuina ovat prosessorien osoittaminen resurssiryhmille ja monipuolisimmat attribuutit. Lifekeeperin ja TruClusterin attribuuttien valikoimat ovat yksinkertaisempia.

tuote	Lifekeeper	MSCS	Sun Cluster	TruCluster
Pisteet	1,5	3,5	3,5	1,5

Taulukko 31. Pisteet hallinnan monipuolisuudesta.

7.2.12 Pisteet yhteensä

Yhteenlaskettujen loppupisteiden erot tuotteiden välillä muodostuivat varsin pieniksi kärkikolmikon osalta. Yhteenlasketuissa pisteissä vertailun voittaja on Microsoftin klusteripalvelu 31,3 pisteellä. Toiseksi sijoittui Sun Cluster 29,3 pisteellä. Kolmantena on Lifekeeper for Windows 28,8 pisteellä ja neljäntenä TruCluster 23,8 pisteellä. Yhteenlasketut pisteet ovat taulukossa 30.

ominaisuus	Lifekeeper	MSCS	Sun Cluster	TruCluster
7.2.1 sovellusten lukumäärä	1	4	2,5	2,5
7.2.2 Solmulaitteiden lukumäärä	4	1,5	3	1,5
7.2.3 klusterin palveluiden näkyvyys yhtenä objektina	3,3	3,3	3,3	3,3
7.2.4 Solmulaitteiden välisen kommunikointiväylän nopeus	1,5	1,5	4	3
7.2.5 Konfiguraatiomuutosten vaikutus tuotantoon	3	3	1	3
7.2.6 Kuormantasaus	2	2	4	2
7.2.7 Lisenssin hinta	3,5	3,5	1	2
7.2.8 Infrastruktuurivaatimukset	4	2	2	2
7.2.9 Tuki maantieteelliselle hajauttamiselle	3	3	3	1
7.2.10 Asennuksen ja ylläpidon helppous	2	4	2	2
7.2.11 Hallinnan monipuolisuus	1,5	3,5	3,5	1,5
yhteensä	***28,8***	***31,3***	***29,3***	***23,8***

Taulukko 32. Pisteet yhteensä.

8 Yhteenveto

Korkean käytettävyyden klusterituotteet asetettiin paremmuusjärjestykseen vertailemalla 11 ominaisuutta ja pisteyttämällä ominaisuudet yhdestä neljään pisteeseen. Vertailun voittajaksi ylsi Microsoftin klusteripalvelu 31,3 pisteellä. Toiseksi sijoittui Sun Cluster 29,3 pisteellä. Kolmantena oli Lifekeeper for Windows 28,8 pisteellä ja neljäntenä TruCluster 23,8 pisteellä.

Microsoftin klusteripalvelun etuina olivat laitevalmistajariippumattomuus, vikatilanteen simulointi ja massa-asennukset. Microsoftin klusteripalvelulla on kehittyneitä ominaisuuksia, esimerkiksi failback ja autoshare subdirectories. Microsoftin klusteripalvelu työkaluilla on hyvä graafinen käyttöliittymä, jonka ansiosta asennus ja ylläpito ovat helppoa. Microsoftin klusteripalvelun heikkoutena ovat Active Directoryn pakollisuus, joka vaatii ylimääräisiä palvelimia toimialueen ohjaukseen, sekä puuttuva tuki dynaamisille levyille.

Vertailun toinen Windows-alustalla toimiva klusterituote Steeleyen Lifekeeper 6.0 ei aseta erikoisia laitevaatimuksia ja on laitevalmistajariippumaton. Muista vertailun tuotteista poiketen Lifekeeperillä jaetun levypinnan käytön voi korvata replikoinnilla, jolloin ei tarvitse investoida kalliiseen ulkoiseen levyjärjestelmään. Replikointivaihtoehto vaatii oman maksullisen ohjelmiston, Lifekeeper Data Replication for Windows. Lifekeeper skaalautuu mainiosti ulospäin, kun klusteriin saa maksimissaan 32 solmulaitetta. Lifekeeperille kelpaavat Windowsin edullisemmat web- ja standard edition lisenssit. Lifekeeperin heikkouksia ovat hankala lisenssisysteemi ja vähäinen lukumäärä klusteritietoisia sovelluksia: vain 10 kappaletta.

Sun Cluster 3.2 mahdollistaa SPARC-käskykannalla maksimissaan 16 solmulaitteen liittämisen klusteriin. x86-käskykannalla maksimilukumäärä on vaatimaton neljä solmulaitetta. Sun Clusterin etuna on prosessorien osoittaminen resurssiryhmille, jota vertailun kilpailevat tuotteet eivät osaa. Sun Clusterilla on kehittyneimmät attribuutit, jolloin resurssien toimintaa pystyy ohjaamaan monipuolisemmin, esimerkiksi kuormantasauksella skaalattavaa data-palvelua. Sun Clusterin heikkouksia ovat hankala

asennusoperaatio ja operoinnin omituisuudet, kuten klusteria käynnistettäessä solmulaitteiden vaadittu käänteinen käynnistysjärjestys sammutukseen nähden. Solariksen paikkauspaketit ovat tarjolla nopeasti vain Sunin kanssa tukisopimuksen tehneille.

HP:n TruCluster Server for Tru64 UNIX Version 5.1B tarjoaa nopean solmulaitteiden välisen kommunikointiväylän memory channel:in, jonka ansiosta klusterin tilamuutokset voidaan havaita nopeasti. 1 Gbit Ethernet on kuitenkin kaventanut etumatkaa memory channel:iin nähden ja Sunin käyttämä SCI on ohittanut suorituskyvyssä memory channel:in. TruCluster ei sovellu kaupunkialueklusteriksi eikä mantereiseksi klusteriksi. TruClusterin komennoista suuri osa on vain komentorivipohjaisia. TruClusterin jatkokehitys on epävarmaa, vaikka HP on luvannut tuen jatkuvan. Vertailun molempien UNIX-järjestelmien suosiota nakertaa asiakkaiden siirtyminen Linuxiin.

Lähteet

Abs03 Absher, J., An Empirical Examination of Current High-Availability Clustering Solutions' Performance. DePaul University, Chicago, IL, USA. March 2003. http://facweb.cs.depaul.edu/research/TechReports/TR03-003.doc [6.10.2006]

BoC01 Bottomley, J., Clements, P., Managing Distributions from the Software Vendor's Perspective. Proceedings of the 5th Annual Linux Showcase & Conference. USENIX Association. Oakland, CA, November 2001. Pages 119 - 126.

BDV01 Bianco, J., Deeths, D., Vargas, E., Sun Cluster Environment, Sun Cluster 2.2. Prentice Hall PTR, 1st edition. April 4, 2001.

Car05 Cartwright, D., SteelEye LifeKeeper review. Techworld. July 25, 05. http://www.techworld.com/applications/reviews/index.cfm?reviewid=312 [18.8.2007]

CGS96 Cardoza, W., Glover, F., Snaman, E., Design of the TruCluster Multicomputer System for the Digital UNIX Environment. Digital Technical Journal Vol. 8 No. 1 1996. Pages 5 - 17.

Chr03 Christensen, E., Guide to Creating and Configuring a Server Cluster under Windows Server 2003 White Paper. Microsoft Corporation. Nov 11, 2003.

DEC97 OpenVMS at 20, Nothing Stops it. Digital Equipment Corporation. 1997. http://h71000.www7.hp.com/openvms/20th/vmsbook.pdf [25.2.2007]

ElR01 Elling, R., Read, T., Designing Enterprise Solutions with Sun Cluster 3.0. Prentice Hall. December 05, 2001.

GSM98 Gamache, R., Short, R., Massa, M., Windows NT Clustering Service. Computer, Oct. 1998, Volume: 31, Issue: 10. IEEE Computer Society. Pages 55 - 62.

HP03 HP Surestore Disk Array XP48, HP e3000 Business Servers Configuration Guide, Hewlett-Packard, 06/2003. [27.1.2007] http://www.hp.com/products1/evolution/e3000/download/52430407e.pdf

HP04 Cluster Interconnect Analysis and Comparison White Paper. Hewlett-Packard Development Company, 05/2004. http://h30097.www3.hp.com/pdf/cluster_interconnect_wp.pdf [4.12.2006]

HP05 HP TruCluster Server V5.1B-3 QuickSpecs. Hewlett-Packard, May 30, 2005. http://h18000.www1.hp.com/products/quickspecs/Division_07-2005/12274_div.PDF [4.12.2006]

HP06 TruCluster Server Online Documentation. Hewlett-Packard, 2006. http://h30097.www3.hp.com/docs/pub_page/cluster_list.html [13.12.2006]

HP07a Transition from Tru64 UNIX to HP-UX 11i. Hewlett-Packard, 2007. http://h30097.www3.hp.com/transition/ [6.8.2007]

HP07b Memory Channel performance. Hewlett-Packard, 2007. http://www.hp.com/cgi-bin/pf-new.cgi?in=http://h20311.www2.hp.com/HPC/cache/274416-0-0-0-121.html#3 [15.10.2007]

Häm02 Hämäläinen, P., Käyttövarmuutta ryvästekniikoilla. Tietokone-lehti 8/2002. Sanoma Magazines, Helsinki, 2002. Sivut 80 - 87.

Lib00 Libertone, D., Windows 2000 Cluster Server Guidebook: A Guide to Creating and Managing a Cluster. Prentice Hall. 2000.

Mat06 Matty, R., Reasons why people are switching from Solaris to Linux. Blog O'Matty. 30 April 2006. http://prefetch.net/blog/index.php/2006/04/30/reasons-why-people-are-switching-from-solaris-to-linux/ [30.7.2007]

Mic07 Listing of products and services by manufacturer Steeleye Technology Inc. MicroAge, 1400 University Drive East, College Station, Texas, USA. http://microagecs.com/qisvmanufacturer/001232.html [21.9.2007]

MS03a Technical Overview of Clustering in Windows Server 2003 white paper. Microsoft Corporation. January 2003. [28.3.2007] http://www.microsoft.com/windowsserver2003/techinfo/overview/clustering.mspx

MS03b Comparing Sun Solaris 9 and Microsoft Windows Server 2003 Technologies. Microsoft Corporation. One Microsoft Way. Redmond, WA, USA. March 2003.

MS04 Server Clusters: Geographically Dispersed Clusters For Windows 2000 and Windows Server 2003. Microsoft Corporation. November 2004.

MSKB07 Windows 2000 and Windows Server 2003 cluster nodes as domain controllers. Microsoft Knowledge Base Article 281662. March 1, 2007. http://support.microsoft.com/kb/281662/en-us [11.3.2007]

MSP01 MCSE Training Kit: Microsoft Windows 2000 Advanced Server Clustering Services. Microsoft Press. Redmond, Washington, USA. 2001.

MST06 Windows 2000 Clustering Technologies: Cluster Service Architecture. Microsoft Technet. http://www.microsoft.com/technet/prodtechnol/windows2000serv/ deploy/confeat/clustsrv.mspx?pf=true [11.9.2006]

Nor07 Nordicmindin tekninen tuki. Nordicmind Oy, Helsinki, Suomi. [25.8.2007] http://www.nordicmind.com/cms/yritys/yhteystiedot/yhteydenottolomake/tekninen-tuki/

Nov07 Novell Cluster Services, System Requirements. Novell, Salt Lake City, Utah, USA. http://www.novell.com/products/clusters/ncs/sysreqs.html [25.2.2007]

Nys07 Nyström, P., Sun Cluster -asiantuntija, TietoEnator Oyj. Haastattelu Helsingissä 15.5.2007.

Ran02 Ranade, D., Shared Data Clusters: Scalable, Manageable, and Highly Available Systems (Veritas series). Wiley Computer Publishing. Indianapolis, USA. 2002.

Ran99 Ranta-aho, M., kurssi S-72.400: Ihminen ja tietoliikennetekniikka. 5. luennon luentomateriaali: Käytettävyys, käyttäjän tarpeiden ja tehtävien huomioiminen ja käyttäjäkeskeisen suunnittelun prosessi. Tietoliikennelaboratorio, Teknillinen Korkeakoulu, Espoo. 1999. http://www.comlab.hut.fi/opetus/400/Suomeksi/luennot.html [27.1.2007]

Ste01 SteelEye Brings Scalable High Availability to Windows 2000 with Packaged Support for Clustered Web and Database Servers. SteelEye Technology Inc., Palo Alto, CA, USA. June 28, 2001.
http://www.steeleye.com/news/press_releases/2001/062801.html [18.8.2007]

Ste03 Implementing Fault Resilient Protection for mySAP in a Linux Environment, Introducing LifeKeeper from SteelEye Technology. SteelEye Technology Inc., Palo Alto, CA, USA. 2003.
http://www.steeleye.com/pdf/literature/wp_mySAP.pdf [25.8.2007]

Ste04 SteelEye LifeKeeper for Windows, Solution Brief. SteelEye Technology Inc., Palo Alto, CA, USA. 2004.

Ste06a LifeKeeper for Windows Planning and Installation Guide. SteelEye Technology Inc., Palo Alto, CA, USA. November 2006.

Ste06b SteelEye Announces LifeKeeper Protection Suite for IBM Director 5.1. SteelEye Technology Inc., Palo Alto, CA, USA. 2006.
http://www.steeleye.com/news/press_releases/2006/020606.html [23.9.2007]

Ste07a LifeKeeper for Windows v6 Release Notes. SteelEye Technology Inc., Palo Alto, CA, USA. January 2007.

Ste07b LifeKeeper Online Product Manual. SteelEye Technology Inc., Palo Alto, CA, USA. http://licensing.steeleye.com/support/documentation.php/04h2/ OnlineProduct-Manual/lksstart.htm [31.1.2007]

Ste07c LifeKeeper Extender, LifeKeeper for Linux FAQs. SteelEye Technology Inc., Palo Alto, CA, USA. http://licensing.steeleye.com/support/faqs.php [25.8.2007]

SUN03 Sun Cluster 3.1 10/03 Concepts Guide, Revision A. Sun Microsystems Inc., Santa Clara, California, USA. October 2003.

SUN06a Sun Cluster Concepts Guide for Solaris OS, Revision A 12. Sun Microsystems Inc., Santa Clara, California, USA. December 2006.
http://docs.sun.com/app/docs/doc/819-2969 [9.5.2007]

SUN06b Sun Cluster Overview for Solaris OS, Revision A. Sun Microsystems Inc., Santa Clara, California, USA. December 2006.

SUN06c Sun Cluster Data Services Planning and Administration Guide for Solaris OS, Revision A. Sun Microsystems Inc., Santa Clara, California, USA. December 2006.

SUN06d Sun Cluster System Administration Guide for Solaris OS, Revision A. Sun Microsystems Inc., Santa Clara, California, USA. December 2006.

SUN06e Sun Cluster Quick Start Guide for Solaris OS, Revision A. Sun Microsystems Inc., Santa Clara, California, USA. December 2006.

SUN06f Sun System Packs for Servers. Sun Microsystems Inc., Santa Clara, California, USA. 2006. http://www.sun.com/service/systempacks/servers.jsp [6.8.2007]

SUN06g Sun Spectrum, Sun System Service Plans for the Solaris OS. Sun Microsystems Inc., Santa Clara, California, USA. 2006. http://www.sun.com/service/serviceplans/sunspectrum/ [6.8.2007]

SUN07a Sun Java Availability Suite. Supported Cluster Configurations for Disaster Recovery. Sun Microsystems Inc., Santa Clara, California, USA. [9.5.2007]

SUN07b Solaris Cluster Delivers New Disaster Recovery and Business Continuity Solutions for Solaris 10 OS. Sun Microsystems Inc., Santa Clara, California, USA. January 9, 2007. http://www.sun.com/aboutsun/pr/2007-01/sunflash.20070109.1.xml [9.5.2007]

SUN07c Sun Java Availability Suite. Premier HA and Disaster Recovery Solution. Sun Microsystems Inc., Santa Clara, California, USA. [9.5.2007] http://www.sun.com/software/javaenterprisesystem/suites/avail_suite_ds.pdf

SUN07d Sun Cluster software product info. Sun Microsystems Inc., Santa Clara, California, USA. [9.5.2007] http://catalog.sun.com/productinfo.xml?site=SOFTWAREUK&catalogue=FC&segment=FC_R&item=FC_SC_CAT&group=2014&fid=5135&sfid=5157&id=12007

SUN07e Solaris Cluster FAQ. Sun Microsystems Inc., Santa Clara, California, USA. http://www.sun.com/software/solaris/cluster/faq.jsp [9.5.2007]

SUN07f DataService Configuration Wizards, Sun Cluster Manager. Sun Cluster Oasis. Feb 05, 2007. http://blogs.sun.com/SC/date/200702 [27.7.2007]

SUN07g Sun Microsystems Documentation. Sun Microsystems Inc., Santa Clara, California, USA. [2.8.2007] http://docs.sun.com/app/docs/prod/sun.cluster#hic

TJT07 Sun löysi kiusallisen puutteen Solariksesta. Tekniikka&Talous-lehti. Talentum Oyj. Helsinki. http://www.tekniikkatalous.fi/doc.te?f_id=171895 [25.7.2007]

Tru02a TruCluster Server, Cluster Technical Overview, TruCluster Server Version 5.1B. Hewlett-Packard Company, Palo Alto, California. September 2002.

Tru02b TruCluster Server, Cluster Installation, TruCluster Server Version 5.1B. Hewlett-Packard Company, Palo Alto, California. September 2002.

Tru02c TruCluster Server, Cluster Administration, TruCluster Server Version 5.1B. Hewlett-Packard Company, Palo Alto, California. September 2002.

Tru02d TruCluster Server, Cluster Hardware Configuration, TruCluster Server Version 5.1B. Hewlett-Packard Company, Palo Alto, California. September 2002.

Tru02e TruCluster Server, Cluster Highly Available Applications, TruCluster Server Version 5.1B. Hewlett-Packard Company, Palo Alto, California. September 2002.

Tru07 Tru64 UNIX Roadmap. Hewlett-Packard Development Company. 4 January 2007. http://h30097.www3.hp.com/pdf/Tru64_Roadmap_Current.pdf [28.4.2007]

VDB98 Vogels, W., Dumitriu, D., Birman, K., The design and architecture of the Microsoft Cluster Service-a practical approach to high-availability and scalability. Fault-Tolerant Computing, Digest of Papers. Twenty-Eighth IEEE Annual International Symposium. 25 June 1998. Pages 422 - 431.

Ver07 Microsoft Windows Server 2003 -lisenssien hintoja. Verkkokauppa.com. [26.3.2007] http://www.verkkokauppa.com/popups/prodinfo.php?id=13754 http://www.verkkokauppa.com/popups/prodinfo.php?id=6328 http://www.verkkokauppa.com/popups/prodinfo.php?id=4496

Wey01 Weygant, P., Clusters for High Availability: A Primer of HP Solutions, 2nd Edition. Prentice Hall. 2001.

Wik07 Wikipedia. High-availability cluster. http://en.wikipedia.org/wiki/HA_cluster [25.2.2007]

Wri02 Wright, J., Gartner Research Note: Compaq TruCluster Server UNIX-based Clustering Solution Gartner Datapro Product Report DPRO-94885, 6 February 2002. http://www.gartner.com/gc/webletter/compaqnh/issue2/article1.html#Table%201 [28.4.2007]

YBP04 Yeo, C., Buyya, R., Pourreza, H., Cluster Computing: High-Performance, High-Availability, and High-Throughput Processing on a Network of Computers. Grid Computing and Distributed Systems Laboratory and NICTA Victoria Laboratory, Dept. of Computer Science and Software Engineering. The University of Melbourne, Australia. 2004.