

Szerverkonszolidáció az SZTE Egyetemi Számítóközpontban

Csóti Zoltán, Szőke Szabolcs, Borús András, Horváthné Jáhni Gabriella,
Meszlényi Zoltán, Soós Tamás, Tóth-Abonyi Mihály, Török Attila, Zombori Zoltán
{csotiz, szoke, borus, jahmig, msz, tom, toth, attilat, zombori}@cc.u-szeged.hu
Szegedi Tudományegyetem Egyetemi Számítóközpont

Tartalomjegyzék

Szerverkonszolidáció az SZTE Egyetemi Számítóközpontban	1
1. Bevezetés	2
2. Előkészítés	3
3. Beszerzés	4
3.1. A beszerzés elvi és adminisztratív kérdései	4
3.2. Beszerzett eszközök	5
3.3. Környezetkialakítás	6
3.3.1. A fizikai környezet kialakítása	6
3.3.1.1. Elektromos hálózat korszerűsítése	6
3.3.1.2. Nagy teljesítményű UPS	6
3.3.1.3. Dízel-áramfejlesztő beszerzése	6
3.3.1.4. Gépterem elrendezése	6
3.3.1.5. Hűtési rendszer	7
3.3.2. Hálózati környezet kialakítása	7
4. Blade	8
4.1. HP Blade telepítése	8
4.1.1. Hálózat	9
4.1.1.1. Ethernet	9
4.1.1.2. SAN	11
4.1.1.3. Menedzsment	11
4.2. Blade-üzemeltetés	12
5. Storage	13
5.1. EMC Celerra NS-480 telepítés	13
5.1.1. Az NS-480 hardverelemei	13
5.1.2. Konfiguráció	13
5.2. EMC Celerra NS-480 üzemeltetés	14
5.2.1. Storage-hibák	14
5.2.2. Egyéb üzemeltetési tapasztalatok	14
6. VMware	15
6.1. VMware vSphere (ESXi) telepítése	15
6.1.1. Hálózat	15
6.1.2. Cluster	16
6.2. VMware vSphere (ESXi) üzemeltetése	17
6.2.1. Vmreg	17
6.2.2. Migráció	17
6.2.3. vSphere 5.0	18
6.2.4. Virtuális gépek, üzemeltetés	18

7. Mentés	19
7.1. Networker telepítés	19
7.1.1. A mentőrendszerrel	19
7.1.2. Logikai hálózati topológia	20
7.1.3. A központi mentőrendszer mentőszoftvere	20
7.1.4. Tartalék mentőszerver	21
7.1.5. A mentési feladatok kialakítása	21
7.2. Networker üzemeltetés	22
7.2.1. A nwrkr1, nwrkr2 mentőszerverek üzemeltetése	22
7.2.2. Mentőeszközök.....	22
7.2.2.1. VTL (HP D2D4112)	22
7.2.2.2. Dell TL4000.....	22
7.2.3. EMC Networker	23
7.2.4. Szalagkezelés Networkerben	23
7.2.5. Networker szerver teljes visszaállítása	23
7.2.6. Hardver hibák	23
8. Fejlesztési lehetőségek	23
8.1. Blade	23
8.2. Storage	24
8.3. VMware	24
8.4. Mentés	24
8.4.1. Tape	24
8.4.2. D2D	24
8.4.3. Networker	24
8.5. Ethernet	25
9. Összegzés	25
Köszönetnyilvánítás.....	25
Függelék:.....	26
Felhasznált irodalom	26

1. Bevezetés

A Dél-alföldi Tudáspólus felsőoktatási infrastruktúrájának fejlesztése című, TIOP-1.3.1.-07/2/2F-2009-0004 azonosító számú uniós projekt keretében a Dél-alföldi régióban folyó matematikai, műszaki, természettudományos és informatikai képzés számára kiemelkedő infrastrukturális háttérrel biztosító, versenyképességet növelő fejlesztések valósultak meg 2009. január 16. és 2012. november 30. között a Szegedi Tudományegyetem és partnerintézményei együttműködésében.

A pályázat „B” komponenseként az oktatási-kutatási infrastruktúrát támogató, a 21. század elvárásainak megfelelő infokommunikációs technológiai fejlesztések történtek. Az alprogram célja az egyetem mindennapi működéséhez, az „üzletmenet folytonosságához” nélkülözhetetlen informatikai hálózat további fejlesztése és korszerűsítése volt. A projekt kiemelt fontosságú részelemei a következők voltak: az informatikai központ fejlesztése, az egyetemi gerinchálózat fejlesztése, az egyetemi épületek aktív eszközeinek és kábelezési rendszereinek korszerűsítése, az egyetemi vezetékek nélküli hálózat fejlesztése, hálózatmenedzsment, szerverkonszolidáció, portál- és üzletiintelligencia-rendszerek fejlesztése.

Egyebek mellett kiemelt figyelmet fordítottunk az egyetemi központi szolgáltatások hardver- és alapszoftverrendszereinek részbeni homogenizációjára: a szerverkonszolidációra, különös figyelemmel a futó szolgáltatásokra és a későbbi tartalmi (TÁMOP) pályázatok várható eszközigényeire.

Az előadásban ismertetjük a szerverkonszolidációs részprojektet: a tervezés, beszerzés, beüzemelés lépéseit, valamint az utóbbi másfél év üzemeltetési tapasztalatait.

2. Előkészítés

A bevezetésben kitűzött célokat megvalósító beszerzés tervezett mértékét (számítási, tárolási és archiválási kapacitás) a jelenleg üzemelő rendszerek kapacitására és a várható pályázatok igényeire alapoztuk.

Először is számba vettük az ESZK által nyújtott szolgáltatásokat:

- ETR (Egységes Tanulmányi Rendszer),
- TÜSZ (Teljes Ügyviteli Szoftver),
- egyetemi webszerver,
- mail firewallok,
- dolgozói és hallgatói levelezőszerverek,
- Coospace és videó streaming szerverek,
- egyéb szerverek (DNS, NTP stb.).

2007 őszén készült egy felmérés, mely alapján a számítógépeink nagy része elavultnak bizonyult:

Vásárlás éve	Számítógépek százaléka
2004 vagy régebben	71%
2003 vagy régebben	60%
2002 vagy régebben	52%

A konszolidáció előtti szerverparkunkban négyfajta operációs rendszer/ hardver-platform kombináció fordult elő:

- Sun Solaris/SPARC,
- Microsoft Windows/Intel,
- Novell NetWare/Intel,
- GNU/Linux/Intel.

A felmérést több lehetséges szállítónak is megmutattuk. A velük való előzetes konzultáció során kialakult egy gyártófüggetlen kép az új rendszerünkről.

A szervereknél a tapasztalataink miatt és költséghatékonysági okokból az Intel-platform és a blade-kivitel mellett döntöttünk. A kívánt HA- és DRS-szolgáltatások csak központi storage mellett valósíthatók meg. A korábbi gyakorlattól eltérően központi mentőrendszer mellett döntöttünk, amely jobban illeszkedik az új infrastruktúrába. A D2D-eszköz a mentőrendszer szabadabb konfigurálását tette lehetővé.

Virtualizációs szoftverként a piacvezető VMware termékét választottuk, mert csak ez nyújtotta az általunk elvárt szolgáltatásokat. E döntést elősegítette, hogy a tervezéssel párhuzamosan a szoftver ingyenes változatának felhasználásával sikerült egy jól működő éles rendszert (a levelezési tűzfalakat) üzembe állítani.

Úgy terveztük, hogy a konszolidációval egy időben a Sun Solaris szervereket GNU/Linux szerverekre migráljuk, és a következő operációs rendszereket fogjuk használni a blade-ben:

- Microsoft Windows,
- GNU/Linux.

A régi eszközparkunk és a beszerzett rendszer főbb paramétereinek számszerű összehasonlítása:

	Processzor	Memória	Merevlemez	Mentés
régi	90	108 GB	9,2 TB	~10 TB
új	112 mag	459 GB	37 TB	80 TB

3. Beszerzés

3.1. A beszerzés elvi és adminisztratív kérdései

2009 nyarának végére a konfiguráció harmadik változatának elkészültével lényegében meghatároztuk azt a műszaki tartalmat, amelyet be kívántunk szerezni. Ez a rendszer egyetlen gyártó konkrét termékeiből állt. Műszaki és anyagi megfontolásokból megkerestünk más cégeket is, hogy tudnak-e alternatív ajánlatot tenni, akár más gyártók eszközeire támaszkodva. A számításba vehető változatok számát nagyban csökkentette az az alapelv, hogy egyetlen szállítóval szándékoztunk szerződést kötni, mert nem akartuk magunkra vállalni a rendszerintegráció feladatát. Ugyanis a teljes rendszer működőképessége volt az elsődleges szempont, hiába tudtuk, hogy külön-külön melyik blade-, storage- vagy backuprendszer felelne meg leginkább az elképzeléseinknek. A kedvező tartalmú és árú változatok számát az is csökkentette, hogy egy lehetséges szállítónak melyik termék gyártójával volt szorosabb kapcsolata, azaz kitől vártunk nagyobb árkedvezményt és magasabb színvonalú támogatást, illetve a cég mérnökeinek melyik eszköztípussal volt gyakorlati tapasztalata.

Végül három, nagyjából azonos műszaki tartalmú ajánlat közül a legolcsóbbat választottuk.

A tárgyalások eredményeképpen 2010. áprilisban adtuk fel a megrendelést. A beszerzés formája az ún. KSZF-es volt, akkor a lehetséges módok között ez tűnt a legegyszerűbbnek és legrugalmasabbnak. A szállító a KFKI (jelenleg T-Systems) lett.

A szállítási szerződés aláírása egy kb. 18 hónapos időszak kezdetét jelentette, amely egy menetrend alapján magába foglalta a következőket:

- a rendszerterv kidolgozása,

- az eszközök telepítése (augusztus végétől 1 hónap, a részleteket lásd az egyes rendszereknekél),
- a rendszergazdák oktatása (a saját rendszerünkön, 7 oktatási nap),
- a rendszer beüzemelése (a részleteket lásd az egyes rendszereknekél),
- a megvalósulási dokumentáció elkészítése.

3.2. Beszerzett eszközök

A beszerzett eszközökre egységesen 5 év garanciát vettünk. A storage-ra 4×7×24-es garanciát. Az összes többi eszközre a NBD (Next Business Day) verziót, mert a szükséges rendelkezésre állást a megvásárolt eszközök redundanciájára támaszkodva tervezzük biztosítani.

Kód	Megnevezés	db
Vmware		
VMware Infrastructure Enterprise Plus		
VS4-ENT-PL-A	Academic VMware vSphere 4 Enterprise Plus for 1 processor (Max 12 cores per processor)	28
VS4-ENT-PL-G-SSS-A (5yrs)	Academic Gold Support/Subscription for VMware vSphere Enterprise Plus for 1 processor	28
VCS-STD-A	Academic VMware vCenter Server 4 Standard for vSphere (Includes Orchestrator and Linked Mode)	1
VCS-STD-G-SSS-A (5yrs)	Academic Gold Support/Subscription for vCenter Server Standard	1
Blade		
HP c-Class Blade c7000 infrastruktúra		
507019-B21	HP BLc7000 CTO 3 IN LCD ROHS Encl	2
499243-B21	HP 2400W High Efficiency Power Supply	12
412140-B21	HP BLc Encl Single Fan Option	12
456204-B21	HP BLc7000 DDR2 Encl Mgmt Option	2
413379-B21	HP BLc7000 1 PH FIO Power Module Opt	2
439034-B23	HP c-Class All FIO 8 lcm 1yr 24x7 Lic	2
AJ822A	HP 8/24c BladeSystem Pwr Pl+ SAN Switch	4
AJ716A	HP 8Gb Short Wave B-series FC SFP+ 1 Pack	3
516733-B21	HP Procurve 6120XG Blade Switch	4
455883-B21	HP BladeSystem 10Gb SR SFP+	5
455889-B21	HP BLc 10Gb LRM SFP+ Opt	10
14x BL490c G6		
BL490c G6		14
498357-B21	HP BL490c G6 CTO Blade	14
509319-L21	HP X5570 BL490c G6 FIO Kit	14
509319-B21	HP X5570 BL490c G6 Kit	14
500658-B21	HP 4GB 2Rx4 PC3-10600R-9 Kit	168
451871-B21	HP BLc QLogic QMH2562 8Gb FC HBA Opt	14
508291-B21	USB Key - HP VMware ESXi Standalone Lic	14
2x BL460c Management		
BL460c G6		2
507864-B21	HP BL460c G6 CTO Blade	2
507798-L21	HP L5520 BL460c G6 FIO Kit	2
507798-B21	HP L5520 BL460c G6 FIO Kit	2
500656-B21	HP 2GB 2Rx8 PC3-10600R-9 Kit	6
507127-B21	HP 300GB 6G SAS 10K SFF DP ENT HDD	4
451871-B21	HP BLc QLogic QMH2562 8Gb FC HBA Opt	2
EMC NAS Storage		
EMC NS4-480 (32x450GB 15K FC HDD, 40x1TB SATAII HDD)		1
DELL T4000 mentőeszköz		
210-19654	PV TL4000 LT04-120 2 FC Drive Silver Base	1
HP D2D		
D2D4112		
EH993A	HP D2D4112 Backup System	1
EH995A	HP D2D4112 12TB Capacity Upgrade Kit	1
EMC Networker Faststart mentőszoftver		
457-100-008	NETWORKER AUTOCHNGR SLOT UPGRD FROM 1-40 TO 1-64	1
457-100-086	NETWORKER FAST START WITH AUTOCHANGER FOR LINUX	1
456-100-699	NETWORKER DISKBACKUP OPT TIER 3: 10TB	1
CE-LGVILTNNFS	NETWORKER-VILT FOR FASTSTART	1

1. ábra: A beszerzett eszközök listája

3.3. Környezetkialakítás

3.3.1. A fizikai környezet kialakítása

A gépterem kialakításáról „Az SZTE Egyetemi Számítóközpont géptermének rekonstrukciója” címmel a 2011-es Networkshopon tartottunk egy előadást. Itt csak néhány fontosabb mozzanatot mutatunk be.

A tervezési folyamatnál figyelembe kellett venni a Szegedre kerülő HPC node igényeit is.

Az új berendezésekkel egy magas rendelkezésre állást biztosító, a későbbi fejlesztéseket is befogadni képes géptermet alakítottunk ki.

3.3.1.1. Elektromos hálózat korszerűsítése

A számítóközpont géptermét ellátó elektromos rendszer teljes cseréje megtörtént. Az elavult berendezések cseréjén és a teljesítmény növelésén kívül a géptermi rackszekrények, illetve az azokban lévő berendezések valódi kettős tápellátását is megvalósítottuk.

3.3.1.2. Nagy teljesítményű UPS

A valódi kettős betáplálás kialakításához a meglévő UPS mellé egy második nagy teljesítményű UPS-t szereztünk be. Az igényfelmérés alapján az APC gyártmányú SY160K160H-PD típust választottuk. Az UPS moduláris rendszerű, N+1 redundancia mellett 144 kVA kimeneti teljesítményt biztosít. A beépített elosztószekrény és szerviz bypass-kapcsolómodul segítségével könnyen illeszthető a gépterem elektromos rendszeréhez.

3.3.1.3. Dízel-áramfejlesztő beszerzése

A biztonságos, kiesésektől mentes üzemeltetéshez szükség volt egy, a gépterem megnövelt kapacitásához illeszkedő áramfejlesztőre is. A meglévő 165 kVA-es áramfejlesztő mellé egy FG Wilson gyártmányú P220HE2 típusú berendezést telepítettünk. Az áramfejlesztőket elláttuk távfelügyeleti rendszerrel is, amely hiba esetén az üzemeltetésnek jelzést küld.

3.3.1.4. Gépterem elrendezése

Korábban a gépteremben a rackszekrények szigeteket alkottak. Az operátorok a gépekkel közös helyiségben tartózkodtak.

A kiválasztott hűtési rendszer miatt folyamatos munkavégzés céljából nem tartózkodhat ember a gépteremben, ezért szükség volt egy külön operátori rész kialakítására. A korábbi géptermet kettéosztottuk. A külső, kisebb rész lett az operátori munkahely. Itt lehet hozzáférni a gépek termináljához.

A belső teret csak a rackszekrények és a hűtők foglalják el. A rackszekrényeket hideg-meleg folyosós rendszerbe szerveztük.

A gépteremben, a HPC node tartozékaként, automata oltóberendezés is telepítésre került.

3.3.1.5. Hűtési rendszer

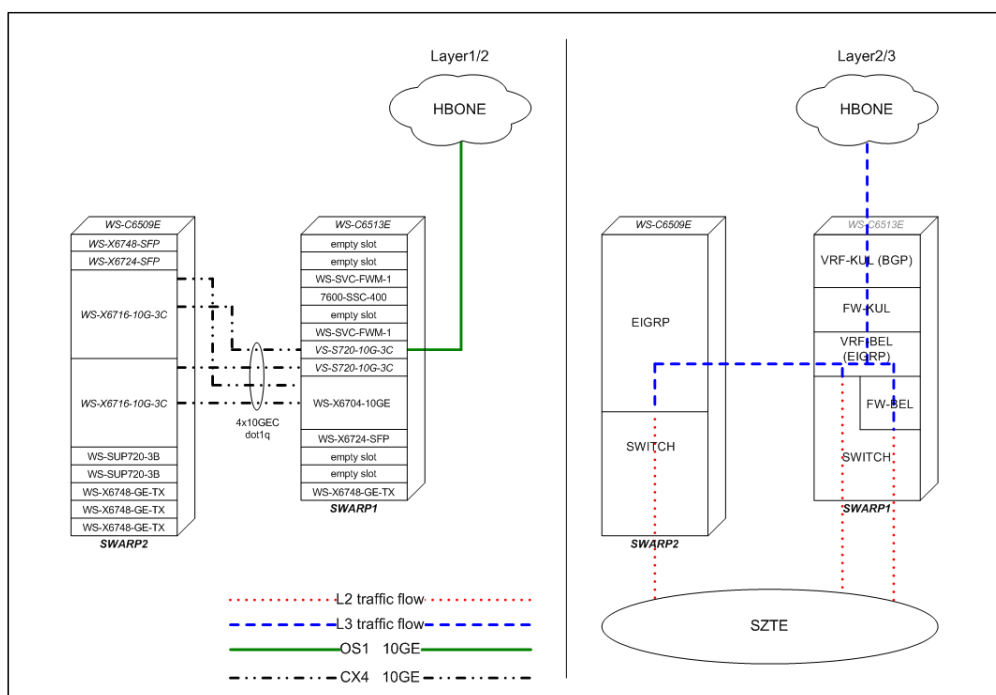
A korábbi rendszert rackszekrényekkel sorolható beltéri hűtőkkel váltottuk ki. 12 db APC ACRC103-at szereztünk be. Ezek jól illeszkedtek a már meglévő APC rackszekrényeinkhez, az új beszerzéseknél pedig ragaszkodtunk az APC típushoz. A beltéri egységek redundáns tápellátással rendelkeznek. Egy berendezés névleges hűtőtéljesítménye 18 kW.

A kültéri hűtők kiválasztásánál fő szempont volt, hogy rendelkezzenek ún. szabad hűtéssel is. Ezért két darab Uniflair gyártmányú ERAF 1221A típusú hűtőt szereztünk be. Ezek egyenként 117 kW névleges és 71 kW szabad hűtési teljesítménnyel rendelkeznek.

3.3.2. Hálózati környezet kialakítása

A TIOP pályázat előtt az SZTE hálózatában a core- és az edge-funkciókat egy L3-as Cisco Catalyst 6500 switch látta el. A megnövekedett sávszélességigényeket kielégítendő és a fentebbi funkciók szétválasztását elősegítendő a pályázatból beszereztünk egy új Cisco Catalyst 6500-as switchet is. A beszerzés lehetővé tette, hogy az így már kettő switchből álló központi rendszerünket további modulokkal bővítsük.

A core-funkciókat ellátó switchet 4×10GE redundáns kapcsolat köti össze a két blade-kerettel.



2. ábra: A központi hálózati rendszer

4. Blade

4.1. HP Blade telepítése

A projekt során két darab HP BladeSystem c7000 Enclosure (blade-keret) került beszerzésre. Összesen 16 félmagas blade-szervert vásároltunk, mindkét keretbe 8-8 db került. Így redundáns kialakítást értünk el, mert a rendszer az egyik keret teljes kiesése esetén is üzemképes marad.

A szerverek konfigurációja:

2 db BL460C G6, egyenként:

- 2× Intel Xeon L5520 (2,26 GHz, 8 MB, 4 core, HT),
- 12 GB RAM/server (6×2 GB, ill. 3×4 GB),
- Qlogic dual 8 Gb FC HBA,
- HP NC532m Dual Port 10 GbE,
- 2×300 GB SAS HDD.

Ezek a szervereken fut a Networker mentőszoftver, ezek kerültek a keretekben az 1. device bay-be. Az operációs rendszer és a mentőszoftver a szerver saját diszkjén van, így az a storage-rendszer nélkül is működőképes.

14 db BL490C G6, egyenként:

- 2× Intel Xeon X5570 (2,93 GHz, 8 MB, 4 core, HT),
- 48 GB RAM/server (12×4 GB),
- Qlogic dual 8 Gb FC HBA,
- HP NC532m Dual Port 10 GbE,
- 2 GB belső pendrive – itt helyezkedik el az ESXi.

ESXi 4.1 került telepítésre rájuk, a keretekben a 2-4., és a 9-12. device bay-ben vannak. Nincs saját merevlemezük, a virtuális gépeket a közös storage-rendszeren tárolják. Ez egyrészt a vMotion miatt szükséges, másrészt nem fordulhat elő az az eset, hogy kiesik egy blade, és a rajta futó virtuális gépek nem lesznek elérhetők, mert azok a lokális diszken voltak.

Az indulás óta sikerült memóriát bővítenünk a szerverek egy részében, így most van 4 szerver 48 GB RAM-mal (2. és 3. bay keretenként) és 10 szerver 72 GB RAM-mal.

Interconnect-modulokból Ethernet- és FC-modul került a konfigurációba, keretenként:

2 db HP ProCurve 6120XG Blade Switch (1-2. interconnect bay)

2 db HP 8/24c SAN Switch Pwr Pk+ BladeSystem c-Class (3-4. interconnect bay)

A szerverek alaplap 1-es Ethernet portja az interconnect bay 1-re, a 2-es Ethernet portja az interconnect bay 2-re, a HBA 1-es portja az interconnect bay 3-ra, míg a HBA 2-es portja az interconnect bay 4-re kapcsolódik.

Mindkét keretben két Onboard Administrator menedzser-modul található (active-standby módban), melyeken keresztül a rendszer felügyelete webböngészőt használva egyszerűen, gyorsan és hatékonyan végezhető.

A 3. ábra az Onboard Administrator szoftverből származik. Előlnézetből a 8 blade-szerver, alattuk a 6 tápegység, hátulnézetből 5 ventilátor, alattuk az Ethernet switchek, a SAN switchek, az Onboard Administrator modulok, majd ismét 5 ventilátor és legalul a tápcsatlakozók láthatók.

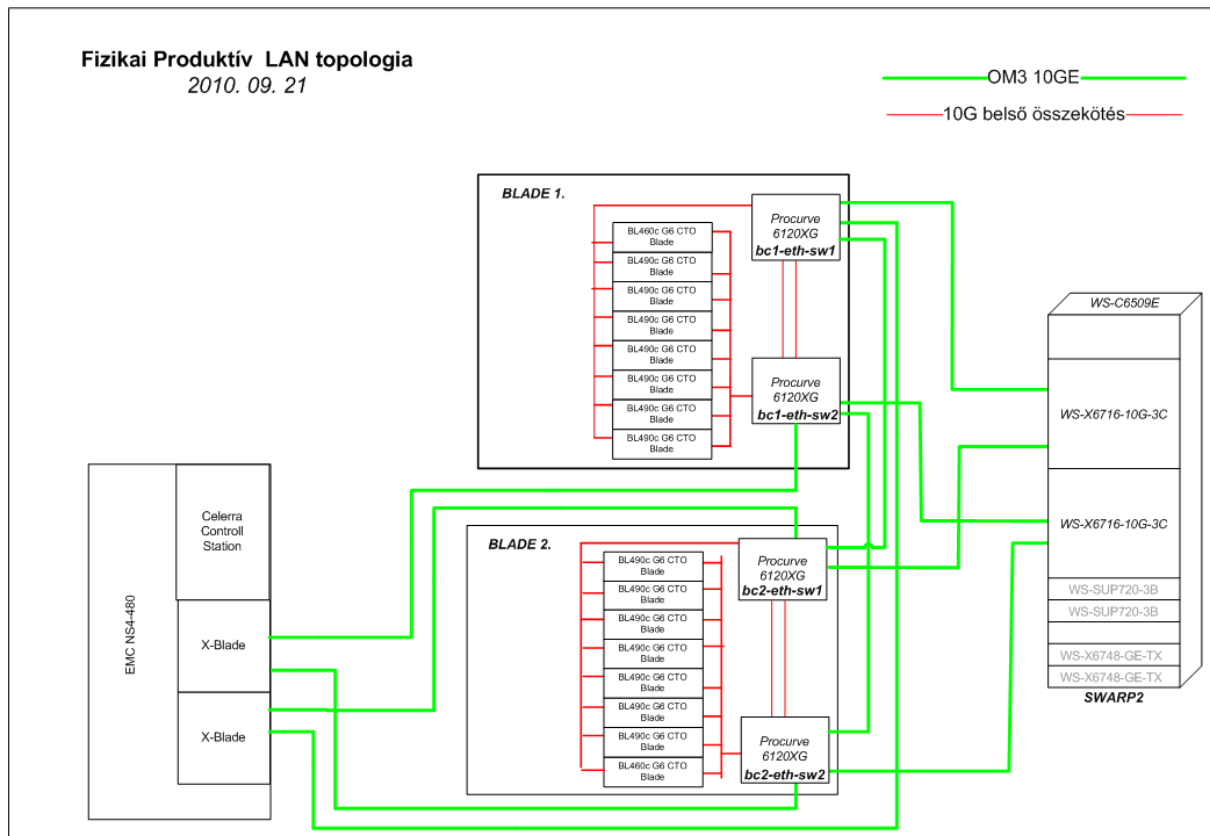


3. ábra: A blade keretek fő alkotóelemei

4.1.1. Hálózat

4.1.1.1. Ethernet

A 4 db HP Procurve 6120XG blade switch (továbbiakban 6120XG) egy-egy dedikált 10 Gbps uplinkkel rendelkezik a központi Cisco Catalyst 6509 switch (továbbiakban C6509) irányába. Az egyes blade-eken belül kettős hátlapi összeköttetésekkel kapcsoljuk össze a 6120XG switcheket, ezek szintén 10 Gbps sebességűek, de egy 20 Gbps-os logikai egységbe összefogva (port trunking) működnek. Mindemellett két keresztirányú 10 Gbps kapcsolat erősíti a magas szintű rendelkezésre állást a két blade-keret között. Minden kapcsolat 802.1Q trunk. Az EMC X-Blade moduljainak minden 6120XG-vel van összeköttetése.



4. ábra: A LAN hálózat produktív részének fizikai összeköttetései

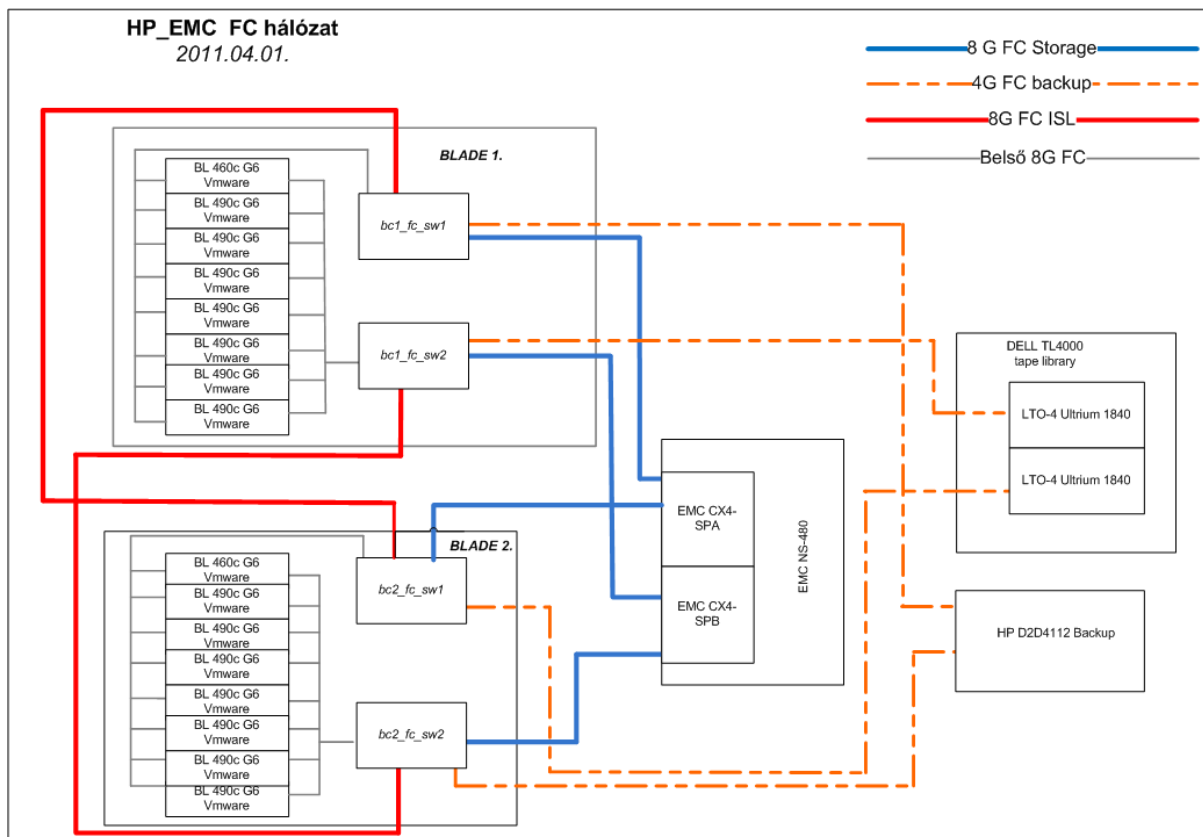
Spanning tree

A központi C6509 switchen a Cisco saját Per-VLAN Spanning Tree Plus (PVST+) protokollja fut, míg a 6120XG switcheken a szabványos IEEE802.1D protokoll. Felmerült lehetőségként a szabványos IEEE802.1s protokoll (MSTP) használata, de ezt jelen projekt keretében nem kívántuk bevezetni. A HP és a Cisco eszközök között a VLAN 1-ben történik meg a közös 802.1D példány futtatása. Ez a HP switchen a teljes switchre, a Cisco esetében csak a VLAN 1-re vonatkozik. Tekintettel arra, hogy a C6509 switch a root bridge, ez az eltérő értelmezés nem okoz problémát. Az elsődleges útvonalak a 6120XG-C6509 közötti uplinkek. Valamely elsődleges uplink meghibásodása esetén a hátlapi 20 Gbps-os kapcsolatra terelődik a forgalom (e link költsége kisebb a 10 Gbps-os keresztirányú kapcsolaténál). Abban a kivételesen ritka, de elméletben lehetséges esetben, amikor a 10 Gbps-os uplink és mindkét hátlapi csatlakozás meghibásodik, a keresztirányú kapcsolatra terelődik a forgalom. Az útvonalak finomhangolása, egyes VLAN-ok más irányokba történő továbbítása a HP oldali IEEE802.1D miatt nem lehetséges, az eszközök megtartása mellett erre csak az IEEE802.1s nyújt lehetőséget. Szintén az IEEE802.1D protokoll sajátosságai közé tartozik a lassú konvergencia, nagyságrendileg akár 30 másodpercre is lehet számítani.

4.1.1.2. SAN

A blade-keretben lévő HP 8/24c SAN switchek egy-egy belső sínre csatlakoznak, ezekre a sínekre kapcsolódnak a blade-szerverek HBA-i.

A két SAN switch 2×8 GB FC (ISL) külső kapcsolattal van összekötve, a storage és a blade-szerverek 8 GB FC, a mentőeszközök 4 GB FC kapcsolattal rendelkeznek a SAN switchek felé.



5. ábra: A SAN rendszer kapcsolatai

SAN-architektúra

A blade-keretekben levő két-két SAN switch egy redundáns Dual Fabric SAN-architektúrát valósít meg. A SAN zónák kialakításánál és a storage LUN-ok prezentációjánál csak a fizikai VMware ESXi (blade) szervereket vesszük alapul, a virtuális gépek nem kapnak LUN-okat.

4.1.1.3. Menedzsment

Biztonsági okokból egy ilyen rendszer menedzsmenthálózatát el kell zárni a külvilágtól. Ezért a rendszermenedzsment IP-címeit két privát subnetből osztottuk ki. Az egyik subnetbe kerültek a virtualizációs szoftver építőelemei. A másik subnetbe az összes többi eszköz (storage, D2D, tape, blade és a mentőszerver) menedzsmentinterfésze.

Fizikai MGMT LAN topologia
2010. 09. 13.

DELL TL4000

BLADE 1.

Enclosure midplane

BL480c G6 CTO Blade

BL480c G6 CTO Blade

BL480c G6 CTO Blade

BL480c G6 CTO Blade

BL480c G6 CTO Blade

BL480c G6 CTO Blade

BL480c G6 CTO Blade

bc1_eth_sw1

bc1_fc_sw1

bc1_fc_sw2

bc1_eth_sw2

Onboard administrator tray

bc1_oe1

bc1_oe2

Dell 710

SWU1

Dell 710

SWU2

WS-C6509E

SWARP2

HP D2D4112 Backup System

BLADE 2.

Enclosure midplane

BL480c G6 CTO Blade

BL480c G6 CTO Blade

BL480c G6 CTO Blade

BL480c G6 CTO Blade

BL480c G6 CTO Blade

BL480c G6 CTO Blade

BL480c G6 CTO Blade

bc2_eth_sw1

bc2_fc_sw1

bc2_fc_sw2

bc2_eth_sw2

Onboard administrator tray

bc2_oe1

bc2_oe2

EMC NS-480

SPA

X-Blade 2

Celerra Control Station

X-Blade 3

SPB

4.2. Blade-üzemeltetés

A napi üzemeltetés során eddig előfordult legjelentősebb tennivalónk a firmware-frissítés volt. Ezt az OA esetén webböngészőn keresztül végeztük el. A szerverekhez a lokális I/O-kábellet és egy USB-elosztóval csatlakoztattunk monitort, USB-s egeret, billentyűzetet és egy bootolható pendrive-ot az új firmware-rel, majd végrehajtottuk az upgrade-et.

Szerverkonszolidáció az SZTE ESZK-ban

5. Storage

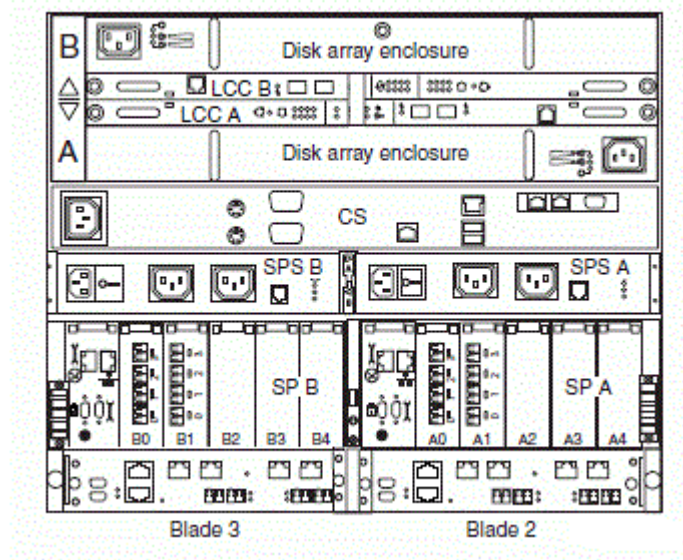
5.1. EMC Celerra NS-480 telepítés

5.1.1. Az NS-480 hardverelemei

Disk array enclosure (DAE – továbbiakban fiók): ezekben helyezkednek el a diszkek, egy fiókba 15 diszk rakható. Jelenleg hat fiókunk van, amelyekben 30 db 450 GB-os FC-, 45 db 1 TB-os SATA- és 9 db 2 TB-os SATA-diszk található.

Control station (CS): Ez biztosítja a NAS-felülethez való hozzáférést mind ssh-n, mind GUI-n keresztül.

Standby power supply (SPS) A és B: a storage processzorok és a legelső (rendszerlemezeket tartalmazó) fiók tartalékolat áramellátására szolgálnak.



7. ábra: Az NS-480 felépítése

Storage processor (SP) A és B: kettő, egymást helyettesíteni képes storage processor, GUI-s hozzáférési felülettel.

Blade 2 és 3: a NAS szolgáltatást biztosító kettő darab X-Blade.

5.1.2. Konfiguráció

A storage-rendszerünk induláskor 1 TB-os SATA- és 450 GB-os FC-diszkeket tartalmazott. Azóta egy bővítésnek köszönhetően 1 és 2 TB-os SATA-diszkekkel sikerült növelni a tárhelykapacitást. A jelenlegi raid groupok a 8. ábrán láthatók.

A storage-ban van még 1 db 450 GB-os FC, 3 db 1 TB-os SATA és 1 db 2 TB-os SATA hot spare diszk, illetve külön tárolva rendelkezünk egy hideg tartalék 450 GB-os FC-diszkekkel.

A 0-s sorszámú FC raid group az NS-480 működéséhez szükséges. A blade-rendszer részére kialakított raid groupoknál RAID 6-ot választottunk magas fokú rendelkezésre állásra törekedve (RAID 6 esetén kettő diszk egyidejű kiesése sem jár adatvesztéssel). Tettük ezt annak ellenére, hogy a szállító külön felhívta a figyelmünket, hogy „a RAID 6 25%-kal gyengébb IOPS teljesítményt jelent írásra, mint a RAID 5”. A raid groupokon LUN-okat hoztunk létre, majd a különböző raid

Raid group sorszáma	Diszk méret	Diszk típus	RAID szint	Diszkek száma/group
0	450 GB	FC	5	4+1
1	450 GB	FC	6	10+2
2	450 GB	FC	6	10+2
3	1 TB	SATA	6	12+2
4	1 TB	SATA	6	12+2
5	1 TB	SATA	6	12+2
6	2 TB	SATA	6	6+2

8. ábra: Raid groupok

groupokon található LUN-okat MetaLUN-okká fogtuk össze, hogy minél egyenletesebb diszkterhelést érjünk el. A MetaLUN-ok kerültek kiajánlásra a blade-rendszer VMware ESXi-t futtató mindenegybes fizikai hostjának.

Az email riasztási rendszerben az „error” és „critical” szintű üzenetek kiküldését állította be a szállító.

5.2. EMC Celerra NS-480 üzemeltetés

5.2.1. Storage-hibák

A rendszer 2010. szeptemberi telepítése óta az eredetileg beszerzett 30 db FC-diszkből 8-at, a 40 db 1 TB-os SATA-diszkből pedig 6-ot kellett cserélni (2013. január közepéig). A 2012. januári bővítéskor beszerzett további 5 db 1 TB-os és 9 db 2 TB-os SATA-diszkek közül eddig egy sem hibásodott meg. Mivel 2012 januárjában és októberében egy raid groupon belüli kettős diszkhiba is előfordult, ez igazolta a RAID 6 használata mellett szóló döntést.

A diszkmeghibásodásokon kívül 2012 novemberében a B storage processor egyik I/O-modulját is cserélni kellett.

5.2.2. Egyéb üzemeltetési tapasztalatok

Az email riasztási rendszert az „error” és „critical” szintűnél alacsonyabb prioritású elemekkel kellett bővíteni, hogy egy diszkcseré a kezdetétől (a proaktív másolás megkezdése a hot spare-re) a végéig (a kicserélt diszk szinkronizálásának befejezése) nyomon követhető lehessen ezen a módon.

A 2012. januári diszkbővítéskor azt vettük észre, hogy az újonnan létrehozott MetaLUN-oknál a rendszer nagyon gyakran lecseréli az elsődlegesen hozzájuk rendelt storage processort („trespass”). Ennek az volt az oka, hogy a bővítés előtt a storage-ról töröltük a tesztelési célból létrehozott LUN-okat, ugyanakkor a VMware oldalon nem végeztük el az elérhető datastore-ok újraszkenelését. Tehát az új MetaLUN-ok létrehozása és a storage-ban történt kiajánlása után a VMware nem azokat a datastore-okat látta, amelyekről tudomása volt. Ennek eredményeképpen lépett fel a gyakori trespass. A VMware-oldalon a datastore-ok újraszkenelése megszüntette ezt a problémát.

6. VMware

6.1. VMware vSphere (ESXi) telepítése

Virtualizációs platformnak a VMware vSphere 4.1-es verziójának Enterprise Plus Edition csomagját választottuk. Ez biztosítja az általunk szükségesnek tartott magas rendelkezésre állás (High Availability, HA), automatikus terheléselosztás (Distributed Resource Scheduler, DRS) és a közös virtuális switch (Distributed Virtual Switch, DVS) funkciókat. A rendszer fizikai alapját a 14 db HP BL490c G6 szerver adja, az adatok (virtuális gépek) tárolása az EMC Celerra NS-480 storage-on történik. A szerverekben nincs saját merevlemez, csak egy 2 GB-os pendrive, amely az ESXi-t tartalmazza.

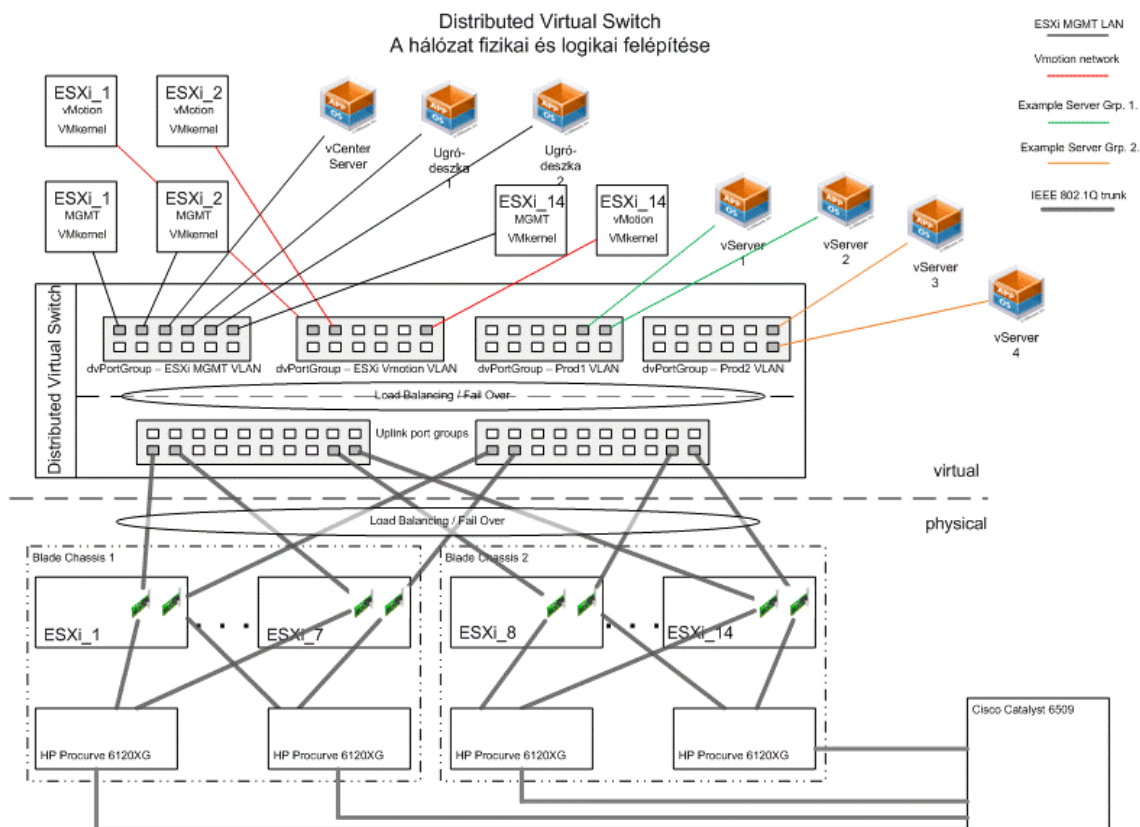
A virtuális infrastruktúra kialakításának első lépéseként telepítésre került a szerverekre az ESXi 4.1-es hypervisor, és megtörtént a hostok alapkonfigurációjának a kialakítása. A rendszer menedzsmenthálózata privát, csak az ugródeszka gépeken keresztül érhető el (l. 4.1.1.3.).

A virtuális gépek (VM-ek) datastore-nak nevezett adattárolási egységekre kerülnek. A storage-rendszeren létrehozott (Meta)LUN-ok kerültek kiejánlásra az ESXi-hostok számára, ezek önállóan, vagy belőlük többet összekapcsolva alkotnak egy datastore-t. A LUN-ok kialakításakor több dolgot kellett figyelembe venni: milyen típusú diszken lesznek, hány diszkből áll az adott raid group, mekkora méretű LUN-t képes kezelni a VMware. Az ESXi 4.1-es verzióban a LUN méretére van egy 2 TB-os korlát, ezért a VMware számára kiejánlott LUN-ok méretét valamivel 2 TB alattira választottuk. A LUN-okat extentekeként összefogva 2 TB-osnál nagyobb datastore-ok is létrehozhatók. Jelenleg a datastore-jaink egy, kettő vagy három LUN-ból állnak. Adminisztratív és biztonsági megfontolásokat, valamint az I/O-igényt figyelembe véve döntjük el, hogy melyik datastore-ra kerüljenek a VM-ek merevlemezei. Például az ETR rendszert futtató virtuális gépek számára kialakítottunk egy fc_etr és egy sata_etr datastore-t.

Következő lépésként a vCenter Servert kellett installálni, ami összefogja, központilag felügyeli az ESXi-eket. Több lehetőség közül azt választottuk, hogy a vCenter Server is egy virtuális gép legyen, az általa menedzselte vSphere-en belül (szállítói ajánlás). Az egyik ESXi-hoston létrehoztunk egy virtuális gépet, majd erre telepítettük az operációs rendszert (MS Windows Server 2008 R2) és a vCenter Server-t, amely az adatai tárolására MS SQL Server 2008R2-t használ. Most már el lehetett kezdeni kialakítani a virtuális infrastruktúránkat.

6.1.1. Hálózat

A hálózat kialakításakor fontos követelmény volt a redundancia biztosítása, valamint a különböző típusú forgalmak (menedzsment, produktív, vMotion) szeparálása. Fizikailag a szerverekben 2 db 10 Gbps-os Ethernet port van, a nic0 az 1. blade switchbe, a nic1 a 2. blade switchbe csatlakozik. A forgalom elkülönítését VLAN-ok használatával oldottuk meg, az ESXi-szerverek, a blade switchek, valamint a központi switch között 802.1Q trunk portok segítségével jutnak el a VLAN-információk a virtuális switchhez. A szerverek két Ethernet adapterét a VMware load balanced failover üzemmódban használja, vagyis nemcsak a terheléselosztás valósul meg, hanem az egyik adapter hibájakor a másik veszi át a forgalmat. A rendszerben egyetlen Distributed Virtual Switchet hoztunk létre két uplink porttal.



9. ábra: Distributed Virtual Switch

Az ESXi-szerverek nic0 portjai a dvUplink1-be, míg a nic1 portok a dvUplink2-be csatlakoznak. A Distributed Virtual Switchben több dvPortGroupot definiáltunk (gyakorlatilag egy portgroup ~ egy VLAN, így mindegyik dvPortGroup saját VLAN ID-vel rendelkezik):

- a menedzsmentforgalom számára, az ESXi-szerverek menedzsment VMkernel portjai ide csatlakoznak,
- a vMotion- (és Fault Tolerance-) forgalom számára, az ESXi-szerverek vMotion VMkernel portjai ide csatlakoznak,
- a produktív forgalom számára több dvPortGroupot, pl.:
 - ETR
 - Mail Firewall
 - Ext-Serv

6.1.2. Cluster

A rendelkezésünkre álló 14 ESXi-ből alakítottuk ki a magas rendelkezésre állást (HA: a host hibája esetén a rendszer az adott gépen futó VM-eket egy másik hoston indítja el) és automatikus terheléelosztást (DRS: a rendszer figyeli az egyes hostok terhelését, és szükség esetén a VM-eket másik hostra mozgatja ~ vMotion) biztosító cluster-t. A clusterben lévő virtuális gépek mennyisége lehetővé tette számunkra, hogy a 14-ből 2 hostot kivegyünk, így most a produktív cluster 12 ESXi-ből áll, és egy 2 hostból álló cluster segíti a tesztelést pl. patchek telepítésekor, de a vSphere 5.0-ra való átállás előkészítésében is nagy szerepet kapott. A produktív clusterben

egyelőre elegendő erőforrás van ahhoz, hogy akár 2-3 fizikai host kiesése esetén is biztosítani tudjuk a szolgáltatásokat.

Mivel a VM-ek emberi beavatkozás nélkül is mozoghatnak az egyes fizikai hostok között, ezért olyan operációs rendszer licenceket kellett beszerezni, amelyek érdemben nem korlátozzák ezt a mozgást. Az ESZK vett 28 db MS Windows Server 2008 R2 Datacenter Edition licencet, ami korlátlan számú Windows Server futtatását teszi lehetővé minden egyes CPU-n, valamint csatlakozott az OpenEDU programhoz, melynek folytán Red Hat licencekhez jutott.

6.2. VMware vSphere (ESXi) üzemeltetése

6.2.1. Vmreg

A virtuális gépek igénylésének és nyilvántartásának megkönnyítésére létrehoztunk egy webes felületet. Ezen a felületen a VM-adminisztrátorok – az ESZK által jóváhagyott regisztráció után – egy formot kitöltve kinyomtathatják a tényleges igénylőlapot (amely természetesen bekerül az adatbázisunkba is), majd a megfelelő személyekkel aláíratva azt vissza kell juttatniuk az ESZK-ba.

6.2.2. Migráció

A szerverek betelepítésekor Intel-platformon Microsoft Windows és GNU Linux, valamint SPARC-platformon Solaris operációs rendszer alatt futó szolgáltatások kerültek a blade-rendszerben létrehozott virtuális gépekre. Egy Novell Netware operációs rendszerű gép migrálása még nem történt meg.

A vSphere-be költözött virtuális gépek több csoportba oszthatók. Súlyát tekintve a legnagyobb a tanulmányi rendszert (ETR) futtató gépek csoportja, de egyetemi oktatók/hallgatók levelezését, a hálózat menedzsmentjét biztosító szerverek, levelezési tűzfalak, webportál is virtualizálásra kerültek. A migráció során nem használtuk ki a fizikai gépek importálásának lehetőségét, ennek részben a megváltozott hardver, az újabb verziójú operációs rendszer és szoftverek, részben a próbaképpen konvertált fizikai gépek nem megfelelő működése volt az oka. A linuxos gépek többségét egyesével, nulláról kezdve telepítettük be, annyira különböznek egymástól. A windowsos gépek esetén elkészült egy template (amit időnként egy windowsos kolléga frissít), és ebből hoztuk létre a windowsos szervereket.

Az eredetileg Solaris alatt futó szolgáltatások CentOS-re, illetve ahol fontos a támogatás, Red Hat-re kerültek. A betelepítéskor oda kellett figyelni, hogy egyazon rendszerprogram nem feltétlenül ugyanúgy paramétereződik Solaris és Linux alatt, attól függetlenül, hogy mindkettő Unix típusú operációs rendszer. A megírt scriptek egy részét ezért módosítani kellett. Az egyedileg fordított programokhoz vagy megfelelő linuxos telepítő csomagot kerestünk, vagy ha ilyet nem találtunk, megpróbálkoztunk a fordítással. Ez sajnos nem mindig járt sikerrel, egy-két meglehetősen régi program (például egy parancssorból is kényelmesen meghívható telefonkönyv) használatáról le kellett mondani.

6.2.3. vSphere 5.0

A rendszert 4.1-es verziójú vSphere-rel telepítették, ám időközben kijött, majd biztonságossá érett az 5.0-ás verzió is, ezért úgy döntöttünk, hogy elvégezzük az upgrade-et az 5.0-ra. Már több hónapja elérhető az 5.1-es verzió is, de az még „túl fiatal”. Így hosszas előkészítés és több tesztcluster (köztük teljesen virtualizált is volt) upgrade-je után végrehajtottuk az átállást, ami a leírásokat követve gond és probléma nélkül zajlott. Jelenleg még hátravan a storage átalakítása, a vmfs3 upgrade-je helyett az új vmfs5 használatát javasolják. Így több más előny mellett megszűnik a 2 TB-os határ a LUN-ok méretére, azaz egy egyetlen LUN-ból álló datastore is elég nagy méretű lehet.

6.2.4. Virtuális gépek, üzemeltetés

A virtuális infrastruktúra menedzselését és felügyeletét a vSphere-adminisztrátorok látják el.

Ha nem is mindennapos, de gyakori feladat új VM-ek létrehozása, esetleg meglevő VM-ek módosítása (ez eddig az esetek többségében bővítést jelentett).

A VM-ek igen változatosak, a legkisebbtől (32 bites, 512 MB RAM, 1 CPU, 8 GB diszk) a jelenleg használt legnagyobbig (64 bites, 32 GB RAM, 8 CPU, több mint 1,5 TB diszk) több különböző konfiguráció van. Közös bennük, hogy a merevlemezük ún. thin provisioning típusú. Ez azt jelenti, hogy létrehozáskor a rendszer nem foglalja le a teljes méretet, hanem egy kis diszkkal indít, és futás közben a szükséges mértékben növeli a diszk valós méretét, maximum a megadott értékig. Így kicsit lassabb lesz az elérés, de a datastore valós méreténél több diszktérületet is ki lehet osztani a virtuális gépek számára.

Több VM létrehozásakor jelezni lehet az esetleges különleges igényeket (pl. két vagy több VM mindig egy fizikai hoston legyen – így gyorsabb lehet közöttük a kommunikáció, vagy biztonsági okok miatt bizonyos VM-ek ne kerüljenek egy ESXi-re).

A VM-ek felügyelete az adott gép rendszergazdájának a feladata, ezért ők korlátozott jogosultságú hozzáférést kapnak a vCenter Serverhez, amelyet vagy vSphere Clienttel, vagy vSphere Web Clienttel tudnak elérni.

Mint minden szoftverhez, a vSphere-hez is rendszeresen készülnek javítások, frissítések.

A patchek telepítése félig automatizáltan, a vCenter Update Manager segítségével történik. Ez jóval kényelmesebb, mint a kézi letöltés, majd telepítés. Az Update Manager automatikusan ellenőrzi az újabb frissítések megjelenését, és az előre definiált baseline-oknak (Critical Host Patches, Non-Critical Host Patches) megfelelően letölti a megjelenő patchek metainformációit. A metainformációk alapján automatikusan képes ellenőrizni, hogy az objektumhoz (host/cluster) hozzákapcsolt baseline mely frissítései telepíthetők a hostokra. A teljesen automatikus telepítés azt jelentené, hogy egy célzottan létrehozott, időzített feladat (Scheduled Task) elvégzi a letöltést és a telepítést is. Ezt azonban a szállító sem javasolta, és mi is biztonságosabbnak tartjuk, hogy a telepítés a vSphere-adminisztrátorok felügyelete alatt történjen. Így nem fordulhat elő, hogy még túl friss, esetlegesen hibás patchet telepítsünk, vagy ha bármiért megakad a telepítés folyamata (pl. host újrabootolásakor), van mód a közbeavatkozásra. A cluster konfigurációját minimálisan módosítani kell a patchelés végrehajtásához (HA-, DRS-funkciók ideiglenes kikapcsolása), de ezt hálózati konfigurálás (pl. a központi switch beállításainak változtatása) esetén is meg

kell tenni. A VM-ek leállítására nincs szükség, az Update Manager elvégzi a VM-ek átmozgatását az éppen frissítendő hostról a cluster többi hostjára. A VM-eken esetlegesen a VMware Tools frissítése igényelhet újraindítást. Nagyobb verzióváltáskor (most utoljára pl. vSphere 5.0 update 2) szükséges lehet a vCenter Server, Update Manager, vSphere Client frissítése is. Ez az MS Windows-ban megszokott módon történik, a DVD-n (ISO-n) lévő megfelelő .exe fájl futtatásával.

A különleges eseményekre (extrém CPU-használat, diszktelítettség stb.) riasztásokat definiáltunk, de a rendszer aktív monitorozása a vSphere Client segítségével nem hagyható el.

7. Mentés

7.1. Networker telepítés

7.1.1. A mentőrendszerrel

A központi mentőrendszer feladata a mentésbe bevont, blade-en belül lévő szerverek fájlrendszerének rendszeres, napi ütemezett mentése (elsődleges mentőeszköz: VTL (Virtual Tape Library), diszkalapú/HP D2D4112), ill. a mentések másolatainak elkészítése (klónozás) az elsődleges mentőegységről fizikai szalagra (másodlagos mentőeszköz: szalagalapú/Dell TL4000).

A mentésben résztvevő eszközök:

- HP D2D4112 – diszkalapú mentőeszköz, 24 db 1 TB-os SATA-diszk, 18 TB nettó kapacitással, redundáns tápellátás:
 - 6 db D2D VTL (D2DBS Generic emuláció, összesen 6×4 db drive, 6×144 slot, FC LTO5 drive).
- Dell TL4000 tape library – szalagalapú mentőeszköz, 48 slot, 2 FC LTO4 drive, redundáns tápellátás.
- Mentőszerver (nwrkr1):
 - hw.: HP BL460C G6 (2×Intel Xeon E5530 2.4 GHz, 12 GB RAM, dual port 8 Gb FC HBA, dual port 10 GbE, 2×300 GB SAS HDD)
 - OS: CentOS 5.5 x86_64,
 - EMC Networker 7.6.4 x86_64.
- Tartalék mentőszerver (nwrkr2): a mentőszerverrel megegyező.
- Kliensek (Linux, Windows).

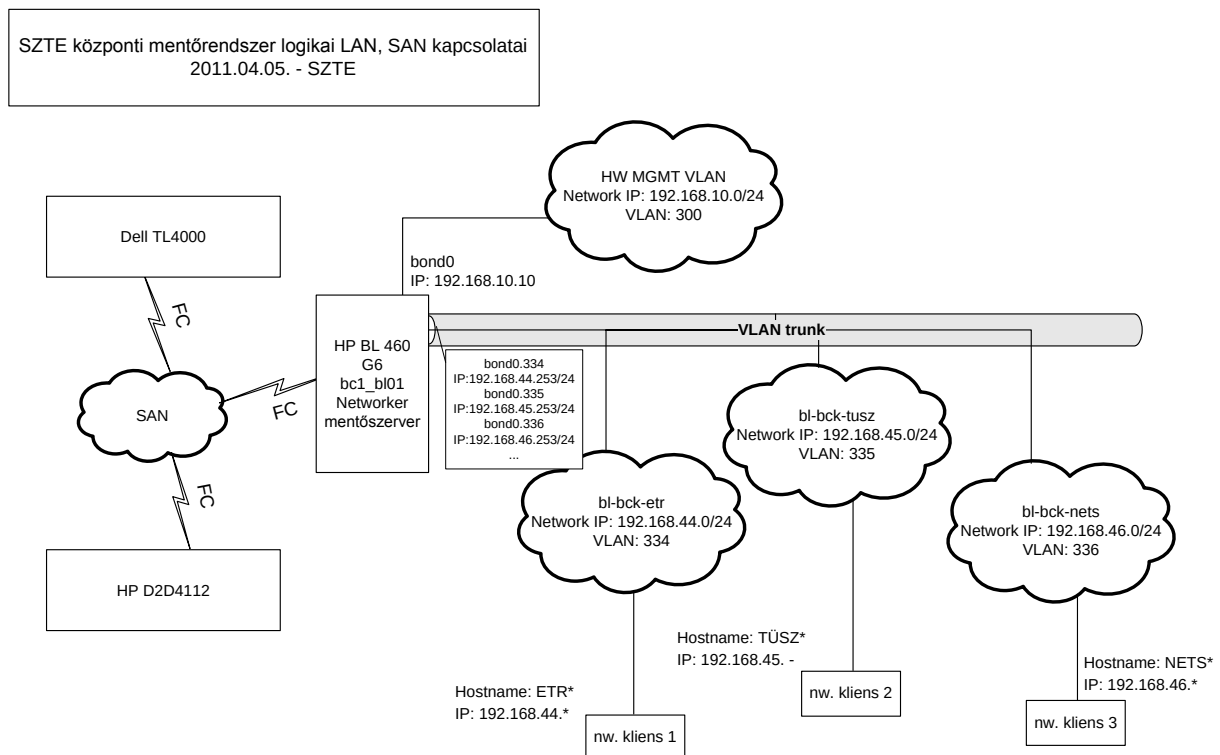
A VTL esetén azért választottuk az LTO5-ös szabványt, mert így nagyobb tárolókapacitást tudunk lefedni ugyanannyi szalaggal. A szalagos mentőegység megvásárlásakor még csak LTO4-es drive-ok és szalagok léteztek, így itt adott volt a választás. Ez a méretbeli különbség a klónozás során nem jelent problémát.

A blade-eszközök leszállításával egyidejűleg a szállító összeszerelte és beüzemelte a mentési alrendszer hardverkomponenseit. A végleges szoftveres telepítést a projekt végére ütemeztük, mivel a mentések kialakításához szükséges volt a mentendő kiszolgálók megléte.

A mentési alrendszer szoftveres telepítése a KFKI Zrt. (jelenleg T-Systems) kollégái közreműködésével történt. A diszkalapú mentőeszközben létrehozott 6 db auto-changert felosztottuk két részre: 4 db-ot jelöltünk ki mentésre és 2 db-ot tartalékként

fenntartottunk. A 4 db autochangeret felosztottuk a mentendő szerverek között, bizonyos csoportosítás alapján, azaz adott szervereket csak adott autochangerre mentünk, ezzel is elkülönítve egymástól a mentéseket. A rendszer kialakítása után több oktatást tartottak az SZTE ESZK kollégái részére a mentési alrendszer használatával kapcsolatban, melynek során mentési mintafeladatokat hoztunk létre.

7.1.2. Logikai hálózati topológia



10. ábra: A központi mentőrendszer logikai kapcsolatai

Az ábra célja – egy-egy mintakliensen keresztül – a mentésbe bevont szerverek LAN-topológiában elfoglalt helyének és kapcsolatainak, ill. a mentőszerver és a mentőeszközök SAN-kapcsolatainak megjelenítése. A VLAN-okban egy-egy mentésbe bevont mintaszervert ábrázoltunk: nw. kliens 1, nw. kliens 2, nw. kliens 3. A mentőszerver két Ethernet interfésszel kapcsolódik a hálózatokhoz, ezeket az OS által nyújtott bonding technikával egy virtuális interfészbe (bond0) összefogjuk: ennek célja a rendelkezésre állás fokozása, ill. terhelésselosztás az interfészek között.

7.1.3. A központi mentőrendszer mentőszoftvere

Az EMC Networker FastStart 7.6.4 mentőszoftver a bc1_c01-re (az elsődleges mentőszerver: nwrkr1) került telepítésre. A mentőszerver a beállított mentési rend szerint napi rendszerességgel, ütemezetten mentéseket készít elsődlegesen VTL-re (biztosítva a gyors visszaállíthatóságot), majd onnan klónozással szalagra másolja

a mentéseket, mivel több esetben bizonyos mentéseket szalagon, páncélszekrényben kell elhelyezni. A mentőeszközök SAN-on keresztül kapcsolódnak a mentőszerverhez. A kliensek mentése LAN-on keresztül történik.

A kliensekről a következő módokon készülhet mentés:

- **Natív kliens-fájlrendszer mentés:**
A központi mentésbe bevont szerverre Networker-klienscsomagot kell telepíteni. A mentés fájl(rendszer) alapú.
- **Networker-modulos mentés:**
A központi mentésbe bevont szerverre Networker-klienscsomagot és az adott alkalmazás Networker-modulját kell telepíteni. A modul lehetővé teszi az adott alkalmazás (Oracle, Exchange, SQL) online, konzisztens mentését.
- **Egyéb szerverek mentése közvetett módon (klienscsomag telepítése nélkül):**
A mentőszerverről kezdeményezett módon [cronjob/mentő (pl. shell)script], ssh kulcsos autentikációval, image-mentést (pl. cpio) készítünk a mentőszerver „/u/backup” könyvtárstruktúrája alá. A „/u” fájlrendszerről a mentőszoftver napi rendszerességgel mentést készít.

Az elkészült mentések visszaállítására több lehetőség is van:

- MS Windows esetén „Networker User” natív windowsos GUI alkalmazással.
- Linux/Unix esetén „nwrecover” Motif GUI alkalmazással.
- Linux/Unix esetén „recover” parancssoros alkalmazással.

7.1.4. Tartalék mentőszerver

A tartalék mentőszerver (nwrkr2) az éles mentőszerver másolataként lett előállítva. A szerverek beállításai többségében megegyeznek.

7.1.5. A mentési feladatok kialakítása

Az SZTE ESZK mentési alrendszert üzemeltető kollégái a blade-rendszerbe költöztetett szerverek adminisztrátorai által megfogalmazott mentési igényeknek megfelelően létrehozták a szükséges mentési feladatokat. Minden esetben először VTL-re készítünk mentéseket, majd ha szükséges, akkor ezekből a mentésekből válogatva klónozzuk a már elkészült mentéseket fizikai szalagra. Azokat a szalagokat, melyeket valamilyen ok miatt cserélni kell vagy páncélszekrényben kell elhelyezni (pl. ETR-mentés), előre kialakított rend szerint cseréljük.

7.2. Networker üzemeltetés

7.2.1. A nwrkr1, nwrkr2 mentőszerverek üzemeltetése

A mentőszerverek (nwrkr1, nwrkr2) adminisztrációs célú elérése történhet közvetlenül HP blade management felület használatával a blade konzolján, vagy távolról (az ugródeszka szerverekről) ssh-val hálózaton keresztül.

A szervereken mint CentOS 5 OS-t futtató hostokon nincs olyan különösebb változó rendszerelem, amely rendszeres napi adminisztrációt kívánna. Az elsődleges mentőszerről (nwrkr) napi rendszerességgel mentések készülnek. A szervereken OS-paraméter módosítása esetén figyelni kell a módosítások szimmetrikus átvezetésére mind a két mentőszerveren (nwrkr1, nwrkr2). A szerverek (nwrkr1, nwrkr2) hardverüzemeltetése a blade-keretek többi szerveréhez hasonlóan történik.

A nwrkr1 mentőszerveren egy script monitorozza a fontosabb rendszerparamétereket, ha ezek elérnek egy adott hibahatárt, akkor a script riasztást küld egy előre definiált email címre.

7.2.2. Mentőeszközök

7.2.2.1. VTL (HP D2D4112)

A virtuális szalagrendszer vezérlését kizárólag az éppen aktív mentőszerver végzi az EMC Networker segítségével. A VTL menedzselése webes GUI használatával történik. Esetlegesen felmerülő hibák esetén a rendszer egy előre definiált email címre küld riasztást. Az eseménynaplókat a webes GUI-ban lehet megnézni.

A mindennapos üzemeltetés része a VTL állapotának és a virtuális szalagok állapotának ellenőrzése.

Szükség esetén a napi üzemeltetés során új szalagokat szoktunk létrehozni, vagy már meglévő szalagot szoktunk törölni, mozgatni vagy szerkeszteni.

7.2.2.2. Dell TL4000

A robotika vezérlését kizárólag az éppen aktív mentőszerver végzi az EMC Networker segítségével, de szükség esetén az eszköz saját webes felületén vagy az eszköz előlapi paneljének használatával is lehet vezérelni. A tape library menedzselése webes GUI használatával történik. A felmerülő hibák esetén a rendszer előre definiált email címre küld riasztást. Az eseménynaplókat a webes GUI-ban lehet megnézni.

A mindennapos üzemeltetés része a tape libraryk és a szalagok állapotának ellenőrzése.

Szükség esetén a napi üzemeltetés során új szalagokat szoktunk címkézni, vagy már meglévő szalagot szoktunk törölni, mozgatni, szerkeszteni, cserélni.

7.2.3. EMC Networker

A Networker adminisztrációja a Networker Management Console (NMC) GUI használatával történik. Az NMC-t böngészőből lehet elérni. A Networker konfigurációs beállításai rendszeres/napi szintű módosítást nem igényelnek. A Networker emailt küld a mentések eredményéről.

NMC GUI használatával az alábbi feladatokat szoktuk elvégezni napi szinten:

- logok ellenőrzése,
- mentések, klónozások monitorozása,
- szalagok állapotának ellenőrzése,
- poolok állapotának ellenőrzése,
- szükség esetén szalag címkézése és poolhoz társítása,
- mentések időzítésének finomhangolása.

7.2.4. Szalagkezelés Networkerben

A Networker automatikusan kezeli a tape library szalagtároló helyeit, a szalagmeghajtókat. Vagyis elvileg nincs szükség a szalagmeghajtókba történő operátori manuális betöltésre vagy onnan való eltávolításra (ezeket a szalagmozgatásokat, robotika-vezérlést a Networker automatikusan elintézi, ha erre szükség van). Ha mégis szükség lenne kézi beavatkozásra (pl. szalagot kivenni a szalagmeghajtóból), akkor ezt az NMC GUI használatával is meg lehet tenni.

7.2.5. Networker szerver teljes visszaállítása

Amennyiben a mentőszerver megsérül, akkor egy teljesen friss OS telepítése és a Networker telepítése után a szalagon vagy D2D-n lévő mentésből visszaállítható a rendszer teljes működőképessége.

7.2.6. Hardver hibák

A rendszer 2010. szeptemberi telepítése óta (2013. január közepéig) a D2D mentőeszközben lévő SATA diszkekből 7 db-ot kellett garanciában kicserélni. Ez nem járt adatvesztéssel, mivel a mentőeszközben RAID 6-ot használunk, és nem egyszerre hibásodtak meg a diszkek.

2012 júniusában az egyik mentőszerver alaplapját is ki kellett cserélni.

8. Fejlesztési lehetőségek

A rendszert horizontálisan és vertikálisan is könnyedén bővíthetjük.

8.1. Blade

A jelenlegi blade-szervereket értelmesen csak memóriával lehet bővíteni.

A két keretbe még 16 db félmagas szervert lehet beszerezni. Jelenleg Gen8-as szerverek kaphatók.

Interconnect-modulokat és mezzanine-kártyákat is vehetnénk a jelenlegi keretekbe és szerverekbe, de ennek nem látjuk szükségét, mert a jelenlegi Ethernet és FC hálózat sávszélessége bőségesen elegendő.

8.2. Storage

A nálunk lévő NS-480 storage-ba a maximális kiépítésénél 480 db HDD-t lehet beszerezni. Mi jelenleg 84 db HDD-t használunk. Háromfajta (SATA, FC és SSD) HDD-t lehet a storage-ba beszerezni többféle kapacitással. (Természetesen gondoskodni kell a megfelelő számú fiókról.)

A pénzügyi feltételeket és az üzemeltetési tapasztalatokat figyelembe véve jelenleg csak SATA- és FC-bővítésben gondolkodunk. Új storage beszerzése 3-4 éven belül nem várható.

8.3. VMware

Bármilyen fejlesztés előtt alaposan tájékozódni kell, mert a VMware árképzése eléggé gyakran változik. (Lásd a vRAM alapú licencek esetét.)

Ha emeljük a processzorok számát a rendszerben, akkor újabb ESXi licenceket kell vennünk.

Ha szét akarjuk két részre szedni a rendszert mondjuk biztonsági okokból, akkor extra költségek merülhetnek fel például a VMware vCenter Site Recovery Manager licence miatt. Ha a két rendszert egymástól teljesen függetlenül akarjuk adminisztrálni, akkor a második VMware vCenter is plusz költségeket jelent.

8.4. Mentés

8.4.1. Tape

A meglévő szalagos egység nem bővíthető tovább (ha szükséges, akkor további LTO4-es szalagokat vehetünk), de a rendszer egy újabb hasonló vagy nagyobb kapacitású szalagos egységgel bővíthető. Ekkor további mentőszoftver-licenceket kell vásárolnunk.

8.4.2. D2D

A meglévő D2D-egység nem bővíthető tovább, de a rendszer egy újabb hasonló vagy nagyobb kapacitású egységgel bővíthető. Ekkor további mentőszoftver-licenceket kell vásárolnunk.

8.4.3. Networker

A jelenlegi licenc 19+1 fizikai szerver lementésére elég. Ha több mint 3 blade-szervert vásárolunk még, akkor licencbővítésre van szükség.

További extra szolgáltatásokkal lehet a Networker-szoftver tudását bővíteni, például natív SAN-mentéssel.

8.5. Ethernet

A kiépített rendszer Ethernet hálózati csatlakozásával kapcsolatban két problémával szembesültünk. Egyrészt a blade-keretek között nincs aktív hálózati kapcsolat a 4.1.1.1. pontban leírt spanning tree problémák miatt, így a köztes forgalom is kijut a Cisco Catalyst 6509 (továbbiakban C6509) core switchre. Másrészt hiba esetén a spanning tree rendkívül lassan konvergál.

Fenti problémák megoldására két lehetőséget vizsgáltunk meg. A HP Procurve 6120XG blade switcheket kicseréljük 4 db HP Virtual Connect Flex-10 modulra, vagy a C6509 és a blade-keretek közé beállítunk egy redundáns tápellátással és supervisorral, valamint legalább 2×4 db 10 Gbps-os porttal rendelkező „előtét”-switchet.

A HP Virtual Connect megoldását teszteltük. A tesztek alapján úgy döntöttünk, hogy inkább az „előtét”-switchre alapozott koncepciót valósítanánk meg.

9. Összegzés

A szerverkonszolidációval sikerült egy egységes, jól kézben tartható, korszerű rendszert kialakítani. Jelentős többletkapacitást terveztünk, így több éven keresztül tudunk megfelelő környezetet biztosítani az egyetem központi szolgáltatásainak.

Külön kiemeljük a mentőrendszert. Ennek segítségével gyorsan menthetők a virtuális gépeken tárolt adatok, melyek szükség esetén bármikor visszaállíthatók.

A rendszer telepítése és üzemeltetése során felhalmozódott tapasztalat jó alapot biztosít a későbbi fejlesztésekhez.

Köszönetnyilvánítás

A szerzők köszönetet mondanak Scherer Ferencnek és Racskó Tamásnak a dolgozat átnézéséért és értékes megjegyzéseikért.

Függelék:

A projektben közreműködő ESZK-s munkatársak:

Név	Feladat	Email
Scherer Ferenc	projekt vezetés	scherer@cc.u-szeged.hu
Csóti Zoltán	tervezés, hardver	csotiz@cc.u-szeged.hu
Szőke Szabolcs	tervezés, menedzsment, SAN, storage	szoke@cc.u-szeged.hu
Borús András	projekt menedzsment	borus@cc.u-szeged.hu
Horváthné Jáhni Gabriella	mentés	jahnig@cc.u-szeged.hu
Meszlényi Zoltán	Ethernet	msz@cc.u-szeged.hu
Soós Tamás	mentés	tom@cc.u-szeged.hu
Tóth-Abonyi Mihály	storage	toth@cc.u-szeged.hu
Török Attila	blade, VMware	attilat@cc.u-szeged.hu
Zombori Zoltán	projekt menedzsment, hardver	zombori@cc.u-szeged.hu

Felhasznált irodalom

- [1] Zombori Zoltán-Csóti Zoltán: Az SZTE Egyetemi Számítóközpont géptermének rekonstrukciója, Networkshop 2011, Kaposvár.
- [2] Szabó András és mtsai.: Virtualizált infrastruktúra megvalósítása VMware platformon, Megvalósulási dokumentum, KFKI Zrt., 2011.
- [3] Varga Zsolt: Központi mentési rendszer, Üzemeltetési kézikönyv, KFKI Zrt., 2011.