Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky

Katedra informačních technologií

Hlavní specializace: Informační technologie

Název diplomové práce:

Zajištění dostupnosti IT infrastruktury pro podporu firemních procesů

Vypracoval: Radek Zima

Vedoucí diplomové práce: Ing. Luboš Pavlíček

Recenzent: Ing. Daniel Dušek

P r o h l á š e n í

Prohlašuji, že diplomovou práci na téma
„Zajištění dostupnosti IT infrastruktury pro podporu firemních procesů“
jsem vypracoval samostatně.
Použitou literaturu a podkladové materiály
uvádím v přiloženém seznamu literatury.

V Praze dne 4.12.2004

P o d ě k o v á n í

Děkuji vedoucímu diplomové práce Ing. Lubošovi Pavlíčkovi

za cenné rady při vedení mé diplomové práce

1.Obsah

Obsah

1.Obsah....4
2.Vysoká dostupnost - úvod do problematiky....7
2.1.Terminologie....7
2.2.Dělení dostupnosti....9
2.3.Příčiny nedostupnosti....11
3.Ekonomické aspekty dostupnosti serverové architektury....14
3.1.Plánování dostupnosti....14
3.2.Business Continuity Planning....14
4.Technologie pro zajištění redundance celých systémů - clustering....18
4.1.Základní terminologie....18
4.2.Modely....19
4.3.Výhody a nevýhody clusteringu....19
4.4.Příklad clusterových řešení....20
4.4.1.Microsoft Cluster Service....20
4.4.2.Serviceguard....23
5.Technologie zajišťující vyšší odolnost serveru a jeho komponent, rychlé zotavení....24
5.1.Ukládání dat....24
5.1.1.Obecné technologie pro připojení diskových zařízení k serveru....24
SAS....26
5.1.2.DAS – direct attached storage....28
5.1.3.NAS – network attached storage....28
5.1.4.SAN – Storage Area Network....29
5.1.5. Technologie diskových polí....30
5.2.Napájení....32
5.2.1.Interní zdroje....32
5.2.2.Nepřerušitelné zdroje napájení (Uninterruptible Power Supplies = UPS)....33
5.2.3.Řízení toku vzduchu....34
5.2.4.Racky a racková příslušenství....34
5.3.Ochrana paměti....35
5.3.1.Dělení paměťových chyb....36

5.3.2.Technologie určené k detekci a opravám paměťových chyb....36
Parita....36
Error Checking and Correcting – ECC....37
Advanced ECC....39
Online spare memory....40
Mirrored memory....41
RAID memory....41
5.4.Zálohování....42
5.5.Imaging, bezzásahové instalace....45
5.6.Vzdálený management a monitoring....46
5.7.Poruchovost jednotlivých komponent....47
6.Organizační a personální zajištění dostupnosti....48
6.1.SLA....48
6.2.Oprávnění....50
7.Případová studie....51
7.1.CES holding....51
7.1.1.Historie....51
7.1.2.Profil....54
Strategie skupiny CES, a.s....54
Cíle společností holdingu Central European Services, a.s....54
7.1.3.Dceřinné společnosti....55
Amden....55
Pro Futuro Consulting....55
LFP Studio....56
CAD Projekt....56
Pythagoras....56
MTEC....56
7.1.4.Serverová infrastruktura, použité technologie....56
Detailní přehled holdingových serverů:....59
7.1.5.Doporučené postupy při oznamování a řešení problémů....68
7.1.6.Popis stavu, návrhy na zlepšení, závěr....69
7.1.7.Poděkování....72
8.Seznam použité literatury....73
9.Abstract....74

9.1.Abstrakt....74
9.2.Abstract....74
10.Seznam tabulek....75
11.Seznam ilustrací....76
12.Terminologický slovník....78
13.Abecední rejstřík....80

2.Vysoká dostupnost - úvod do problematiky

Bez kvalitní IT infrastruktury se v současné době neobejde žádná společnost a to napříč všech odvětví ekonomiky. Některé části této infrastruktury jsou dobře viditelné, nicméně ovlivňují pouze práci jedince (pracovní stanice, lokální tiskárny, …) a jejich případný výpadek sníží produktivitu pouze svého uživatele. Zde zpravidla neinvestujeme větší částky do zajištění dostupnosti a spíše se věnujeme tomu, aby v případě výpadku bylo možné vadnou komponentu rychle vyměnit. Jiné části běžně viditelné nejsou (servery, centrální storage, síťová infrastruktura, skupinové tiskárny …), nicméně jejich selhání mohutně ovlivní produktivitu napříč celé společnosti a zpravidla s sebou nese nemalé náklady. Správné řízení dostupnosti vyžaduje nejen ekonomické znalosti, které umožní rozhodnout do kterých oblastí IT infrastruktury investovat, je třeba znát i technické pozadí tak aby bylo možné spočítat nebo odhadnout pravděpodobnostní rozdělení jednotlivých typů případných problémů a technologií, které s takovýmto výpadkem dokážou vyrovnat.

V současné době lze u nás pozorovat trend, který známe již několik let ze zahraničí, tj. příliv investic do vyšší dostupnosti. Souvisí to nejen s vyšší ekonomickou silou českých společností, ale i s rostoucí produktivitou zaměstnanců a tím vyšších škod způsobených nedostupností.

Cílem mé práce bude uvést čtenáře do problematiky vyšší dostupnosti, shrnout dostupné technologie pro vyšší odolnost jednotlivých komponent serveru pro výpadku, seznámit se základy clusteringu. Po prostudování tohoto textu bude mít čtenář přehled o vyšší dostupnosti, umožní mu to lépe porozumět této problematice ve vlastní firmě nebo u zaměstnavatele a zjednoduší vyjednávací pozici s dodavateli těchto technologií.

2.1.Terminologie

Pro efektivní studium tohoto textu je nutné se shodnout na použité terminologii, vybral jsem takové pojmy, které jsou v odborné literatuře nejvíce rozšířené a nejvýstižnější:

- High Availability – vysoká dostupnost:
  - vysoká dostupnost je vlastnost systému, která zajistí jeho dostupnost při výpadku kterékoli jeho části [2]

- Definice v angličtině: High Availability is a system design protocol and associated implementation that ensures a certain absolute degree of operational continuity during a given measurement period [7]
- Snaha minimalizovat jak plánovaný, tak neplánovaný downtime; eliminace „single points of failure“

- Business continuity – zajištění nepřetržitého fungovaní společnosti
- Jednotlivé skupiny, do kterých je možné technologie a postupy vyšší dostupnosti rozdělit:
  - **fault prevention – prevence**; u nejčastějších poruch bývá efektivní učinit taková opatření, která zabrání vzniku výpadku (monitoring, …)
  - **fault tolerance – odolnost proti výpadku dílčí komponenty**; pokud prevence nezabrání závadě, používáme technologie, které se s takovouto závadou transparentně vyrovnají (RAID, redundantní ventilátory, …)
  - **rapid recovery – rychlé zotavení**; pokud se nepodaří preventivními opatřeními zabránit vzniku závady, žádná technologie není schopná se se závadou vyrovnat a dojde k výpadku, je nutné co nejrychleji uvést systém do původního stavu (ASR, bezzásahové instalace, …)
- **downtime –** nedostupnost služby
- **uptime –** dostupnost služby, upravidla se snažíme, aby dostupnost služby byla co nejblíže 100%

Dostupnost v nejjednodušším světle slova smyslu je čas, kdy systém pracuje normálně. Downtime je čas, kdy službu, kterou systém nabízí svým klientům (uživatelům, dalším aplikacím, zákazníkům, …) není možné využívat (dále budeme používat pojem služba a to v tomto širším pojetí – nebude-li explicitně uvedeno jinak).

Určitě si umíme představit situace, kdy není možné tolerovat výpadek IT infrastruktury. Typické příklady obsahují prostředí, kde jsou ohroženy lidské životy:

- služba řízení letového provozu (těžko ospravedlníme havárii letounu způsobenou závadou na pevném disku),

- řízení vesmírných letů (dejte si ještě jeden oblet zeměkoule, měníme obrazovku …)
- dispečink záchranného systému (počkejte chvilku na telefonu, restartujeme rádiovou ústřednu …)

Takovýmto prostředím se říká „mission critical“, a vyžadují 100% dostupnost provozovaných aplikací.

Vyšší než standardní dostupnost ovšem vyžadují i společnosti, kde lidské životy bezprostředně ohroženy nejsou, nicméně v případě výpadku dochází k velkým ztrátám (zpravidla označujeme jako business critical prostředí):

- bankovní domy, systémy pro elektronické bankovnictví, ověřovací centra platebních karet
- akciové burzy

Vysoká dostupnost bývala v minulosti doménou největších společností. A to nejen z důvodu dostupných financí na nákup technologií, především to byly negativní zkušenosti s dopadem případného výpadku na výkon společnosti. V současné době se i menší a střední firmy zamýšlejí nad nepřetržitým provozem, neovlivněným výpadky v infrastruktuře.

2.2.Dělení dostupnosti

Při měření dostupnosti zpravidla uvádíme procento času, kdy je možná službu využívat. Nejčastěji se setkáme s takovýmto odstupňováním:

Uptime v %	*Downtime v %*	*Nedostupnost ročně*	*Nedostupnost týdně*
98 %	2 %	7,3 dne	3 hodiny, 22 minut
99 %	1 %	3,65 dne	1 hodina, 41 minut
99,8 %	0,2 %	17 hodin, 30 minut	20 minut, 10 vteřin
99,9 %	0,1 %	8 hodin, 45 minut	10 minut, 5 vteřin
99,99 %	0,01 %	52,5 minuty	1 minuta
99,999 %	0,001 %	5,25 minuty	6 vteřin
99,9999 %	0,0001%	31,5 vteřiny	0,6 vteřiny

Tabulka 2.1: Procentuální dělení dostupnosti

Z této tabulky je možné si udělat představu, co na první pohled nepřehledná čísla znamenají. Dostupnost služby „pět devítek“ (tj. 99,999%) značí 5 minut nedostupnosti ročně … 5 minut

zpravidla nestačí ani k pouhému restartu systému (např. při aplikování bezpečnostních záplat), jedná se již o vysoce dostupný systém. Oblastem 95% - 99% říkáme standardní dostupnost, až po 99,99% vysoká dostupnost a rozmezí 99,999% - 100% nazýváme nepřetržitá dostupnost.

Nesmíme zapomenout, že ve výše uvedené tabulce mluvíme o nedostupnosti IT infrastruktury, nicméně 5 minut nedostupnosti IT neznamená 5-ti minutový výpadek obchodní činnosti. Po výpadku je např. nutné zajistit přepsání informací, které nebylo možné zapsat do CRM z papírových poznámek, zatelefonovat některým zákazníkům s omluvou, apod. [8]

Procento dostupnosti spočítáme jako $dostupnost = \frac{mtbf}{(mtbf + mtrr + mttd)}$, kde

MTBF ... mean time between failure, tj. střední doba mezi chybami (typicky s normálním rozložením)

MTRR … mean time to recover, střední doba opravy

MTTD … mean time to detect, střední doba zjištění závady

Další typické členění je na jednotlivé úrovně dostupnosti (více detailů o jednotlivých pojmech najdete v kapitolách 4 a 5).

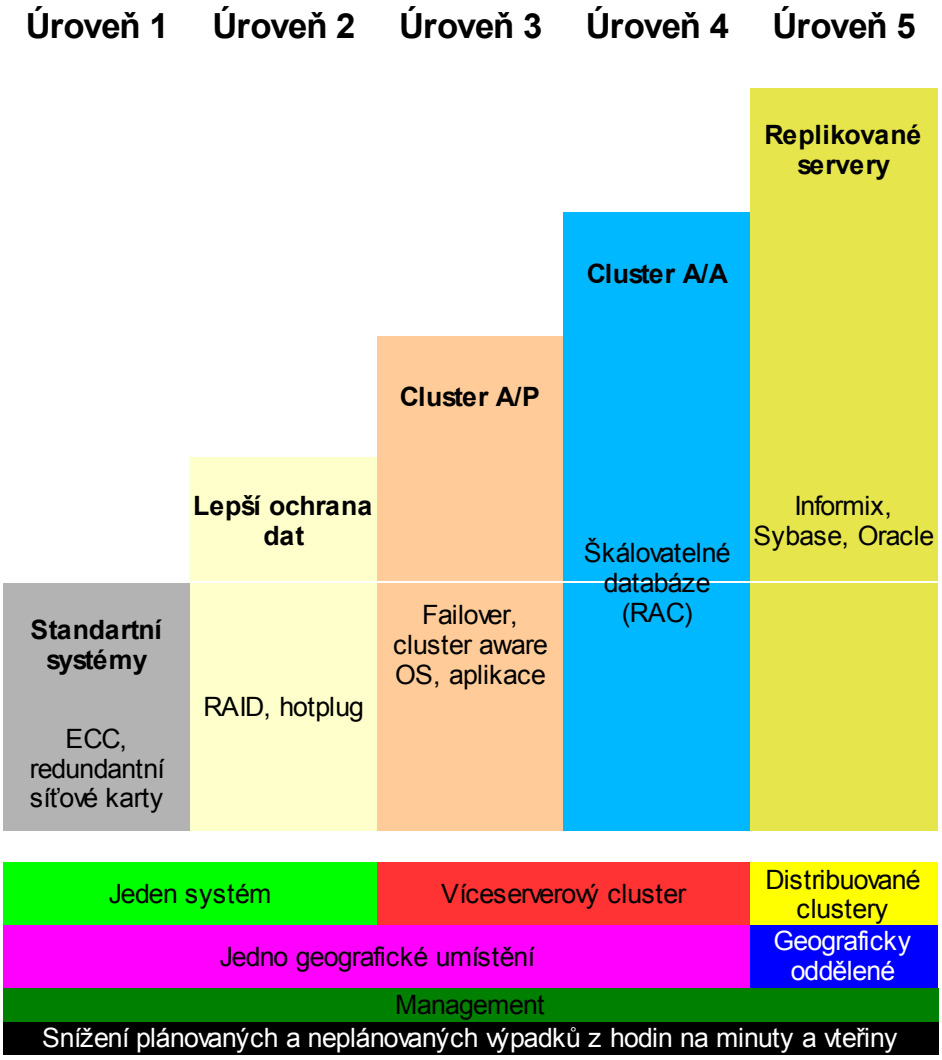
- Úroveň 1 – standardní server (některé redundantní komponenty, ochrana paměti ECC, …)
- Úroveň 2 – systémy s RAIDovými disky, hotplugové komponenty, …
- Úroveň 3 – vysoce dostupný cluster, využívající failover, sdílení zdrojů (disky, síť, …), např. Microsoft Windows Cluster Service

Ani tato úroveň nedosahuje 100 % dostupnosti, v případě závady proběhne tzv. failover – tj. aplikace se musí spustit na jiném node v clusteru, během této operace je služba krátkou dobu nedostupná.

- Úroveň 4 – clusterové řešení, kde aplikace běží zároveň na více nodech, tj. v případě závady neprobíhá failover, není nutné aplikace znovu spouštět (např. RAC – Real Application Cluster společnosti Oracle)
- Úroveň 5 – řešení, které se vyrovná i s katastrofou na úrovni celé serverovny, budovy, města … (tj. jaderný útok, povodně, válka, politický převrat, …)

V tomto případě je již nutné, aby jednotlivé komponenty systému byly i dostatečně geograficky vzdálené.

Pro přehlednost uvádím názorný přehled jednotlivých úrovní dostupnosti, vyšší úroveň pochopitelně obsahuje i technologie ze všech nižších úrovní.



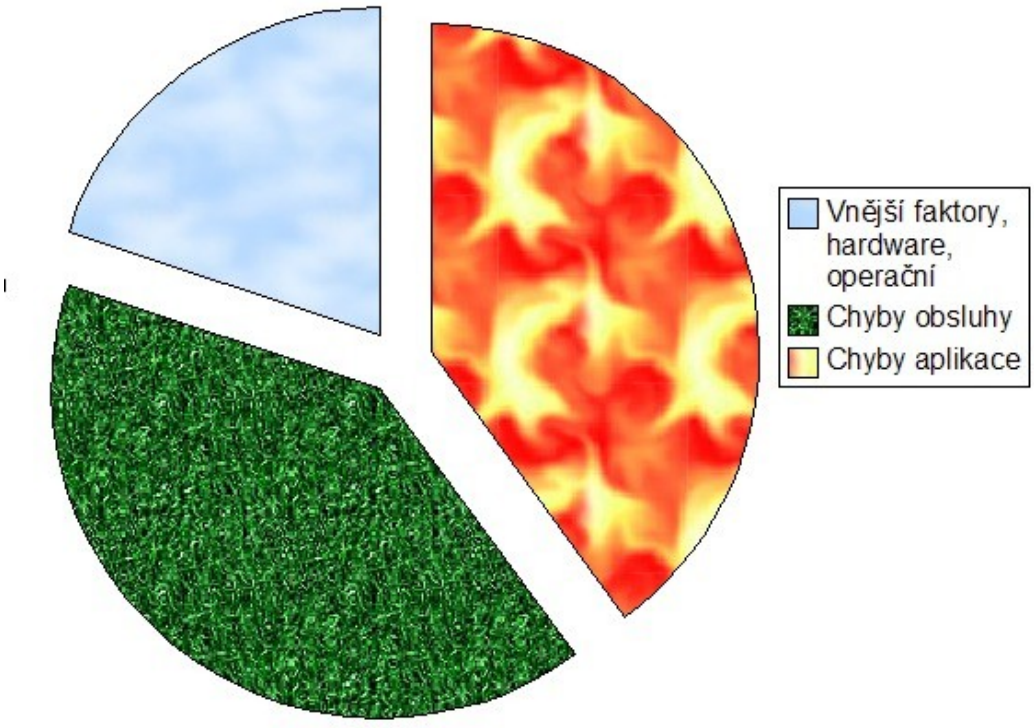
Obr. 2.1: Přehled úrovní dostupnosti

2.3.Příčiny nedostupnosti

Jak k nedostupnosti dochází? Nejčastějšími příčinami jsou dle IDC

- **hardwarové chyby** – závady způsobené vadným vybavením
- **softwarové chyby –** chyby způsobené OS, aplikacemi či ovladači, příp. spojením těchto komponent

- **plánované servisní úkony –** upgrade hardware (přidávání paměti, diskové kapacity, výměna procesorů za rychlejší), upgrade software (hotfix v OS, nová verze aplikace)
- **chyby obsluhy –** lidské chyby (tzv. lidský faktor),
  - špatný zásah do konfigurace
  - neúmyslné vypnutí serveru (zní nepravděpodobně, ale stát se to opravdu může a to nejen mě, překliknout se o jednu položku v comboxu při pokusu o restart serveru po aplikací bezpečnostní záplaty je skutečně jednoduché)
  - vypnutí špatného serveru (představte si serverovnu se 100 servery; zajistit, aby byl vypnut správny server není triviální)[1]



Obr. 2.2: Příčiny nedostupnosti

1 Jestlipak víte, jak takové situaci předejít ? Někteří renomovaní výrobci serverů používají na čele i zádi serveru speciální LED, kterou je možné softwarově ovládat a označit si stroj. Pokud nic takového nemáme k dispozici můžeme použít ovládání CD mechaniky a server, do kterého je nutné přidat nové paměťové moduly si označit vysunutým CDčkem

- **prolomení bezpečnosti –** neautorizovaný uživatel se dostane k ovládání systému, kde může provádět konfigurační zásahy, mazat soubory, ...

Docílení vyšší než standardní dostupnosti vyžaduje nasazení specializovaných technologií a postupů, které jsou přiblíženy v kapitolách 4 a 5 (předporuchová záruka, nepřetržité zdroje napájení, ECC paměti, RAID disky, redundantní komponenty, clustering, SAN, …).

3.Ekonomické aspekty dostupnosti serverové architektury

3.1.Plánování dostupnosti

Při plánování dostupnosti IT infrastruktury je potřeba se zamyslet nad následujícími aspekty a potenciálními „body“, kde může dojít k závadě. V potaz je třeba brát jakoukoliv komponentu, spojení, činnost správce sítě, software, atd. který může zastavit některý ze systémů.

- **Hardware serveru –** paměťový subsystém, procesor, I/O, napájení, lokální storage
- **Síťová infrastruktura –** síťové switche, návrh topologie, struktura routování a bridgování
- **Síťová storage infrastruktura –** topologie SAN, vícecestné routování (tzv. multipath)
- **Operační systém –** schopnost operačního systému obsluhovat výjimky, chybnou alokaci paměti (memory leaks), zatížení procesoru, schopnost aplikovat bezpečnostní záplaty
- **Middleware, aplikační služby –** chování aplikačních služeb, využití paměti, využití procesoru, rychlost spouštění, reakce na změnu nastavení (restart většiny Windowsových aplikacá vs. řízení pomocí signálů v Unixovém prostředí)
- **Ukládání dat** - škálovatelnost (ve smyslu flexibility rozšíření – jak snadno a s jakým dopadem na dostupnost je možné zvýšit kapacitu diskového subsystému), ochrana dat (RAID, SAN, geografické zálohovaní)

3.2.Business Continuity Planning

V současné době organizace věnují stále větší pozornost plánování kontinuity činností (BCP neboli Business Continuity Planning). Pro každý živý organizmus, a tedy i pro každou dobře fungující organizaci, je starost o vlastní přežití na prvním místě. Mimořádná událost může zlikvidovat know-how a obchodní prestiž budovanou po mnoho let. Stále větší provázanost IT s obchodními procesy zároveň znamená nutnost řešit celé spektrum specifických hrozeb. Oblast

BCP postihuje všechny tyto aspekty a nastavuje firemní procesy tak, aby hrozby byly odstraněny nebo alespoň minimalizovány. [9]

Plánování kontinuity činností lze definovat jako souhrn aktivit zaměřených na snížení rizika vzniku havárie a omezení dopadů havárie na kritické podnikové procesy. Důležité je si uvědomit, že plánování kontinuity není jen plánem reakce na krizovou událost, ale obsahuje také důležitý preventivní aspekt. Jedním z hlavních výstupů tohoto procesu je plán zachování kontinuity činností (Business Continuity Plan). Kvalitní plány kontinuity činností jsou schopny minimalizovat následky mimořádných událostí a zároveň umožňují a urychlují uvedení provozu do normálního stavu.

Typy nepředvídaných událostí jsou popsány v kapitole 2.11. Následkům těchto hrozeb se organizace snaží čelit mnoha způsoby, což je také jedním z důvodů pro investice nemalých finančních částek do různých druhů pojištění. Ale i sebelepší pojistka má svá omezení – je to statické řešení, které může pouze zpětně kompenzovat finanční ztrátu a těžko zabrání vzniku škody. Pokud má však organizace zavedený BCP, je díky němu schopna nejen minimalizovat škody přímo v průběhu krizové události, ale navíc je schopná jim i účinně předcházet.

Pokud se blíže podíváme na plánování kontinuity v oblasti IT, jsme postaveni před mnoho specifických problémů. Mimořádná událost (havárie) v oblasti IT nemusí nutně znamenat přírodní katastrofu. Výpadek proudu, virový útok, selhání hardwaru, chyba administrátora nebo chyba technologií poskytujících infrastrukturní služby na výpočetních sálech může napáchat velké škody. Praxe nás přesvědčuje o tom, že i nejlepší technologie nasazené na zvládání konkrétních hrozeb mohou z nejrůznějších příčin selhat.

Prakticky každý výpadek služeb informačního systému dnes znamená ztrátu produktivity, příležitostí, pokles zisku, případně úbytek zákazníků. Navíc díky širokému využití internetu jsou tyto výpadky velmi rychle zjistitelné ze strany zákazníka a partnerů. Pro služby poskytované IT musí být tedy nutně vypracován specifický systém řízení a plánování kontinuity činností s tím, že infrastruktura IS by měla být přizpůsobena těmto potřebám.

Zjednodušeně řečeno, jedním z hlavních přínosů zvládnutého řízení kontinuity činností je to, že v okamžiku havárie budou všichni zodpovědní pracovníci vědět co dělat. Přesné návody jak postupovat, koho kontaktovat, jaké služby a v jakém pořadí zachovat, mají v tento okamžik nedocenitelnou hodnotu. Plán kontinuity zpracovaný na základě důkladných analýz umožní

v momentě havárie rychle a přesně identifikovat nejkritičtější procesy a aktiva organizace, a soustředit tak hlavní energii právě na jejich obnovu či zachování. [10]

Potřeba tohoto druhu plánování je také zakotvena v českých právních normách:

- například Zákon č. 240/200 Sb. o krizovém řízení ukládá státním organizacím povinnost mít připravené krizové plány;
- opatření ČNB 2/2004 zase požaduje, aby banky měly pro případy havárie informačních systémů připraveny „ ... postupy, které vedou k obnovitelnosti činností a informačních systémů významných z hlediska fungování banky“.

Řízení a plánování kontinuity činností je komplexní problém, jehož řešení vyžaduje holistický přístup k porozumění fungování organizace. Tato oblast by neměla být považována za výhradní doménu IT oddělení, ale naopak by měla být vedením vytyčena jako jeden ze strategických cílů organizace, čímž mu zajistí patřičnou podporu. Nyní se blíže podívejme na to, jak vypadá životní cyklus tohoto procesu.

Nejprve musí proběhnout důkladná analýza prostředí organizace. V průběhu analýzy jsou identifikována aktiva, jejich zranitelnosti a jednotlivé hrozby vůči těmto aktivům. Dalším krokem je provedení BIA analýzy (Business Imapct Analysis), jejímž úkolem je především identifikace možných rizik, odhad a kvantifikace následků havárie na organizaci a určení požadavků na strategii obnovy. Přínosem BIA analýzy je určení klíčových obchodních procesů, vyčíslení nedostupnosti klíčových služeb po určitou dobu, a určení minimální úrovně podpory ze strany informačního systému, kterou organizace potřebuje k alespoň omezenému provozu.

Dále se v této fázi často provádí analýza stávajících možností obnovy (Recoverability Assesment), ve které je zjištěna stávající připravenost IT prostředí na řešení mimořádné situace. Zvláště u větších organizací je pro vypracování těchto analýz vhodné využít služeb nějaké renomované externí firmy, která má nejen know-how potřebné pro jejich provedení, ale zároveň vnese díky své nezainteresovanosti kritický pohled na jednotlivé vazby a procesy uvnitř organizace.

Další fází je vlastní návrh strategie řešení problému zachování kontinuity činností. Každý uživatel (člen vedení) by jistě chtěl mít zajištěnu 100 % dostupnost služeb, většinou ale jen do té doby, než se dozví cenu za takové řešení. Ztráty z nedostupnosti služeb exponenciálně rostou s ubíhajícím časem, zatímco náklady na obnovu provozu se rapidně zvyšují se zkracujícím

se časovým limitem pro obnovu. Výsledky BIA analýzy v korelaci s cenou (v poměru k ceně) řešení slouží jako dobré podklady pro volbu vhodné investiční strategie do této oblasti.

Při návrhu strategie obnovy je nutno se vyhnout dvěma extrémům. Jedním extrémem je „pouze technologické“ řešení, kdy robustní nasazení konkrétní technologie, bez zvážení širšího organizačního, technického a strategického kontextu, může znamenat zbytečně promrhané investice. Druhý extrém spočívá v „pouze papírovém plánování“, kdy plány obnovy sice obsahují samé dobré nápady, ale nikomu není jasné, jaké konkrétní důsledky z toho pro něj vyplývají.

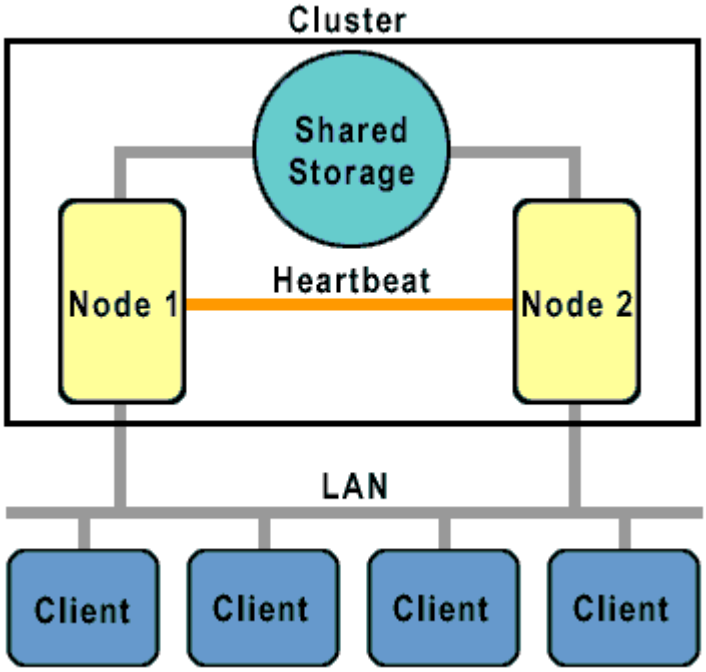
Kvalitní zpracovaní plánů kontinuity činností by mělo být strategickým cílem jakékoliv organizace - od velkých nadnárodních společností až po malý či střední podnikatelský subjekt. I když se masivnost nasazení konkrétních technologií bude v organizacích různého typu lišit, je nutné zachovat při navrhování plánů kontinuity stěžejní principy životního cyklu řízení kontinuity. Patří mezi ně zejména kvalitně provedená analýza, testování a pravidelná údržba.

4.Technologie pro zajištění redundance celých systémů - clustering

4.1.Základní terminologie

Před bližším výkladem clusterových technologií osvětlím základní používanou terminologii a představím jednoduchý clusterový model.

- **cluster** – řešení obsahující dva a více serverů (tzv. nodů), sdílené úložiště dat a zajišťující provoz služby i v případě výpadku celého serveru; z hlediska klientů je cluster monolitický (je jim jedno kolik serverů ve skutečnosti službu poskytuje, hlavně aby fungovala)
- **cluster service** – služba, kterou cluster poskytuje (databáze, webová aplikace, …)
- **node** – člen clusteru, typicky jeden server
- **resource** – zdroj přiřazený clusteru (síťová karta, disk/LUN na poli)
- **interconnect** – spojení mezi nódy určené pro komunikaci mezi jednotlivými členy clusteru, používáme buď průmyslově standardní technologie (ethernet), nebo speciální clusterové spojení (nízkolatentní Myrinet apod.)
- **failover –** proces, při kterém dojde k přesunutí aplikace z jednoho node na druhý; může být automatický nebo manuální (ručně provedený z administrátorské konzole nebo CLI)
- **failback –** aplikace je přesunuta zpět na původní node, na kterém běžela před failoverem, tak aby bylo zajištěno optimální využití členů clusteru po opravení závady
- **heartbeat** – spojení mezi jednotlivými nody, pomocí kterého je možné zjistit, zda nedošlo k výpadku stroje (může být sériové, síťové, …)



Obr. 4.1.: Jednoduché schéma clusteru

Typický active/passive cluster se dvěma nody naleznete na obrázku, cluster obsahuje dva servery, které vystupují na síti jako jedno TCP/IP zařízení (tj. jeden přístupový bod pro klienty). Používají sdílené úložiště dat (na něm je umístěný databázový engine nebo webová aplikace).

4.2.Modely

- shared nothing – jednotlivé nody používají sdílené úložiště dat, k jednotlivým logickým diskům může přistupovat vždy jen jeden nod
- shared disk – takováto konfigurace potřebuje systém, které bude pro jednotlivé nody uzamykat části sdíleného disku, tak aby si nody nepřepisovali data navzájem, takovéto aplikaci se zpravidla říká Distributed Lock Manager
- shared everything – nody clusteru sdílí mimo disků i procesory a paměť, tzv. „single system image“; sem patří např. NonStop systémy společnosti Hewlett-Packard nebo TruClustery od společnosti Digital (později Compaq, dnes HP)

4.3.Výhody a nevýhody clusteringu

Existují pochopitelně rizika, před kterými nás cluster z principu věci neochrání. Napadení virem nebo červem, chyba v aplikaci, lidská chyba.

Hlavní výhodou clusteringu (koneckonců i důvod pro zařazení do této práce) je zajištění vyšší dostupnosti zdrojů, které cluster nabízí (veškeré komponenty v serveru máme zdvojené – u dvounodového clusteru). Nezapomínáme ale ani na snadnou možnost škálovatelnosti, rozšiřitelnosti, tak jak stoupá množství členů clusterů (nodů), stoupá i možnost jejich rozšiřování (u dvounodového clusteru dvounásobný počet socketů pro procesory, šachet pro disky, …). Dalšími výhodami clusteringu je centrální administrace a u active/active řešení i sdílení zátěže. Clustering nás nechrání jen proti hardwarovým chybám některých komponent (procesor, disky, paměť, sběrnice, …), ale i proti závadám v aplikaci či OS. Umožní nám aplikovat service packy, upgradovat aplikace či instalovat nové aplikace zatímco stávající služby jsou stále využívány klienty.

4.4.Příklad clusterových řešení

4.4.1.Microsoft Cluster Service

Microsoft jako jedna z vedoucích společností na trhu IT má vlastní clusterové řešení pro svou rodinu serverových operačních systémů, v současnosti (rok 2006) se jedná o serverové operační systémy Windows Server 2003 a službu Microsoft Cluster service.

Jedná se o plnohodnotnou clusterovou aplikaci, která zajistí detekci jakékoliv hardwarové chyby a obnovení běhu aplikace na jiném nodu v clusteru (failover). Po výměně nebo opravě vadné komponenty spustí aplikaci opět na původním nodu (failback).

Službu je možné konfigurovat a spravovat nativními prostředky OS Windows 2003, včetně možností vzdálené správy a konfigurace služeb či aplikací, které chceme na clusteru provozovat. Microsoft Cluster Service je pochopitelně možné zaintegrovat do doménových adresářových služeb (Active Directory – AD), nejdůležitější vlastností je přidání nového objektu do AD a sice virtuálního počítače. Na „zdravotní stav“ clusteru a jeho komponent je možné se zeptat pomocí rozhraní WMI, což umožňuje sledovat cluster pomocí standardních management aplikací.

Windows Server 2003 existuje v šesti verzích:

- Windows Server 2003 Standard Edition
- Windows Server 2003 Enterprise Edition
- Windows Server 2003 Datacenter Edition

- Windows Server 2003 Web Edition
- Windows Storage Server 2003
- Windows Small Business Server 2003

Pro zajištění vyšší dostupnosti pomocí služby Microsoft Cluster service je nutné použít buď verzi Enterprise Edition nebo Datacenter Edition. Obě podporují 8 nodů v clusteru, což poskytuje administrátorům poměrně velkou flexibilitu při rozhodování na kterém nodu má běžet jaká aplikace a dovolí vytvářet sofistikovaná pravidla jak rozmístit aplikace při výpadku některého z nodů. Starší verze Windows 2000 Advanced Server podporovala pouze 2 nody, Datacenter Server 4 nody. Clusterová služba je součástí instalace operačního systému (nikoliv přídavná aplikace jako např. u Windows Server 2000) a obsahuje jednoduchého instalačního průvodce, který zjednoduší instalaci clusteru.

Vedle klasické clusterové služby můžeme použít i „Network Load Balancing“, kdy je více strojů identifikováno na síti stejnou IP adresou a příchozí požadavky jsou rovnoměrně rozkládány na více strojů.

Clusterová služba obsahuje následujicí komponenty:

- Node Manager – zajišťuje vlastnictví zdrojů (viz. níže) jednotlivými nody, má přehled o dostupnosti nodů
- Resource Manager – rozhraní mezi clusterovou službou a jejími zdroji (viz. níže), clusterová služba používá Resource Manager pro přístup k jednotlivým dynamickým knihovnám (DLL) zdrojů
- Failover Manager – společně s resource managerem zaštiťuje restarty a failovery
- Checkpoint Manager – provádí aktualizaci registrových klíčů a stavů jednotlivých aplikací
- Configuration Database Manager – synchronizuje konfigurace nastavení clusteru mezi jednotlivými nody (nody, zdroje, skupiny zdrojů, …)
- Event Processor
- Membership Manager – organizuje členství nodů ve skupinách

- Event Log Replication Manager – synchronizuje záznamy souvisejícíc s událostmi v clustery mezi nody
- Global Update Manager
- Object Manager
- Log Manager

Zdroj (angl. Resource) je fyzická nebo logická komponenta, kterou je možné spravovat. Jako příklad uvádím pevné disky, síťová jména či IP adresy, databáze, webové adresy, aplikace, prostě cokoliv co můžeme přepínat mezi stavy online/offline. Zdroj může být používán současně jen jedním nodem. Zdroje jsou organizovány dle typu:

- fyzické – např. disky, ...
- logické – aplikace, sdílené adresáře, …

Zdroje je možné seskupovat do skupin, které pak tvoří samostatnou jednotku reprezentující logický počítač, který je možné přesouvat z jednoho nodu na druhý a která poskytuje určitou službu. Typická skupina obsahuje síťové jméno, IP adresu, fyzický disk a aplikaci. Mezi skupinami zpravidla definujeme závislosti, tak abychom měli jistotu, že jednotlivé komponenty budou zapínány a vypínány ve správném pořadí.

Př. sdílení souborů je závislé na fyzickém disku a síťovém jménu

Každý zdroj se může nacházet v jednom z následujících stavů:

- offline
- started
- online
- pending
- failed

Výchozím stavech všech zdrojů je „offline“ - tedy zdroj je vypnutý, stejně jako všechny zdroje na něm závislé a není možné ho využívat ani klienty.

4.4.2.Serviceguard

Vedle řešení pro operační systémy MS Windows nalezneme na trhu i řešení pro jiné OS. Jako příklad uvádím Hewlett-Packard Serviceguard. Jedná se specializované řešení, které pomůže s ochranou aplikací před hardwarovými i softwarovými chybami. Umožní připojení více nodů do clusterů, které poskytují vysoce dostupné služby klientům. Tj. i v případě hardwarové či softwarové závody aplikace stále slouží.

Serviceguard sleduje zdravotní stav různých komponent na každém nodu a rychle reaguje na závady tak, aby byla eliminována nebo minimalizována nedostupnost aplikace. Serviceguard chrání uživatele detekcí a reakcí na závady následujících komponent:

- procesor
- systémová paměť
- procesy OS a aplikací
- síťové adaptéry

Podporuje jak active/active clustery, tak standby clustery. Je možné provádět tzv. rolling upgrade. Rolling upgrade je procedura, při které provádíme upgrade operačního systému nebo aplikace postupně nod po nodu (vždy dáme nod do „offline“ stavu, tak aby aplikace na něm běžící byli přesunuty na jiný stroj a poté provedeme upgrade).

Serviceguard je primárně určen pro servery HP ProLiant a HP Integrity s operačními systémy Linux nebo HPUX. Podporuje čtyři servery a SCSI/FC připojené externí diskové pole.

5.Technologie zajišťující vyšší odolnost serveru a jeho komponent, rychlé zotavení

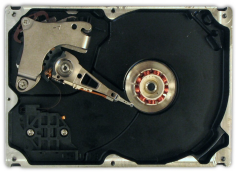
5.1.Ukládání dat

Jedna z klíčových komponent na kterou je třeba se zaměřit při přemýšlení o vyšší dostupnosti je oblast ukládání dat (v anglické literatuře je většinou používán pojem „*storage*“). V tomto textu budu pod pojmem ukládání dat předpokládat technologie pro fyzické uložení dat na nosiče na magnetickém principu. Metody pro uložení dat rozpoznáváme tři:

- DAS
- NAS
- SAN

5.1.1.Obecné technologie pro připojení diskových zařízení k serveru

Datové rozhraní je většinou jedno z následujících:



Obr. 5.1.: Pevný disk

- Paralelní SCSI
- Paralelní IDE/ATA
- Serial ATA
- FATA
- Sériové implementace SCSI
  - FibreChannel
  - FireWire / IEEE1394
  - SAS

V serverovém prostředí, kde vyžadujeme vyšší dostupnost diskového subsystému používáme v drtivé většině případů rozhraní SCSI, a to jak paralelní tak sériovou implementaci.
Od IDE/ATA se SCSI liší mj. v těchto ohledech:

- Vyšší MTBF – řádově vyšší spolehlivost SCSI disků, daná spolehlivější mechanikou použitou v SCSI discích (dané použitými výrobními postupy a materiály na osičky, ložiska a magnety pro ramena), SCSI disky jsou dimenzované pro 100% zatížení (24 hodin denně, 7 dní v týdnu a 100% využití pro čtení/zápis). Střední doba mezi chybami u „enterprise“ disků se na základě statických měření zpravidla uvádí kolem 1,5 miliónů hodin (pro porovnání – „entry-level“ disky používané např. v desktopech nebo u nekritických dat mají MTBF cca 500 tisíc hodin – je počítáno se zátěží 8 hodin denně, 5 dní v týdnu a 20-30% využití pro čtení/zápis).

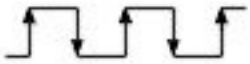
- Rozdíly v protokolu usnadňují „enterprise“ použití:

  - hotplug - možnost manipulace s diskem za chodu (výměna vadného disku nevyžaduje restart serveru, je možné přidat disk a potažmo zvýšit kapacitu pole bez restartu, tj. Systém jako takový a aplikace, které na něm běží výměna nijak nepostihne)

  - disconnect-reconnect – SCSI disk se po přijmutí pokynů od řadiče odpojí od sběrnice a umožní komunikaci ostatním disku na stejném SCSI kanálu (sběrnici)

  - TCQ – disk je schopen přijmout více příkazů a sekvenčně je vykonávat zatímco je odpojený od sběrnice (skutečnost, která přispívá k velké výkonnosti diskových polí v porovnání s jednotlivými disky)

V současnosti (rok 2005) využíváme převážně protokol standartu SCSI3, Ultra 320, který zajišťuje dostatečnou propustnost (320 MB/s – sdílená kapacita na jednu SCSI sběrnici je takto vysoká, protože je použita 80MHz sběrnice a technologie „double transition clock“, nebo-li data jsou přenášena na obou hranách taktu) a díky signálování LVD – Low voltage differential (nahrazuje starší rozhraní SE – Single ended) máme větší odolnost proti rušení, tedy lepší odezvu pro klienty (ať už to jsou zaměstnanci, zákazníci nebo další aplikace) a delší kabely. SCSI standart dnes povoluje 12 metrů dlouhou kabeláž při použití LVD rozhraní, delší metalické kabely bylo možné používat s rozhraním HVD – High Voltage Differential, které používalo pro signálování 5V, čím byla zajištěna lepší odolnost proti rušením – to dovolovalo až 25 metrů

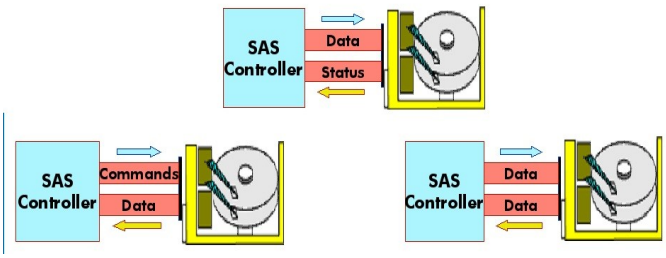


Obr. 5.2.: DDR

dlouhé kabely. Daní za to bylo
a) větší pořizovací cena a to mechanik, disků i řadičů,
b) riziko poškození při připojení ne-HVD zařízení – 5V na signálové sběrnice je může poškodit.

SAS

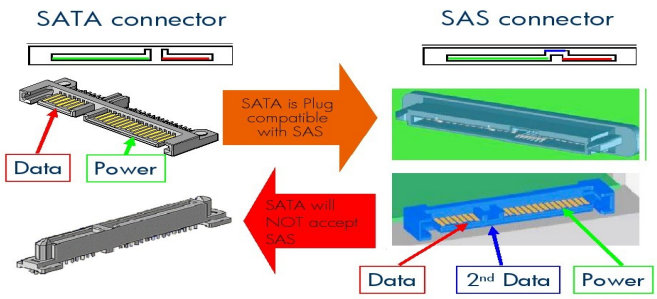
Principiální potíže se zvyšováním frekvence u paralelního SCSI vedou hlavní výrobce výpočetní techniky k opuštění této technologie a její nahrazení sériovou implementací – SAS. Protokol Ultra 640 sice existuje v „papírové“ podobě, ale jeho praktické nasazení není ekonomicky výhodné. Tento protokol měl využívat frekvenci sběrnice 160MHz a „double transition clock“, při celkovém počtu 34 signálních kanálů (kvůli diferenciálnímu signálování tedy 68 drátů) však narážíme na potíže s vzájemným ovlivňováním (tzv. crosstalk) a složitou synchronizací jednotlivých datových kanálů. Při současné technologické úrovni je odstranění těchto potíží neekonomické (již Ultra 320 bylo velkou výzvou pro vývojáře a noční můrou pro výrobce jak disků, tak řadičů či backplanů). Přímým následovníkem Ultra320 je tedy SAS první generace. Proč je vůbec nutné zabývat se vyšší propustností protokolu pro připojení diskových zařízení? Jaké výhody má sériové rozhraní?



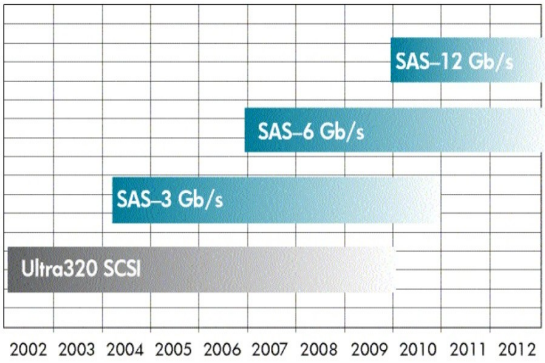
Obr. 5.3 SAS - full duplex

- neustále se zvyšující transakční efektivnost pevných disků (cca 30% ročně)
- menší množství drátů v kabelech, tj. levnější kabeláž, menší a levnější konektory - stávající paralelní konektory mají cca 80 pinů, sériové cca 12 (včetně napájení a uzemnění)
- jednodušší „backplane“ v serverech
- vyšší propustnost

SAS je hybridní technologie, která vznikla spojením nejlepších vlastností technologií stávajících. Množina příkazů použitá pro komunikací s diskovými zařízeními je převzata ze standartu SCIS 3, zapouzdřování do „framů“ nutné pro sériovou komunikaci bylo inspirováno FibreChannelem a elektrické signálování je identické se Serial ATA. První generace SAS má propustnost 3 gigabity za vteřinu na jeden kanál, který je vždy dedikován pro komunikaci dvou zařízení, tj. architektura Point-To-Point (rozdíl oproti stávajícím paralelním sběrnicím, kde na jednom kabelu mohlo být připojeno více zařízení – na paralelním SCSI podle typu použitého protokolu až 16 – u „narrow“ protokolů adresovaných třemi bity 8 zařízení včetně řadiče). SAS je možné snadno rozšířit pomocí tzv. „expanderů“, což jsou levné switche, které umožní na jeden kanál připojit více disků při zachování point-to-point topologie (v podstatě podobný princip jako u switchů ethernetových). Expandery umožní připojení nejen jednoho serveru k více diskům, ale i více serverů ke stejným diskům. SAS podporuje full duplex, kdy je možné komunikovat s diskem v obou směrech zároveň (např. posílat příkazy řadiči v disku a zároveň přijímat data nebo zapisovat data a zároveň přijímat potvrzení o úspěšném zápisu, viz. obrázek). Nezanedbatelnou výhodou je i možnost připojit přes SAS rozhraní SATA pevné disky, rozdělit tedy data na kritická a nekritická, a pro každé používat odpovídající pevné disky v jednom serveru nebo jednom diskovém poli (tj. levné disky např. pro zálohování nebo referenční data, ke kterým nepřistupujeme tak často a dostupnost 8x5 je vyhovující).



Obr. 5.4 SAS vs. SATA interoperability



Obr. 5.5 Budoucnost SAS

SAS již nemá designová omezení, která by bránila jeho vývoji, současný výhled na budoucnost SASu je znázorněn na obrázku.

Z pohledu fyzické velikosti disku se dnes můžeme setkat s konvenčními 3,5“ disky, typicky o rychlostech 10.000 nebo 15.000 otáček za minutu a pomalu se začínají prosazovat i disky 2,5“, zatím nejčastěji o rychlosti 10.000 otáček za minutu. Menší disky mají výhodu hlavně v nižší spotřebě a menší prostorové náročnosti, což jsou jedny z důležitých kritérií při vybavovaní datových center. U 2,5“ disků je nutné nezaměňovat s disky notebookovými, které kromě velikosti nemají s těmito „enterprise“ zařízeními nic společného.

5.1.2.DAS – direct attached storage



Obr. 5.6 Diskové pole

Nejlevnější a nejobvyklejší metoda pro ukládání dat je DAS. Jedná se o diskové zařízení přímo přípojené k serveru pomocí datového rozhraní. Jako implementace DASu vystupují jak interní disky v serveru, tak přímo připojené diskové pole, které nemá vlastní řadič. Pochopitelně je možné disky v DASové konfiguraci dále chránit různými technologiemi pro zvýšení odolnosti dat proti výpadkům.

5.1.3.NAS – network attached storage

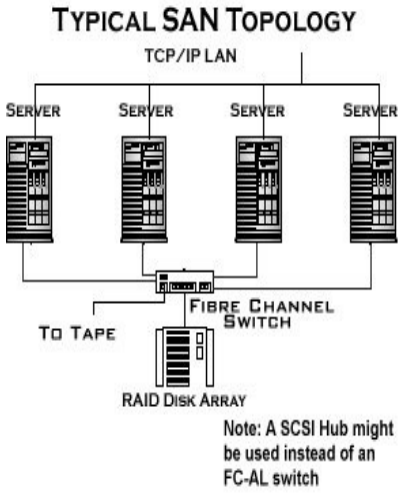
NAS je produkt, jehož smyslem je jednoduše a rychle přidat do IT infrastruktury zvýšenou diskovou kapacitu. Moji představu dokonale splňují takové produkty, jejichž prodejní cena je stejně vysoká jako v nich obsažené disky. Implementace spočívá v zapojení napájení a připojení do datové sítě. Po této síti je již zařízení schopno pomocí IT průmyslových standartů (CIFS, FTP, NFS, …) nabízet diskové služby. NASový produkt obsahuje jak diskovou část (zpravidla interní

nebo externí police pro disky, případně kombinaci obojího) tak i serverovou část. Řada renomovaných výrobců používá pro NAS řadové servery s upraveným OS (Linux, Windows Storage Server 2003) a dostatečnou diskovou kapacitou, součástí NAS produktu je tak většinou i řada technologií pro zajištění vyšší dostupnosti, často však nebývá splněn požadavek nízké ceny.

5.1.4.SAN – Storage Area Network

Na rozdíl od NAS, je třeba si pod zkratkou SAN představit nikoliv produkt, ale řešení. Řešení, které se snaží oddělit síťový provoz samotných aplikací od provozu, který má na starosti ukládání dat. Řešení, sestávající se z celé řady komponent, a zpravidla rozdělené na několik vrstev, o kterých pojednám později.. Z dosud představených přístupů SAN nejlépe vyhovuje požadavkům kladeným na dostupnost storage subsystému většího řešení. Je možné mnohem flexibilněji řešit dostupnost jednotlivých komponent řešení.

Jako první vrstva SAN řešení zpravidla vystupuje vrstva klientská, je jedno zda klienty jsou jiné severy v síti, aplikace zaměstnanců, aplikace, ke kterým přistupují přímo zákazníci či partneři společnosti.



Obr. 5.7 Typická SAN

Do druhé vrstvy patří servery, servery které pro klientskou vrstvu představují „přístupový“ bod pro klienty k samotné storage síti. Servery musí být vybavené technologií, která zajistí jejich konektivitu do storage sítě. V současné době je nejrozšířenější technologií FibreChannel, v serverech nalezneme speciální karty (nejčastější označení je HBA – Host Bus Adapter), které se většinou doplňují výměnnými konektory (GBIC – Gigabit interface converter). Protože se k přenosům používají sériová média (optické kabely) a principiálně paralelní protokol (SCSI), je nutné na úrovni HBA provést dvě konverze, konverzi elektrického signálu na optický (pokud tedy nepoužíváme levnější, ale méně výkonné metalické kabely) a konverzi paralelní vs.. sériovou. Součástí GBIC je i laser, který emituje světelný signál, na laseru a s ním spojeného typu kabelu je závislá maximální délka a propustnost. V současnosti používáme tyto typy laserů a kabelů:

- multimode – vlákno je široké 50 micronů (v Americe se používá 62,5 micronů), stínění 125 micronů, GBIC modulům pak říkáme shortwave, kabely mohou být 300 – 500 m dlouhé

- singlemode – vlákno je pouze 9 micronů široké (pouze jeden paprsek), používáme longwave GBIC, kabely 10 – 100 km



Obr. 5.8 GBIC

V nižších vrstvách jsou už samotné diskové police (zpravidla vybavené řadiči, které jsou schopné lépe disky využívat (viz. kapitola o RAID). Samotné disky jsou zpravidla připojeny ke řadičům metalickými spoji (buď paralelní SCSI nebo FibreChannel v topologii Arbitrary Loop. V této nejnižší vrstvě bývají typicky i řešení pro zálohování, tj. páskové mechaniky nebo robotiky – zařízení, která pro zvýšení zálohovací kapacity používají větší množství médií (magnetických pásek, magnetooptických nosičů) a speciální „robot“ pásky mění před čtecími/zápisovými hlavami. Čtecích/zápisových hlav může být větší množství (typicky dvě), tím zařídíme lepší dostupnost zálohy – je možné obnovovat data z nosičů a zároveň nenarušit zálohovací plán. Takovéto řešení si můžete představit např. jako Jukebox.

V neposlední řadě obsahuje SANové řešení síťové prvky, které se starají o konektivitu jednotlivých komponent vyjmenovaných výše. Typické síťové prvky jsou kabely (pasivní) a switche či huby (aktivní). Kabely mohou být metalické (méně časté – používáme typicky na backplanech v diskových policích) a optické (viz. výše). Switche a huby jsou analogické k ethernetovým (až na cenu pochopitelně :-)

Co se týče rychlosti, nejčastěji se setkáme s 1 Gb/s, 2 Gb/s a 4 Gb/s. Větší rychlosti (10 Gb/s) se běžně nepoužívají (rok 2006) a jejich praktické využití si představuji spíše v meziswitchové komunikaci.

5.1.5. Technologie diskových polí

S ukládáním dat úzce souvisí i technologie diskových polí. Diskové pole je skupina pevných fyzických disků, které se vůči operačnímu systému a aplikacím tváří jednotně jako logický disk. Typy diskových polí, na které u zákazníka můžete narazit, jsou dva – softwarový RAID nebo hardwarový RAID. Softwarový RAID je sice levnější, ale vzhledem k tomu, že veškeré operace dělení dat na bloky (striping), jejich distribuce na všechny disky a počítání parity musí provádět operační systém, procesor a paměti, diskové operace velmi zpomalují a zatěžují hostitelský

server. Naproti tomu hardwarový RAID zajistí, že veškeré činnosti spojené s ochranou dat provádí externí řadič vybavený procesorem.

Smysl sdružovaní fyzických disků do polí je zajistit:

- odolnost dat proti výpadku
- vyšší výkon
- vyšší kapacitu

Odolnost pole proti výpadkům zajistíme použitím paritních nebo zrcadlených typů RAIDu, tj. 5 nebo 1 (1+0). Vyšší výkon spočívá ve schopnosti více fyzických disků pracovat simultánně, vyšší kapacita pole je jasná, dnes nekoupíme desetiterrabytový pevný disk, chceme-li takovouto kapacitu, musíme disky sdružovat do polí.

Přehled nejpoužívanějších typů RAIDu:

Typ RAIDu	*ochrana dat ?*	*min. počet disků*	*režie (kolik disků nevyužijeme)*	*poznámka*
RAID 0	ne	0	0%	nejrychlejší RAID
RAID 1	ano	2	50%	používáme při dvou discích, chceme-li ochranu proti výpadku
RAID 1+0	ano	4	50%	používáme při čtyřech a více discích, nejrychlejší RAID s ochrannou dat
RAID 5	ano	3	1 disk	kompromis mezi výkonem, ochrannou dat a využitím diskové kapacity

Tabulka 5.1: Nejpoužívanější typy RAIDu

5.2.Napájení

Důležitým tématem při designu datového centra je zajištění dostatečného přísunu elektrické energie a její zabezpečení proti výpadku. Při řešení problematiky napájení se zamýšlíme nad těmito oblastmi:

- napájení uvnitř serveru – interní zdroje el. proudu
- dodávka energie pro zdroje – UPS
- zajištění energie pro datové centrum – generátory

5.2.1.Interní zdroje

Úkolem zdrojů v serveru je zajistit elektrickou energie pro vnitřní komponenty serveru a v odpovídajícím napětí a proudu (typicky 12V, 5V a nižší napětí pro procesory a paměti, např. 3,3V či 1,6V). Zdroj převádí napětí v našich zeměpisných šířkách typicky z 220V, v severní Americe ze 110V (většina moderních zdrojů je tzv. „autosense“, nebo-li zdroj se spokojí s čímkoliv mezi 100V-240V – na výstupu je vždy potřebné napětí bez ohledu na vstup). Některé specializované systémy (např. Ultratenké, žiletkové – v anglické terminologii používáme výraz „blade“) používají i jiné vstupní napětí (Hewlett-Packard pracuje ve svých bladových řešeních se 48 stejnosměrnými volty – v normální serverovně nic takového pochopitelně nenajdete, nicméně toto napětí se většinou používá u telekomunikačních společností, tedy společností, které tyto state-of-the-art technologie nejčastěji nakupují).

Vyšší dostupnost zdrojů se zpravidla provádí možností zapojit do systému druhý (nebo třetí) zdroj, který přebírá odpovědnost za dodávku energie při výpadku primárního zdroje. Redundance bývá řešena buď systémem N+1 (např. server potřebuje pro správnou funkci dva zdroje, pro zajištění redundance přidáváme třetí) nebo systémem 2N (pro správnou funkci jsou nutné dva zdroje, pro zajištění redundance přidáváme další dva). Moderní zdroje jsou schopny při plně funkčnosti (tj. v nedegradovaném režimu) sdílet zátěž, kdy pracují všechny zdroje včetně náhradních a nižším zatížením se tak prodlužuje jejich životnost. Většina renomovaných výrobců používá zdroje hotplugové, tak aby výměnu vadného zdroje za funkční bylo možné provést při provozu.

Zdrojům je nutno věnovat zvýšenou pozornost, a to proto, že podle mých zkušeností se jedná o druhou nejporuchovější komponentu serveru.

5.2.2.Nepřerušitelné zdroje napájení (Uninterruptible Power Supplies = UPS)



Obr. 5.9 UPS

Nepřerušitelné zdroje napájení (dále jen UPS) patří mezi neopomenutelné komponenty při plánování vyšší dostupnosti IT infrastruktury. Mezi hlavní přednosti těchto „inteligentních baterií“ patří:

- čas pro korektní ukončení systému při výpadku elektrické energie, případně jeho překonání v plně provozuschopném stavu
- konzistentní dodávka elektrické energie (stále stejné napětí i při jeho kolísání v rozvodné síti)
- ochrana zdrojů před poškozením v případě přepětí a náhodných špiček
- možnost spravovat funkce náhradního zdroje (management UPS)

Většina UPS umí regulovat podpětí a přepětí v rozsahu +/- 25% elektronicky bez asistence baterií (anglicky „boost“ nebo „buckle“ regulace), někteří výrobci jsou schopni vyrovnat výkyvy až +/- 35%. Častou možností je přidání další baterie pro prodloužení doby, po kterou je UPS schopna zásobovat server elektrickou energií. UPS vybavená kartou podporující SNMP umožňuje upozorňovat správu sítě o výpadku energie, či nutnosti vyměnit baterie, častou vlastností dražších zařízení je i plánovač pro postupné vypínání a zapínání jednotlivých portů, to umožní zajistit konzistenci dat (vypneme napájení do serverů dříve než do diskových polí) či prodloužení běhu kritický strojů na úkor méně důležitých zařízení (firewall s routerem běží do poslední chvíle, intranetový web server se vypne jako první). Při postupném zapínání zabráníme výpadku pojistek v serverovně (po zapnutí bývá okamžitý příkon až o 50% vyšší než při běžném provozu – to určitě znáte i z domova, kde před nahozením vypadlých jističů také vypínáte televizi a žehličku). Management se provádí buď přes zabudovaný webserver (v tomto případě tedy klasickým internetovým prohlížečem) nebo přes sériové rozhraní pomocí RS232 kabelu a použitím aplikace typu Minicom nebo HyperTerminal. Připojené servery jsou upozorněny přes RS232 nebo USB rozhraní o plánovém vypnutí napájení, tak aby operační systém mohl provést korektní vypnutí stroje, nebo alespoň kritické akce typu sesynchronizování souborové cache či korektní ukončení databázového serveru. Pro správný výběr UPS se dají použít různé webové či excelové kalkulátory jednotlivých výrobců:

- http://www.apc.com
- http://www.upssizer.com

5.2.3.Řízení toku vzduchu

Na rozdíl od desktopů či levnějších serverů, většina velkých strojů renomovaných značek nenechává chlazení jednotlivých komponent jednoúčelovým ventilátorům přímo na procesorech či pevných discích. Zodpovědnost za správné chlazení a řízení proudění vzduchu uvnitř serveru má na starosti celé šasi stroje. To bývá vybavené několika ventilátory, zpravidla rozdělených do zón, které mají chladit (např. procesorová zóna, I/O zóna, paměťová zóna), pro každou zónu je definovaný minimální počet ventilátorů, které jsou nutné pro správné chlazení. Typicky bývá o jeden (tzv. N+1) ventilátor v každé zóně navíc, tak aby v případě poruchy (nejčastěji zadření ložiska), bylo v dané zóně stále dostatečný počet ventilátorů. Systém dá zpravidla vědět administrátorovi (SNMP, mail, …), že došlo k závadě (typicky i s informací o objednacím čísle nového ventilátoru). I vizuálně pak bývá možné zjistit (LED) , že se serverem není něco v pořádku.

5.2.4.Racky a racková příslušenství

Pro jednodušší zajištění chlazení serverů, jejich fyzické bezpečnosti, připojení zálohovacích zařízení apod. se umísťují servery do speciálních skříní, tzv. racků. Ty se prodávají v různých velikostech, dobré je ověřit, zda mají dostatečnou hloubku pro námi používaná (nebo plánovaná) zařízení, výška racku se udává v tzv. „U“ (units, jednotky). Tato jednotka se používá i v popisu jednotlivých serverů (př. do 42U racku se vejde 21 „dvouúčkových“ serverů – zařízení, jejichž výčka je 2U). Dále doporučuji zamyslet se nad nutným chlazením serverů, zpravidla jsou počítány tak, aby vepředu nasávaly studený vzduch, proto bývá nutné používat racky s perforovanými dveřmi (min. 65%).



Obr. 5.10: Racky existují v různých velikostech

Pro rozvod elektrické energie do serverů používáme tzv. PDU – Power Distribution Unit. Ty se starají o konsolidaci kabelů v zadní části racků, obsahují zpravidla i pojistky proti energetickým špičkám, které by mohly poškodit zdroje serverů a některé z nich jsou dokonce ! typu „dual input“, mají dva vstupy, tak aby bylo možné do jednotky zapojit dva na sobě

nezávislé zdroje energie (např. standartní rozvod + rozvod z generátoru) , docílíme tím AC redundance.



Obr. 5.11: Power Distribution Unit - se dvěma vstupy

Nedílnou součástí serverovny jsou i zařízení pro lokální přístup k serverům, tzv. konzolové (někdy se udává zkratka KVM) switche – zařízení, která odbourávají nutnost připojit ke každému serveru separátní KVM (K=keyboard, klávesnice; V=video, obrazový výstup; M=mouse, myš). Můžeme nalézt jak switche analogové, tak digitální. Analogové switche trpí mnoha neduhy, na obrazovce se objevují stíny (obraz má „duchy“), některé operační systémy (Windows Server 2000) se špatně smiřují s odpojením a znovu připojením PS/2 zařízení. Proto se dnes zpravidla setkáme s digitálními switchy, které neustále emitují signál na PS/2 konektorech, tak aby si OS myslel, že má k sobě stále připojenou klávesnici a myš. Některé konzolové switche fungují jako nezávislé IP zařízení, na které je možné se vzdáleně přihlásit a „lokální“ práci se servery provést vzdáleně.

5.3.Ochrana paměti

Chyba paměti (v tomto odstavci budeme pojmem „paměť“ myslet operační paměť systému, tzv. RAM) kritickým způsobem ovlivní chod systému. Není možné si do proměnné „A“ uložit číslo 10 a v okamžiku kdy ji potřebujeme, přečíst číslo 74. Chyby v paměťovém subsystému nejsou v současné době ty nejčastější, s kterými je možné se v IT setkat (disky, zdroje, ventilátory v pravděpodobnosti chyb vítězí), nicméně pokud k nim dojde, následky jsou vážné. Výroba paměťových čipů je sice stále dokonalejší, nicméně množství paměti v serverech roste, pro každý přidaný GB do serveru zvyšujete pravděpodobnost paměťové chyby o 3%. Dalším problémem je snižování napětí na sběrnicích (v posledních letech poměrně populární), nižší napětí je pochopitelně náchylnější k rušení. Příčiny paměťových chyb mohou být různé, nejčastěji se hovoří o:

- elektromagnetickém rušení,
- kosmickém záření,
- fyzický defekt

5.3.1.Dělení paměťových chyb

Paměťové chyby nejčastěji dělíme podle následujících dvou hledisek:

- počet bitů zasažených chybou
- schopnost reprodukovatelnosti chyby

Podle počtu zasažených bitů rozpoznáváme typicky kategorie chyb:

- jednobitových – při zápisu nebo čtení je přehozena hodnota jednoho bitu, př. zapisujeme hodnotu 8 (tj. 1000), ale při uložení dojde k přehození druhého bitu a uložíme číslo 10 (1010)
- vícebitových – ukládáme hodnotu 12 (tj. 1100), ale uložíme 9 (1001)

Podle schopnosti reprodukovatelnosti:

- měkké chyby – objevují se nahodile; měření společnosti IBM ukazuje, že k nim dochází zhruba třikrát ročně (na jednom stroji s 4GB paměti).
- tvrdé chyby – trvalý defekt kondenzátoru v čipu, případně trvalý defekt ve spojení modul <--> paměťový řadič; uložení čísel 12 (1100) a 13 (1101) dopadne vždy stejně, uložíme 12 (1100), protože vada v kondenzátoru způsobí, že cokoliv chceme zapsat do prvního bitu (tj. v příkladu výše ten první vpravo), vždy přečteme „0“). Tvrdé chyby se nelze zbavit jinou cestou než vyměnit modul, paměťovou desku, paměťový backplane či základní desku (záleží pochopitelně na architektuře konkrétního stroje).

V drtivé většině případů se můžeme setkat s paměťovými chybami jednobitovými a měkkými.

5.3.2.Technologie určené k detekci a opravám paměťových chyb

Technologií a algoritmů určených k opravám a detekci chyb v paměťovém subsystému existuje velké množství, popíšeme si několik nejdůležitějších.

Parita

Ochrana paměti pomocí paritní informace je jednou z nejstarších a nejjednodušších. Při ukládání dat vypočítáme jednobitovou paritní informaci (typicky vypočítanou pomocí logické funkce „exkluzívního nebo“ (častěji se uvádí zkratka xor) proti nule nebo jedničce – podle toho,

zda chceme používat lichou nebo sudou paritu). Tuto paritní informaci uložíme společně s daty a při čtení pak snadno zjistíme, zda nedošlo k paměťové chybě.

Př. ukládáme číslici 138 (1000 1010), vypočítaná parita je 1, takže uložíme 9 bitů 1000 1010 1. Dojde k jednobitové chybě a při čtení paměťový řadič přečte 1000 1110 1, tj. číslici 142. Na běh systému by to mělo kritický dopad (představte si aplikaci, která se stará o zobrazování akciových kurzů na tabulích ve Wall Street). Paměťový subsystém je ovšem vybaven kontrolou parity, počet jedniček při čtení není sudý (ve výše uvedeném příkladě je to 5), přečtená data nejsou správná! Dojde k vygenerování NMI a zastavení běhu systému. Pro vyšší dostupnost sice žádný velký pokrok, ale zabránili jsme alespoň sebevraždám obchodníků.

Výhody parity ? Výpočet a kontrola je jednoduchá, tedy rychlá, tedy levná. Základní nevýhodou je nemožnost opravit tuto chybu (pouze detekce) a možnost nezachytit některé typy vícebitových chyb (viz. výše uvedený příklad, představte se, že přečteme vzorek 1000 1110 0, tj. opět číslici 142, nicméně vzhledem k tomu, že chyba se dotkla i uložené paritní informace, memory controller na nic nepřijde).

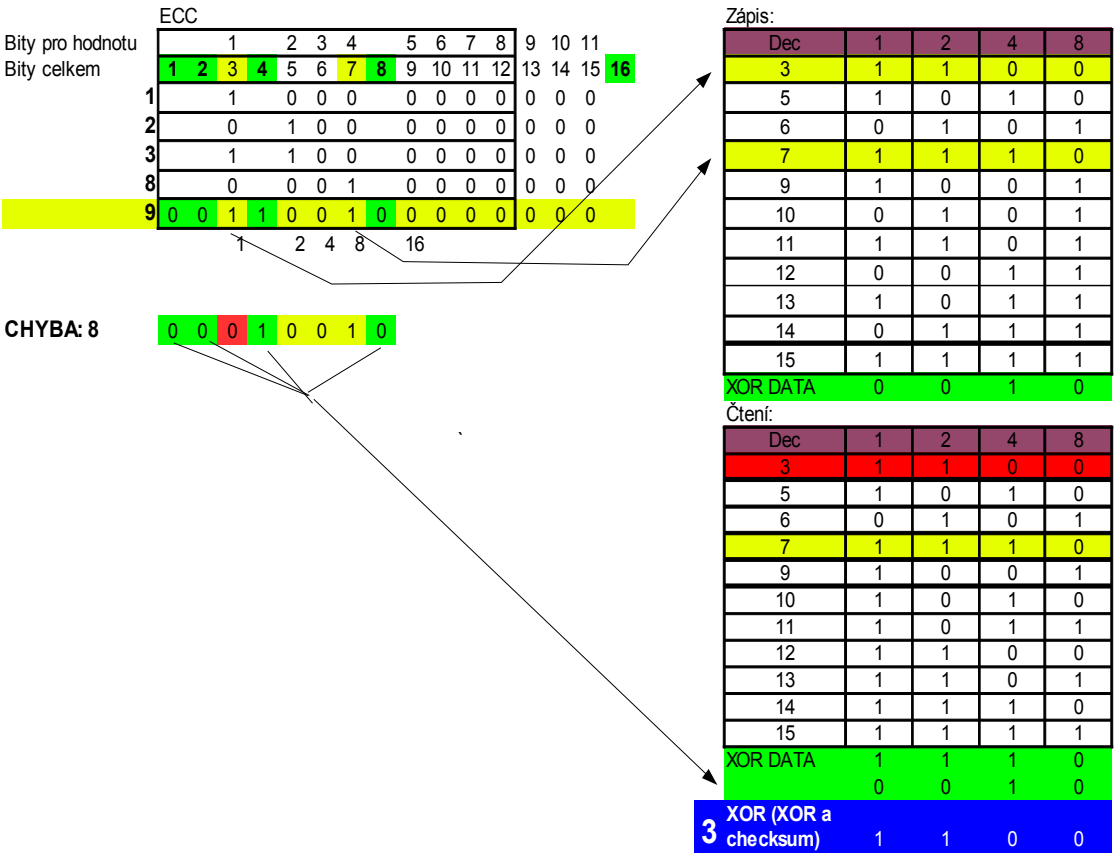
Error Checking and Correcting – ECC

Vzhledem k výše popsaným nedostatkům se s kontrolou pomocí parity už většinou nesetkáte (někdy se používá u datových přenosů, kde je možné přenos opakovat – to ovšem v případě paměťových operací není možné), všechny nové paměťové moduly však podporují technologii ECC. ECC je schopna spolehlivě detekovat až čtyřbitové chyby, a co je ještě důležitější, jednobitové chyby je možné opravit. ECC je založeno na Hammingově kódu. Při zápisu je vygenerován kontrolní součet, kterým je proložen vzorek ukládaných dat a společně uložen. Při čtení se kontrolní součet přepočítá a porovná s uloženým. V případě situace, že součty nesouhlasí, memory controller se pokusí zjistit, na které pozici došlo k chybě a dotčený bit otočí (tj. z nuly jedničky, z jedničky nulu) a správně spočítaná data se pošlou zařízení, které o ně požádalo (zpravidla procesor). Při porovnávání kontrolních součtů se použije populární xor , pokud vyjde nulový, došlo k vícebitové chybě, je vygenerováno NMI a systém se zastaví.

Př. Pro názornou ukázku jsem připravil tabulku, v které je proces výpočtu, zápisu, čtení a opravy naznačen, je názorně ukázáno, jak tento algoritmus funguje, jedná se o modelový příklad, prosím nekamenovat za drobná zjednodušení, která mají přispět k lepšímu pochopení problému (např. zápis binárních čísel zleva apod.) Představme si situaci kdy ukládáme číslici 9 v systému, který pracuje s 11-ti bitovými čísly (pro jednoduchost – v praxi se pochopitelně

s takto nesmyslným systémem nesetkáte, zpravidla pracujeme s 64-bitovými streamy dat). Číslice 9 je reprezentována binárně jako 1001 (v tabulce žlutý řádek označený jako 9, bity obsahujícíc samotná data jsou žlutá políčka, označení bitů najdete v prvním řádku „bity pro hodnotu“). Výpočet kontrolníního součtu je v pravé sekci tabulky (označeno jako „Zápis“). Stanovíme binární podobu čísla sloupce, ve kterém se vyskytují „nastavené“ bity (ty, které obsahují jedničku). Jako číslo sloupce použijeme druhý řádek v levé horní tabulce (tj. „bity celkem“ - číslujeme i sloupce, které budou obsahovat kontrolní součet). V našem případě se jedná o bity ve sloupci 3 a 7 (datové sloupce 1 a 4, tj. 1+8=9). Přes čísla sloupců (3 a 7) provedeme logické xor, tj. výsledkem je hodnota 0010 (zelený řádek v tabulce „Zápis“). Tímto kontrolním součtem proložíme datové bity a výsledek uložíme do paměti (datové bity žluté, ECC zvýrazněný zeleně v řádku 9 levé tabulky) – 00110010000000.

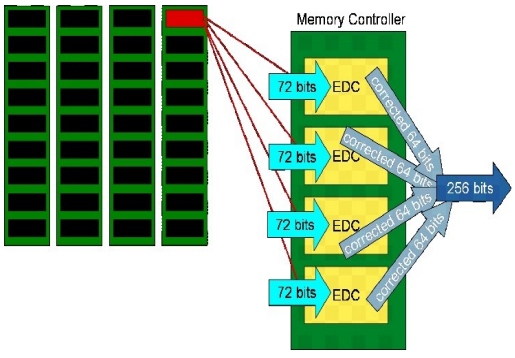
Při čtení dojde k jednobitové chybě, ze sloupce 3 místo uložené jedničky přečteme nulu, přečtená datová hodnota je tedy 1000 = 8 (řádek „CHYBA: 8“). Na to, že došlo k chybě přijdeme ihned, stejným způsobem jako při zápise spočítáme kontrolní součet (spodní pravá tabulka označená jako „Čtení), v tomto případě 1110 což evidentně nesedí s uloženým kontrolním součtem 0010. Pomocí oblíbeného xor přes kontrolní součty získáme číslici 0011 (v tabulce modrý řádek vpravo dole – v obrácené notaci 1100), což je v decimální podobě číslice 3. A co značí číslice 3 ? No přeci sloupec v kterém došlo k jednobitové chybě, stačí převrátit přečtenou hodnotu (v našem případě z nuly na jedničku) a získáme číslici 1001, neboli 9 ! Heuréka, nejen že jsme odhalili, že k chybě došlo (což můžeme ihned zalogovat a v případě, že na některém modulu dochází k příliš velkému množství měkkých opravitelných chyb můžeme se např. rozhodnout k výměně modulu – někteří výrobce toto dokonce nabízejí zdarma v rámci záruky), ale jsme schopni s kontrolního součtu i opravit přečtenou hodnotu a dostat se k původním datům !



Tabulka 5.2: Schéma funkčnosti ECC

Advanced ECC[2]

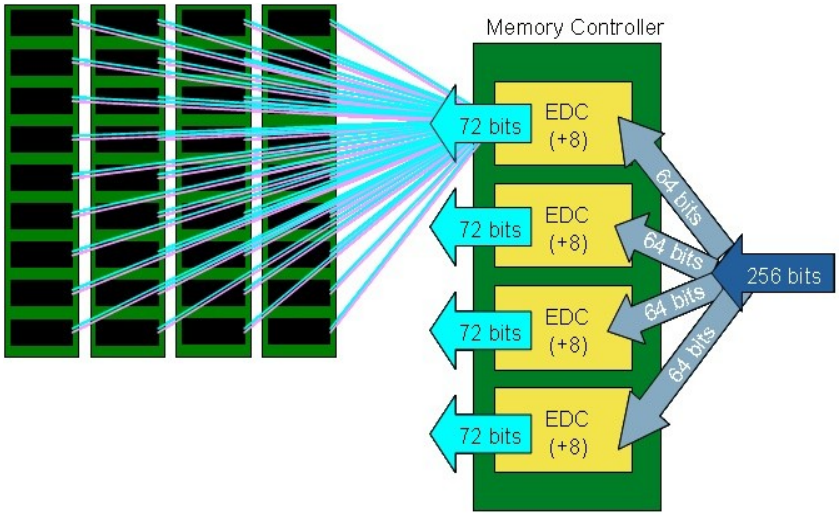
Některé nedostatky standartního ECC (hlavně nemožnost vypořádat se s vícebitovými chybami) vedly k vývoji lepší technologie, Advanced ECC. Advanced ECC je schopna se transparentně vyrovnat s vícebitovými chybami při použití standartního ECC algoritmu. Klíčová myšlenka spočívá v tom, že když už dojde k nepravděpodobné vícebitové chybě, dojde k ní v rámci jednoho DRAMu na paměťovém modulu. U Advanced ECC je proto každý 72 bitový vzorek dat (64 dat proložených kontrolním součtem) umístěno nikoliv na jeden DIMM, ale na všechny (max. 8) DIMMů v systému. Dojde-li k výpadku čipu na DIMMu, rozloží se závada na několik jednobitových chyb, které je možné spočítat v rámci klasického ECC. Tato technologie se dnes standartně



Obr. 5.12: AECC - výpadek DRAMu

2 Tato technologie bývá někdy nazývána „chipkill“ - pojem vystihuje podstatu věci a sice možnost překonat výpadek celého DRAMu, neboli „chipu“

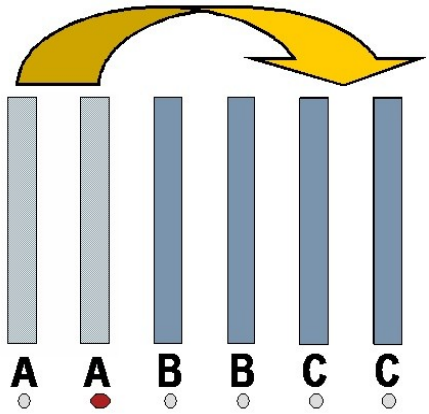
používá v serverových systémech, kde není problém osadit základní nebo paměťovou desku potřebným množstvím modulů (na desktopu tato technologie pochopitelně použít nelze, protože není kam data rozmisťovat).



Obr. 5.13: Zápis při použití Advanced ECC

Online spare memory

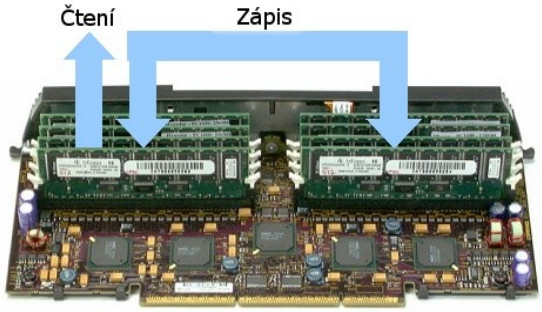
Nestačí-li nám výše uvedené technologie a chceme být imunní proti širšímu spektru paměťových chyb, příchází na řadu „online spare“. Při použití online spare pamětí vyčleníme jeden „bank“ (skupina DIMMů pracujících dohromady – použití dual channel) pamětí jako spare, paměťové moduly v tomto banku nejsou používané (tj. OS ani aplikace je nevidí a nemohou používat) a na řadu přijdu v okamžiku, kdy paměťový řadič usoudí, že na některém z produkčních DIMMů dochází k příliš velkému množství měkkých chyb. V tomto případě memory controller provede přesun banku obsahující vadný nebo podezřelý modul na spare. My si tak můžeme vybrat, kdy provedeme výměnu za nový (o víkendu, večer) a navíc odstraníme problém se zvýšenou latencí způsobenou nutností memory controlleru dopočítávat za pomoci ECC vadný bit.



Obr. 5.14: Přesun banku při zapnutém online spare

Př. Na obrázku je znázorněna situace, kdy bank C je označen jako spare a v pravém DIMMu v banku A došlo k překročení povoleného množství měkkých chyb, memory controller tak zahájil přesun dat z banku A do banku C.

Mirrored memory



Obr. 5.15: Zrcadlená paměť - použití jedné paměťové desky

Technologie zápisu do paměti, při které se každý blok dat zapíše dvakrát, se nazývá zrcadlená paměť (mirrored memory). Tato technologie nás chrání i proti výpadkům paměti, které již neopraví ECC nebo AECC (např. závada celého paměťového modulu).

Speciálním typem zrcadlené paměti jsou takové implementace, které umožňují výměnu vadného modulu za chodu (tzv. „hotplug“). Principem hotplugové operace je zajistit vypnutí napájení do komponenty, kterou chceme ze systému vyjmout (nebo naopak přidat – výměna je vlastně sekvence těchto dvou operací). V praxi se to dělá tak, že moduly umisťujeme na speciální paměťové desky, kterých je v systému více, min. dvě. Zápis pak probíhá vždy na obě paměťové desky a máme možnost jednu desku ze systému vyjmout. Při této konfiguraci je nejen možné vyměnit vadný modul (příp. celou paměťovou desku) za chodu systému, navíc můžeme i přidávat paměť za chodu (rozšíření fyzické paměti tak nemusí vyžadovat vypnutí systému), jedná se o další významný krok ke zvýšení dostupnosti serveru. Nutno podotknout, že je pochopitelně nutné používat operační systém, který paměť přidanou za chodu rozpozná (např. Windows Server 2003, Linux, HPUX, AIX; v současnosti stále využívané Windows Server 2000 toto nepodporují, potřebují k rozpoznání paměti restart, což hotplugovým paměťem přeci jen trochu ubírá na kouzlu).

RAID memory

V současné době je nejlepším dostupným systémem pro ochranu paměti implementace RAIDových paměťových modulů či desek. Zpravidla se používá pro výpočet paritní informace algoritmus RAIDu 4 (na rozdíl od diskových polí nedochází k úzkému hrdlu na paritním paměťovém modulu, protože vždy pracujeme s celou cache-line, takže požadavky na jednotlivé moduly jsou identické). RAIDové paměti mají nejlepší využití paměťových modulů, zpravidla 80 nebo 75 procent (u zrcadlených je to pouze 50%). Na druhou stranu je potřeba sofistikovaný

(čti drahý) a hlavně rychlý (na úrovni paměťového subsystému nesmíme dopustit zvyšování latence) paměťový řadič, který se objevuje jenom v chipsetech nejvýkonnějších serverů a proto je rozšíření tohoto typu pamětí poměrně nízké.

5.4.Zálohování

Zálohování patří do kategorie technologií sloužících pro rychlé zotavení v případě potíží, které nejsou zachyceny a transparentně vyřešeny některou technologií nebo postupem patřící do kategorie „fault tolerance“, příp. „fault prevence“.

Podle cílové komponenty, na kterou se záloha provádí rozlišujeme mezi těmito typy zálohování:

- zálohování na magnetické nebo magnetooptické mechaniky
- zálohování na disky

Cílové zařízení zpravidla vybereme na základě požadavků od zákazníka na výkonnost (rychlost zálohování a obnovy, schopnost práce bez obsluhy, …) a cenu. Požadovanou rychlost zálohování spočítáme jednoduše z množství dat, která bude nutné zazálohovat a množstvím času (tzv. „zálohovací okno“), které na zálohování máme:

$$Požadovaná\ přenosová\ rychlost\,(GB/hod) = \frac{(Množství\ dat\ v\ GB)}{(zálohovací\ okno\ v\ hodinách)}$$

V potaz je nutné vzít i propustnost kanálu směrem k zálohovacímu zařízení – tj.:

- rychlost I/O sběrnice (PCI-X, PCIe),
- propustnost síťové karty, optické karty, SCSI/ATA adaptéru,
- propustnost síťové infrastruktury a jejich prvků (SAN switche, ethernet),

aby se nám nestalo, že máme sice nejnovější Ultrium ale nejsme schopni jej zásobovat daty. Páskovací mechaniky je nutné zásobovat takovou rychlostí, která zajistí, že budou zapisovat nepřetržitě na své nominální rychlosti (tzv. streaming). V opačném případě neustálé zapínání a vypínání páskové mechaniky značně sníží rychlost a životnost pásek.

Před samotným zálohováním je nutné připravit zálohovací (rotační) schéma, jak často a jaké typy dat budeme zálohovat. Zálohy dělíme na plné a částečné, částečné pak dále na diferenciální a inkrementální. Diferenciální záloha ukládá nové nebo změněné soubory od poslední plné zálohy, inkrementální od poslední zálohy. Výhoda diferenciální zálohy je rychlejší obnova na úkor většímu množství dat, inkrementální záloha ukládá menší množství dat (pouze to co se změnilo od poslední zálohy), v případě obnovy však musíme aplikovat všechny zálohy od poslední plné zálohy a to postupně ve správném pořadí. Nečitelnost jakékoliv pásky v tomto řetězu způsobí nemožnost obnovit data.

Den	*Soubory na disku a verze*	*Diferenciální záloha*	*Inkrementální záloha*
Pondělí	a1,b1,c1		
Úterý	a1,b2,c1	b2	b2
Středa	b2,c2	a-,b2,c2	a-,c2
a1	soubor „a“ - první verze		
c2	soubor „c“ - druhá verze (tj. proběhla změna)		
a-	informace o vymazání souboru „a“		

Tabulka 5.3: Rozdíly mezi diferenciální a inkrementální zálohou

Často se používá zálohovací schéma dědeček-otec-syn (Grandfather-father-son – GFS), kdy jednou měsíčně provedeme plnou zálohu, kterou archivujeme mimo budovu, jednou týdně (na začátku nebo na konci) provedeme plnou zálohu uchovávanou v budově a každý den provedeme inkrementální nebo diferenciální zálohu. Při použití tohoto schématu se vždy můžeme vrátit k jakémukoliv stavu v historii a je zachován optimální poměr mezi množstvím zálohovaných dat a rychlostí případné obnovy. Každý týden máme referenční plnou zálohu provedenou např. o víkendu a jsme chránění i proti katastrofě na úrovni budovy (neobnovíme sice vše, ale ztráta dat a uchovávaných pásek onsite neznamená ještě likvidaci firmy).

Dalším typem záloh (uvádí např. [1]) je Hanojská věž (Tower of Hanoi). Schéma Hanojské věže vychází z logické hry, které má svůj původ v Číně. Hra je postavena na přesunutí pěti kotoučů z jednoho kolíku na některý další, a to s minimálním počtem tahů. Přitom hráč nikdy nesmí mít v ruce více než jeden kotouč a nikdy nesmí dát větší kotouč na menší. Je dokázáno, že nejmenší počet tahů je 31.

Metoda Hanojské věže využívá pro zálohování pět media setů:

- Media set **A** je použit každý jiný den

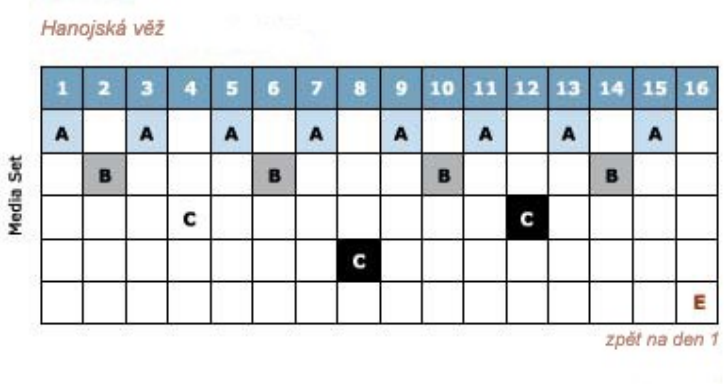
- Media set **B** je použit každý čtvrtý den
- Media set **C** je použit každý osmý den
- Media sety **D a E** jsou použity střídavě každý šestnáctý den

Naplánování schématu **Tower of Hanoi** je následující:

Záloha začíná na media set „A“ a pak následně pokračuje v každý jiný den. Další záloha probíhá na media set „B“ (ne však v den kdy proběhla záloha na media set „A“) a následně se opakuje každou čtvrtou zálohu. Media set „C“ nezačíná v den zálohy media setů „A“ a „B“ a opakuje se každou osmou zálohu. Pro média sety „D“ a „E“ je nastavena politika následovně. První záloha nezačíná v den zálohy media setu „A“, „B“ a nebo „C“ a opakuje se každou šestnáctou zálohu.

Předností tohoto schématu je především možnost přidání nového média setu a tím získání větší historie zálohy (obdobně jako u GFS). Častěji používané media sety obsahují novější kopie souborů, zatímco méně používané media sety obsahují starší verze souborů.

Toto schéma je celkem obtížné na manuální správu. Proto je pro toto schéma doporučeno používat zálohovací software s možností plánování celého procesu (např. NetVault 7.1) a hlavně za použití páskového autoloaderu (např. Tandberg, autoloader SLR140) nebo vhodnější řešení s více sloty a to páskové knihovny pro dostatečný počet media setů zahrnující pásky pro zálohování, archivaci a disaster recovery řešení. Stejně jako schéma Grandfather-Father-Son umožňuje řešení Tower of Hanoi periodické vyjmutí media setu za účelem archivace.



Obr. 5.16: Hanojská věž, [1]

Nejjednodušší zálohou dle [1] je systém **Round Robin** (schéma s jednou páskou pro každý den).

Nejjednodušší schéma rotace pásek získáme, když vyhradíme pro každý den pracovního týdne jednu pásku. Pásky jsou označeny (pondělí, úterý, středa, čtvrtek, pátek). Každý den je provedena na danou pásku plná záloha všech dat vyhrazených pro zálohování. Tato rotace umožňuje obnovu dat s maximálním posunem zpět - jeden týden. Schéma je vhodné pro použití v menších společnostech s využitím interní nebo externí páskové mechaniky a nebo s využitím zařízení NAS s vytvořenou VDL, (virtual disk library) která může sloužit jako primární úložiště dat. Toto řešení je vhodné také tam kde je možné provádět každý den plnou zálohu a časový posun o 1 týden zpět je dostačující.

Nestačí pochopitelně pouze pravidelné zálohování, je potřeba připravit i plán co dělat v případě nutnosti provést obnovu. Tento plán musí být přístupný všem zainteresovaným osobám a musí obsahovat:

- popis zálohovacího schématu (abychom věděl, kde začít a které zálohy je nutné dodatečně obnovit pro dosažení cílového stavu)
- vyzkoušený postup, jak obnovu provést (osvědčil se kompletní postup se screenshoty a popisy kam kliknout – nikdy nevíme, kdo konkrétní bude obnovu provádět a čím přesnější postup máme, tím méně kvalifikovanou osobu tím můžeme pověřit)
- naplánovat ukládání pásek – kam vozit offsite zálohy (k administrátorovi domů) a kam ukládat onsite pásky (do žáruvzdorného trezoru)
- postupy pro obnovu prostředí (nastavení síťových parametrů, konfigurace switchů, apod.)
- umístění instalačních médií a licenčních informací

5.5.Imaging, bezzásahové instalace

Pro urychlení zotavení po havárii se často nepoužívá obnova způsobem nové instalace OS, aplikací a obnova dat, ale snahou je tuto činnost zjednodušit, zrychlit a zautomatizovat.

První cestou je tzv. imaging, tj. zazálohujeme obraz diskového oddílu (partition), tj. včetně OS, aplikací i dat. Výhodou je velmi rychlé obnovení systému, nevýhodou časová náročnost zálohování i skutečnost, že daný stroj musí být „offline“ (zpravidla nabootovat z CD, diskety nebo sítě přes PXE).

Druhá cesta je automatizovaná instalace, když využijeme schopnosti většiny moderních instalalátorů OS připravit si odpovědi na jednotlivé otázky během instalace do speciálního souboru, tak abychom při instalaci již nemuseli na nic „klikat“. Ve Windows se tento soubor jmenuje unattend.txt, u RedHatových distribucí Linux kickstart.cfg.

5.6.Vzdálený management a monitoring

Nezbytnou podmínkou při zajišťování vyšší dostupnosti je monitorování serverů. K tomu dnes používáme řadu protokolů, nejčastěji SNMP (simple network management protocol), který dovolí používat různé aplikace různých výrobců pro monitorování různých zařízení.

Administrátor musí být vždy první kdo se o závadě dozví, tak aby mohl co nejrychleji zjednat nápravu. Obecné přístupy k monitorování jsou dva:

- pull monitoring
- push monitoring

Každý má svá pro a proti, ideální je pochopitelně kombinace obého. Pull monitoring znamená, že periodicky zjišťujeme stav monitorovaných síťových elementů (nejen servery, ale i síťové prvky, tiskárny, důležité stanice, …). Výhodou je, že v okamžiku provedení kontroly máme jistotu, že stroj funguje. Nevýhodou je zbytečně generovaný síťový provoz a fakt, že se o závadě dozvíme až při refreshi stavu. Push monitoring je druhý přístup, zařízení samo o sobě pošle do centrální management aplikace zprávu, že má nějaký problém. Výhodou je „real-time“ monitorování a eliminace zbytečného provozu. Na druhou stranu některé závady zabrání zařízení, aby informovalo management aplikaci, že se něco děje[3]

Pokud zjistíme závadu, řešení výrazně urychlí možnost administrátora vzdáleně se na stroj přihlásit a provést bližší hledání chyby. K tomu používáme softwarové vzdálené konzole (RDP – Remote Desktop Protokol společnosti Microsfot; VNC) nebo hardwarové konzole. Speciální hardwarové karty nám umožní sledovat, jak server bootuje, vzdáleně přistupovat k BIOSu apod.

---

3 Představte si situaci, kdy dojde k závadě síťové karty. Operační systém (nebo speciální mng. agenti u renomovaných výrobců) vygeneruje trap, popisující závadu a detaily, které se k tomu vztahují (např. která síťová karta selhala. Nicméně vzhledem k tomu, že server již nemá síťovou konektivitu, informace se k nám vůbec nedostane. Jiný případ je např. vyhoření stroje, po zásahu bleskem již nezbyde nic, co by nás mohlo o havárii informovat.

Karta zpravidla obsahuje vlastní VGA, tak aby procesor měl přístup k obrazu, nezávislé TCP/IP spojení do serveru a zabudovaný webserver.

5.7.Poruchovost jednotlivých komponent

Nejporuchovější komponentou v serveru jsou pevné disky, zdroje a ventilátory. Příčinou jsou pochopitelně mechanické komponenty. Z elektronických součástí jsou to paměti (zde však k závadě zpravidla nedochází v průběhu životnosti, paměťové moduly někdy bývají nefunkční již při dodání – tzv. dead on arrival).[4]

4 Zdrojem jsou rozhovory a vyplněné dotazníky servisními techniky společností Gesto Computers, Autocont CZ a Informatika České Spořitelny.

6.Organizační a personální zajištění dostupnosti

6.1.SLA

Jak jsme si mohli všimnout v kapitole 2, velká část nedostupnosti systému je způsobená chybou obsluhy, proto bychom ve společnosti měli klást důraz i na organizační zabezpečení chodu infrastruktury.

Často používaným prostředkem pomáhajícím zajistit dostupnost služeb IT infrastruktury je definovaní tzv. SLA – Service Level Agreement. Service Level Agreement, zkráceně SLA, je dohoda o úrovni poskytovaných služeb. Umožňuje definovat kvalitu a garance IT služeb, například maximální povolený výpadek, odezvu aplikace nebo dostupnost a rychlost reakce technické podpory. Obsahuje minimální hranice, nad jejímiž úrovněmi jsou dodávané služby ještě akceptovatelné.

Podle [11] by SLA měli obsahovat následující tři části:

1. Základní specifikace, podmínky a pravidla
2. Tvrdé metriky
3. Měkké metriky

Do základních specifikací patří:

- Kategorie příjemců.
- Přesné vymezení počtu a umístění příjemců dané kategorie.
- Popis služeb
- Objem poskytovaných služeb (aplikace ve vztahu k počtu zpracovávaných dokladů v dané periodicitě apod.)
- Poskytovatel - bližší určení (uvádí se, má-li smysl, resp. je-li uživatelů více)
- Měření - postup, způsob, periodicita, odpovědnost a vykazování výsledků

- Ověřování - postup, způsob, periodicita, odpovědnost a vykazování výsledků ověřování správnosti měření
- Určení a způsobu realizace podpory (kupř. fyzicky na místě, vzdáleně apod.)
- Návazné podpůrné služby spojené s danou službou (kupř. tréning na místě, resp. na učebně)
- Cena služby
- Platební podmínky
- Pravidlo pro změny služby
- Práva a povinnosti obou stran - podmínky součinnosti
- Ostatní podmínky pro realizaci SLA (bezpečnost, právo informovanosti, odpovědnost za vady a škody apod.)

Tvrdé metriky jsou definovány jako:

- Dostupnost (v % vyjádřený skutečný čas disponibility aplikace na daném zařízení uživatele ve vztahu k celkovému efektivnímu fondu pracovní doby za určenou časovou jednotku)
- Běžná a maximální přípustná (kritická) doba odezvy na požadavek - tzv. incident (v členění na jednotlivé typy požadavků, jako je např. hlášení poruchy aplikace , poruchy HW, přemístění koncové stanice, apod.)
- Běžná a maximální přípustná (kritická) doba řešení požadavků (v členění na jednotlivé typy požadavků).
- Průměrná a mezní odezva aplikace v rámci služby.

Mezi měkké metriky patří další kvalitativní ukazatele (zápisy a akceptace při přebírání aplikace či nové konfigurace, ústní hodnocení spokojenosti, …)

6.2.Oprávnění

Jak už jsem zmiňoval výše, častá příčina nedostupnosti služby je chyba způsobená lidským faktorem, jedním z prvků, které se snaží tuto příčinu odstranit je omezit uživatelům možnost provádět zásahy ovlivňující chod infrastrukturních zásahů. Uživatelům je systémem práv (využíváme zpravidla adresářové služby společnosti Microsoft – systém Active Directory, případně adresářové služby společnosti Novell – eDirectory) umožněno přistupovat pouze k takovým službám a prostředkům, které potřebují ke své práci. Např. uživatel smí tisknout na tiskárně na svém patře, ale nesmí již manipulovat s frontou (mazat dokumenty apod.). Zabezpečení konzolového přístupu k serverům a povolení přístupu pouze pro administrátory je samozřejmostí. Pro přidělování práv zpravidla bývá nastavený proces s definovaným postup, jak o práva žádat, určení pracovníků, kteří práva schvalují a postup pro samotné přidělení práv.

7.Případová studie

7.1.CES holding

7.1.1.Historie

Central European Services, a.s. (původně Ibi-ubi, a.s.) byla založena v roce 1996. Central European Services, a.s. se stala matkou postupně vznikajících dceřiných firem. Se změnou jména společnosti Ibi-ubi, a.s. na Central European Services, a.s. byla definována nová a jediná úloha mateřské akciové společnosti jako správce majetkových účastí v dceřiných společnostech.

V lednu 1997 vznikla Pro Futuro Consulting s.r.o. (PFC), jejíž činností bylo školení firem prodávající produkty Hewlett-Packard v rámci tehdy vznikajícího školicího programu HP STAR. První rok činnosti PFC byl charakterizován pracovním nasazením malého týmu lidí, kteří během jednoho roku dokázali zahájit program HP STAR a vytvořit jeho dobrou pověst, zřídit a vybavit nové kanceláře a učebny pro další růst firmy a vytvořit potřebné finanční zdroje pro další roky činnosti společnosti.

Rok 1998 byl zcela ve znamení přijímání nových zaměstnanců a rozvíjení činnosti, tedy poskytování školení, prezentací a konzultací firmám působícím v oblasti IT. Během roku 1998 se podařilo zahájit činnost školení servisních techniků HP produktů a definovat další tři činnosti, které měla firma rozvíjet v dalším roce své činnosti.

V roce 1999 se PFC zkonsolidovala a rozvinula nové projekty, například kompetenční středisko řešení pro enterprise segment na platformě Microsoft a Hewlett-Packard, nebo tisková řešení a document management. Úspěšný prodej služeb se příznivě projevil na hospodaření firmy. Během roku 1999 podařilo se také vybudovat plně funkční PFC ve Varšavě v Polsku.

V roce 2000 PFC začala zúročovat své znalosti a zkušenosti na zahraničních trzích. A již do konce roku 2001 obsloužila svými školicícmi aktivitami klienty z 36 zemí Střední a Východní Evropy, Blíkého východu a Afriky.

Úspěšný vývoj společnosti byl potvrzen v roce 2002 její transformací ze společnosti s ručeným omezením na společnost akciovou a v roce 2003 získáním mezinárodního certifikátu systému řízení jakosti dle mezinárodní normy ISO 9001:2000 u společnosti Bureau Veritas. V roce 2004 firma

významně zvýšila svoji podnikatelskou efektivitu prostřednictví optimalizace interních procesů a zdrojů.

Společnost Amden s.r.o. byla založena v polovině roku 1997 za účelem zastupování společnosti Hewlett-Packard v oblasti prodeje jejího spotřebního materiálu. Amden s.r.o. dále získala od společnosti Hewlett-Packard zakázku na vybudování a provozování demonstračního centra pro plottery. V listopadu roku 1997 byla společnost Amden převzata akciovou společností Central European Services, a.s., která na ni převedla svou vlastní činnost v oblasti přímého marketingu. Tím ze společnosti Amden vznikla marketingová agentura poskytující služby v oblasti podlinkových marketingových aktivit.

Během roku 1998 společnost získala další zakázky na administraci dlouhodobých marketingových programů, k jejichž zajištění bylo potřeba přijmout další pracovníky. Během roku 1999 řešil Amden hlavní problémy v oblasti personální, a to proškolováním zaměstnanců a jejich částečnou obměnou. Na podzim došlo i k rozhodnutí o založení společnosti Amden v Polsku a dalším rozvoji společnosti Amden na Slovensku.

Na aktivity společnosti Amden měla v roce 1999 velký vliv spolupráce se sesterskou společností Pythagoras, která pro Amden začala vyvíjet aplikace v oblasti e-marketingu, konkrétně v oblasti nazývané CRM (Customer Relationship Management).

Z důvodů rozšíření portfolia činnosti Amden zvýšení počtu klientů a poptávky po službách společnosti Amden byly v průběhu let 2000-2003 založeny společnosti Amden v Maďarsku a na Ukrajině.

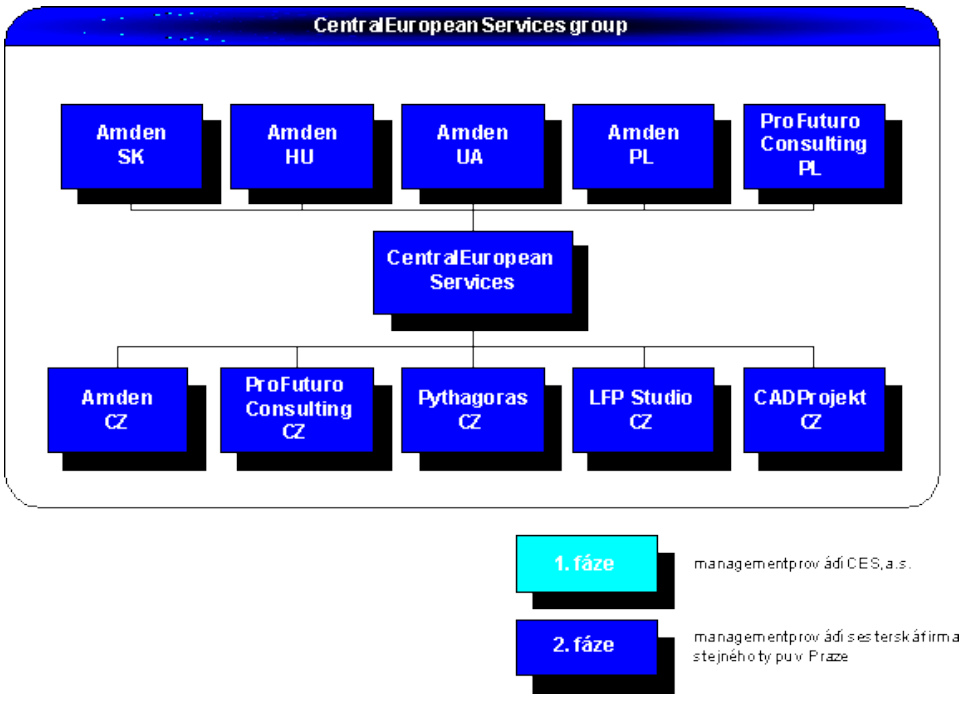
Společnost CAD Projekt s.r.o. byla založená na jaře 1997 za účelem poskytování kresličských prací projektovým a architektonickým firmám a ateliérům. V roce 1997 společnost investovala do pracovních stanic, plotteru, zřízení sítě LAN a do vybavení kanceláří, které poskytovaly dostatečný prostor pro práci až 10 zaměstnanců.

CAD Projekt se postupně rozvíjel, na podzim roku 1998 získal velkého klienta, projektovou kancelář se zahraničním know-how, tedy firmu, která pochopila výhodnost služby nabízené společností CAD Projekt a začala ji rozsáhle využívat. V důsledku toho byl celý rok 1999 poznamenán přijímáním zaměstnanců, jejich postupnou stabilizací, vytvářením provozních pravidel, ceníku a formulováním strategie. Ke konci roku došlo k vyprofilování CAD Projektu na firmu poskytující HW a SW vybavení a vyškolené kresliče a projektanty projektovým a architektonickým kancelářím na zpracování zakázek.

Společnost LFP Studio byla založena na jaře roku 1998. Klienty LFP Studia jsou reklamní agentury, výstavářské firmy a vlastníci reklamních ploch, pro které se zhotovují velkoformátové tisky na různé druhy papírových, umělohmotných a textilních médií.

Na jaře roku 1999 byla do holdingové struktury začleněna společnost Pythagoras s.r.o. Jednalo se o velmi malou společnost, která však měla náskok v oblasti vývoje intranetových a internetových aplikací. Cílem začlenění bylo docílení symbiotických efektů, kdy aplikace vyvíjené společností Pythagoras budou využívané sesterskými společnostmi holdingu, (převážně Amdenem) pro něž budou zdrojem dalších obchodních příležitostí. Během roku 1999 došlo k vybudování nové kanceláře Pythagora a rozšíření počtu zaměstnanců. Ke konci roku byla dokončena profilace vyvíjeného SW na aplikace z oblasti Customer Relationship Managementu.

V průběhu let 2002 až 2004 společnost podstatně rozšířila portfolio svých Internet a Intranet řešení.



Obr. 7.1: Struktura holdingu CES

7.1.2.Profil

Strategie skupiny CES, a.s.

Využívat příležitostí k investování do společností, které slibují budoucí úspěšné zhodnocení takovéto investice

- zakládat nebo kupovat dceřiné firmy a rozvíjet je v rámci holdingového prostředí CES, a.s.
- vypracovat a uplatnit dlouhodobý plán rozvoje dceřiných společností včetně případného vyhledání jejich budoucího strategického partnera

Usilovat o teritoriální rozšíření dceřiných společností, pokud je to možné a vhodné.

CES, a.s, je investiční společností, jejímž cílem je dlouhodobé zhodnocení investic, majetkových účastí v dceřiných firmách.

Cíle společností holdingu Central European Services, a.s.

Profit

Cílem je dlouhodobé dosahování maximálního možného profitu při zachování našich ostatních cílů. Profit je jediným možným měřítkem smysluplnosti naší práce a měřítkem síly naší společnosti.

Zákazník

Usilujeme o kontinuální zlepšování kvality a užitečnosti našich služeb a produktů dodávaných zákazníkům. Uvědomujeme si, že není důležité, zda zákazník má či nemá pravdu, důležité je, zda je spokojený.

Obory podnikání

Koncentrace úsilí a vyhledávání nových obchodních příležitostí do oborů, které souvisí s tím, v čem již společnost podniká.

Růst

Usilování o růst, protože je podmínkou přežití. Růst rozšiřuje firemní možnosti a možnosti profesního růstu.

Zaměstnanci

Všichni se společně podílí na úspěších a neúspěších společnosti. Úspěch společnosti závisí na zvyšování schopností zaměstnanců, a to získáváním zkušeností, školením, studiem a kvalitním výběrem nových spolupracovníků. Základem spokojenosti zaměstnanců v práci je uspokojení z výsledků. Předpokladem společné práce je vzájemná důvěra.

Organizace

Základem organizace je:

1. týmová spolupráce,
2. maximální využívání všech zdrojů, které poskytuje společnost,
3. přidělení pravomocí a následné převzetí odpovědnosti za realizaci vytyčeného cíle.

7.1.3.Dceřinné společnosti

Amden

Společnost Amden je marketingovou agenturou, která poskytuje marketingové služby zaměřené na podporu prodeje a tvorbu marketingových systémů. Provádí analýzu marketingových činností a procesů, jejich vyhodnocení a navržení vhodných marketingových programů, které také realizuje a následně vyhodnocuje jejich účinek. Dokonale zná potřeby zákazníků svých klientů a výborně se orientuje v jakémkoli odvětví trhu. Sídla má na území České a Slovenské republiky, Polska, Maďarska a Ukrajiny. Mezi její nejvýznamnější klienty patří společnosti Hewlett-Packard, Olympus, Motorola, Expert&Partner Computer 2000 Group, eD´system, Xerox, Mediasoft, Intergraph, ComputerCity, CSc Computer Services CZ, Financial Times, Procter&Gamble, Autodesk, Diderot, Esso, Tesco

Pro Futuro Consulting

Společnost poskytuje hi-tech konzultace, prezentace a školení firmám zabývajících se výpočetní technikou a jejich klientům pro produkty a služby společností Microsoft, Hewlett-Packard, SuSE. Společnost poskytuje konzultace, vyvíjí a instaluje SW a HW aplikace v oblasti tiskových řešení a v oblasti managementu sítí.

Mezi nejvýznamnější klienty patří Hewlett-Packard, Microsoft, distributoři a dealeři HP a MS, Komerční banka, ČSOB Pojišťovna a.s., NKÚ, UniControls a.s., TATRA BANKA a.s., Kaufland.

V oblasti školení má klienty nejen z České a Slovenské republiky, ale z dalších 36 zemí Střední a Východní Evropy, Blízkého východu a Afriky.

LFP Studio

Společnost je výrobcem fotorealistických velkoformátových tisků tištěných ink-jet technologií na širokou škálu materiálů. Zabezpečuje předtiskovou přípravu, tisk i následné zpracování - laminaci, kašírování na různé druhy podkladu a formátování. Hlavními klienty jsou reklamní a výstavářské agentury a majitelé reklamních ploch.

CAD Projekt

Společnost poskytuje kresličské a projekční služby ve stavebnictví formou pronájmu HW, SW a vyškolených kresličů a projektantů architektonickým studiím a projekčním kancelářím. Mezi nejvýznamnější klienty patří Skanska-stav, Loxia, AED Project.

Pythagoras

Společnost vyvíjí a prodává intranetové a internetové aplikace zaměřené na e-commerce, vzdělávání po internetu a Customer Relationship Management. Aplikace jsou založeny na intranetové technologii na platformě Microsoft, což významně zjednodušuje správu, instalaci aplikací, jejich ovládání, licencování a propojení s MS Office produkty. Mezi nejvýznamnější klienty patří Compaq Computer (dnes již součástí HP), Hewlett-Packard, Autodesk, Messe Duesseldorf, Tesco Stores.

MTEC

Distribuce a prodej komínových systémů ASV. V současné době jediná ne-IT společnost v holdingu.

7.1.4.Serverová infrastruktura, použité technologie

V holdingu je serverová infrastruktura tvořená následujícími stroji:

- INTERNATOR
- INFORMATOR
- MODERATOR
- DISTRIBUTOR
- TRANSFORMATOR
- PRIMATOR
- NAVIGATOR
- ALLOCATOR
- ACTIVATOR

(Detaily o serverech jsou uvedeny níže)

Některé dceřinné společnosti, mají další stroje pro zabezpečení specifických úkolů nesouvisejících s holdingem a ostatními sesterskými společnostmi:

- ALIGATOR – vývojový server společnosti Pythagoras
- MAIL.CES.CZ – mailserver pro managera zahraničního rozvoje
- GWCES – server, ke kterému se nikdo nehlásí, proběhne řízené vypnutí a přetransformování na jiné použití
- PROJECTOR – sdílení nefiremních (uživatelských/soukromých) dat – Pythagoras
- MTEC-FAX – faxserver pro MTEC
- SYLVAN-SERVER – server pro testovací centrum (Pro Futuro Consulting)
- PFC-TERMINATOR – quality management systém – nikdo nepoužívá, bude vypnut a přetransformován na jiné použití
- PFC-HPSTAR – primary domain controller (PDC) pro učebny

- PFC-HCY-MTRACK – parsování tiskových úloh, monitorování a reporting tisků – Pro Futuro Consulting
- PFC-HCY01 – server, ke kterému se nikdo nehlásí, proběhne řízené vypnutí a přetransformování na jiné použití
- PFC-PROPRINT – server, ke kterému se nikdo nehlásí, proběhne řízené vypnutí a přetransformování na jiné použití
- ZETOR – server, ke kterému se nikdo nehlásí, proběhne řízené vypnutí a přetransformování na jiné použití

Níže nabízím detailnější pohled na jednotlivé stroje a nejdůležitější informaci, tj. co který stroj nabízí. Vlastnosti serveru, které je nutné znát pro reporting, případně další rozhodování na strategické i taktické úrovni, jsem vybral na základě vlastních zkušeností a [4].

- Typ obsahuje výrobce (pokud je znám) a modelové či produktové číslo
- procesor označuje typ procesoru a jeho frekvenci
- RAM – velikost fyzické paměti
- Použité disky – počet a velikost disků v serveru, souvisí s ochrannou disků (viz. níže)
- ochrana disků – zda je použita nějaká technologie pro ochranu disků, pokud ano, tak jaký typ RAIDu a jak jsou jednotlivé logické disky namapovány v operačním systému
- IP adresy
- OS – jaký typ operačního systému je nainstalován
- Odpovědná osoba – kdo za provoz serveru zodpovídá (a tudíž ho zná, ví jak se přihlásit apod.)
- umístění – fyzické umístění, servery zpravidla není jednoduché nalézt (nejsou-li všechny v serverovně)
    - AMD – serverovna v Amdenu
    - SC – nová serverovna u učeben

    - BENCH PFC – místnost za přepážkou v kanceláři PFC

- Funkce – jaké funkce server poskytuje, jaká je jeho role v infrastruktuře
- Seznam SW – jaký server je na serveru nainstalován
- Poznámka – pro zapsání relevantních informací, pro které jsem nevymyslel kolonku

Detailní přehled holdingových serverů:

INTERNATOR	
Typ (výrobce)	HP NetServer LC 2000 U3
Procesor	Pentium 3, 1GHz
RAM	1024MB (2x128, 1x256, 1x512, volný 0)
Použité disky	4x 18GB
Ochrana disků (RAID, typ, mapování v OS)	RAID 5 (C,D) C - 5,9GB D - 44,9GB
IP adresy	192.168.10.60
OS	Windows 2000 Server
Odpovědná osoba	P. Hovorka
Umístění	AMD
Funkce	SQL File Server Symantec AntiVirus Server Money - sprava sitovych klicu Active Directory
Seznam SW	BackupExec Remote Agent, BackupExec Open File option HP NetServer Agents MS SQL Server 2000 MS Exchange 2000

INTERNATOR	
	Windows 2000 Admin. Tools Windows 2000 Support Tools Instant TopTools Active Directory Migration Tool Intel PRO Network Adapters and Drivers, Intel PROSet II MS Office 2000 SR-1 Premium MS Proxy Client NetRaid Assistant Package Acrobat Reader 5.0, WinRar Remote Desktop Connection Symantec - AMS Server, System Center, Quarantine Console Snap-in, AntiVirus Server, AntiVirus Central Quarantine, LiveReg, LiveUpdate 2.0, LiveUpdate Administration Utility
Poznámka	

INFORMATOR	
Typ (výrobce)	HP NetServer LC 2000
Procesor	P3 800MHz
RAM	1024MB (2x128, 1x256, 1x512, volný 0)
Použité disky	5x 18 GB
Ochrana disků (RAID, typ, mapování v OS)	RAID 5 (nad třemi disky) (C,D) RAID 1 (dva disky) (E) C - 5,9GB D - 27,3GB E - 16,9 GB
IP adresy	192.168.10.70

INFORMATOR	
OS	Windows 2000 Server
Odpovědná osoba	P. Hovorka
Umístění	AMD
Funkce	Domain Controller File Server - Install Home Folders SQL Internator SQL BackUp Active Directory Dynamics
Seznam SW	BackupExec Remote Agent, BackupExec Open File option HP NetServer Agents, HP Server Agents MS SQL Server 7.0 MS Exchange 2000 Instant TopTools NetRaid Assistant Package Acrobat Reader 4.0, WinZip Remote Desktop Connection Intel PRO Network Adapters and Drivers, Intel PROSet for Wired Connections Symantec - AntiVirus Client, LiveUpdate 1.8 APC PowerChute Business Edition - Agent, Console, Server BDE5 Great Plains eEnterprise 5.50 Windows 2000 Support Tools Windows 2000 Admin. Tools
Poznámka	

MODERATOR	
Typ (výrobce)	Compaq ProLiant ML350 G3
Procesor	Xeon 2GHz
RAM	768MB (1x256, 1x512, 2 volné sloty)
Použité disky	2x 36GB
Ochrana disků (RAID, typ, mapování v OS)	RAID 1 C - 33,9GB M - 33,9GB
IP adresy	192.168.10.71 192.168.13.2 62.44.6.46
OS	Windows 2000 Server
Odpovědná osoba	P. Hovorka
Umístění	AMD
Funkce	Exchange Mail Symantec Mail Filter AZ Pro
Seznam SW	BackupExec Remote Agent Compaq DMI Insight Web Based Agents HP - Survey Utility, Management Agents Insight Management Agent JRE SE 1.42_03 MS Baseline Security Analyzer 1.2 MS Exchange 2000 MS Office 2000 Premium MS SQL Server Desktop Engine

MODERATOR	
	Remote Desktop Connection Symantec - Mail Security for Microsoft Exchange, Mail Sec. Console 4.0 + 4.5 for MS Exchange, AntiVirus Client, LiveUpdate 2.0 Version Control Agent 2.0 Proliant performance Analyzer PROSPEKS-IT AZ.PRO Adobe Acrobat 5.0 Windows 2000 Support Tools WinImage
Poznámka	

DISTRIBUTOR	
Typ (výrobce)	HP NetServer
Procesor	P3 866MHz
RAM	512MB (2x256, 1x512, 1xvolný slot)
Použité disky	18GB 18GB
Ochrana disků (RAID, typ, mapování v OS)	není
IP adresy	192.168.10.41
OS	Windows 2000 Server
Odpovědná osoba	P. Hovorka
Umístění	SC
Funkce	Backup
Seznam SW	Veritas Backup Exec 8.6

DISTRIBUTOR	
	MS Office 2000 Premium Adobe Acrobat 5.0 Windows 2000 Support Tools HP Autobackup Console HP Toptools
Poznámka	

TRANSFORMATOR	
Typ (výrobce)	HP NetServer TC 3100
Procesor	P3 1266MHz
RAM	1024MB
Použité disky	2x 18GB
Ochrana disků (RAID, typ, mapování v OS)	RAID 1 (C)
IP adresy	192.168.10.63 62.44.6.29
OS	Windows 2000 Server
Odpovědná osoba	P. Hovorka
Umístění	AMD
Funkce	Web server
Seznam SW	IIS
Poznámka	

PRIMATOR	
Typ (výrobce)	HP LC3

PRIMATOR	
Procesor	P3 1266 Mhz
RAM	256MB
Použité disky	18 GB
Ochrana disků (RAID, typ, mapování v OS)	není
IP adresy	192.168.10.69 62.44.6.35
OS	RedHat Linux 7.2
Odpovědná osoba	J. Vanous
Umístění	AMD
Funkce	firewall gateway
Seznam SW	iptables
Poznámka	

ACTIVATOR	
Typ (výrobce)	HP LC3
Procesor	P3 XEON 1266 MHz
RAM	512MB (2x 256)
Použité disky	18GB
Ochrana disků (RAID, typ, mapování v OS)	není
IP adresy	192.168.10.68 62.44.6.48
OS	Windows NT 4.0

ACTIVATOR	
Odpovědná osoba	J. Vanous
Umístění	AMD
Funkce	VPN
Seznam SW	
Poznámka	

ALIGATOR	
Typ (výrobce)	Proliant ML330
Procesor	733MHz
RAM	192MB (128, 64)
Použité disky	9GB
Ochrana disků (RAID, typ, mapování v OS)	není
IP adresy	192.168.10.64
OS	Windows 2000
Odpovědná osoba	J. Vanous
Umístění	AMD
Funkce	Vývojový server pro Pythagoras
Seznam SW	SQL
Poznámka	

MAIL.CES.CZ	
Typ (výrobce)	HP NetServer E 800
Procesor	P3 1GHz
RAM	128MB

MAIL.CES.CZ	
Použité disky	2x 36GB
Ochrana disků (RAID, typ, mapování v OS)	RAID 1
IP adresy	62.44.6.4
OS	Debian Sarge 2.6.6
Odpovědná osoba	P. Hovorka
Umístění	BENCH PFC
Funkce	mail (pouze pro managera zahraničních aktivit)
Seznam SW	
Poznámka	

GWCES	
Typ (výrobce)	HP NetServer LC II
Procesor	P2 266MHz
RAM	64MB
Použité disky	2x 2GB
Ochrana disků (RAID, typ, mapování v OS)	RAID 1
IP adresy	192.168.10.1
OS	Slackware 9.1 2.6.6
Odpovědná osoba	J. Vanous
Umístění	BENCH PFC
Funkce	záložní předkonfigurovaná gateway a firewall, trvale vypnutá
Seznam SW	
Poznámka	

MONITOR	
Typ (výrobce)	HP NetServer LH +
Procesor	2xP166
RAM	128MB
Použité disky	2x 2GB
Ochrana disků (RAID, typ, mapování v OS)	RAID 1
IP adresy	192.168.10.47
OS	Windows 2000 Server
Odpovědná osoba	P. Hovorka
Umístění	BENCH PFC
Funkce	monitoring
Seznam SW	HP TopTools
Poznámka	

7.1.5.Doporučené postupy při oznamování a řešení problémů

Velké nedostatky vidím ve způsobu oznamování a řešení problémů na IT. Je nutné zefektivnit proces zadávání a řešení problémů s IT a tím nejen dosáhnout finančních, kapacitních a časových úspor, ale hlavně vyšší spokojenosti uživatelů – zákazníků IT správy a zprůhlednění objemu a náročnosti IT služeb poskytovaných v rámci holdingu. Současně dojde k zprůhlednění úhrad za odvedené práce a zákazníkům bude předáván jasně definovaný seznam provedených prací, které jsou jím přímo potvrzené a odsouhlasené po jejich dokončení a uzavření. Přestože primárně jsem se zabýval serverovou infrastrukturou, řadu poznatků je možné využít i v hlášení a řešení problémů s koncovými zařízeními, tam kde vidím takovou příležitost jsem to explicitně uvedl.

Je nutné definovat způsob úhrad za IT služby pro dceřinné společnosti a to následovně :

- paušální úhrady

- placené služby

Do paušálních úhrad náleží následující :

- správa serverů
- správa mailu
- antivirová ochrana
- správa diskových prostorů
- standardní backup
- správa tisků
- provoz internetu

Kromě těchto serverově orientovaných služeb by mělo do paušálních úhrad spadat i update antiviru na stanicích, práce na kabeláži a aktivních prvcích sítě, správa tiskáren, správa a aplikace updatů na stanicích, provoz telefonů a ústředny.

Do placených služeb náleží ad-hoc prováděné zásahy zpravidla nesouvisející s infrastrukturou (instalace a konfigurace OS a aplikací na stanicích, správa, změny uživatelských účtů v AD, konfigurace tisků na stanicích, připojení do sítě, poradenství při nákupu HW a SW a helpdesk uživatelům.

7.1.6.Popis stavu, návrhy na zlepšení, závěr

Jednotlivé stroje a vybavení technologiemi vyšší dostupnosti odpovídají situaci holdingu v minulosti, kdy jednotlivé servery byly dokupovány a vybavovány ad-hoc. V posledních 5-ti letech nedošlo k výraznějšímu výpadku v infrastruktuře holdingu, která by si vynutila pozornost vedoucích pracovníků holdingu. Funkce, které jsou poskytovány infrastrukturou a jejich závislosti na jiných funkcích (uvedeny v závorce – mimo elementárních služeb jako DHCP a spol.) jsou následující:

- SQL
- File Server

- Symantec AntiVirus Server
- Money - sprava sitovych klicu (SQL)
- Active Directory
- Domain Controller + DHCP + DNS
- File Server – adresář s instalacemi
- File Server – domovské adresáře
- SQL BackUp
- Dynamics (SQL)
- Exchange Mail (Symantec)
- Symantec Mail Filter
- AZ Pro

Neplánovaný výpadek výše uvedených služeb nepřináší žádné přímé škody, může v některých případech snížit produktivitu práce, ale vzhledem k pravděpodobnostem poruchy jednotlivých komponent (podloženo statistikou z provozu posledních 10-ti let) a potenciálním škodám, neplánuje vrcholný management žádné větší investice do zajištění vyšší dostupnosti IT infrastruktury. Poměrně jednoduše vyčíslitelné jsou škody vzniklé výpadkem pevného disku ve všech použitých serverech. Výměna disku a obnova serveru ze zálohy zabere IT správci 3 hodiny času, u funkcí „domovské adresáře“ a „Dynamics“ se škody navyšují o ztracenou práci zaměstnanců. Jednoduchým a levným řešením je použití řadičů HP NetRaid pro stroje Hewlett-Packard NetServer (případně Smart Array Controller pro stroje společnosti ProLiant – Compaq, dnes HP) a vytvoření logických disků chráněných RAIDem poskytujícím ochranu dat (RAID 5, RAID 1, RAID 1+0). Všechny stroje s výše uvedenými funkcemi (file server s domovskými adresáři a IS Dynamics) již mají tuto technologii nasazenou. Single point of failure v podobě neredundantních disků (nejčastější HW závada) nalezneme ve strojích ALIGATOR, ACTIVATOR, PRIMATOR a DISTRIBUTOR. Použité funkce, které mají velký vliv na chod společnosti jsou zálohování, VPN a firewall/gateway. V druhém případě se jedná o funkci, která nebude-li poskytována odřízne holdingovou LAN od okolního světa, zde bylo správně rozhodnuto o zvýšení dostupnosti připravením a předkonfigurováním záložního stroje (zpravidla

označováno jako „cold standby“), který stačí pouze zapnout. U stroje ACTIVATOR bych ovšem doporučil hw RAID nasadit, mnoho zaměstnanců přistupuje k síťovým diskům a poště přes VPN a to zejména v době, kdy není přítomný správce (časový posun, práce večer, o víkendech) a na vzdálenou práci se spoléhá.

Další postřeh je neexistující monitoring serverů, přestože je dedikovaný server (MONITOR) pro tuto důležitou funkci, jakým sledování zdravotního stavu serverů nepochybně je, nejsou výstupy ze systému TopTools využívány (a aplikace není nakonfigurována). To velkou měrou ovlivňuje rychlost řešení případné závady, kdy se IT správa o závadě dozví až hlášením od jiného zaměstnance či dokonce od někoho z okolního světa. Doporučuji zvážit obnovení funkčnosti aplikace TopTools (vzhledem k ukončení vývoje asi není nejlepší volba) nebo nasazení jiné management aplikace, vzhledem k použitým serverům doporučuji nasazení aplikace Hewlett-Packard Systems Insight Manager (HPSIM) , náklady spojené s nasazením aplikace nejsou vysoké (aplikace je distribuována zdarma ke všem ProLiant serverům – tj. v holdingu např. MODERATOR), implementace je dle bestpractice společnosti HP v podobně rozsáhlé infrastruktuře otázkou 4 hodin. HPSIM dokáže provádět jak Push monitoring (tj. zařízení sama dají pomocí některého mng. protokolu – SNMP, WMI – sami vědět management aplikaci, že něco není v pořádku) tak Pull monitoring (tj. management aplikace periodicky zjišťuje stav sledovaných zařízení). Na zjištěné změny stavu je možné zareagovat např. zasláním e-mailu správci sítě nebo spuštěným skriptu (zaslání SMS). Přístup k některým funkcím aplikace je možné přidělit i vyšším managerům, tak aby měli rychlý přehled, zda vše funguje tak jak má. Pochopitelně je možné použít i jiný typ mamangement aplikace, existuje řada takových, které jsou zdarma k dispozici, při testování HPSIM se mi však velmi líbilo do jakých detailů nás aplikace při zjištěné závadě zavede (způsobeno pochopitelně tím, že jak aplikace, tak servery mají stejného výrobce).

V průběhu analýzy bylo nalezeno několik strojů, které nemají žádné využití a žádnou funkci v současné době neposkytují. Proběhne mailing po zaměstnancích holdingu s dotazem na potřebnost těchto serverů a po negativní nebo žádné reakci budou vypnuty. Po uplynutí jednoho měsíce budou využity pro jiné úkoly.

Až na výše uvedené nedostatky je IT infrastruktura v holdingu připravena odolávat většině závad uvedených v teoretické části, kritické služby jsou dostatečně zabezpečeny proti výpadku a dobře je zajištěn proces obnovy funkce po havárii.

7.1.7.Poděkování

Děkuji představitelům holdingu a zástupcům IT správy za podporu při psaní této diplomové práce. Speciálně (dnes již bývalému) členovi představenstva mateřské organizace Central European Services, a.s. - Ing. Martinovi Hákovi, Csc. a IT managerovi Petrovi Hovorkovi.

8.Seznam použité literatury

[1] Jirka Novák, Agora plus a.s. - Definice a rotace záloh

[2] Actinet – Vysoká dostupnost

[3] Matthew Eastwood, Jean S. Bozman – Server Clustering: An End-User Perspective

[4] Faulkner Information Services – Choosing a High-Availability Server

[5] Marcus Evan – Blueprints for High Availability

[6] Iain Campbell – Reliable Linux: Assuring High Availability

[7] http://wikipedia.org – High Availability

[8] http://wikipedia.org – myth of 5 nines

[9] Jan Trupl, Jan Zeman, Trask – Plánování kontinuity činností

[10] ISO/IEC 27001:2005 (dříve BS 7799-2:2000), International Organization for Standardization

[11] Učeň, P. a kol, Grada - Metriky v informatice. Jak objektivně zjistit přínos informačního systému.

9.Abstract

9.1.Abstrakt

Cílem práce je podat ucelený přehled o aktuálních technologiích sloužících k zajištění dostupnosti IT infrastruktury v podniku. Autor má zkušenosti nejen s konkrétními technologiemi a přístupy k zajištění vyšší odolnosti jednotlivých komponent serveru vůči výpadku (ukládání dat, napájení, řízení toku vzduchu, ochrana paměti, …) , ale i se softwarovými produkty, které zajistí redundanci celého serveru. Tento přístup byl otestován i v případové studii, kde proběhla analýza stávající situace ve vybrané společnosti a nalezeny funkce poskytované jednotlivými komponentami infrastruktury. Kromě detailního zmapování situace byla nalezena i slabá místa a doporučeno řešení.

9.2.Abstract

The goal of this thesis is present to reader an overview of actual technologies used to achieve higher availability of IT infrastructure in corporation. Author has experience and knowledge of both particual technologies and approaches to achieve higher resistance to failure of individual server components (disc storage, power and cooling, memory protection, …) , but also with software products used to guarantee redundancy of whole server. This attitude was tested in study case, where analysis of current set-up in chosen corporation was performed and functions of individual infrasctructere's components was determined. Based on this mapped envinronment author found out weak points and recommended actions for solving.

10.Seznam tabulek

Seznam tabulek

Tabulka 2.1: Procentuální dělení dostupnosti.....9

Tabulka 5.1: Nejpoužívanější typy RAIDu.....31

Tabulka 5.2: Schéma funkčnosti ECC.....39

Tabulka 5.3: Rozdíly mezi diferenciální a inkrementální zálohou.....43

11.Seznam ilustrací

Seznam ilustrací

Obr. 2.1: Přehled úrovní dostupnosti....11

Obr. 2.2: Příčiny nedostupnosti....12

Obr. 4.1.: Jednoduché schéma clusteru....19

Obr. 5.1.: Pevný disk....24

Obr. 5.2.: DDR....25

Obr. 5.3 SAS - full duplex....26

Obr. 5.4 SAS vs. SATA interoperability....27

Obr. 5.5 Budoucnost SAS....28

Obr. 5.6 Diskové pole....28

Obr. 5.7 Typická SAN....29

Obr. 5.8 GBIC....30

Obr. 5.9 UPS....33

Obr. 5.10: Racky existují v různých velikostech....34

Obr. 5.11: Power Distribution Unit - se dvěma vstupy....35

Obr. 5.12: AECC - výpadek DRAMu....39

Obr. 5.13: Zápis při použití Advanced ECC....40

Obr. 5.14: Přesun banku při zapnutém online spare....40

Obr. 5.15: Zrcadlená paměť - použití jedné paměťové desky....41

Obr. 5.16: Hanojská věž, [1]....44

Obr. 7.1: Struktura holdingu CES.....53

12.Terminologický slovník

Termín	*Význam*
Termíny přejaté z terminologického slovníku katedry informačních technologií	
LAN	Lokální -> komunikační sít (Local Area Network). Vyznačuje se (na rozdíl od rozlehlé sítě -> WAN) tím, že uživatel (organizace, podnikatelský subjekt) je typicky shodný s provozovatelem sítě (síť je v majetku uživatele), má menší vzdálenosti mezi uzly sítě a (často) vyšší přenosové rychlosti.
middleware	V počítačovém průmyslu je termín middleware používán pro označení jakéhokoli programu, který slouží pro spojení jiných, zpravidla už existujících programů, a nebo jako prostředník mezi nimi. Typicky jde o propojení na různé databázové systémy nebo komunikaci mezi komponentami. Systematické propojování aplikací pomocí middleware se označuje jako integrace podnikových aplikací (Enterprise Application Integration, EAI). Úloha middleware neustále roste tak, jak roste úloha integrace od integrace aplikací v rámci podniku, přes integraci podniků až po integraci služeb.
Operační systém	Množina programových modulů, které slouží jako rozhranní mezi uživateli a jejich aplikačními programy a technickými prostředky počítačového systému, s primárním cílem vytvořit efektivní, bezpečný a snadno využitelný počítačový systém.
Počítačová síť	Představuje obecně systém vzájemně propojených počítačů, terminálů a periferních zařízení, komunikujících prostřednictvím komunikačního subsystému sítě, přenosových médií a aktivních komunikačních prvků. Lze je klasifikovat dle rozsahu pokrytí (lokální, metropolitní, rozsáhlé), dle uspořádání prvků v systému (hvězda, kruh, sběrnice), dle účelu (v informačních systémech, v průmyslových aplikacích), dle rychlosti přenosu dat (klasické, vysokorychlostní), dle použitých technologií (drátové – metalické, optické, bezdrátové – pozemní směrové spoje, družicové spoje).
Další termíny	

Termín	***Význam***
nod	člen clusteru, zpravidla server, na kterém je možno spouštět aplikace, vypínat a zapínat
Nepřerušitelné zdroje napájení (UPS)	Komponenta, která se postará o napájení serveru (stanice, diskového pole) v případě výpadku napájení. Obsahuje baterii a elektroniku pro řízení nabíjení a informování obsluhy o výpadku energie
cluster	řešení obsahující dva a více serverů (tzv. nodů), sdílené úložiště dat a zajišťující provoz služby i v případě výpadku celého serveru; z hlediska klientů je cluster monolitický (je jim jedno kolik serverů ve skutečnosti službu poskytuje, hlavně aby fungovala)
downtime	čas, po který není možné využívat službu poskytovanou IT infrastrukturou (nezáleží na tom, zda
MTBF	mean time between failure – statistické číslo, uvádějící průměrnou dobu, po kterou zařízení funguje

13.Abecední rejstřík

Abecední rejstřík

backplane 26

Business Continuity Planning 14

business critical 9

cluster 10, 18, 19, 20, 79

downtime 8, 79

ECC 10, 13, 37, 38, 39, 40, 41

failback 18, 20

failover 10, 18, 20, 21

GFS 43, 44

heartbeat 18

High Availability 7, 8, 73

mission critical 9

MTBF 10, 25, 79

Pull monitoring 46, 71

Push monitoring 46, 71

RAID 8, 10, 13, 14, 30, 31, 41, 58, 59, 60, 62, 63, 64, 65, 66, 67, 68, 70, 71

SAN 13, 14, 24, 29, 30, 42

SLA 48

SNMP 33, 34, 46, 71

storage 24

Tower of Hanoi 43, 44

UPS 32, 33, 79

xor 36, 37, 38