

WWW SERVER V PŘÍLIVU UŽIVATELŮ INTERNETU

Jaromír Ocelka

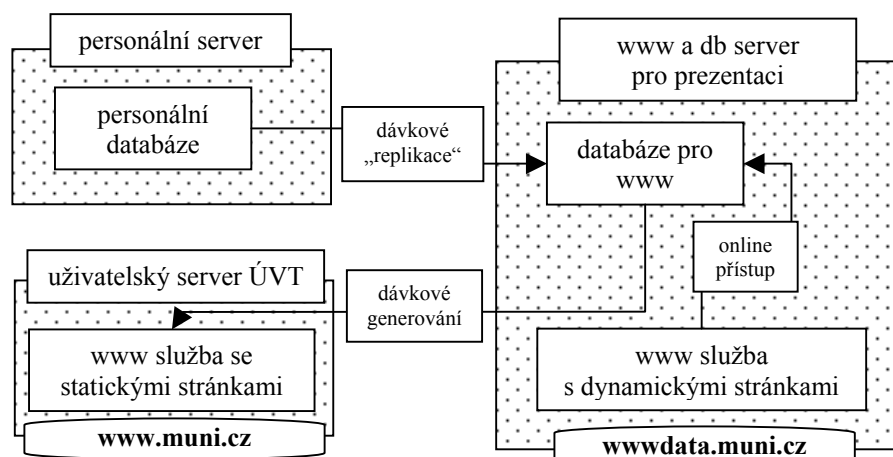
ÚVT MU, Botanická 68a, 602 00 Brno, ČR, ocelka@ics.muni.cz

Abstrakt

Internet se dostává do povědomí čím dál tím většího počtu osob, a tak jsou www servery, které svým výkonem stačily ještě před několika lety, v dnešní době přetížené. Lidé si více zvykají hledat informace namísto v tištěných dokumentech přímo na www stránkách, takže případný výpadek informačního serveru má v dnešní době větší následky. WWW servery vysokých škol jsou tímto trendem postiženy možná ještě více, neboť každý uchazeč o studium na vysoké škole je obeznámen s Internetem nejpozději na střední škole. Článek popisuje postupný vývoj www prezentace Masarykovy univerzity v Brně z pohledu dostupnosti a výkonu od počátku až do současné doby, kdy je ve fázi implementace prototypového řešení clusteru serverů, realizovaného operačním systémem Windows 2000 Advanced server a doplněného vlastními řešeními.

Vznik

Na přelomu let 1996/1997 se MU rozhodla zrealizovat zastřešující centrální www server, který by jednotně prezentoval základní informace o univerzitě (viz [1], [2], [5] a [6]). Větší část těchto informací bylo rozhodnuto (z důvodů automatické aktualizace) čerpat z personální databáze univerzity – strukturu fakult, seznamy zaměstnanců a studentů. Celkem se jednalo o cca 13 tisíc stránek (přes 10 tisíc studentských a 2 tisíce zaměstnaneckých stránek, stránky pracovišť atd.), které bylo nutné prezentovat virtuálním návštěvníkům univerzity. Díky tomuto velkému počtu byla zamítnuta varianta vytvoření (byť automatizovaného) všech stránek na disk a bylo nutné použít technologii dynamických stránek www serveru, které generují obsah stránek na požádání. Jelikož tato technologie nebyla v době vzniku univerzitní prezentace ještě řádně prověřena provozem, bylo rozhodnuto zkombinovat oba přístupy a nejdůležitější část prezentace z preventivních důvodů generovat do statických stránek, prezentovaných jiným www serverem. Jednalo se o celkem 300 statických stránek – hlavní stránku univerzity a základní stránky o fakultách a pracovištích. Další otázkou bylo, zda přímo využívat personální databázi. Z bezpečnostních důvodů zvítězila varianta odděleného databázového serveru, který tímto zabrání případným distribuovaným útokům na personální databázi.



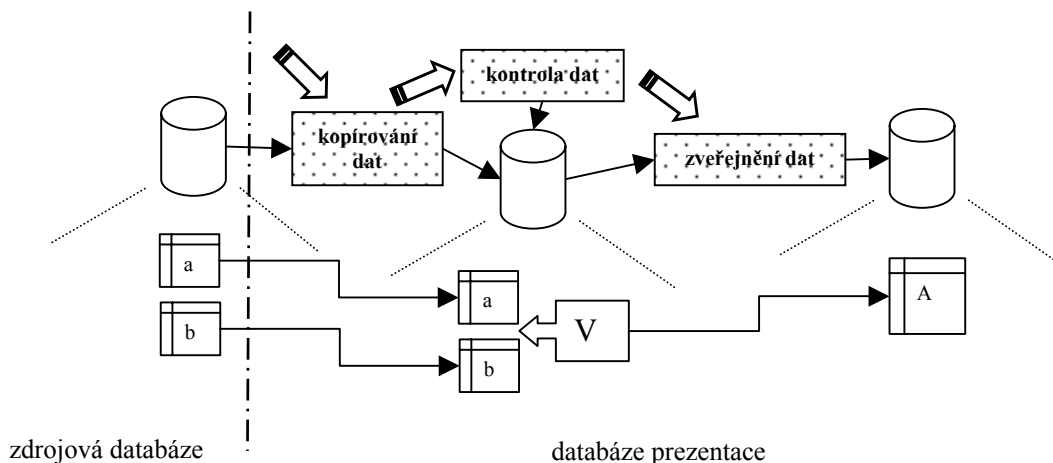
Obr. 1: Technické schéma www prezentace

Jelikož uživatelský server řídí operační systém SunOS, je pro www službu (pro statické univerzitní stránky s adresou www.muni.cz) použit software Apache. Pro dynamické stránky a vlastní podpůrnou databázi byl pořízen jediný server Intel Pentium Pro s operačním systémem Windows NT 4.0, www službou IIS 3.0 (veřejnosti přístupným pod hlavičkou wwwdata.muni.cz) a databázovým strojem MSSQL 6.5 (podrobnosti viz [2]). Přenosy dat z personální databáze se prováděly pomocí textových souborů s oddělovači a po naimportování do prezentační databáze byly k dispozici ASP skriptům pro on-line www požadavky na adresu wwwdata.muni.cz. Statické stránky www.muni.cz se vytvářely perl skripty z předpřipravených dat (textové soubory s daty a části HTML kódu) vyexportováním již z prezentační databáze. Perioda dávkového přenosu dat z personální databáze a generování statických stránek byla z počátku cca 1 x za měsíc, a to hlavně kvůli poloautomatickým generovacím skriptům (mezi jednotlivými kroky probíhala ruční kontrola).

Automatizace datových toků

Aktualizaci údajů z personální databáze a postupem času i z dalších datových zdrojů (viz [3]) bylo nutné provádět častěji, neboť s rostoucím počtem uživatelů Internetu (a tedy i návštěvníků prezentace MU) přicházelo více stížností na neaktuálnost dat. Dokud však byla aktualizace poloautomatická, nebylo možné ji provádět o mnoho častěji z důvodů „lidské síly“. Bylo tedy nutné vyvinout mechanismy, které by zajistily bezpečnou, plně automatickou aktualizaci dat a generování statických stránek.

Nejvhodnějším řešením takového problému je transakční zpracování. Komplikací ovšem bylo, že data pro www prezentaci byla získávána netraskčními přenosy a také to, že data, která jsou v pořádku ve zdrojové databázi, obecně nemusí vyhovovat pro www prezentaci. Např. personální databáze nemusí nutně evidovat ke každému názvu pracoviště anglický překlad, avšak pro plně dvojjazyčnou www prezentaci je anglický název nezbytností. Poslední podstatnou úlohou bylo zajistit transakční vygenerování html souborů na disk – tj. řešení odolné proti výpadku sítě při generování stránek apod. Výsledné řešení skládající se ze dvou modulů (modul *příjem dat* – viz obr. 2 a modul *generování stránek www.muni.cz* – viz obr. 3) se v drobných obměnách provozuje dodnes.



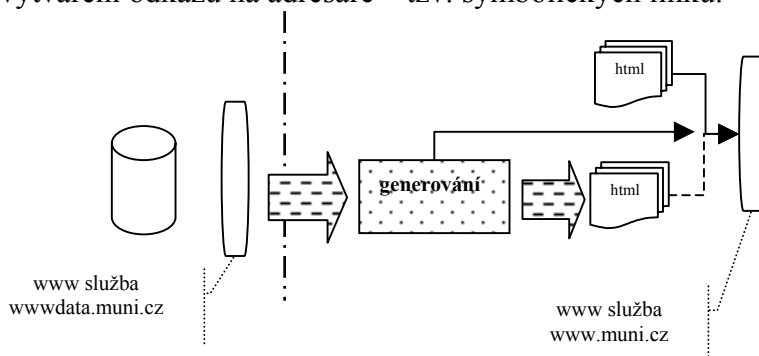
Obr. 2: Modul „příjem dat“

Modul pro příjem dat (viz obr. 2) nejprve získá data a uloží je do databáze vyhrazené pro importovaná data. Zde se data překontrolují, zda obsahují všechny údaje, které jsou z pohledu univerzitní www prezentace nezbytné, provedou se statistická srovnání na změnu počtu položek od posledního importu, atd. V případě, že data těmto kontrolám vyhovují, jsou převedena do ostré databáze. V první verzi byl přenos dat (uložených v textových souborech s oddělovači) ze zdrojového serveru uskutečňován pouze pomocí protokolu ssh. Později s přechodem na novou verzi www databáze, bylo možné využít distribuované propojení databází (www a personální) a přenos dat převést na plně transakční pomocí SQL příkazů INSERT – SELECT. Ve speciální databázi pro import je pro každý datový přenos jedna či více tabulek (např. a, b), které svým jménem a strukturou odpovídají zdrojovým tabulkám. V případě, že je struktura dat v ostré www databázi jiná, jsou nad původními tabulkami vytvořeny pohledy (např. V), které slouží pro transformovaný pohled na data v takové struktuře, v jaké se nacházejí v cílové databázi.

Časové údaje o jednotlivých krocích příjmu dat (import, verify, publish) jsou zaznamenávány. Tyto údaje pak slouží při řešení problémů, jsou použity pro automatickou kontrolu, zda nějaký typ dat není příliš starý (tedy dlouho neaktualizovaný) a také jsou časové údaje o úspěšném provedení importu – z pohledu uživatele: o poslední aktualizaci – zveřejňovány na informačních www stránkách.

Modul pro vygenerování hlavních stránek (viz obr. 3) je postaven na faktu, že z důvodů rychlosti a stability chceme mít některé *stránky* v okamžiku požadavku uživatele již vygenerovány. Postupně se tedy změnil způsob sestavování těchto stránek až do dnešního stavu, kdy je každá takováto *celá* stránka dynamická a jednou za den se uloží na oddělený server www.muni.cz. Aby se stránky mohly uložit automatizovaným způsobem, bylo nejprve nutné zavést jejich systematickou evidenci (viz [4]). Tato evidence je v procesu generování použita a nejprve je z ní sestaven seznam všech stránek (včetně variant parametrů a cílových jmen souborů) určených ke generování. Na základě tohoto seznamu specializovaný program kladé na www službu serveru wwwdata.muni.cz požadavky na jednotlivé stránky a jejich obsah ukládá na disk serveru www.muni.cz. Proces generování takto každodenně vytvoří více než 500 stránek za několik desítek minut. Rozšířením wwwdata.muni.cz na www cluster bylo možné do procesu generování vložit prvky paralelismu, které ve své podstatě kladou více www požadavků najednou – toto vedlo ke zmenšení času téměř na polovinu.

Jak již bylo řečeno, proces generování stránek musí být řešen „transakčně“. Není tedy přípustné, aby byly uživatelům prezentovány některé stránky již nové a některé staré. Podobně je nutné zajistit korektní data v případě, že je generování přerušeno (výpadek sítě, ...). Stránky se tedy generují nejprve do speciálního adresáře, který je mimo kořen www služby a teprve v případě úspěchu je kořen www služby přeměřován do nových adresářů. Jelikož je www služba www.muni.cz provozována na UNIX systému, je pro tuto „výměnu“ využito možnosti vytváření odkazů na adresáře – tzv. symbolických linků.



Obr. 3: Modul „generování stránek www.muni.cz“

Celý import dat, skládající se z uvedených modulů, však celkově není transakční v důsledku samostatnosti obou modulů. Po provedení importů dat do databáze již dynamické stránky ukazují nová data, ale předgenerované stránky www.muni.cz po několik minut ještě ne. Jelikož se tyto změny dějí v noci a stránky na www.muni.cz jsou voleny vhodně, nečiní tento fakt žádné problémy.

Rozmach návštěvnosti

S rozšířením Internetu do škol a domácností začalo přibývat návštěvníků veřejné prezentace univerzity. V dobách vzniku prezentace byli nejčastějšími návštěvníky samotní zaměstnanci a studenti školy, postupem času k nim přibývali i uchazeči a zájemci o studium a další zájemci o nejrůznější typy informací.

Téměř od vzniku www prezentace jsou udržovány záznamy o návštěvnosti, díky nimž bylo možné konfrontovat hardwarové možnosti nejenom s aktuální vytižeností, ale i s předpokládanými navýšeními v budoucnu. Na www službě wwwdata.muni.cz bylo koncem roku 1998 zaznamenáno v průměru 90 požadavků¹ na stránky za hodinu. Za rok 2002 se tento ukazatel zvýšil na hodnotu 400 a podle údajů z prvního čtvrtletí roku 2003 trvale roste (aktuální stav je cca 600 požadavků za hodinu). Tento trend však může být ovlivněn případnými delšími návštěvami uživatelů, proto je nutné nezanedbat také počet různých uživatelů za jednotku času. U veřejné prezentace není možné rozlišit požadavky různých uživatelů, proto se počet uživatelů nahrazuje počtem různých klientských stanic rozlišených jejich IP adresou. Rostoucí řada hodnot vyjadřující průměrný počet IP adres za hodinu je od roku 1998 následující: 9, 14, 23, 45, 50 a v letošním roce vykazuje poslední člen řady navýšení na hodnoty vyšší než 80. Jeden a tentýž uživatel se však může k www prezentaci v průběhu času vracet. Nemůžeme tedy tvrdit, že počet uživatelů (IP adres) byl za celý rok 2002 $50 \cdot 24 \cdot 365$ tj. 438 tisíc. Ze statistik roku 2002 je vypočten celkový počet 94 tisíc různých IP adres. V letech 1998-2001 byla hodnota (v tis.) této statistiky: 18, 26, 42, 76. Za první čtvrtletí roku 2003 byla již dosažena polovina stavu roku 2002. Různých statistik by se dalo počítat jistě více, některé jsou ve formě grafu součástí prezentace samotné.

¹ Ze statistiky jsou vypuštěny požadavky na obrázky.

Analýza architektury systému

Při posuzování, zda posílit či neposílit hardware serveru, se statistiky o návštěvnosti dávaly do souvislosti se statistikami o vytížení systému. Zde je nutné poznamenat, že server s www službou www.muni.cz „trpí“ zvýšením počtu návštěvníků mnohem méně, neboť jeho jedinou činností v reakci na www požadavek je vrátit obsah souboru.

V případě, že bychom při návrhu hardware chtěli systém takový, aby vyhověl všem uživatelům (pro zjednodušení jen ČR) Internetu, kterých bylo v roce 2002 28% dospělé populace – tj. cca 2,4 miliónu uživatelů, došli bychom k velice vysokým finančním nárokům. Pokud by totiž měl každý uživatel navštívit pouze tři www stránky prezentace denně, znamenalo by to 7,2 miliónu požadavků za den – tj. cca 83 požadavků za sekundu. Zde je ovšem nutné poznamenat, že nemůžeme počítat s rovnoměrným rozložením návštěvností v různou denní dobu – ze statistik aktuální prezentace vychází průměrná minutová návštěvnost ve špičce oproti běžnému dennímu průměru až 17-ti násobně vyšší. Celkově bychom v našem ideálním případě potřebovali, aby www služba s databází zvládala vyřídit až 1400 požadavků za sekundu, což například z pohledu síťového připojení znamená mít server připojen přímo k pátevní síti gigabitovou linkou.

Univerzitní www server není naštěstí žádaný všemi uživateli Internetu, takže je zbytečné připravit se na všechny uživatele a je lépe vycházet z aktuálních statistik (proto je jim věnována taková pozornost) a z nich získaných trendů vývoje. Z aktuálních údajů o návštěvnosti veřejné www prezentace MU je průměr 10 požadavků za minutu, což jsou 2 promile výše uvedeného hypotetického² příkladu. Zde stojí za zmínku také případ, který se udál na jedné komerční TV, kdy její ředitel sdělil divákům adresu stránky a teprve poté chtěl ukázat obsah stránky na obrazovce. Již ji neukázal, neboť server byl přetížen vysokým počtem návštěvníků. Svým neuvědomělým postupem zvýšil ředitel v jednom okamžiku návštěvnost serveru více než stonásobně. V nedávné době byl obdobný problém se zveřejněním seznamů spolupracovníků STB.

Jelikož www služby musí umět zpracovávat požadavky uživatelů paralelně a v případě dynamických stránek je tato nutnost přenesena i na databázový stroj, je vhodné, aby celkový systém podpořil paralelní zpracování více procesory. Mimo jiné i z tohoto důvodu byl na počátku roku 2000 pořízen nový server se dvěma procesory Intel Pentium III 500 MHz, 256 MB paměti a jedním SCSI diskem o kapacitě 8 GB. Pro provoz serveru byl zvolen operační systém Windows 2000 Server s www službou IIS 5.0 a databází MSSQL 7.0. Aby nedošlo k zahlcení serveru zvyšujícími se počty uživatelů, docházelo v souladu se zvyšujícími se hodnotami důležitých faktorů k postupnému rozšiřování hardwaru. Paměť byla rozšířena na celkovou kapacitu 512MB a postupně byly přidávány další SCSI disky až do počtu 4, aby bylo možné zrcadlením disků rozložit zátěž databáze a dosáhnout co nejmenšího počtu neplánovaných výpadků (což se v roce 2002 ukázalo jako velmi užitečné).

Jelikož se v roce 2002 průměrný počet požadavků od pořízení nového serveru více než zdvojnásobil, bylo nutné vymyslet další strategii, abychom stále stačili tempu uživatelů a nenastal krizový moment, kdy by prezentace nemohla ve špičkách plnit svou úlohu. Samozřejmě, že jednou z možností je každé cca 2-3 roky pořizovat nový hardware. Tím však okamžitě vyvstane otázka: „Co s odpadem?“ a navíc je jediný server rizikový z hlediska

² Příklad není až tak moc hypotetický v případě úspěšného Internetového deníku.

možné hardwarové poruch. Řešením je tedy místo hardwarové škálovatelnosti použít škálovatelnost softwarovou – tj. několik méně výkonných serverů sdružit do clusteru.

Clusterové řešení

Plán rozšířit www prezentaci z jednoho počítače na clusterové řešení se zrodil na jaře roku 2002. Cílem bylo dosáhnout jak vyšší odolnosti proti výpadku hardwaru, tak zvýšit výpočetní schopnosti a možnosti škálovatelnosti celého systému v souvislosti s neustálým přílivem uživatelů.

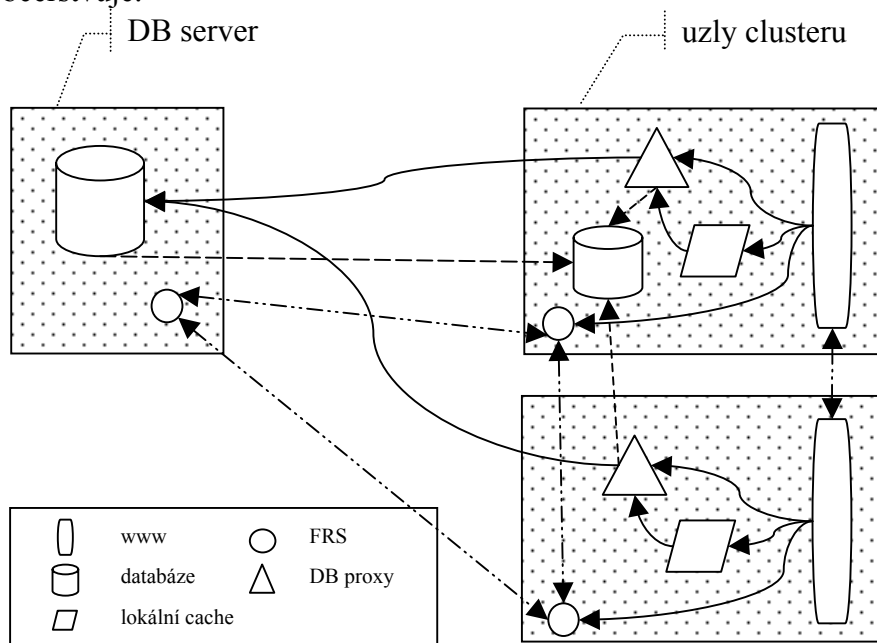
WWW prezentace se skládá ze dvou částí – www služby a databáze. Vzhledem k tomu, že není nutné v samotné www službě udržovat stavové informace, postačuje pro realizaci mechanismus load balancingu, tj. rovnoměrného rozdělování www požadavků mezi více serverů s identicky konfigurovanou www službou. Mohlo by se zdát, že bude problémem intranet www prezentace (viz [5]) kvůli němuž jsou po dobu přihlášení uživatele v paměti udržovány jisté informace. Pomocí tzv. affinity je však možné na protokol https, který je pro intranet používán, definovat, že požadavky z jednoho počítače budou vždy vyřizovány jedním uzlem clusteru. Problémem zůstává cluster databází, kde load balancing nepostačuje z důvodů například on-line modifikací z intranetu, ... Řešením by bylo použít mocnější nástroj – tedy „pravý“ clustering databází.

Pro řešení load balancing postačuje instalace OS Windows 2000 Advanced Server, avšak pro clusterování databází je nutné navíc pořídit MS SQL 2000 v Enterprise verzi a doplnit diskové úložiště speciálně sdílené mezi servery. Realizace clusteru databází by tedy způsobila výrazný finanční nárůst celé investice jak v položce softwarové, tak hardwarové. Bylo proto tedy zvoleno řešení, kdy je v clusteru pouze www služba, přičemž snížení zátěže databáze bude řešeno pomocí lokálních datových cache nejvíce používaných dat a zajištění případného hardwarového výpadku databázového serveru bude realizováno čistě softwarově pomocí záložní databáze a automatické detekce výpadku.

Ze stávajícího serveru byla tedy přemístěna www služba na dva nové servery, které dohromady tvoří dva uzly www clusteru. Protože na uzlech není uložena žádná informace primárně a výpadek jednoho z nich není kritický, bylo rozhodnuto, že hardwarem těchto serverů budou v podstatě obyčejné pracovní stanice: Intel Pentium 4 1,8 GHz, 1xIDE disk 20 GB, 384 MB RAM. Navíc jsou tyto servery osazeny dvojicí síťových karet, z nichž jedna je použita pro load balancing www služby a druhá je určena pro administrativní přístup a systémovou komunikaci služby load balancing. Operační systém na obou uzlech je Windows 2000 Advanced Server, na datovém serveru byl ponechán Windows 2000 Server.

Programové vybavení se skládá z hotových produktů firmy Microsoft – tj. MSSQL 7.0, IIS 5.0, network load balancing, file replication service (dále FRS) a vlastních modulů vytvořených pro účely www prezentace – na systémové úrovni se jedná o *lokální cache* a *databázovou proxy*. FRS bylo zvoleno pro účely automatické replikace ASP skriptů mezi uzly www clusteru, přičemž jako „primární“ úložiště je dál používáno původní úložiště z dob jediného serveru. Ke konfiguraci www služby nedochází příliš často, a proto byla pro „replikaci“ konfigurace zvolena varianta „lidské síly“. Aby se zabránilo výpadku celého systému v případě výpadku databáze, byla na jeden uzel clusteru nainstalována záložní databáze, která se periodicky občerstvuje z primární databáze. Pro zajištění automatického přepnutí v případě výpadků vznikla *databázová proxy*, která si udržuje aktuální přehled o funkčních databázích, a protože jsou veškeré požadavky směřovány na TCP port této proxy, jsou požadavky dále s její pomocí přesměrovány do primární resp. záložní databáze. Pro odlehčení databázové zátěže byl vyvinut modul *lokální cache*, který při svém startu načte

z databáze základní a většinou neměnné číselníky (pracovišť, adres, stránek, ...) a z nich případně vypočítá složitější struktury (např. hierarchii pracovišť). Tyto údaje periodicky z databáze obcerstvuje.



Obr. 4: Schéma cluster řešení

Celkové řešení (viz obr. 4) bylo již několikrát prověřeno, jednak umělými pokusy a také nutnými reálnými situacemi. Při havárii jednoho z disků primární databáze a jeho výměně běžel systém po automatickém přepojení po nutnou dobu na záložní databázi. Dalším nezanedbatelným přínosem je instalace záplat a oprav, které je možné provádět za plného provozu postupně – nejprve vypojením prvního uzlu a instalací oprav na něm a následně provedením téže akce na druhém uzlu.

Závěr

Jednotlivá parciální technologická řešení mají drobné nedokonalosti. Například intranet www prezentace není možné používat při výpadku primární databáze z důvodu možných on-line zápisů. Aktualizace databáze není plně transakční – v případě neúspěchu vygenerování stránek pro www.muni.cz zůstává stará verze se včerejšími daty, ač dynamické stránky zveřejňují data nová. V současné době jsou však tyto nedokonalosti zanedbatelné, ať už pohledu poměru počtu uživatelů veřejné prezentace a intranetu prezentace, nebo z pohledu havárie databáze v kombinaci s faktem, že se v datech od minulého dne něco zásadního změnilo.

Současné řešení není konečné a o jeho dalším rozvoji budou převážně nepřímo rozhodovat návštěvníci www prezentace univerzity. Nyní jsou ve spolupráci s odborníky na síťovou infrastrukturu vedeny úvahy u zdvojení síťového switchu tak, aby v případě jeho výpadku mohly být požadavky od uživatelů vyřízeny alespoň jedním uzlem clusteru za použití, když ne jinak, minimálně záložní databáze. Díky již navržené robustní síťové topologii univerzity by mohl být každý switch napojen na páteřní linky jinou síťovou cestou. Paralelně s těmito úvahami se připravuje návrh pořízení nového databázového serveru, přičemž stávající by „osvobodil“ jeden z uzlů www clusteru od záložní databáze. Kromě záložní databáze by mohl tento „vysloužilý“ server v případě náhlého zvětšení návštěvnosti www stránek provozovat na dočasnou dobu třetí uzel www clusteru. Jelikož se www cluster vykazuje stabilitou, uvažuje se také o přesunu www služby www.muni.cz na počítače clusteru.

Literatura:

1. Kohoutková, Jana. Masarykova univerzita na internetových WWW stránkách. Zpravodaj ÚVT MU: bulletin pro zájemce o výpočetní techniku na Masarykově univerzitě. ISSN 1212-0901, 1997, roč. 7, č. 3, s. 10-13
2. Valach, Zdeněk. Personální databáze zaměstnanců MU pro WWW. Zpravodaj ÚVT MU: bulletin pro zájemce o výpočetní techniku na Masarykově univerzitě. ISSN 1212-0901, 1997, roč. 8, č. 1, s. 10-13
3. Procházková, Šárka, Ocelka, Jaromír. Internetová prezentace MU v Brně. In UNINFOS 2000. Zborník príspevkov. Nitra: SPU v Nitre, 2000. ISBN 80-7137-713-9, s. 186-189
4. Ocelková, Šárka, OCELKA, Jaromír. Automatizace univerzitní prezentace. In UNINFOS 2001. Zborník príspevkov. Zvolen: Vydavateľstvo TU vo Zvolene, 2001. ISBN 80-228-1062-2, s. 124-128
5. Ocelka, Jaromír. www.muni.cz po dvou letech. Zpravodaj ÚVT MU: bulletin pro zájemce o výpočetní techniku na Masarykově univerzitě. ISSN 1212-0901, 1998, roč. 9, č. 1, s. 11-14
6. Ocelková, Šárka. Webová prezentace MU po 4 letech. Zpravodaj ÚVT MU: bulletin pro zájemce o výpočetní techniku na Masarykově univerzitě. ISSN 1212-0901, 2002, roč. 13, č. 1, s. 4-8