Modernizace hardwarového vybavení IS MU

Michal Brandejs, Jan Kasprzak, Miroslav Křipač, FI MU

Jeden z důležitých projektů, které byly realizovány v letošním roce v rámci vývoje Informačního systému Masarykovy univerzity (IS MU), se týkal modernizace hardwarového vybavení. Ačkoliv výsledky projektů obvykle běžný uživatel zakusí velmi brzy po jejich realizaci, ať už se jedná o nové aplikace nebo o zlepšení funkčnosti aplikací stávajících, dopady výměny serverů často nejsou na první pohled vůbec patrné. Přesto může nově navržená a připojená infrastruktura systému přinést velký užitek a nebo naopak naprostý krach. V tomto článku se čtenářům pokusíme přiblížit technické aspekty povýšení základních serverů IS MU, ke kterému došlo během letošního léta. Volně tak navážeme na náš předchozí článek [1], který popisoval stav hardwarového vybavení v roce 2004.

Motivace

Důvodů pro rozsáhlé změny uvnitř systému je hned několik. První z nich je doslova překotný vývoj aplikací IS MU, které pokrývají stále širší spektrum univerzitních činností, a které musí reagovat na nové a vyšší nároky. Tento vývoj je doprovázen nejen většími nároky na koncové uživatele, kterým však zároveň usnadňuje práci, ale také na systém samotný. Každý nový nástroj, který IS MU poskytuje, znamená přirozeně další zátěž pro servery, které jej realizují.

Další příčinou růstu nároků na systém je stále rostoucí počet uživatelů. Nejedná se pouze o celkový počet uživatelů, kteří se systémem mohou pracovat, ale také o vzrůstající počet těch, kteří využívají stále více služeb ke své každodenní činnosti. Zatímco před lety běžný učitel vystavoval známky a zadával zkušební termíny, dnes mohou učitelé běžně se studenty diskutovat, zadávat různé druhy studijních materiálů, testovat znalosti on-line za pomoci počítače nebo naskenováním a automatickým ohodnocením písemných testů apod. Došlo také ke zrušení papírových indexů a obecně k zavedení snazší elektronické administrativy do běžné výuky.

Vyšší nároky na systém vedly v uplynulých třech letech k postupnému vyčerpání kapacity původního hardwarového vybavení. Zatímco pro běžný provoz uprostřed semestru bylo ještě možné stávající zátěž úspěšně obsloužit, během období se zvýšenými nároky – zejména na začátku a konci výuky – začalo přibývat situací, kdy systém nebyl schopen zvládnout obsloužit všechny uživatele v požadované kvalitě bez zbytečné prodlevy způsobené čekáním na některou z klíčových komponent. Již během roku 2006 se začaly projevovat náznaky tohoto vyššího zatížení.

Prvním krokem v takové situaci je vždy ladění výkonnosti systému, a to jak na úrovni aplikací tak na úrovni systémových komponent, tedy zejména aplikačního a databázového serveru. V průběhu provozu se většinou ukáže, že některé součásti mohou obsahovat výkonnostní rezervy, které typicky přinesou zlepšení celkové odezvy systému v kritických situacích.

V další fázi rostoucí zátěže systému přichází v úvahu změna aplikací tak, aby byla omezena funkčnost pouze na služby, které jsou v provozní špičce nezbytně nutné. Tím dochází k dočasnému omezení méně významných služeb jako je volná diskuse v Plkárně, inzerce nebo některé výpočetně náročné statistiky, které nevyžadují bezprostřední reakci a je tedy možné je odložit na klidnější dobu.

S postupným rozvojem systému a zvyšováním počtu zátěžových špiček se však musí dostatečně dopředu naplánovat také samotné povýšení kapacity, které vzhledem k celkové složitosti a finanční náročnosti přichází na řadu jako poslední krok. V našem případě bylo plánovaní modernizace hardwarového vybavení zahájeno v létě roku 2006, tedy ještě v době, kdy běžný uživatel zvýšené zatížení nijak významně negativně nepociťoval. Bylo však důležité jednak zamezit zmiňovaným omezením a jednak naplánovat kapacity systému pro nové aplikace, zejména v oblasti e-learningu, ale i plánovaného prodeje kurzů a vzdělávání pro širokou veřejnost pomocí tzv. Obchodního centra, které mají významně rozšířit činnost univerzity a tím i zatížit IS MU.

Cíle

Prvním z cílů povyšování hardwarové infrastruktury IS MU bylo proto zrychlení jednotlivých operací. Zejména u velmi náročných výpočtů, které se provádějí v reálném čase, může zvýšení rychlosti výpočtu jednoho dotazu zvýšit pohodlí práce se systémem. Například v oblasti složitého vyhodnocování přístupových práv, které musí být dynamické a naprosto přesné, se zvýšení rychlosti výpočtu projeví při každodenní práci v podstatě se všemi základními agendami systému.

Důležitějším přínosem zvýšení kapacity systému je však celková propustnost systému. Tedy vlastnost, která zajistí, že každý jednotlivý dotaz je obsloužen pokud možno stejně rychle bez ohledu na to, kolik dalších různých dotazů systém zpracovává. Přestože obě vlastnosti se navzájem prolínají, zajištění celkové prostupnosti není obvykle snadné a vyžaduje složitější architektury celého systému.

Vedle navýšení kapacity, a tím i rychlosti a propustnosti systému, byla důležitým faktorem také stabilita nového řešení, která je kriticky důležitá pro bezproblémový chod a tím i spokojenost uživatelů. Přestože počítačové systémy obecně nedosahují takových spolehlivostí, jako o mnoho let starší inženýrská řešení, snahou vývojářů systému je nabídnout službu, která bude dostupná kdykoliv odkudkoliv stejně samozřejmě, jako například elektrická energie.

Použitá architektura

Základem pro provoz systému IS MU je webový přístup, to znamená, že webový prohlížeč pro přístup k IS MU používá každý uživatel od namátkově přistupujících studentů kombinovaného studia až po administrativní pracovníky s rutinní každodenní prací se systémem. Prohlížeč se pomocí redundantního spojení připojuje do fyzicky oddělené sítě několika desítek počítačů IS MU. Jednotlivé dotazy jsou v rámci této sítě rozeslány mezi jednotlivé aplikační servery, které jsou vzájemně zastupitelné a obsahují v sobě jak zpracování HTTPs požadavků, tak samotnou aplikační funkčnost. K tomuto účelu používá IS MU cluster běžných jednoprocesorových serverů s operačním systémem Linux a webovým serverem Apache, na který je navázáno vlastní aplikační prostředí využívající programovací jazyk Perl. Výhodou tohoto řešení jsou zejména velmi nízké pořizovací a provozní náklady, kdy každý server lze jednoduše odpojit pro případnou údržbu, ale i povýšit výkon jeho jednotlivých komponent tak, jak to známe z běžných kancelářských počítačů. Efektivita celého řešení se navíc ještě zvyšuje tím, že aplikační servery, které mohou mít zapojeny až čtyři velkokapacitní pevné disky, slouží také jako obrovské distribuované úložiště pro celou řadu dat včetně objemných studijních materiálů, videí a studentských prací.

Navýšení celkového výkonu na aplikační úrovni je rovněž poměrně jednoduché, neboť vzhledem k tomu, že aplikační servery téměř nesdílí žádná data, dojde k navýšení výkonu pouhým přidáním dalších uzlů. Přestože výkon jednotlivých serverů aplikačního clusteru není nikterak vysoký, celková propustnost může být ohromující. Tím dochází k poměrně značné úspoře zejména v oblasti cenných investičních prostředků.

Naproti tomu databázová část realizuje sdílení všech dat zpracovaných v systému. Změny, které byly zavedeny pomocí aplikace běžící na jednom aplikačním serveru, musí být bezprostředně k dispozici všem ostatním serverům tak, aby mohlo být bezpečně realizováno zpracování všech kritických transakcí. Právě výkon databázové části je z toho důvodu kritickou stránkou architektury celého řešení.

Pro navýšení výkonu databázové vrstvy lze v současné době použít v zásadě dva přístupy. První z nich je, podobně jako v předchozím případě, založen na distribuci databázové zátěže do clusteru několika menších nezávislých uzlů, které dohromady poskytují potřebnou propustnost. Ze zkušeností nasazování databázových clusterů v rámci IS MU se však ukazuje, že režie spojená se zajištěním konzistence všech dat napříč uzly databázového clusteru výrazně ovlivňuje výkon celého řešení. Navíc sofistikované softwarové řešení, které databázové clustery představují, může zvýšením složitosti celého systému přinést řadu nových chyb a problémů při

provozu, které snižují stabilitu takového řešení. V neposlední řadě je cena za licence pro využití této funkcionality tak vysoká, že při reálném nasazení v rozsáhlé infrastruktuře přestává být konkurenceschopná.

Druhým způsobem pro navýšení výkonnosti systému pro on-line transakční zpracování na databázové úrovni je využití systému se sdílenou pamětí, kdy procesy obsluhující jednotlivé požadavky přistupují ke všem údajům jednotně, přičemž komunikace je realizována operačním systémem a hardwarově. Právě tento přístup byl zvolen pro další rozvoj IS MU, a to ze dvou důvodů: efektivita dosažení požadovaného výkonu a celková stabilita řešení. Tento způsob se například hojně využívá ve velkých bankovních ústavech uvnitř rozsáhlých finančních systémů.

Technická realizace

Konkrétních řešení, která nabízí daný způsob zpracování vysokého výkonu, je v současné době na trhu více. My jsme se omezili pouze na ta, která jsou založena na databázovém software Oracle Database, který IS MU využívá a pro který je optimalizován. Zároveň je prostředí IS MU omezeno na systémové úrovni na operační systémy unixového typu, což nevylučuje žádného z dodavatelů velkých hardwarových řešení, pouze reálně omezuje některé typy serverů.

Na základě upřesněných podmínek pak byl v rámci výběrového řízení vybrán systém *Altix 450* společnosti SGI, který nabídl nejvyšší výkon při zachování nízké ceny a dodržení podmínek záruční i pozáruční podpory, které v sobě zahrnovaly velmi důležitý servis včetně dostupnosti náhradních dílů.

Systém Altix 450 je založen na modulární architektuře, která umožňuje propojit více nezávislých komponent do jednoho systému tak, že výměna jednotlivé komponenty v případě poruchy nebo povýšení je podobně snadná, jako tomu je v případě zmiňovaných clusterů. Zároveň však systém propojení jednotlivých komponent, který je založen na proprietárních kabelech pro komunikaci mezi procesory navzájem a procesorem a pamětí (jedná se o architekturu typu NUMA), nevyžaduje nejprve zapojení složitější infrastruktury, která by dále zvyšovala cenu. Další výhodou je použití procesorů typu Intel Itanium 2, které oproti RISCovým procesorům lépe kopírují cenovou hladinu levných komoditních serverů.

Altix 450 sestavený pro IS MU je tak v současné době složen ze 4 modulů, které dohromady obsahuji 3 nezávislé servery, plně propojitelné do jednoho systému. To znamená, že v případě výpadku kterékoliv komponenty bude možné systém provozovat až do její výměny nejméně na 66 % výkonu. Běžné komponenty jsou pak obvykle na centrálním skladu běžně dostupné, což v praxi díky otevřenosti hranic Evropské unie znamená, že dojde k jejich výměně následující pracovní den po nahlášení.

Celkový výkon databázového systému je nyní 52 procesorových jader (jedná se o dvoujádrové procesory), přičemž všechna procesorová jádra mohou využít celých 104 GB operační paměti. Redundantně propojené diskové pole s 16 výkonnými disky obsahuje přibližně 2 TB hrubé kapacity, která je v současné době rozložena zejména pro navýšení výkonu přístupu k diskům. Samotná databáze je optimalizovaná tak, aby podstatná její část byla permanentně přístupná v paměti serveru tak, že k přístupu na disk dochází asynchronně při ukládání změn.

Zajímavosti o novém serveru

Přemýšlivého čtenáře jistě napadá, zda nejsou tři roky na provoz jednoho systému příliš krátká doba. Tedy zda nebylo výhodnější pouze navýšit výkon stávajícího serveru při zachování většiny dosud fungujících komponent. V praxi se však ukazuje, že obchodní politika firem dodávajících tato řešení přeje spíše nákupu nového systému, než povyšování stávajícího. Jinými slovy výkon dosažený novým systémem za stejných nákladů je často výrazně vyšší, než při povyšování starší technologie.

Zároveň nákupem nového systému otevíráme možnosti využít výhodnějších nabídek od jiných dodavatelů. Nejdůležitějším důvodem pro nákup nového serveru však byla skutečnost, že IS MU doposud nedisponoval dostatečnou výpočetní kapacitou pro případ zničení celého serveru. V takovém případě jsou data sice bezpečně umístěna na jiném místě v Brně a jejich obnova

by nebyla problém. Provozovat systém takového rozsahu na běžně dostupném hardwarovém vybavení ale není možné, a proto by došlo k delší odstávce systému. Navíc pokud by došlo k chybě z důvodů nepokrytých zárukou dodavatele (požár, zaplavení apod.), trvalo by dodání nového systému až několik týdnů. Vyšší míra záruky, kterou výrobci pro tento případ také nabízí, ve skutečnosti obvykle znamená, že dodavatel udržuje na skladě přesnou konfiguraci téhož stroje ještě jednou, což také zákazník zaplatí. Ukazuje se tedy výhodné využít pro tento případ původní hardware umístěný v jiném místě tak, aby byl zároveň dostupný pro případ katastrofy a zároveň mohl být dostupný pro další úlohy nebo vývoj nových aplikací.

Server byl vyroben na zakázku ve Spojených státech a na MU dorazil v červnu 2007. Jeho provoz byl spuštěn o víkendu 11. srpna 2007.

Operační systém serveru je Linux, stejně jako na 85 % nejvýkonnějších počítačů na světě.

Celková cena za popisovaný hardware nepřesáhla 65,- Kč na jednoho aktivního uživatele systému a rok, což v daném rozsahu a ve srovnání s obdobnými systémy je velmi příznivá cena.

Server již zaznamenal rekordní výsledky ve všech měřených hodnotách. Například dosáhl 46 000 operací za pět minut, 2 100 000 operací za den, 5 300 uživatelů v jeden okamžik a 27 000 uživatelů v jednom dnu. Server dosud nebyl zatížen špičkovým provozem ani z poloviny.

Dosažená dostupnost celého systému byla 99,989 % což představuje 100 % dostupnost hardwaru a operačního systému a jeden výpadek v řádu několik minut způsobený chybou v databázovém softwaru.

Server se stal jednou z nejvýznamnějších dodávek společnosti SGI v dané oblasti za poslední roky.

Výhled

Ukazuje se, že zvolená cesta tzv. vertikální škálovatelnosti výkonu na databázové úrovni je efektivním způsobem řešení architektury rozsáhlých on-line informačních systémů. Předpokládáme, že při zachování současného tempa zavádění nových služeb a podsystémů (jako je systém na odhalování plagiátů s propojením do archivu závěrečných prací na národní úrovni, Obchodní centrum pro administrativu placené výuky apod.) bude stávající hardware dostatečný nejméně následující tři roky.

Ruku v ruce s navýšením výkonu na databázové straně však došlo, a zřejmě bude ještě docházet, k modernizaci aplikačních součástí, kterou lze, vzhledem k charakteru jednotlivých dodávek, provádět jednodušeji postupně, podle aktuálních potřeb. Celkově tedy doufáme, že stávající architektura bude dostatečně pevným základem pro další rozvoj univerzity v řadě oblastí.

Literatura

[1] M. Brandejs, J. Kasprzak, M. Křipač. *is.muni.cz na novém hardware*. Zpravodaj ÚVT MU. ISSN 1212-0901, 2004, roč. XV, č. 1, s. 5-7. □