

# HYBNÁ SÍLA WEBU

## Jak si vedou webové databáze?

**S databázemi se dnes setkáváme na každém kroku, počínaje jednoduchou tabulkou v Excelu, přes vyhledávání v Googlu, až třeba při ověřování zadaných údajů při výběru peněz z bankomatu. Lze říci, že v moderním světě jsou takřka všudypřítomné a v některých případech si jejich existenci ani neuvědomujeme.**

**N**ároky na databáze přitom stále rostou, a to nejen pokud jde o velikost či rychlost zpracování dat. Dokládá to průzkum konzultační společnosti Winter, která již několik let po sobě vyhláší vítězné databázové systémy v několika kategoriích. I v roce 2005 došlo k překonání nových milníků. Například největší komerční databáze na světě (Oracle), kterou provozuje společnost Yahoo! na platformě UNIX už letos překročila velikost 100 TB, a na platformě Windows (SQL server) vede UPS. Také jiné ukazatele rostou – největší počet záznamů (řádků) uložených v databázi překročil letos 2,8 bilionu a největší dosažené zatížení je více než 1 miliarda zpracovaných SQL dotazů za

hodinu (databáze UPS). Již z tohoto výčtu je patrné, že nároky a vlastnosti jednotlivých databázových systémů se natolik liší, že je obtížné najít pro ně společného jmenovatele.

### Webové databáze

Speciální kategorií jsou webové databáze. Téměř žádný web, který pravidelně aktualizuje svůj obsah, se dnes neobejde bez integrované databáze. Statické weby jsou minulostí a webové stránky jsou dnes převážně orientované na data – tyto weby umožňují interaktivní prohlížení, nebo dokonce vkládání či jiné úpravy dat. Naopak i data z dříve interních databází jsou dnes alespoň částečně přístupná přes webová rozhraní. Existuje několik způsobů, jak mohou weby data uchovávat. Mezi nejprimitivnější patří nestrukturovaná data nebo data uložená v jednoduchých

typech souborů (např. text). Jejich údržba je náročná a jsou pro frekventované weby nepoužitelné. O něco lepší variantou je ukládání v podobě strukturovaných datových objektů (např. XML), ale pokud má systém plnit i jiné funkce než pouhý sběr dat, je řešením databázový server, který disponuje nástroji pro správu a práci pomocí dotazovacího jazyka (nejčastěji SQL).

### Platformy a servery

Pro vybudování funkční webové databáze je nutné zvolit vhodnou kombinaci technologií a produktů. Je potřebný operační systém, webový server, SQL server a skriptovací prostředí. Pokud jde o operační systémy, v drtivé míře se nasazují dvě platformy: komerční Microsoft (Windows 2000 nebo 2003 atd.) nebo různé open source linuxové distribuce (popř. UNIX).

Výhodou Windows a produktů Microsoft je snadná instalace, vzájemná kompatibilita, kvalitní rozhraní (ODBC) a technická podpora, nevýhodou je samozřejmě cena a uzavřenost vůči změnám. Na druhé straně stojí open source software, jenž je „skoro“ zadarmo i při nákupu řešení pro komerční

účely. Jeho výhodami jsou rychlost, stabilita, bohatství funkcí a aktivní komunity uživatelů. Ale protože v nákladech na vybudování a údržbu databáze v organizaci není zahrnuta jen cena za hardware a software, i pro open source produkty platí, že je třeba zaplatit odborníky na vývoj a správu. Na tom staví argumentace Microsoftu, který se zaštiťuje podporou zahrnutou v ceně licence a celkově nižšími náklady na vlastnictví (TCO) daného řešení.

Nejrozšířenějším webovým serverem současnosti je jednoznačně Apache, který si své prvenství drží už od roku 1996. Je oblíbený zejména díky svému výkonu, možnostem konfigurace, škálovatelnosti a rozsáhlé dokumentaci. Nedávno se objevila verze – Apache 2.2.0, která přišla s mnoha vylepšeními včetně inteligentního filtrování, kešování, vyvažování zátěže, podpory soborů a autentifikace, což dále zvýší jeho oblibu i pro databázové aplikace.

Podle průzkumu Netcraft Web Server Survey z listopadu 2005 Apache používá více než 70 % webů připojených k internetu, celkem skoro 53 milionů, kdežto servery Microsoftu obsadily až druhé místo s 20 procenty a 15 miliony instalací. Tato dominance ale neplatí všude, například z tisíce největších amerických společností používá server Apache pro své webové stránky jen 22,7 procenta, kdežto konkurenční Microsoft IIS celých 53,7 procenta (výzkum Port80 Software, 2005).

Při výběru samotného databázového systému je potřeba vzít v úvahu více kritérií – kompatibilitu se standardy SQL jazyka, podporu transakcí, dále rychlost a výkon, kapacitní výkon (tzn. jaká data

## Největší databáze světa (2005)

Společnost/organizace	Velikost databáze (GB)	Databázový systém	Platforma	Dodavatel databáze
Yahoo!	100 386	Oracle	UNIX	Oracle
AT&T	93 876	Daytona	UNIX	AT&T
KT IT-Group	49 397	DB2	UNIX	IBM
AT&T	26 713	Daytona	UNIX	AT&T
LGR – Cingular Wireless	25 203	Oracle	UNIX	Oracle
Amazon.com	24 773	Oracle RAC	Linux	Oracle
Anonymní	19 654	DB2	UNIX	IBM
UPS	19 467	SQL Server	Windows	Microsoft
Amazon.com	18 558	Oracle RAC	Linux	Oracle
Nielsen Media Research	17 685	Sybase IQ	UNIX	Sybase

Zdroj: Winter Corporation, 2005

je schopen ukládat, maximální velikost jedné tabulky atd.). Dalším významným měřítkem je škálovatelnost, to znamená, zda je systém schopen rozšíření při nárůstu požadavků (tzn. využití víceprocesorových systémů, možnost vytvářet clustery, gridy atd.) Existují různé kombinace komerčních a freewarových produktů pro vybudování webové databáze, ale jejich vyjmenování překračuje rámec tohoto článku. Mezi možné varianty (a nejčastěji používané) patří například tyto: produkty a nástroje Microsoft (Windows, MS IIS, MS SQL Server a scriptovací prostředí ASP) nebo Linux a další open source software (např. kombinace Slackware Linux, server Apache, databáze MySQL a Perl nebo PHP).

### Linux versus Windows

Na linuxové platformě má nyní největší podíl databáze MySQL, která je považována za rychlou a robustní databázi, ale s jednou

### MySQL

Nejpoužívanější databázi v open source světě je databáze MySQL, která se nasazuje hlavně v kombinaci s již zmíněným serverem Apache a je přímo obsažena i v některých linuxových distribucích. MySQL je prověřenou a rozšířenou databázovou platformou, která je známá svoji rychlostí při vyhledávání a zobrazování dat. Pokud však uživatelé nebo aplikace často mění záznamy, dochází při zápisu a editaci dat k přetížení a zpomalení systému. Nejen proto není oblíbená v korporátní sféře, byť na poli webových aplikací malých a středních firem drtivě vede. Nepomáhá jí ani horší zázemí a změny, kterými často prochází.

### MS SQL

MS SQL server patří už několik let mezi nejpoužívanější platformy pro vývojáře databázových aplikací a současně patří i do

## Web, který pravidelně aktualizuje svůj obsah, se dnes neobejde bez integrované databáze

### MySQL bez InnoDB?

Před nedávnem koupil databázový gigant Oracle poměrně bezvýznamnou firmu InnoDB. InnoDB vyvíjela pro MySQL AB jednu z klíčových komponent, storage engine pro online transakce InnoDB. A Oracle se netváří, že by chtělo na spolupráci s MySQL pokračovat.

MySQL AB zatím situaci řeší za zavřenými dveřmi, je ale patrné, že mu nezbyvá mnoho východisek. Kód InnoDB sice může používat v souladu s GPL, pokračovat ve vývoji ale pravděpodobně není v jeho silách. Z tétož důvodu nejspíš nemůžeme očekávat ani to, že MySQL postaví zcela novou storage engine. Nezbude mu zřejmě nic, než se poohlédnout jinde.

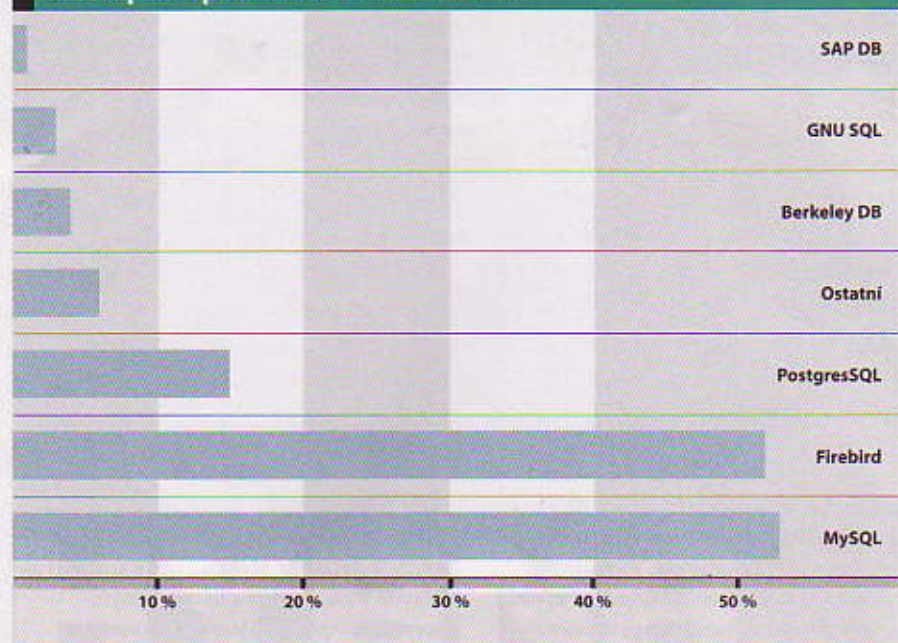
velkou nevýhodou – nedotaženou podporou transakcí (s novými verzemi se to již zlepšuje, ale na konkurenci ještě stále ztrácí) a méně přívětivým uživatelským rozhraním. Na druhém místě je pravděpodobně open source projekt PostgreSQL (nejnověji ve verzi PostgreSQL 8.1), který transakce podporuje a je také použitelný na Windows, Linuxu i UNIXech. V celkovém podílu open source databází je sice těsně za MySQL systém Firebird, ale ve využití jako databáze pro webservery bude pravděpodobně až za PostgreSQL.

Microsoft dodává svůj MS SQL server v kombinaci s MS IIS a ASP, který se vyvinul v prověřený systém, oblíbený zejména ve firemním prostředí. Výhodou linuxové platformy jsou samozřejmě menší náklady, a to nejenom na nákup softwarového řešení, ale i hardware.

rodiny produktů .NET. Uplatňuje se zejména u webových databází středních velikostí, i když zvládne i velké implementace (v těch ovšem kraluje Oracle). Výhodou je zajištěná kompatibilita s dalšími produkty firmy Microsoft.

Microsoft letos na podzim uvolnil verzi MS SQL 2005, která přináší množství nových funkcí například z oblasti Business Intelligence (BI). Měla by vykazovat vyšší bezpečnost, dostupnost a škálovatelnost i produktivitu vývojářů. Rozšíření se týká replikačních služeb, možnosti integrace s dalšími systémy (vyjmutí, transformace a nahrání dat) včetně databází Oracle. Komplexní správu i vývoj zajišťují integrované nástroje. Novinkou jsou funkce BI včetně data miningu, analytických a reportovacích nástrojů, s cílem zajištění

## Zastoupení Open Source databázi na trhu



dostupnosti dat pro všechny pracovníky v organizacích. SQL Server 2005 podporuje zrcadlení dat, které je důležité pro zajištění dostupnosti v případě poruchy primárního systému. Administrátorům je rovněž umožněno vytvářet okamžité snímky databáze pouze pro čtení, což šetří čas i zdroje. Rovněž je vylepšena funkce rychlé obnovy, takže se nyní uživatelé mohou připojit k obnovující databázi dříve.

## Další produkty

Dalším pojmem ve světě databází jsou produkty firmy Oracle (naposledy Oracle Database 10g). Tato oblíbená databáze s dlouhou historií se v nové verzi chlubí především možností rozložení výpočetních operací do gridu, tedy více geograficky vzdálených počítačů, což šetří náklady a snižuje výdaje na nákup nových technologií. Tato databáze

je dostupná pro většinu operačních systémů. Mezi další komerční produkty s významnými podíly na trhu patří Sybase ASE (Adaptive Server Enterprise), která je tradičně považována za robustní a výkonnou platformou, na které běží některé největší databáze světa. Klasikou relačních databází je rovněž databáze DB2 od IBM, s možností provozování na prakticky všech operačních systémech včetně Windows.

Zajímavou alternativou jsou objektové databáze, jako například Caché od InterSystems nebo Objectivity, které jsou oblíbené zvláště v některých oborech. Odstraňují totiž neuhdy relačních databází, tím že mohou trvale ukládat přímo objekty a při tvorbě aplikací není třeba dělat objektově-relační mapování, tedy provádět dekompozici a konverzi objektů do relačních tabulek jako u většiny relačních systémů.

Výhody objektových databází se projevují při růstu složitosti ukládaných dat. U relačních databází v tomto případě roste také exponenciálně počet operací (a tím i potřebný výpočetní výkon), kdežto u objektových je tento nárůst pouze lineární. Tyto databáze dnes přitom podporují i SQL a jsou díky dostupným nástrojům připraveny i pro nasazení jako webové aplikace. □