

**Česká zemědělská univerzita v Praze**

**Provozně ekonomická fakulta**

**Katedra informačních technologií**



**Teze diplomové práce**

**Optimalizace databáze pro CMS Drupal**

**Václav PEKÁREK**

**Vedoucí práce: Ing. Pavel Šimek, Ph.D.**

© 2015 ČZU v Praze

## Souhrn

Tato diplomová práce se zabývá optimalizací databáze MySQL pro systém správy obsahu Drupal, popisuje vývoj tohoto systému a analýzu přístupu CMS Drupal k databázi. Tato práce také zahrnuje zhodnocení postupů vedoucích k optimálnímu využití databáze a srovnání jejich efektivity v simulovaném zatížení.

**Klíčová slova:** CMS, redakční systém, Drupal, optimalizace, databáze, MySQL.

## Úvod

Redakční systémy umožňují kompletní správu obsahu webu. Umožňují uživatelsky přijatelnou cestou bez znalosti programování a značkovacích jazyků publikovat nejen texty, ale i jiné typy obsahu. Umožňují instalovat doplňující části, doplňky již realizovaných řešení, jako jsou fotogalerie, pokročilá diskuzní fóra, nebo celé elektronické obchody.

Pro vývojáře představuje použití redakčních systémů rozdělení vývoje do více částí za použití šablon a doplňků. Komunita uživatelů a vývojářů určitého redakčního systému se podílí na společném vývoji mnoha doplňků, jejichž znovupoužitelnost zefektivňuje vývoj celé řady systémů.

Systém Drupal patří mezi nejznámější trojici redakčních systémů WordPress a Joomla! Architektura CMS Drupal je z výše jmenovaných nejrobustnější, jeho využití je jak v malých projektech, tak i v rozsáhlých systémech s mnoha uživateli. Umožňuje nejen vytvořit pro méně zkušeného uživatele relativně snadno osobní stránky nebo blog, ale i velmi složitou webovou aplikaci pro státní správu, např. portál městského úřadu (Město Limerick, Irsko), univerzitní web (Dublin City University), nebo stránky Bílého domu.

CMS Drupal se svým charakterem open-source, osvědčil v oblasti podporující technologie, otevřená data, znalosti a otevřený přístup v rezortu zemědělství. Specifická distribuce AgriDrupalu představuje nástroj, který je např. vhodný pro využití v rozvojových zemích. Zcela odlišné použití systému Drupal představuje portál Zemědělských univerzitních novin (iZUN.eu).

Diplomová práce je zaměřena na problematiku využití databáze systémem Drupal. V případech použití s vysokým vytížením systému je právě databáze pro redakční systém hlavním uložištěm dat. Efektivní chod databáze představuje jeden z klíčových aspektů pro

celkový výkon redakčního systému. CMS Drupal je velmi často provozován s open-sourceovou databází MySQL, na jejíž optimalizaci se zaměřuje tato práce.

## **Cíl práce**

Diplomová práce je tematicky zaměřena na problematiku využití databáze CMS Drupal. Hlavním cílem práce je optimalizace databáze MySQL pro práci web content management systému Drupal.

Dílčí cíle práce jsou:

- analýza CMS Drupal a jeho přístupu k databázi MySQL
- vypracování přehledu vývoje CMS Drupal

## **Metodika**

Práce čerpá ze zdrojů odborné literatury a informačních zdrojů zveřejněných na internetu, uvedených v seznamu použité literatury. Praktická část práce je zaměřena na analýzu CMS Drupal a zhodnocení možností optimalizace práce s databází MySQL. Pro analytickou část, ověření navrhovaných optimalizačních nástrojů a jejich porovnání, bude využito zátěžových testů CMS Drupal. Na základě syntézy teoretických poznatků a výsledků praktické části budou formulovány závěry diplomové práce.

V rámci studia odborných informačních zdrojů v teoretické části bude navázáno na zpracované téma bakalářské práce *Zabezpečení a správa přístupů k databázi MySQL*.

Komplexní analýza CMS Drupal bude zahrnovat analýzu možností optimalizace zdrojového kódu systému souvisejícího s přístupem k databázi. Analýza bude také v některých částech obecně souviset s výkonem systému jako celku. Dále bude zahrnovat zhodnocení možností zvýšení výkonu, respektive odezvy v souvislosti s počtem uživatelů modelového webového serveru.

V práci se vychází z instalace CMS Drupal 6 a databáze MySQL verze 5.6 na platformě LINUX/Ubuntu, není-li uvedeno jinak.

## **Vlastní práce**

Vývojový cyklus zahrnuje právě podporovanou verzi Drupalu 7 v režimu dlouhodobé podpory. Po jeho ukončení bude následovat jen podpora kritických

bezpečnostních aktualizací. Dokončovaná verze 8 bude zahrnovat již zcela objektový přístup. Vývojový cyklus přináší každých 6 měsíců uvedení nové verze a pravidelné bezpečnostní aktualizace. V souvislosti s uváděním nových verzí autor zdůrazňuje omezení ve zpoždění dostupnosti aktualizace nových modulů.

V práci byly zhodnoceny možnosti optimalizace databáze a zdůrazněny možnosti vhodné pro použití v prostředí středně velké webové aplikace. Autor rozčlenil způsoby optimalizace podle způsobu správy webového serveru. V práci jsou rozčleněny postupy optimalizace v případě hostingu se značně omezenými právy správy, kdy je podle provedené analýzy doporučeno použití dostupného modulu (*DB Maintenance*). V případech plné správy serveru byly zhodnoceny postupy optimalizace v operačním systému LINUX - serveru MySQL s příklady použití programu *myisamchk*.

V rámci analýzy úložných enginů byla zkoumána možnost využití jiných typů než je primárně doporučováno pro CMS Drupal 6 v databázi MySQL. Zejména byla zkoumána možnost změny úložného enginu MyISAM pro ukládání paměti cache systému Drupal. Možnost využití velice rychlého úložného typu MEMORY se neprokázalo jako možné z důvodu specifických požadavků vzhledem ke struktuře dat a jejich objemu, systém Drupal využívá např. datové typy *longblob* a *text*, které analyzovaný úložný engin neumožňuje využít. Byla vyloučena i možnost změny datových typů na straně samotné aplikace.

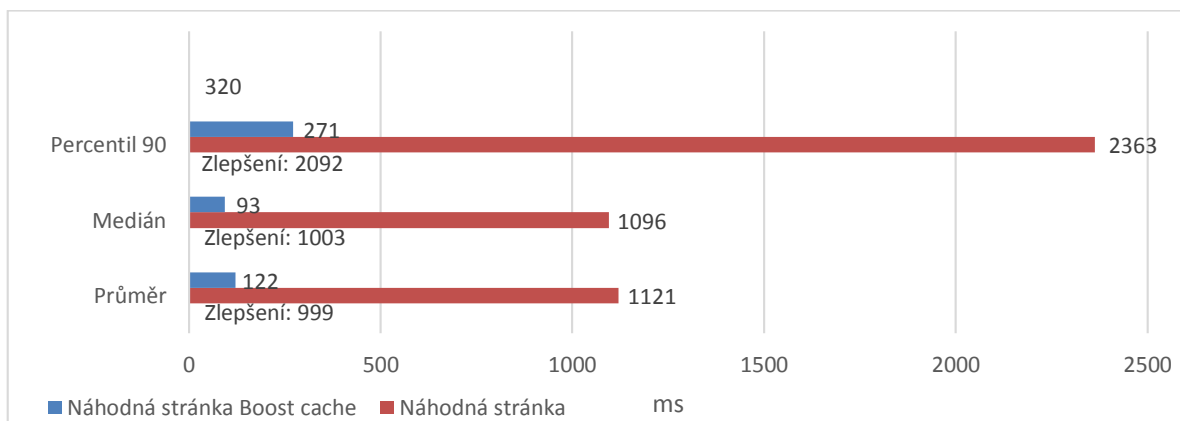
Byly analyzovány možnosti využití replikace databáze a zhodnoceny konkrétní typy replikace asynchronní a semisynchronní. Podle způsobu přístupu k tabulkám nebo jejím částem v průběhu replikace bylo doporučeno použití typu asynchronní replikace v režimu MIXED. V případě použití transakcí, zvláště v případě vyšších verzí od CMS Drupal 7, autor došel k závěru využít semisynchronní replikace, která vede k vyšší spolehlivosti a konzistentnosti dat. Byla také analyzována možnost využití MySQL Clusteru, která zvyšuje spolehlivost až na teoreticky možných 99,999 % (přesněji přepočteno na výpadek serveru maximálně celkem 5 min/ročně). Toto řešení však obnáší vyšší hardwarové nároky. Na druhou stranu kromě vyšší spolehlivosti umožňuje dynamicky za chodu měnit počet uzlů v clusteru, a tím reagovat na možný nárůst či pokles očekávaného výkonu. Přes zjištěné výhody, které jsou ovšem velice specificky úzce zaměřeny jen na některé případy, pro které je využití MySQL Clusteru výhodné, není toto řešení vhodné pro CMS Drupal.

Širší uplatnění pro zvýšení výkonu databáze CMS Drupal přináší využití replikace ve spojení s load-balancerem. Problematika rozložení zátěže společně s přímým připojením k replikovaným serverům byla popsána se zaměřením na způsob rozdělení a směrování dotazů pro zápis a čtení. Analýza rozložení zátěže vedla k položení otázky, zda se přiklonit pouze ke škálovatelnosti, anebo zvážit náročnější přístup, který vede k efektivnímu využití replikace a škálovatelnosti.

Řešení, jak plně využít přínos replikace v optimalizaci pro vysoký výkon databáze MySQL v kombinaci s CMS Drupal 6, je prostřednictvím Pressflow 6, které dokáže efektivně pracovat s replikačními servery.

Srovnání naměřených výsledků profilování CMS Drupal 6 a Pressflow 6 prokázalo značné přepracování jádra systému Pressflow 6 a jeho odlišnost. Naměřené hodnoty v základní instalaci obou sledovaných systémů vykazovaly obdobnou celkovou dobu trvání cca. 53μs.

Aplikace modulu Boost využívající paměť cache uloženou místo v databázi v souborovém systému vedla při zátěžovém testu jen k mírnému nárůstu zpoždění oproti výchozí paměti cache systému Drupal. Při 10 konkurenčních připojeních byl změřen 90. percentil 82ms.



**Obrázek č. 1: Porovnání doby načtení náhodné str. anonym. uživatelem bez paměti cache a s paměti cache modul Boost, 30 konkurenčních připojení.**

Ucelený pohled na rozdíl při použití paměti cache a modulu Boost při vyšším počtu 30 konkurenčních připojení je uveden na *obrázku č. 1: Porovnání doby načtení náhodné str. anonym. uživatelem bez paměti cache a s paměti cache modul Boost, 30 konkurenčních*

*připojení*. V porovnání s 10 konkurenčními připojeními byl nárůst doby načtení náhodné stránky zvýšen pouze o 189ms (90. percentil).

## **Závěr**

Redakční systém CMS Drupal je profesionální nástroj, který se rozvíjí za podpory silné komunity. Pro aktuálně podporovanou verzi Drupalu 7 je dostupný režim dlouhodobé podpory. Dokončovaná verze 8 bude zahrnovat již zcela objektový přístup. Vývojový cyklus přináší každých 6 měsíců uvedení nové verze a pravidelné bezpečnostní aktualizace.

V práci byly podrobně popsány a zhodnoceny metody optimalizace databáze MySQL v souvislosti s redakčním systémem Drupal 6. Byly analyzovány možnosti využití doplňkových modulů k optimalizaci využití databáze jak pro přihlášené uživatele, tak i pro anonymní uživatele. Dále také byly zhodnoceny přímo moduly optimalizující a provádějící údržbu databáze MySQL.

Pro anonymní uživatele byly provedeny výkonnostní testy při použití různých přístupů k paměti cache systému Drupal 6. Bylo dosaženo obdobných výsledků při použití standardní cache a speciální souborové cache modulu Boost.

Analýzou systému Drupal a struktury databáze byly zamítnuty změny typu uložiště databáze MySQL, a to i v případě jednotlivých částí databáze. Není možné využít pro zvýšení výkonu databáze jiné typy uložišť, nebyla prokázána ani vhodnost úložného typu MySQL Cluster. Autorem bylo ale doporučeno využití replikace databáze ve spojení se speciální distribucí Drupalu Pressflow 6. Ta je vhodná pro použití v aplikacích vyžadujících zvýšení výkonu. Je plně kompatibilní s CMS Drupal 6 a umožňuje podporu integrace s technologiemi na aplikační a prezentační vrstvě serveru vedoucí ke zvýšení výkonu.

Oba systémy Drupal 6 a Pressflow 6 byly porovnány využitím profilování a následně základním zátěžovým testem byl ověřen jejich výkon. Oblast optimalizace databáze je velmi úzká a prostor pro výraznější zvýšení výkonu poskytuje právě oblast prezentační a aplikační, které byly v práci také analyzovány. Autor dále upozornil na možné bezpečnostní riziko při použití distribuce Pressflow, která obsahuje zásah do kódu jádra systému.

## Zdroje

DUBOIS, P. *MySQL Developer's Library*. Boston: Addison-Wesley, 2009.

ISBN 978-0-672-32938-8.

HODGDON, J. *Programmer's Guide to Drupal*. O'Reilly Media, Inc., 2012.

ISBN 978-1-449-34331-6.

KOFLER, M. *Mistrovství v MySQL 5*. Brno: Computer Press, a.s., 2007.

ISBN 978-80-251-1502-2.

SCHWARTZ, B., ZAITSEV, P., TKACHENKO, V., ZAWODNY, J., LENTZ, A.,

BALLING, D. *MySQL profesionálně - optimalizace pro vysoký výkon*. Brno: Zoner software, a.s., 2009. ISBN 978-80-7413-035-9.

TREVOR, J. *Drupal 6 Performance Tips*. Birmingham: Packt Publishing, 2010.

ISBN 978-1-847195-84-5.