

**Česká zemědělská univerzita v Praze**

**Provozně ekonomická fakulta**

**Katedra informačních technologií**



**Teze diplomové práce**

**Ukládání dat pomocí XML nebo MySQL databáze**

**Zdeněk Srb**

© 2015 ČZU v Praze

# Ukládání dat pomocí XML nebo MySQL databáze

---

## Data storage comparison between XML and MySQL database

### Souhrn

Tato práce se snaží přiblížit dvě technologie ukládání dat, relační model a možnost využití značkovacího jazyka XML. Relační model je v práci zastoupen systémem řízení báze dat MySQL a možnosti značkovacího jazyka zastupuje systém řízení báze dat eXist. Oba systémy jsou open source řešením. Pro pochopení základních problematik je nejdříve vysvětlen základní princip ukládání dat při využití obou technologií, následně jsou blíže představeny oba systémy. Stručně je popsán jejich vznik, vývoj a architektura. Dále jsou popsány některé často používané nástroje a postupy pro výkonnostní testování systému řízení báze dat a je definována základní metodika pro testování systémů MySQL a eXist. Na závěr jsou předvedeny výsledky testů, ze kterých jsou vyvozeny závěry práce.

**Klíčová slova:** XML značkovací jazyk, MySQL relační databáze, XPath a XQuery jazyk, XSLT, Data Query Language, metodika testování, srovnávací testy

Systemy řízení báze dat tvoří důležitý prvek při manipulaci s datovou základnou, která je základním kamenem většiny webových stránek, služeb, aplikací a podnikových informačních systémů. Tyto systémy v sobě mohou uchovávat různorodá data, ať už se jedná o názory jedinců v podobě databáze uchovávající v sobě data z webových diskuzí a fór, přes počty zákazníků v určitých zemích nebo kontinentech v případě nadnárodních společností, až po lékařské záznamy a osobní data obyvatel celého státu. Během posledních čtyřiceti let se systémy řízení báze dat staly všudypřítomnou součástí informačních technologií. Bylo vyvinuto několik modelů, které definovaly základní předpoklady pro operace s danou datovou základnou. Mezi nejznámější modely patří hierarchický model, síťový model, relační model, dokumentově orientovaný model, objektově orientovaný model a objektově relační model. Tato práce se zaměřuje na problematiku systémů řízení báze dat zastupující relační model a dokumentově orientovaný model. Jako zástupce těchto modelů byly zvoleny následující open source řešení.

- MySQL je open source systém řízení báze dat. Jde o velmi spolehlivý a rychlý systém, který je stále více využíván v oblastech, kde panovali komerční produkty od společností Oracle nebo Microsoft (Dubois, 2014, s. 1,2). Díky různým utilitám obsaženým v balení s MySQL, je správa poměrně snadná. Díky velkému množství rozhraní pro programování aplikací (API), je snadné vytvořit vlastní software pro komunikaci s MySQL.
- eXist si klade za cíl, aby splňovaly požadavky na široké uživatelské základně, tudíž jde o velmi bohatý systém na rozšíření a specifické rysy ve své třídě. Systém řízení báze dat eXist byl navržen v průběhu let a vyvíjen tak, aby uspokojil potřeby uživatelů. Uživatelská základna je velmi pestrá, najdeme zde studenty a profesory na univerzitách, kteří provádějí jazykově zaměřené projekty, nebo velké mezinárodní vydavatele, jenž pracují s miliony dokumentů, až po vývojáře, kteří chtějí rychle vytvářet dokumenty a datově orientované webové aplikace (Siegel, Retter, 2014, s. 10).

Cílem bylo tyto dva systémy řízení báze dat porovnat tím, že bude vytvořena metodika testování, která umožní následné srovnání, i když oba systémy vycházejí z odlišných modelů. Výsledky nemají za úkol najít vítěze a poraženého, ale vyzdvihnout silné a slabé stránky daného řešení.

Teoretická část práce se zabývá základními teoretickými východisky relačního modelu systému řízení báze dat a dokumentově orientovaného modelu systému řízení báze dat využívajícího značkovací jazyk XML. Stručně jsou popsány základy relačního modelu a dvanáct Coddových pravidel. Po představení teoretických východisek je blíže představen systém řízení báze dat MySQL, jakožto zástupce relačního modelu. Krátce jsou zmíněny autoři, historie, licencování a logická architektura databázového systému. Pozornost je věnována systému úložných enginů, kterými systém řízení báze dat MySQL disponuje, pomocí nich dosahuje tento databázový systém velké variability ve spravování datové základny. Blíže jsou popsány vybrané dva úložné enginy – MyISAM a InnoDB. Stručný přehled je věnován i možnosti indexování uložených dat. Přiblíženy jsou základní indexy a jejich funkce. Práce se nadále věnuje značkovacímu jazyku XML. Hlavní pozornost je věnována jazykům XPath a XQuery, které slouží pro navigaci a vyhledávání v XML dokumentech. Následuje popis systému řízení báze dat eXist, nativní XML databázi. Velmi stručně je zmíněn autor, historický vývoj, licencování a logická architektura tohoto systému. Na závěr teoretické části práce jsou popsány indexovací schopnosti v systému eXist.

V Analytické části bude vysvětlena základní problematika testování systémů řízení báze dat. Teoreticky budou nastíněny důvody, strategie, možná řešení a nástroje pro tvorbu srovnávacích testů. Obecně problematika výkonnostního testování systému řízení báze dat je velice specifická a vyžaduje velkou znalost testovaného systému, všechny zde zmíněné mechaniky jsou spíše chápány jako základní postupy. Tento výklad je rozšířen o několik základních softwarových programů, které se využívají při testování systému řízení báze dat MySQL a pro testování XML dokumentů a dotazů v jazyce XQuery. V této části práce jsou dále provedeny praktické testy na systémech řízení báze dat MySQL ve verzi 5.6 a eXist ve verzi 2.2. Metodika testování spočívá ve tvorbě datové základny, která bude naplněna náhodnými daty. Tuto základnu následně pomocí systému řízení báze dat MySQL podrobíme systému dvanácti dotazů – testů a změříme jejich dobu trvání. Dotazy byly navrženy tak aby pokryly základní funkce dotazovacího jazyka SQL. Aby bylo možné výsledky nějakým způsobem porovnávat mezi systémy MySQL a eXist, bude následně datová základna převedena do podoby XML dokumentu a uložena do databázového systému eXist. SQL dotazy budou převedeny pomocí jazyků XPath a XQuery do takové formy, kterou je možné reprodukovat na systému řízení báze dat eXist, při zachování stejných

výsledků. Opět bude změřena délka trvání těchto dotazů. Analytická část je zakončena postupem měření a následnou statistickou úpravou naměřených hodnot.

V části výsledky a diskuze jsou nejprve prezentovány výsledky všech provedených měření na obou systémech řízení báze dat. Uvedeny jsou číselné hodnoty v tabulkovém formátu i grafická podoba v podobě spojnicových grafů. Následně je provedeno srovnání hodnot testů pro oba databázové systémy. Srovnání je také nejprve uvedeno číselně v tabulkách a následně v grafické podobě jako spojnicové grafy. Následně jsou z prezentovaných výsledků vyjádřeny závěry testování, které ukazují na silné a slabé stránky použitých databázových modelů v testovaných systémech. Na závěr je uvedena malá prognóza možného budoucího vývoje zmiňovaných systémů řízení báze dat.

Závěr práce rekapituluje popsaná témata a cíle práce.

### Seznam použitých zdrojů:

1. FAWCETT, Joe, QUIN, R. E. Liam, AYERS, Danny. *Beginning XML*. 5th Edition. John Wiley & Sons, Inc., 2012. 868 s. ISBN: 978-1-118-16213-2.
2. HOSOYA, Haruo. *Foundations of XML processing*. United States of America by Cambridge University Press, 2011. 240 s. ISBN: 978-0-511-90402-8.
3. SCHWARTZ, Baron, ZAITSEV, Peter, TKACHENKO, Vadim. *High Performance MySQL*. 3rd Edition. O'Reilly Media, 2012. 826 s. ISBN: 978-1-449-31428-6
4. DUBOIS, Paul. *MySQL Cookbook*. 3rd Edition. O'Reilly Media, 2014. 866 s. ISBN: 978-1-449-37402-0
5. DUBOIS, Paul. *MySQL*. 5th edition. Pearson Education Inc., 2013. 1176 s. ISBN: 978-0321833877.
6. HOSKINS, Dorothy. *XML and InDesign*. 1st Edition. O'Reilly Media, 2013. 148 s. ISBN: 978-1449344160
7. DAVIES, Alex. *High Availability MySQL Cookbook*. Pack Publishing Ltd., 2010. 276 s. ISBN: 978-1-847199-94-2.
8. ARENAS, Marcelo, BARCELO, Pablo, LIBKIN, Leonid, MURLAK, Filip. *Relational and XML Data Exchange*. Morgan & Claypool Publishers, 2010. 106 s. ISBN: 9781608454129.
9. POWELL, Gavin. *Beginning XML database*. John Wiley & Sons, Inc., 2007. 499 s. ISBN-13: 978-0-471-79120-1.
10. KOFLER, Michael. *Mistrovství v MySQL 5*. Computer Press, a.s., 2007. 808 s. ISBN: 978-80-251-1502-2.

11. RUSSELL, Dyer. *MySQL in a Nutshell*. 2nd Edition. O'Reilly Media Inc., 2008. 566 s. ISBN: 978-0-596-51433-4.

12. OPPEL, Andy. *SQL bez předchozích znalostí*. Computer Press, a.s., 2008. 242 s. ISBN: 978-80-251-1707-1.

13. SIEGEL, Erik, RETTER, Adam. *eXist*. O'Reilly Media Inc., 2014. 581 s. ISBN: 978-1-449-33710-0