

# Posudek oponenta diplomové práce

Název diplomové práce: *Vyhodnocení výkonu a migrace nerelačních distribuovaných datových skladů*

Student: *Miroslav OUJESKÝ*

Téma hodnocené práce je zaměřeno na problematiku datových skladů a vysoce dostupných datových úložišť. Práce si klade za cíl porovnat relační databázi MySQL s nerelačními datovými úložišti v situacích s velmi vysokým objemem dat. Záměrem práce je také analyzovat datové struktury systému Webnode a navrhnout a částečně realizovat proces migrace některých kritických částí tohoto systému do nerelačního datového úložiště.

Práce je rozdělena do dvou částí. V první (teoretické) části autor postupně analyzuje známá datová úložiště. Zaměřuje se na škálování SQL databází a detailně rozebírá architektury a vlastnosti NoSQL databází s ohledem na distribuci zátěže a dosažení vysokého výkonu. Ve druhé (praktické) části se práce věnuje provedení výkonnostních testů a výkonnostnímu srovnání všech hodnocených databází při různých typech zatěžování. Dále zde autor navrhuje migraci konkrétních elementů systému Webnode z prostředí databáze MySQL do prostředí několika NoSQL úložišť.

Co se týká formální stránky, lze považovat práci za přehlednou, zmínit lze pouze některé drobné prohřešky proti typografickým pravidlům – dle normy ČSN 01 6910 se odsazují počátky všech odstavců, kromě prvního v sekci. Jinak je ale práce typograficky na velmi vysoké úrovni.

Z hlediska obsahového je práce poměrně zdařilá, byť vytknout lze v případě hodnocení výkonu serveru MySQL použití storage engineu MyISAM i když je nejméně od konce r. 2011 známo, že storage engine InnoDB byl velmi výrazně optimalizován na výkon a dosahuje o jeden řád lepších výsledků než MyISAM. Dále diplomant provádí řadu měření a srovnávání výkonu SQL a NoSQL databází, ale v práci chybí nezbytné výchozí stanovení a ověření hypotézy o limitech výkonu relačního databázového systému. Zjištění těchto parametrů je vždy v praxi nezbytné, aby bylo možné určit okamžik, od kdy se vyplatí investovat úsilí vývojového týmu do implementace NoSQL řešení. Ostatně diplomant sám ověřil, že metody práce a přístupu k relační a nerelační databázi se liší, což má vliv na aplikační architekturu systému, který je nutné migrovat. Proto je nutné migraci z SQL do NoSQL prostředí pečlivě zvažovat.

Celkový dojem z práce je ale velmi dobrý. Zadání bylo splněno a byly zjištěny a ověřeny důležité vlastnosti a výkonnostní parametry SQL a NoSQL datových úložišť. Práci doporučuji k obhajobě, a i přes výše uvedené nedostatky ji hodnotím **stupněm A (výborně)**.

K obhajobě práce doporučuji zodpovězení následující otázky:

- Jaký je limitní počet záznamů v systému MySQL v situaci, kdy je databáze zatěžována v poměru 73% čtení / 27% zápis. Za limitní latenci považujte 750ms na transakci. Situaci testujte na jediné tabulce, ve které bude 10 sloupců VARCHAR(1024), 5 sloupců UNSIGNED INT, 2 sloupce DATETIME, 1 sloupec FLOAT, primární klíč UNSIGNED INT. Indexujte 1 sloupec VARCHAR na prvních 10 znaků a 4 INT sloupce. Toto hodnocení proveďte pro storage enginey MyISAM a InnoDB.
- Jednou větou vysvětlete, proč není vhodné indexovat sloupec typu FLOAT?
- Jmenujte nějaké nevýhody shardingu (horizontální distribuce dat databáze do více uzlů v clusteru)?

V Brně dne 29.5. 2012

.....  
Ing. Lubomír Cvrk, Ph.D.