

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy v Praze

Autor práce Martin Adam
Název práce Optimization of Processing of Data Files in System DIRAC
Rok odevzdání 2015
Studijní program Informatika **Studijní obor** Správa počítačových systémů

Autor posudku RNDr. Martin Svoboda, Ph.D. **Role** Oponent
Pracoviště KSI MFF UK

Prosím vyplňte hodnocení křížkem u každého kritéria. Hodnocení *OK* označuje práci, která kritérium vhodným způsobem splňuje. Hodnocení *lepší* a *horší* označují splnění nad a pod rámec obvyklý pro bakalářskou práci, hodnocení *nevyhovuje* označuje práci, která by neměla být obhájena. Hodnocení v případě potřeby doplňte komentářem. Komentář prosím doplňte všude, kde je hodnocení jiné než *OK*.

K celé práci

	lepší	OK	horší	nevyhovuje
Obtížnost zadání	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Splnění zadání	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rozsah práce ... <i>textová i implementační část, zohlednění náročnosti</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Komentář Cílem hodnocené bakalářské práce bylo rozšířit možnosti práce s metadaty a datasety v distribuovaném souborovém systému DIRAC a zvážení přechodu k vybrané NoSQL databázové technologii pro ukládání těchto dat namísto tradičních relačních databázových systémů, v tomto případě MySQL.

Pozornost byla konkrétně věnována rozšíření vyjadřovací síly dotazů nad metadaty souborů (dvojicemi klíč-hodnota, asociovanými jak přímo se soubory, tak děděnými mezi adresáři), kde byly uživatelům nově nabídnuty standardní boolské výrazy. Ty pak jsou interně nejenom z důvodu zvýšení efektivity jejich vyhodnocování (v dynamickém prostředí) převáděny do DNF, normalizovány, kontrolovány a optimalizovány. Druhá hlavní část práce spočívala v praktickém porovnání možností systémů Apache Cassandra, MongoDB a Elasticsearch, z nichž byl vybrán a integrován poslední zmíněný.

Zadání práce bylo v plném rozsahu splněno. Přestože rozsah implementační části je malý, student ve vztahu k oboru práce nepochybně prokázal schopnost pochopit a upravit zdrojový kód cizího systému, a to citlivým způsobem zaručujícím mimo jiné i požadavky na zpětnou kompatibilitu nebo skutečnou nasaditelnost řešení v praxi. Výstupy práce a jejich podoba jsou navíc motivovány probíhajícím fyzikálním výzkumem.

Textová část práce	lepší	OK	horší	nevyhovuje
Formální úprava ... <i>jazyková úroveň, typografická úroveň, citace</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Struktura textu ... <i>kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</i>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Analýza	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vývojová dokumentace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uživatelská dokumentace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Komentář Text vlastní práce je napsán kvalitní angličtinou s minimem chyb, je plynulý a čtenářům velmi dobře srozumitelný. Z hlediska struktury obsahuje práce všechny očekávané části, předkládané informace jsou navíc uvedeny stručně, systematicky a s jasným záměrem. Přesné vyjadřování je navíc doplněno o argumentaci učiněných rozhodnutí a nastínění alternativ. Pozitivem je také dobrá práce s odbornými informačními zdroji. Škoda jen, že se v práci nevěnovalo více prostoru detailům statického zpracování výrazů dotazů, tedy potenciálně zajímavé části z hlediska její netriviálnosti.</p>				

Implementační část práce	lepší	OK	horší	nevyhovuje
Kvalita návrhu ... <i>architektura, struktury a algoritmy, použité technologie</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kvalita zpracování ... <i>jmenné konvence, formátování, komentáře, testování</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Stabilita implementace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Komentář Jak již bylo uvedeno v úvodu posudku, hodnocená práce nepředstavuje standardní softwarový projekt, a tedy ani není možné standardními měřítky hodnotit implementační část. Rozšíření existujících komponent systému se však jeví jako řešení navržená a integrovaná velmi citlivě, a to nejenom ve vztahu k praktickým požadavkům, ale i celkové architektuře a kontextu systému DIRAC. Výběr technologie Elasticsearch a její integrace je navíc podpořen provedenými a dobře komentovanými experimenty a umožňuje další potenciálně zajímavá rozšíření. Aplikaci samotnou jsem sice osobně kvůli technickým překážkám v přístupu do výzkumné sítě nevyzkoušel, při předvádění práce studentem na osobní schůzce se však aplikace chovala stabilně.</p>				

Celkové hodnocení Výborně
Práci navrhuji na zvláštní ocenění Ne

Datum 18. ledna 2016

Podpis