



POSUDEK OPONENTA BAKALÁŘSKÉ PRÁCE

Jméno studenta: Lukáš Trumm
Název práce: Paralelizace výpočtů pomocí modelu MapReduce
Autor posudku: Ing. Pavel Janečka
Cíl práce: Navrhnout a implementovat paralelizovaný výpočet statistiky přístupů na web pomocí MapReduce. Popsat principy tohoto přístupu.

Povinná kritéria hodnocení práce	Stupeň hodnocení (známka)			
	A	C	E	F
Práce svým zaměřením odpovídá studovanému oboru	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vymezení cíle a jeho naplnění	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování teoretických aspektů tématu	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zpracování praktických aspektů tématu	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Adekvátnost použitých metod, způsob jejich použití	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Hloubka a správnost provedené analýzy	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Práce s literaturou	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Logická stavba a členění práce	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Jazyková a terminologická úroveň	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Formální úprava a náležitosti práce	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vlastní přínos studenta	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Využitelnost výsledků práce v teorii (v praxi)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Dílčí připomínky a náměty:

Autor vhodně definuje české ekvivalenty k mnoha klíčovým termínům v kapitole 2, v dalších kapitolách se jich však nedrží. Obdobně kapitola 2.3 popisuje páry klíč-hodnota, ale v definice funkcí jsou uvedeny anglické ekvivalenty k,v (key-value).

V úvodu práce je uvedeno, že byl projekt Hadoop veden nejprve společností Yahoo a poté přešel pod Apache, kapitola 3 popisuje historii projektu jinak.

Celkové posouzení práce a zdůvodnění výsledné známky:

Autor práce si kladl za cíl navrhnout a implementovat paralelizovaný výpočet statistiky přístupů na web pomocí MapReduce. Text práce a přiložené zdrojové kódy dokazují, že autor cíl práce splnil.

Jazyková a formální úroveň práce i odbornost textu jsou na dobré úrovni, autor se však několikrát nevyhnul první osobě. Téma práce je na bakalářskou práci možná až příliš široké, v textu tedy chybí větší rozvedení distribuovaných souborových systémů (kterým je věnována pouze jedna stránka textu práce), kdy autor sám uvádí, že se jedná o velmi důležitou součást problematiky. Zároveň v práci kompletně chybí kategorizace vhodnosti problémů pro paralelní výpočty pomocí MapReduce a popis možné heuristiky dovolující převést úlohu z neparalelizovaného zpracování na zpracování paralelizované pomocí MapReduce – pokud existuje.

Přiložené zdrojové kódy jsou přehledné, dokumentované a dobře čitelné. Autor vhodně doplnil přílohy a množství poznámek obsahující doplňující informace k implementaci a její instalaci a konfiguraci. Poznámky jsou ve formátu Markdown, který je čitelný i v čistě textové podobě, jejich převod do PDF (či jiného vhodného formátu) je však nasnadě.

Zvolená metodika pro zpracování vybraného problému pomocí skriptu a porovnání výsledku se zpracováním pomocí MapReduce je velmi neobjektivní. Nástroje využité ve skriptu (awk, sort, a další) jsou úzce specializované s mnoha lety vývoje. Jsou však srovnávány s kódem autora, které nebyl nikterak dále optimalizován (alespoň to práce neuvádí). Výsledky porovnání je tak nutné brát s rezervou. Míra škálovatelnosti jednotlivých řešení je však i přesto z výsledků dobře zřetelná.

Otázky k obhajobě:

V kapitole 3.1.4 uvádíte že framework Hadoop 2 podporuje i jiné typy úloh než jen MapReduce. O jaké úlohy se jedná?

Práci doporučuji k obhajobě.

Navržená výsledná známka: B - výborně-velmi dobře

V Praze, dne 7. ledna 2016

podpis