

Posudek doktorské disertační práce

Předkladatel: Zbyněk Falt

Název: Towards Efficient Parallel Data Processing on Modern Hardware

Předložená práce se zabývá paralelními výpočty, tedy jednou z nejaktuálnějších oblastí výzkumu na poli informačních technologií. Autor přispívá k rozšíření poznatků v této oblasti zejména návrhem a implementací jazyka Bobolang, který zjednodušuje programování některých paralelních aplikací.

Samotný text je rozdělen do šesti logických částí. Úvod motivuje celou práci, definuje její cíle, vymezuje příspěvek autora a nastiňuje obsah dalších kapitol. Rozsáhlý prostor je dále věnován platformě Bobox. Je rozebrána obecná struktura a, zejména, realizovaná rozšíření. Další kapitola se věnuje navrženému jazyku Bobolang. Autor představuje pokročilé prvky jazyka a ukazuje, jak přispívají ke zjednodušení paralelního programování. Následuje kapitola věnovaná paralelním algoritmům zpracování dat se zaměřením na databázové aplikace. Pátá kapitola se potom detailněji věnuje speciálnímu případu paralelního stroje SPARQL, který pracuje s velkou pamětí. Předkladatel ukazuje způsob reprezentace jednotlivých prvků a implementaci konkrétních operátorů a porovnává rychlost odpovědí s jinými systémy. Závěrečná část shrnuje dosažené výsledky a nastiňuje možné směry budoucího vývoje.

Na práci je nutné ocenit především skutečnost, že se autorovi podařilo proniknout do složité oblasti, která zasahuje do několika oblastí informatiky, a shrnout získané poznatky do uceleného textu. Pozitivně hodnotím také autorův přínos v oblasti plánovače úloh systému Bobox, který tento systém posunuje směrem k praktické použitelnosti pro systémy se sdílenou (a velkou) pamětí. Za zmínku stojí v neposlední řadě i dobrá orientace Zbyňka Falta v problematice SPARQL – výsledky uvedené v páté kapitole naznačují vysokou výkonnost v případě složitých dotazů.

Výhrady lze mít naopak k ne zcela jasnému vymezení vlastního vědeckého přínosu, jdoucího nad rámec profilování a inženýrské práce na existujícím systému. Obdobně je z mého pohledu nedostatečně diskutována praktická použitelnost výsledků – dnešním trendem je práce s výrazně většími objemy dat, než které běžně zvládají jednotlivé stroje se sdílenou pamětí (architektura NUMA je samozřejmě výhodou) a v pojednání o paralelním zpracování je také téměř ignorován současný prudký rozvoj práce s proudy dat v distribuovaných systémech. Porovnání s jinými systémy a vysvětlení rozdílů by si také zasloužilo mnohem větší pozornost, navržený systém a algoritmy jsou někdy lepší, někdy horší, nezdá se, že by vždy prokazatelně zlepšovaly současné metody. V oblasti plánovače by mne např. zajímalo porovnání se systémem DAGuE/ParSEC, který, ač není v práci zmíněn, je pro téma velmi relevantní. U SPARQL potom citelně chybí srovnání s OWLIM a dalšími systémy zaměřenými na zpracování v paměti.

Text se bohužel nevyhnul triviálním gramatickým chybám, např. *Algorithms which ensures...*, *The summarize, the main contributions...*, *I declare that I carried out this *master* thesis...*

Zbyněk Falt zjevně odvedl během svého doktorského studia hodně práce a prokázal schopnost prezentovat výsledky v mezinárodním kontextu. Pokud bude v rámci obhajoby přesvědčivě ukázána novost a relevance výsledků v porovnání s výše uvedenými systémy, navrhuji předloženou práci přijmout jako disertační a při splnění dalších podmínek udělit předkladateli titul doktor.