

Posudek oponenta bakalářské práce

Student: Sasák Tomáš
Téma: Optimalizace spouštěcích konfigurací k-Wave úloh (id 21798)
Oponent: Jaroš Marta, Ing., UPSY FIT VUT

- 1. Náročnost zadání** **průměrně obtížné zadání**
Cílem práce bylo navrhnout a implementovat heuristiky, které naleznou vhodné spouštěcí parametry (počet výpočetních zdrojů, délka alokace zdrojů/výpočetní čas) pro danou úlohu na daném superpočítači. Tyto heuristiky pracují s výkonnostními metrikami získanými s předchozích běhů daných úloh. Téma této práce považuji za průměrně obtížné.
- 2. Splnění požadavků zadání** **zadání splněno**
Všechny body práce byly splněny.
- 3. Rozsah technické zprávy** **přesahuje obvyklé rozmezí**
Technická zpráva svým rozmezím **přesahuje běžný počet normostran**, avšak je napsána velice **kvalitně**. Více jak polovina práce je věnována detailnímu popisu návrhu řešení, implementaci a testování, což hodnotím velice kladně.
- 4. Prezentací úroveň předložené práce** **85 b. (B)**
Logicky je technická zpráva stavěna dobře a jednotlivé kapitoly na sebe navazují. Pouze implementační část je čtenářky docela náročná vzhledem k množství informací a detailnímu popisu jednotlivých funkcí. Pro příště by možná bylo vhodné text doplnit více jednoduchými obrázky, které by čtenáři pomohly ho snadněji pochopit.

Mám pouze pár poznámek:

- U obrázku 8.2 na straně 55 postrádám pojmenování jednotlivých grafů a bližší vysvětlení.
- Popis obrázků a tabulek je většinou tvořen větami - ty je nutné ukončit tečkou.
- Grafy na obrázcích 9.1 a 9.2 (strany 69 a 73) - nejde vidět žádný rozdíl. Pro příště doporučuji použít pro obě osy grafu logaritmické měřítko.
- Tabulky 9.1 - 9.3 (strany 70 - 71) - uvedené časy jsou v sekundách a vztahené na celou simulaci nebo pouze na jeden simulační krok? Postrádám jednotky a v popisu grafu by bylo dobré zmínit, co znamená číslo před dvojtečkou (počet jader).
- Bylo by dobré ukázat grafy, jak se chová úloha, která má vhodné rozměry a úloha, která má naopak rozměry nevhodné pro výpočet.

V teoretické části jsou vysvětleny použité metody aproximace. Postrádám jen popis Akima spline a PCHIP interpolace.

- 5. Formální úprava technické zprávy** **80 b. (A)**
Z typografické stránky se mi technická zpráva jeví v pořádku. Někde se vyskytují dobré překlapy (chybějící čárka nebo mezera navíc).

Mám pouze drobná upozornění:

- Na straně 18 jsou pouze 3 řádky. Pro příště bych doporučila text upravit, aby se celý vešel na předchozí stránku.
- Na stranách 19 a 26 jsou citace osamoceny na novém řádku.
- Vyhněte se použití znaků > a < přímo ve větě a nahraďte je slovním popisem.
- Vyhněte se výrazům jako "najoptimálnější", "až taký silný rozdiel", "nie je vela" nebo "zopar bodmi".

Dále bych akorát upozornila na chybějící odkazy na obrázky v textu. Je nutné ale poznamenat, že student obrázky do textu vložil tak, aby přímo navazovaly na příslušný text.

- 6. Práce s literaturou** **75 b. (C)**
Práci s literaturou hodnotím dobře. Student pracoval jak se články, tak s webovými zdroji. Vzhledem k tomu, že práce se zaměřuje na úlohy využívající **k-Wave**, **postrádám citaci** na tento toolbox.
- 7. Realizační výstup** **90 b. (A)**

Kód je přehledně strukturován a dokumentován pomocí Doxygen značek. Popsány jsou třídy, funkce a proměnné. U dokumentace funkcí je drobná chyba, kdy se dokumentace vkládá bezprostředně před funkcí a ne do jejího těla. Samotný kód uvnitř funkcí bych však místy doporučila komentovat více. **Práce byla vytvořena v rámci skupiny SC@FIT. Zdrojové kódy však porušují politiku této skupiny (1) pojmenováním funkcí a (2) špatnou hlavičkou v souborech.** Doxygen dokumentace vygenerována nebyla.

Soubor readme popisuje, jak aplikaci spustit. Doporučila bych ještě doplnit popis, co jednotlivé moduly/soubory obsahují.

Kód je funkční a výstup aplikace je velice přehledný.

8. Využitelnost výsledků

Práce má veliký potenciál v oblasti vysoce výkonného počítání (HPC). Efektivní spuštění úloh vyžaduje znalosti experta a hledání vhodných spouštěcích parametrů představuje nemalou zátěž. Tato práce by jednak mohla urychlit rutinní spouštění těchto úloh, představuje možnost volby mezi celkovým výpočetním časem úlohy a cenou jejího výpočtu nebo možnost zavedení časových a cenových omezení.

9. Otázky k obhajobě

- Můžete vysvětlit, proč v případech, kdy pro danou velikost úlohy existuje v databázi více měření, průměrujete výpočetní časy těchto měření? Proč nepočítáte třeba střední hodnotu?
- Jak se plánuje úloha, pro kterou plánovač nemůže použít žádná historická data v databázi?
- Mohl byste procentuálně vyjádřit, jak se liší naměřená data (výpočetní čas pro daný počet jader a velikost úlohy) od těch aproximovaných? Jak jsou od sebe body vzdáleny?
- Jak byste vyhodnotil, který aproximátor je v daném případě nejvhodnější, aby plánování a spuštění úlohy mohlo probíhat automatizovaně?
- Píšete, že by vaše aplikace mohla být v budoucnu používána nástrojem k-Dispatch. k-Dispatch je ale postaven na Pythonu 3.5. Je vaše aplikace zpětně kompatibilní, popř. jaké změny by bylo nutné implementovat?

10. Souhrnné hodnocení

90 b. výborně (A)

Téma bakalářské práce má vysoký potenciál v oblasti plánování efektivního spuštění úloh na superpočítačích. Technická zpráva je zpracována velice pečlivě a svojí délkou přesahuje běžné rozmezí bakalářské práce, což ale není nijak na škodu. Převážná část se věnuje návrhu aplikace, řešeným problémům, implementaci a testování. Kód je přehledně strukturovaný a komentovaný. Realizační řešení je funkční.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 18. června 2020

Jaroš Marta, Ing.
oponent