

Obliczenia wielkiej skali PROJEKT 3 OMP studia dzienne rok ak. 2023/2024 (drugi termin)

Temat projektu dotyczy analizy efektywności przetwarzania równoległego realizowanego w komputerze równoległym z procesorem wielordzeniowym z pamięcią współdzieloną (projekt można realizować w grupach maksymalnie dwuosobowych, proszę o zgłoszenie składu grup przez email przed przystąpieniem do realizacji projektu)

W skład projektu wchodzi punkty 1-4

- 1) Przygotowanie kilku wersji kodu zgodnie z wymaganiami wersji zadania.
 - 2) Analiza efektywności każdej wersji kodu:
 - a) **analiza podziału pracy na wątki**, proszę za pomocą rysunków (dla każdego wariantu kodu) określić zadania realizowane przez poszczególne wątki i obszary danych wejściowych i wyjściowych przetwarzanych przez jeden wątek (uwzględnienie dyrektyw OpenMP)
 - b) analiza warunków **lokalnego czasowo dostępu do danych** w pamięci podręcznej:
 - lokalność czasowa występuje gdy ograniczymy do minimum (w danej wersji kodu) liczbę pobrań danych z pamięci operacyjnej do pamięci podręcznej,
 - uwzględnienie pamięci podręcznej L3 jako miejsca lokalnego przechowywania danych,
 - na podstawie:
 - wielkości mnożonych macierzy,
 - wielkości pamięci podręcznej procesora L3,
 - liczby jednocześnie uruchomionych wątków OpenMP i
 - organizacji przetwarzania - dane wspólne i **jednocześnie używane dane prywatne** wątków
- określenie liczby niezbędnych etapów wieloetapowych metod mnożenia macierzy (metoda 4 i 6 pętlowa). Każdy z etapów ma charakteryzować się lokalnością czasową dostępu do danych i minimalizować ilość danych pobieranych z pamięci głównej (por. wykład z zapewniania lokalności dostępu w zagadnieniu mnożenia macierzy),
- przeprowadzone eksperymenty mają dotyczyć przetwarzania równoległego realizowanego przy wykorzystaniu środowiska Openmp przy zapewnieniu zrównoważenia pracą rdzeni procesora,
 - miarą efektywności przetwarzania jest przyspieszenie przetwarzania równoległego i prędkość bazująca na czasie obliczeń i złożoności algorytmu mnożenia macierzy,
 - eksperymentom mają zostać poddane kody mnożenia macierzy z **optymalną kolejnością** zagnieżdżenia pętli kody 3,4 i 6 pętlowe (por. wykład z zapewniania lokalności dostępu w zagadnieniu mnożenia macierzy),
 - należy wyznaczyć przyspieszenie wynikające z:
 - równoległości przetwarzania dla metod 3,4 i 6 pętlowej,
 - zapewnienia efektywnego wieloetapowego przetwarzania z podziałem na etapy bazującym na podziale jedno i dwuwymiarowych danych wejściowych i wyjściowych (metoda 4 i 6 pętlowa).
- 3) Eksperyment:
 - a) Proszę wykonać dla kodów kompilowanych w trybie optymalizacji kodu -O3
 - b) Należy wykonać obliczenia mnożenia macierzy kwadratowych o rozmiarach 2048x2048 , 3072x3072 4096x4096 zmiennych typu float.
 - 4) Przygotowanie sprawozdania.

Sprawozdanie zawierać powinno:

Forma: sprawozdanie z numeracją stron, rysunków, tabel, wzorów, tabele i rysunki muszą posiadać podpisy określające w sposób jednoznaczny i pełny zawartość tabeli i rysunku.

Sprawozdanie z projektu i eksperymentu obliczeniowego

Wstęp:

- Nazwa zaliczanego przedmiotu;
- Imiona i nazwiska autorów sprawozdania, numery indeksów, numer grupy dziekańskiej i termin zajęć laboratoryjnych.
- Terminy: wymagany i termin rzeczywisty oddania sprawozdania.
- Krótki opis treści realizowanego zadania.

- Adres email kontaktowy do autorów sprawozdania.

Punkt 1 sprawozdania:

- Opis wykorzystanego systemu obliczeniowego:
 - oznaczenie używanego procesora i jego parametry: liczba procesorów fizycznych, liczba procesorów logicznych, liczba uruchamianych w systemie wątków, oznaczenie typu procesora, wielkość i organizacja pamięci podręcznych procesora.
 - Nazwa systemu operacyjnego, nazwa kompilatora i sposób uruchomienia kompilacji.
 - wersja systemu operacyjnego, wersja i nazwa oprogramowania użytego do przygotowania kodu wynikowego i przeprowadzenia testów.

Punkt 2 sprawozdania:

- Prezentacja przygotowanych wariantów kodów z wyjaśnieniem przewidywanego przebiegu przetwarzania. W sprawozdaniu powinny znaleźć się kluczowe elementy kodów charakterystyczne dla poszczególnych wersji kodu. Kolejne wersje kodu powinny zostać oznaczone w sposób skrótowy jednoznaczny i nawiązujący do wariantu użytego algorytmu i sposobu podziału pracy wątków. Oznaczenia te zostaną użyte w dalszej części sprawozdania: w prezentacji wyników eksperymentów i wniosków związanych z poszczególnymi wariantami kodu. W opisie przebiegu przetwarzania (dla poszczególnych wariantów kodu) należy uwzględnić:
 - zagadnienie **podziału** pracy – wielkość zbioru zadań, który powstaje w przetwarzaniu równoległym,
 - sposób **przydziału** zadań do procesorów – kiedy i w jakiej liczbie zadania są przydzielane do procesorów,
 - omówić dyrektywy i klauzule Open MP użyte w kodzie i ich znaczenie dla poprawności przebiegu obliczeń,
 - omówić w sprawozdaniu analizę lokalności odwołań do pamięci opisaną w zakresie projektu
 - omówić występujące w kodzie potencjalne problemy efektywnościowe:
 - konflikty polegające na unieważnianiu kopii danych (czy będą występować i jak często będą występowały podczas obliczeń),
 - synchronizacja – czy i gdzie w kodzie występuje i jaki ma wpływ na czas obliczeń.

Punkt 3 sprawozdania:

Prezentacja wyników i omówienie przebiegu eksperymentu obliczeniowo-pomiarowego:

- omówienie jakie kodu były testowane, jakich wielkości instancji przetwarzanie dotyczyło,
- tabela z wynikami z czytelnym określeniem jakiego uruchomienia: testowany wariant kodu i parametrami instancji; (warto aby tabela była jedna – zbiorcza zorientowana poziomo na stronie dla lepszego wykorzystania miejsca) i w sposób czytelny prezentowała kluczowe parametry oraz miary wyliczone:
 - czas przetwarzania
 - przyspieszenie przetwarzania równoległego dla badanego wariantu kodu równoległego – parametr ten jest ilorazem czasu przetwarzania **najlepszego** dostępnego przetwarzania oraz czasu przetwarzania równoległego, dla którego przyspieszenie jest wyznaczane.
 - prędkość przetwarzania liczona jako liczba operacji mnożenia i dodawania wartości tablic przypadająca na jednostkę czasu przetwarzania; liczba operacji mnożenia i dodawania (bez uwzględnienia operacji na zmiennych sterujących pętlami) dla mnożenia tablic o wielkości $N \times N$ elementów wynosi $2 \cdot N \cdot N \cdot N$ i wynika z konieczności wyliczenia $N \cdot N$ elementów składowych tablicy wynikowej, a każdy z takich wyników wymaga N operacji dodawania i N operacji mnożenia.
 - efektywność przetwarzania równoległego jako iloraz przyspieszenia przetwarzania równoległego i liczby użytych w przetwarzaniu procesorów fizycznych czyli rdzeni.
- wartości parametrów powinny być prezentowane w formacie pozwalającym na ich czytelne porównanie - wartości wraz z jednostkami prezentowanych wartości: przykładowo: czas przetwarzania 0,0053 s, prędkość obliczeń 1,23E+05 1/s (liczba zbadanych liczb na sekundę obliczeń).

Punkt 4 sprawozdania:

Wnioski. W wnioskach należy:

- porównać jakość rozwiązań problemu przy użyciu różnych wymaganych wariantów kodu oraz zaproponowanej liczbie etapów przetwarzania w metodach 4 i 6 pętlowych zależnej od wielkości instancji problemu, liczby wątków i wielkości pamięci podręcznej procesora L3,
- we wnioskach uwzględnić różnice w spodziewanej **ilości danych pobieranych do pamięci podręcznej** wynikającej z liczby etapów lokalnego czasowo dostępu do danych i ilości danych używanych w poszczególnych etapach przetwarzania,

- określić, które podejście okazało się najlepsze/najgorsze pod względem prędkości przetwarzania, czy ta przewaga jest niezależna od liczby użytych procesorów i rozmiaru testowanej instancji. Jakie mogą być prawdopodobne przyczyny takich wartości miar efektywności przetwarzania. ,

W próbie uzasadnienia przyczyn wysokiej lub niskiej efektywności poszczególnych podejść do realizacji kodu proszę odwoływać się w sposób jednoznaczny do **poszczególnych wartości** miar w tablicy wyników, poszczególnych **linii kodu** omawianej wersji kodu (znaczenia występujących w tych liniach konstrukcji językowych), rysunków określających zakres danych używanych przez poszczególne wątki w poszczególnych fazach obliczeń. Proszę unikać sformułowań ogólnych typu szybki, wolny, lepszy, gorszy oraz uwzględnić znaczenie wartości parametru - np. różne znaczenie ma dwukrotny wzrost wartości przyspieszenia przetwarzania równoległego wynoszącego 5 a inny dla przyspieszenia równego 0,3. ,

Przygotowanie sprawozdania.

Forma: dokumentacja w formie elektronicznej (plik sprawozdania (format pdf) i plik archiwum (format zip) z kodami źródłowymi) wymagana jest w terminie do 10 sierpnia 2024. W sprawozdaniu zawierającym powyżej wymienione elementy składowe:

- powinny być ponumerowane strony,
- każdy obiekt sprawozdania poza tekstem ciągłym: listing kodu, rysunek, wzór, tabela powinien posiadać numer kolejny i podpis określający w sposób jednoznaczny i pełny przedstawioną w obiekcie zawartość.

Podsumowanie i ocena realizacji projektu

Po sprawdzeniu merytorycznej zawartości zostanie wystawiona ocena, możliwe jest skierowanie sprawozdania nie spełniającego powyżej opisanych wymagań do uzupełnienia i/lub rozmowa z autorami na temat przebiegu eksperymentu i zawartości sprawozdania. Ewentualne uzupełnienia sprawozdania powinny zawierać zaznaczone dodane elementy (sprawozdanie należy uzupełnić w terminie 7 dni).

Nieuzasadnione opóźnienie w oddaniu sprawozdania obniża ocenę przyznaną za jakość sprawozdania. Obniżają o ½ stopnia ocenę następujące opóźnienia - po terminie, po tygodniu, po 2 tygodniach, po 4 tygodniach).

Konsultacje:

Trudności praktyczne i koncepcyjne związane z realizacją projektu mogą Państwo zgłaszać przez email.

Literatura:

Wykłady i literatura przedmiotu – szczególnie te dotyczące problem sekwencyjnego i równoległego mnożenia tablic kwadratowych i OpenMP