CS211 HW4

Hugo Wan

862180666

12/2/2024

Q1

NP (not parallelizable) = 95% = 0.95.

CP (can be parallelized) = 1 – NP = 1 – 95% = 5% = 0.05.

P = 10 processors. S is the desired speed up.

S = 1 / [CP + (NP / P)] = 1 / (0.05 + 0.95 / 10) = **6.897 maximum speed up**.

Q2

S (desired speedup) = 10. P is the processors.

CP = 6% = 0.06. NP = 1 – CP = 1 – 0.06 = 0.94 = 94%.

S = 1 / [CP + (NP / P)] → 10 = 1 / (0.06 + 0.94 / P) → 0.06 + 0.94 / P = 1 / 10

→ 0.94 / P = 1 / 10 – 0.06 → 0.94 / P = 0.04 → 0.04 P = 0.94

→ P = 0.94 / 0.04 = 23.5 → 23.5 + 1 = **24 processors** (round up).

Q3

Assume we can have infinite processors, then NP / P = NP / infinite, which is close to 0. S (desired speedup) = 50.

So, S = 1 / [CP + (NP / P)] → 1 / [CP + 0] = 1 / CP

Plug in S = 50 when S = 1 / CP → 50 = 1 / CP

CP = 1 / 50 = 0.02 = **2%, the maximum fraction of the computation**.

Q4

S = 9.   P = 10.   NP = 1 – CP.

S = 1 / [CP + (NP / P)] → S = 1 / {CP + [(1 – CP) / P)]}

→ 9 = 1 / {CP + [(1 – CP) / 10)]}

→ CP + [(1 – CP) / 10)] = 1/9

→ 10 CP + (1 – CP) = 10/9

→ 9 CP + 1 = 10/9

→ 9 CP = 10/9 – 1

→ 9 CP = 1/9

→ CP = 1/81 = 0.0123 = **1.23%, the maximum fraction of the computation**.

Q5

PT (parallel time) = 242 seconds.   CT (cleanup time) = 9 seconds.

AT (active time) = 233 seconds.   P (processors) = 16.   SS = Scaled Speedup.

SS = (CT + AT * P) / PT

→ SS = (9 + 233 x 16) / 242

→ SS = (9 + 3728) / 242

→ SS = 3737 / 242 = **15.442 scaled speedup**.

Q6

P = 40.   CP = 99% = 0.99.   NP = 1 – CP.   SS = Scaled Speedup.

SS = P * NP

→ SS = P * (1 - CP)

→ SS = 40 * (1 – 0.99) = **39.6 scaled speedup**.

Q7

S1 = 9. P1 = 10. S2 = 90. P2 = 100. S = 1 / {CP + [(1 – CP) / P)]}.

9 = 1 / (CP + (1 – CP) / 10)	vs.	90 = 1 / (CP + (1 – CP) / 100)
(CP + (1 – CP) / 10) = 1/9	vs.	(CP + (1 – CP) / 100) = 1/90
10 CP + 1 – CP = 10/9	vs.	100 CP + 1 – CP = 100/90
9 CP = 1/9	vs.	99 CP = 10/90 = 1/9
CP1 = 1/81	vs.	CP2 = 1/891

CP1 / CP2 = 1/81 / 1/891 = 11 times.

So, they cannot achieve the same speedup because the one with 100 processors has a speedup of 11 times faster than the one with 10 processors.

Q8

PT = 1000 seconds when P = 1.

PT = 500 seconds when P = 4.

PT = ? when P = 16? S = 1 / {CP + [(1 – CP) / P)]}

- S (4) = PT (1) / PT (4) → S (4) = 1 / {CP + [(1 – CP) / 4)]

→ CP = (1/S(4) – 1/4} / (1 – 1/4)

- S (4) = PT (1) / PT (4) = 1000 / 500 = 2

→ CP = (1/2 – 1/4) / (1 – 1/4) = (0.5 – 0.25) / 0.75 = 0.25 / 0.75 = 0.3333.

- S (16) = 1 / {CP + [(1 – CP) / 16)]

→ S (16) = 1 / [0.3333 + (1 – 0.3333} / 16)

→ S (16) = 2.6669

- T (16) = T (1) / S (16)

→ T (16) = 1000 / 2.6669 = **374.9672 seconds**.

Q9

Isoefficiency relations:

a. N >= sqrt (Cp)
b. N >= sqrt (C * sqrt (p) * log p)
c. N >= sqrt (C * sqrt (p))
d. N >= sqrt (Cp * log p)
e. N >= Cp
f. N >= p^C where 1 < C < 2
g. N >= p^C where C > 2

Scalability: c > b > a > d > e > f > g.

Q10

T seq = 2 N^3.   T comm (N, P) = 16 N^2 log_2 P.   Target Speedup: S = 256.

Total Memory: M (N) = 24N^2.   P = 1024.   1 GB = 10^9 bytes.

- T par (N, P) = (2 N^3 / P) + 16 N^3 log_2 P.
- S max (P) = T seq (N) / T par (N, P) = (2 N^3) / [(2 N^3 / P) + 16 N^2 log_2 (P)]

  → S max = (2 N^3) / [(2 N^3 / 1024) + 16 N^2 log_2 (1024)]

  → S max = (2 N^3) / [(2 N^3 / 1024) + 16 N^2 * 10]

  → **S max = (2 N^3) / [(2 N^3 / 1024) + 160 N^2] as the maximum speed up**.
- S = 256 and S max = (2 N^3) / [(2 N^3 / 1024) + 160 N^2]

  → 256 = (2 N^3) / [(2 N^3 / 1024) + 160 N^2]

  → 2 N^3 = 246 * (2 N^3 / 1024) + 160 N^2]

  → 2 N^3 = 512 N^3 / 1024 + 256 * 160 N^2

  → 2 N^3 = 1/2 N^3 + 40960 N^2

  → 3/2 N^3 = 40960 N^2

  → 3 N = 81920

  → N = 27306.7

  → N = 27306+ 1 = **27307 minimum problem size** (round up).