Kevin Lin

CSC346

Lab 3 – K-Means Clustering

**Page 1:** kMeansLarge   
**Page 2:** kMeansMedium  
**Page 3:** Questions

kmeansLarge.txt

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Processors | Clusters | Run 1 | Run 2 | Run 3 | Average |
| 1 | 2 | 0.594467 | 0.590893 | 0.567043 | 0.58413 |
| 1 | 5 | 1.065788 | 1.056082 | 1.061712 | 1.06119 |
| 1 | 10 | 9.49087 | 1.89607 | 9.255212 | 6.88071 |
| 1 | 25 | 14.09474 | 21.13538 | 26.75197 | 20.66069 |
| 2 | 2 | 0.440893 | 0.439584 | 0.448009 | 0.44282 |
| 2 | 5 | 0.678668 | 0.676309 | 0.675571 | 0.67684 |
| 2 | 10 | 2.513349 | 2.53665 | 4.689296 | 3.24643 |
| 2 | 25 | 12.45543 | 11.75484 | 13.44054 | 12.55027 |
| 4 | 2 | 0.401569 | 0.572605 | 0.432722 | 0.46896 |
| 4 | 5 | 0.505988 | 0.495042 | 1.1716 | 0.72421 |
| 4 | 10 | 2.138256 | 3.343904 | 2.519585 | 2.66724 |
| 4 | 25 | 6.266256 | 8.501845 | 7.254783 | 7.34096 |
| 8 | 2 | 0.382603 | 0.442711 | 0.410102 | 0.4118 |
| 8 | 5 | 0.534241 | 0.471517 | 0.481738 | 0.49583 |
| 8 | 10 | 3.008609 | 0.580959 | 3.932124 | 2.50723 |
| 8 | 25 | 4.999785 | 5.232899 | 4.850849 | 5.02784 |
| 16 | 2 | 0.59264 | 0.51611 | 0.379949 | 0.49623 |
| 16 | 5 | 0.583839 | 0.554914 | 0.632435 | 0.59039 |
| 16 | 10 | 1.463452 | 1.83673 | 1.502491 | 1.60089 |
| 16 | 25 | 4.657302 | 3.202278 | 4.65931 | 4.17296 |

kmeansMedium.txt

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Processors | Clusters | Run 1 | Run 2 | Run 3 | Average |
| 1 | 2 | 0.078289 | 0.189914 | 0.114439 | 0.12754 |
| 1 | 5 | 0.84967 | 0.915533 | 0.704074 | 0.82309 |
| 1 | 10 | 2.960898 | 2.148134 | 1.670461 | 2.25983 |
| 1 | 25 | 3.412331 | 5.664185 | 4.251154 | 4.44255 |
| 2 | 2 | 0.092004 | 0.102739 | 0.065813 | 0.08685 |
| 2 | 5 | 0.838743 | 0.897029 | 0.936182 | 0.89065 |
| 2 | 10 | 1.208376 | 1.613071 | 1.863732 | 1.56172 |
| 2 | 25 | 2.673548 | 2.146778 | 2.936422 | 2.58558 |
| 4 | 2 | 0.098105 | 0.110723 | 0.089617 | 0.09948 |
| 4 | 5 | 0.54425 | 0.749153 | 0.790488 | 0.69463 |
| 4 | 10 | 1.640648 | 0.755787 | 1.006525 | 1.13432 |
| 4 | 25 | 1.931826 | 1.509673 | 1.754017 | 1.73183 |
| 8 | 2 | 0.185269 | 0.098392 | 0.130858 | 0.13817 |
| 8 | 5 | 0.137844 | 0.160189 | 0.133651 | 0.14389 |
| 8 | 10 | 1.748473 | 1.309633 | 0.92476 | 1.32762 |
| 8 | 25 | 1.644154 | 1.381303 | 1.437702 | 1.48771 |
| 16 | 2 | 0.280268 | 0.2329 | 0.229979 | 0.24771 |
| 16 | 5 | 1.403694 | 2.290159 | 1.181543 | 1.62513 |
| 16 | 10 | 1.145926 | 1.956281 | 1.702421 | 1.60154 |
| 16 | 25 | 1.856818 | 1.310485 | 1.625432 | 1.59757 |

Questions:

1. Even though you only have a small sample size (two different values for S, the dimension upon which you are distributing the work), how does the speedup change for fixed P as the S parameter changes? Explain why you believe you are seeing such changes in the speedup.

Answer:

With fixed processor count, the bigger the sample size, the bigger the speedup, which becomes more apparent as P becomes bigger (like 8 or 16). This is because as the problem size grows (with more data to process), the relative cost of overhead becomes less.

|  |  |  |
| --- | --- | --- |
| Processors | kmeansMedium | kmeansLarge |
| 2 | 1.718202492 | 1.646234703 |
| 4 | 2.565234463 | 2.814439801 |
| 8 | 2.986166659 | 4.109257653 |
| 16 | 2.780817116 | 4.951087477 |

*Speedup, for clusters = 25*

1. How does the speedup change when you increase P for a given file? Explain why you believe you are seeing such changes in the speedup.

Answer:

Increasing the number of processors also increases the speedup. In a random sample of both 25 clusters each, an increase from 1 to 8 processors gave a speedup of 2.81 and an increase from 1 to 16 processors gave a speedup of 4.9 (kmeansLarge.txt). In another test, a P1->P8 increase gave a 2.56 speedup, and a P1->P16 increase gave 2.78 (kmeansMedium.txt).

However, because of Amdahl’s law, unless the serial parts of the program can be completely erased, we will eventually see a point where speedup will no longer increase (hits a peak), even with the number of processors still increasing.

1. How does the efficiency change when you increase P for a given input file? Explain why you believe you are seeing such changes in the speedup.

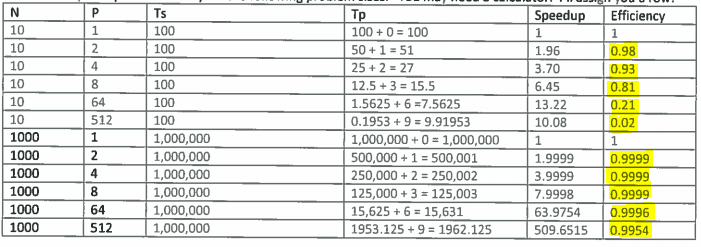
Answer:

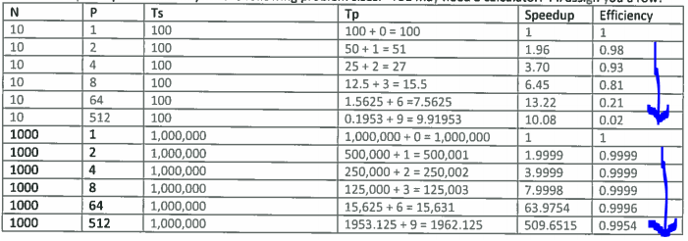
As the number of processors increase, the efficiency decreases.

|  |  |  |
| --- | --- | --- |
| Processors | Speedup | Efficiency |
| 2 | 1.646235 | 0.823117 |
| 4 | 2.81444 | 0.70361 |
| 8 | 4.109258 | 0.513657 |
| 16 | 4.951087 | 0.309443 |

*Speedup and efficiency for clusters = 25, kmeansLarge.txt*

This is consistent with what we have learned in class, on lecture 7’s worksheet.





I think this might be because the overhead (creation, management, synchronization, communication, mutual exclusion) will start to show as the number of processors increase (as opposed to none of communication, mutex, … etc. worries if only 1 processor is being used).