ELEN4020: Data Intensive Computing
Tutorials

May 20, 2019

Question 1

(i.) What is Flynn's toxonomy of parrallel programs. Briefly explain each toxonomy.

(ii.) Consider a 4-dimensional hypercube computer architecture of p processors. What is:

(a) the diameter?

(b) the connectivit?,

(c) the bisection width

(iii.) Repeat the above for a 3-Dimensional cube of p processors?

Question 2

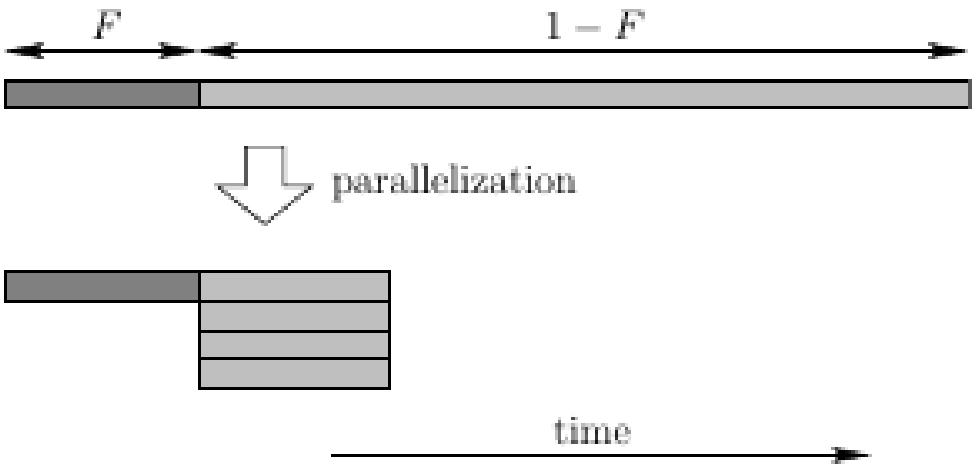

Figure 1: Sequential and parallelizable parts of a calculation

Amdahl's law states that if F is the fraction of a calculation that is sequential (it cannot benefit from parallelization), and $(1 - F)$ is the fraction that can be parallelized, then the maximum speed-up that can be achieved by using p processors is

$$S_{max} = \frac{p}{1 + (p - 1)F}$$

(i.) Derive Amdahl's law.

(ii.) Assuming you can use as many processors as you want. For the type of calculation sketched in Figure 1, is there a limit to the speed-up you can achieve? What does that mean for the parallel efficiency of computations using large numbers of processors?

Question 3

a) How will you define a *Thread* in parallel computing?

b) Give four characteristic features of a thread.

c) Many languages have a row major memory layout. For example:

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

is stored as [1, 2, 3, 4, 5, 6, 7, 8].

d) Given the row major storage layout after multiplying the matrices $C = A \times B$ where

$$C = A \times B = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} \times \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{bmatrix}$$

e) Write an openMP function:

void matrixMult(. . .)

to compute the matrix multiplication of $C = A \times B$ where A is a $2 \times k$ matrix and B is a $k \times 2k$ using p threads.

Question 4

Assuming we have an $A = N \times N$ upper triangular matrix as illustrated in Figrue **??** held by process of rank 0.

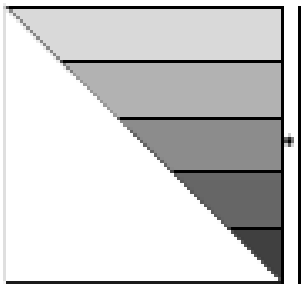

Figure 2: An upper triangular matrix

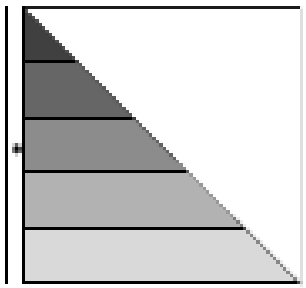

Figure 3: A lower triangular matrix

We wish to transform it into a lower triangular matrix as illustrated in Figrue 3 by N processes.

Write an MPI program to achieve this with N processes.

Question 5

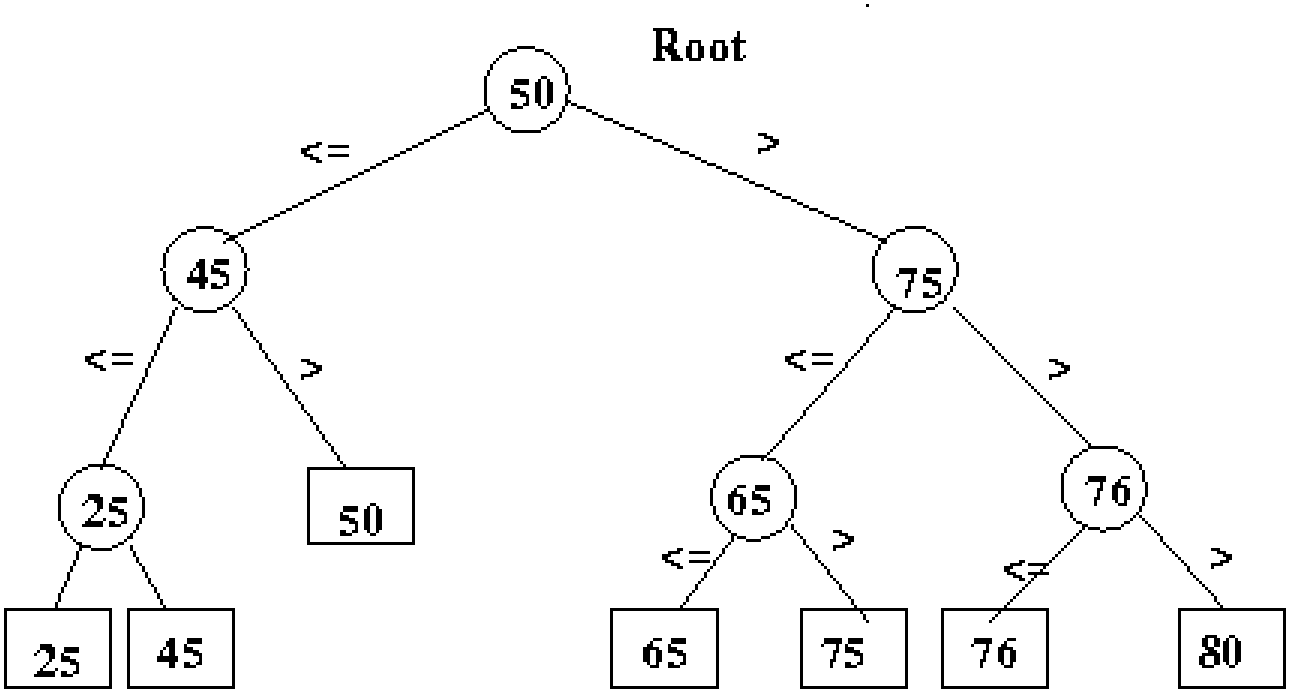

Figure 4: Binary Search Tree Traversal, Insertion and Deletion

Consider the special case binary search tree in which all keys are stored in a leaf node. The internal nodes only store keys to direct the search path. Processes traverse the tree to determine if a key exists; insert keys if the key does not exist, and delete keys if a key exists. Sketch algorithms in pseudo code for openMP processes that:

- perform simple lookup; OpenMP_Find_key(...);
- or perform an insertion; OpenMP_Insert_key(...);

You may assume that the tree structure is in shared memory.

Question 6

Write an MPI program that does collective parallel IO by writing, the content of an $N \times N$ matrix $A[N][N]$ in row-major order in a file.

Question 7

Write an OpenMP program to compute the value of π, by:

i.) numerical integration of a function

$$f(x) \quad = \quad \int_0^1 \frac{4}{1+x^2} \quad = \quad \pi$$

ii.) By Monte-Carlo simulation

Question 8

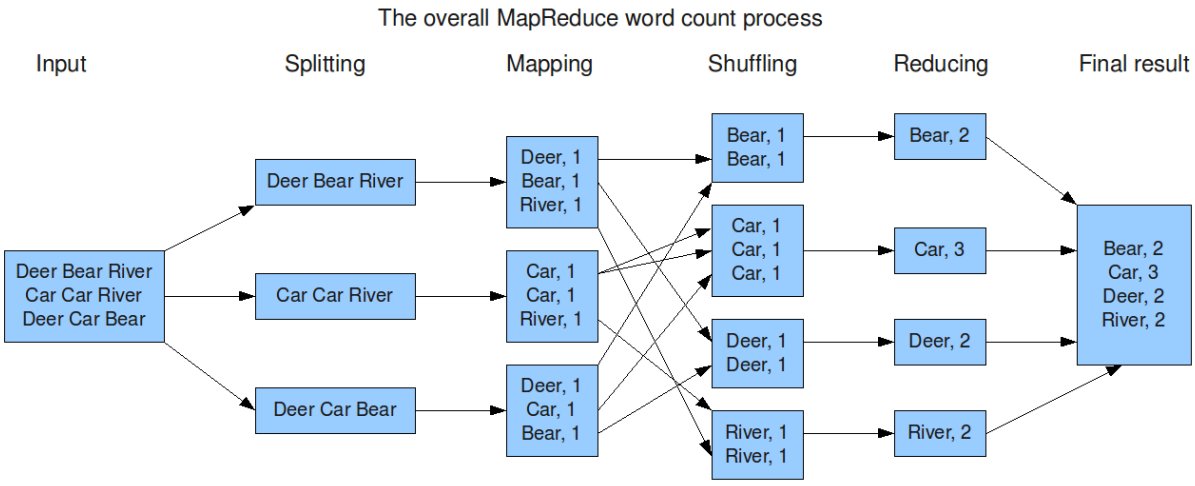

Figure 5: Illustration of MapReduce Word Count Process

a) The Figure 5 illustrates the MapReduce process to count the number of distinct words in a text of three lines:

```
Deer Bear River
Car  Car River
Deer Car Bear
```

You may assume that we have 3 nodes that are used for generating the mapping. Each node is sent a line of text as shown. There are 4 nodes that do the reduction with the final result sent to an output node. Your goal is to compute the co-occurrences of all pairs of adjacent words. The *Count(A, B)* gives the number of occurrences of A and B in sequence. The pair (A, B) should be considered the same as (B, A) and as such counted together. Describe the process of using MapReduce technique, with an illustrative diagram similar to the one shown above, to generate the counts of the pairs of words for the text below. Assume the same node arrangement is used as in the case of a single word count.

```
Deer Car River Deer
Bear Car Car Deer River
Deer Car Bear
```

b) Big Data is typically characterized by 3Vs. What does this mean?

c) Briefly describe five technologies appropriate for processing big-data.

d) Give 4 characteristic features of NoSQL databases.