Lab: Parallel and Big Data

Test / Assignment Worksheet

In this assignment you will define key concepts and apply parallel programming skills to develop a parallel program to process spatial data. Parallel programming will be essential in the future as more and more computing systems are transformed from single core to multi-core or many-core architectures, which is further amplified as spatial data shifts from "traditional" spatial data to spatial big data.

Answer 1

Define spatial big data (Select the best answer):

A. Data that describes very big spatial objects such as buildings and bridges
B. Spatial data that exhibits at least one of the three V's
C. Spatial data that exhibit at least two of the three V's
D. The spatial location of big dataset

Answer: B

Answer 2

Spatial big data challenge CyberGIS in which ways (check all that apply)
A. Geovisualization
B. Network
C. Algorithms and Methods
D. Database
E. Computational systems

Answer: A,B,C,D,E (all of them)

Answer 3

With respect to the three V's of big data, which is most related to the speed at which data is generated
A. Volume
B. Variety
C. Velocity

Answer: C

Answer 4

With respect to the three V's of big data, which is most related to massive amounts of data
A. Volume
B. Variety
C. Velocity

Answer: A

Answer 5
Identify the two common parallel programming paradigms
A. Shared memory and message passing
B. Data parallelism and spatial domain decomposition
C. Message passing and functional parallelism

D. Data parallelism and functional parallelism

Answer: D

Answer 6
Amdahl's law is generally used to:
A. Determine the amount of data parallelism in a program
B. Predict the maximum theoretical speedup in a program
C. Predict the maximum message passing in a program
D. Determine the maximum amount of shared memory in a program

Answer: B

Answer 7
Write your own parallel operation in a file named lastname_localmaximum.py using PCML (replace lastname with your last name). Specifically, create a LocalMaximum operation that will return the maximum value at each location given one or more input layers. Refer to the paper written by C. Dana Tomlin for additional details on cartographic modeling operations. Refer to the pcml/lib/LocalOperationPrimitives.py file in the PCML directory or see the example LocalSum operation below for example operations implemented in PCML. Copy and paste your operation into the textbox below.

```
from PCML import *
@localoperation

def LocalSum(self, locations, subdomains):

    # LocalSum will return the sum of values at each location for two+ layers.

    val = 0

    for loc in locations:

        val += loc['v']

    return val
```

Answer: Self graded

Answer 8
Within the file named lastname_localmaximum.py call your LocalMaximum operation using two input layers:
data/data_f.asc
data/data_m.asc
and save the results to the variable *layero* . Print out the data for layero using the *print_data()* method. Refer to examples.py for additional information on how to call a PCML operation with input layers and how to print the data of a layer. Copy and paste the results into the textbox below.

Answer: Self graded

Answer 9

Use the 'time' command to run your Python program.

```
time python lastname_localmaximum.py
```

The time command returns three different times. The real time it took to run your program, the user portion of the time, and the system portion of the time to run your program. We will focus on the real time it took to run your program. Now modify your program and set the following PCMLConfig.num_procs variable to be 1 before you call your LocalMaximum operation.

PCMLConfig.num_procs = 1

Time your Python program again and record the real time. Now change num_procs to 2 and time your Python program. Now change num_procs to 4 and time your Python program.
Answer 3 text boxes
Real time when num_procs = 1 : Textbox 1
Real time when num_procs = 2 : Textbox 2
Real time when num_procs = 4 : Textbox 3

Answer 10
Did using more processing cores speedup your parallel program? What problems could influence the speedup of your parallel program? Write 3-5 sentences in the textbox below.

Answer: Self graded