Xi'an Jiaotong-Liverpool University
西交利物浦大学

XJTLU Entrepreneur College (Taicang) Cover Sheet

Module code and Title	DTS205TC High Performance Computing
School Title	School of AI and Advanced Computing
Assignment Title	Assessment 2 – Lab Report
Submission Deadline	April 22nd, 2024 @ 23:59
Final Word Count	NA
If you agree to let the university use your work anonymously for teaching and learning purposes, please type **"yes"** here.	

I certify that I have read and understood the University's Policy for dealing with Plagiarism, Collusion and the Fabrication of Data (available on Learning Mall Online). With reference to this policy I certify that:

- My work does not contain any instances of plagiarism and/or collusion.
- My work does not contain any fabricated data.

By uploading my assignment onto Learning Mall Online, I formally declare that all of the above information is true to the best of my knowledge and belief.

Scoring – For Tutor Use	
Student ID	

Stage of Marking	**Marker Code**	**Learning Outcomes Achieved**					**Final Score**
		A	**B**	**C**	**D**	**E**	
1st Marker – red pen							
Moderation – green pen	**IM Initials**	The original mark has been accepted by the moderator (please circle as appropriate):					Y / N
		Data entry and score calculation have been checked by another tutor (please circle):					Y
2nd Marker if needed – green pen							

For Academic Office Use			**Possible Academic Infringement (please tick as appropriate)**	
Date Received	**Days late**	**Late Penalty**	☐ **Category A**	Total Academic Infringement Penalty (A,B, C, D, E, Please modify where necessary)
			☐ **Category B**	_____
			☐ **Category C**	
			☐ **Category D**	
			☐ **Category E**	

Assignment Project Quiz Exam Essay Help
WeChat: cestbon 6888
Email: accoder overseas@163.com

Xi'an Jiaotong-Liverpool University
西交利物浦大学

DTS205TC High Performance Computing

Individual Lab Report

Due: April 22nd, 2024 @ 23:59

Weight: 50%

Maximum Marks: 100 marks

Learning Outcome

This assessment tests your ability to:

A. Demonstrate understanding of the concepts used in modern processors for increasing the performance.

B. Demonstrate optimization techniques for serial code.

C. Understand and apply parallel computing paradigms.

D. Write optimized programs designed for high-performance computing systems.

Overview

The purpose of this assignment is to evaluate the student's learning outcome base on their weekly lab exercises. All tasks should be done individually.

Avoid Plagiarism

- Do ***not*** submit work from other teams/individuals.
- Do ***not*** share code/work to students other than your own team members.
- Do ***not*** read code/work from other teams/individuals, discussions between teams should remain high level.
- Do ***not*** use code from the Internet, Generative AI, or published textbooks.

1. Lab Reports

Lab 1 (15 marks)

In a hierarchical storage system, the cache hit rate has a significant impact on program performance. Different cache strategies will result in different cache hit ratios. Now, we have recorded 2 datasets, containing CPU access requests to memory for a period of time. They both have 10,000 items from addresses 0 to 63. We will simulate the process of the CPU reading and

Assignment Project Quiz Exam Essay Help
WeChat: cstbon_688
Email: accoder_overseas@163.com

caching data from the memory through a program. We care about how many cache hits occur under a particular policy.

Please run the program in the attachment to compare the hit rates of different strategies. Please answer:

1) Why are the hit rates of the two strategies different on the two data sets? Please perform a visual analysis of the data and explain why in text.

2) Please design and implement a strategy so that when the cache size is no greater than 5, it can achieve better results than the existing two strategies on the trace2 data set.

Lab 2 (15 marks)

By measuring the bandwidth and delay between each two directly connected nodes, we obtained three computer network's performance, which are saved in the attachment. Examples of bandwidth data are:

NaN	86	16	48	92
86	NaN	43	6	93
16	43	NaN	66	13
48	6	66	NaN	53
92	93	13	53	NaN

Among them, the data in the i-th row and j-th column represent the bandwidth from node i to node j. Examples of latency data are:

NaN	0.86	0.165	0.483	0.921
0.86	NaN	0.429	0.057	0.925
0.165	0.429	NaN	0.658	0.133
0.483	0.057	0.658	NaN	0.533
0.921	0.925	0.133	0.533	NaN

Among them, the data in the i-th row and j-th column represent the delay from node i to node j. Answers are now requested:

1) What type of topology are these networks? Please plot the network structure and judge its type.

2) For each network, program to calculate the shortest delay and maximum throughput from node 1 to node 5, and which nodes are passed through to achieve them? (Any programming language and algorithm library can be used)

Lab 3 (30 marks)

In statistics, bootstrapping[1] is a resampling method that involves repeatedly sampling with replacement from the original data set to estimate the distribution of a population parameter. This approach allows for the estimation of parameter characteristics without assuming the

[1] https://en.wikipedia.org/wiki/Bootstrapping_(statistics)

Assignment Project Quiz Exam Essay Help
WeChat: cestbon_688
Email: acoder_overseas@163.com

distribution of the population parameter. It is particularly useful when dealing with small sample sizes or unclear population distributions.

Below is a parallel version of the bootstrapping program implemented using mpi4py. Its purpose is to calculate the "variance of the sample mean". Formally speaking, we need to first generate a set of datasets by resampling with replacement $\{\widetilde{\mathrm{D}}^{(1)}, \widetilde{\mathrm{D}}^{(2)}, \ldots, \widetilde{\mathrm{D}}^{(\mathrm{N})}\}$, then take their mean $\widetilde{\mathrm{M}} = \{\widetilde{\mathrm{m}}^{(1)}, \widetilde{\mathrm{m}}^{(2)}, \ldots, \widetilde{\mathrm{m}}^{(\mathrm{N})}\}$, and then calculate their variance $\mathrm{Var}(\widetilde{\mathrm{M}}) = \mathrm{E}[(\widetilde{\mathrm{M}} - \mathrm{E}[\widetilde{\mathrm{M}}])^2]$.

```
import numpy as np
from mpi4py import MPI
import random

np.random.seed(205)

# environment info
comm = MPI.COMM_WORLD
R = comm.Get_rank()
P = comm.Get_size()

# number of tasks
N = 10
D = 100

assert N % P == 0

# resampling function
def bootstrap(data):
    samples = np.empty((0, data.shape[0]))
    for i in range(round(N / P)):
        samples = np.vstack((samples, np.random.choice(data, size=len(data), replace=True)))
    return samples

if R == 0:  # master

    # generate data -- this is only for test!
    data = np.arange(D)
```

Assignment Project Quiz Exam Essay Help
WeChat: cestbon_688
Email: accoder_overseas@163.com

```
    # send data
    for i in range(1, P):
        comm.send(data, dest=i)

    # receive samples
    samples = bootstrap(data)  # do my part
    for i in range(1, P):
        samples = np.vstack((samples, comm.recv(source=i)))

    result = np.var(np.mean(samples, axis=1))

    # output
    print(f'proc {R}:var={result}')

else:  # slave
    # recv data
    data = comm.recv(source=0)

    # resample
    samples = bootstrap(data)

    # send back
    comm.send(samples, dest=0)

    print(f'proc {R}:done')
```

The tasks of this experiment:

1) Estimate the communication time used by this procedure. Assume that the dataset is an array containing D pieces of float type data and needs to be resampled N times. The transmission delay is L=0 s, and the transmission bandwidth is B bytes/s. The program contains a total of P processes. N is always divisible by P.

2) Modify the part enclosed by the red box to minimize the amount of communication while ensuring the correctness of the results. Explain the mathematical basis for your modification, give the code, and give the new communication time calculation formula.

NOTE: Only estimation is required, no actual measurement is required; only point-to-point, non-blocking communication, i.e. send/recv, can be used.

Assignment Project Quiz Exam Essay Help
WeChat: cestbon_688
Email: ac_coder_overseas@163.com

Lab 4 (30 marks)

Continue with the previous experiment. When the data set is too large, due to the limited memory capacity of each process (assuming that it cannot store more than 2*D/P float values), they can each load/generate a part of the original data, and they cannot directly exchange them between processes. Only small amounts of intermediate results can be passed. Based on the previous experiment, we further assume that each process can complete S samples per second. It is assumed here that operations other than sampling take no time; D is always divisible by P.

Please redesign the program to shorten the overall running time as much as possible while ensuring the correctness of the results. Please explain the mathematical basis for your modification, and provide the code and running time calculation formula (including calculation and communication; if randomness exists, take the expectation).

NOTE: Only estimation is required, no actual measurement is required; only point-to-point, non-blocking communication, i.e. send/recv, can be used.

Lab 5 (10 marks)

In functional programming, the focus is on using functions to manipulate and transform data, rather than relying on modifying states or performing mutable operations. In python, similar style operations are also provided for lists, such as map, reduce, sorted and itertools. Since they do not modify individual elements but always create new lists, they can be safely parallelized and are ideal for big data analysis.

Now we have the data of a directed graph whose record format is a dict. The key is the node, and the value is the destination node pointed by the node through the edge. For example

```
{0: [1, 4, 6, 7, 12, 13, 16, 18, 19],
```

represents the existence of directed edges from node 0 to 1, 4, 6... Please complete the following program and print the 10 nodes that are pointed to the most.

NOTE: 'for' and 'while' loops cannot be used here, only map, reduce, sorted and itertools can be used to implement equivalent programs.

```
from pickle import load

with open("./links.pkl", 'rb') as f:
    links = load(f)
```

Assignment Project Quz Exam Essay Help
WeChat: cestbon688
Email: accoder_overseas@163.com

```
# ============= ========================

#                    insert your code here

# ============= ========================

# output top 10 nodes
# linkin_count: an ordered list of tuples that represent (node id, times of linked in)
for node, count in linkin_count[:10]:
    print(f"Page {node} is referred {count} times")
```

2. Submission

You must submit the following files to LMO:

1) A report named as Your_Student_ID.pdf.
2) A directory containing all your source code, named as Your_Student_ID_code.

NOTE: The report shall be in A4 size, size 11 font, and shall not exceed **8** pages in length. You can include only key code snippets in your reports. The complete source code can be placed in the attachment.

Assignment Project Quiz Exam Essay Help
WeChat: cestbon_6822
Email: accoder_overseas@163.com

Xi'an Jiaotong-Liverpool University
西交利物浦大學

Assignment Project Quiz Exam Essay Help
WeChat: cestbon_688
Email: accoder_overseas@163.com

Xi'an Jiaotong-Liverpool University
西交利物浦大学

Appendix

A. MARKING CRITERIA

The tasks in this assessment can be divided into 3 categories:

- ✓ Charts Presentation & Analysis;
- ✓ Essay;
- ✓ Programs.

Criteria(%)	Exemplary (100)	Good (75)	Satisfactory (50)	Limited (25)	Very Limited (0)
Programs	Demonstrated correctly implemented code that produces correct output. Excellent coding quality follows best practices.	The program runs correctly and gives the expected results. However, special cases are not fully considered, or the program performs redundant calculations.	Program basically works correctly for major functionality, however, with some conceptional problems.	The program implements some minor functionality, or incorrectly implements major functionality. There is a certain degree of misunderstanding about the requirements of the questions.	Program works incorrectly with limited attempt or irrelevant to the task.
Charts Presentation & Analysis	Excellent quality of report with clear structure, clear logic, concise writing, pleasing visual aids.	Most of the results in the chart are correct, but there is a certain degree of sloppy or wordy in the overview and analysis.	Moderate quality of report with basic structure, where writing and visual aids can be improved.	Only some of the results in the chart are correct, or some of them are not filled in. The analysis of the results was obviously biased.	Limited or no attempt of report.
Essay	Provides a detailed, accurate description of the methods. Provide comprehensive comparison between the methods,	The analysis provided demonstrates that the student's understanding of the various methods is correct and that they have	Provides adequate description of the methods. Comparison is provided with some level of details, however, with some obvious	There are obvious deviations in the understanding of the main methods, and it fails to reflect the ability to independently design algorithms. The	Limited or no description of methods. Limited comparison provided.

Assignment Project Quiz Exam Essay Help
WeChat: cestbon688
Email: accouter overseas@163.com

	including pros and cons, performance analysis.	the ability to solve problems independently. Although there are certain flaws, or incomplete.	mistakes.	description of the problem is vague, or the thought is incomplete.	

Assignment Project Quiz Exam Essay Help
WeChat: cestbon_688
Email: accoder_overseas@163.com