



HIGH PERFORMANCE AND PARALLEL COMPUTING REPORT

CAB401: High Performance and Parallel Computing

OCTOBER 27, 2017
MICHAEL CARTWRIGHT
N8869871

Table of Contents

Paragraph	Page No.
1 Introduction	1
1.1 Software Architecture.....	1
1.1.1 Time	1
1.1.2 Initialise	1
1.1.3 Input Filename and Read Input File	1
1.1.4 Compare All Bacteria	2
1.1.5 Bacteria Class	3
1.1.6 Compare Bacteria	5
2 Bioinformatics – Genome Similarity Using Frequency Vectors	7
2.1 Software, Compiler, Hardware and Techniques.....	7
2.1.1 Software	7
2.1.2 Compiler.....	7
2.1.3 Hardware	7
2.1.4 Techniques	7
2.2 Analysis of Code and Parallelisation Process	8
2.2.1 Successful Parallelisation	8
2.2.2 Further Analysed Parallelisation Attempts	9
2.2.2.1 Two Functions at the Same Time in Main.....	9
2.2.2.2 Parallelisation Bacteria Class	9
2.2.2.3 Parallelisation Comparing Two Bacteria	10
3 Code Implemented.....	11
3.1 Code Changes in Compare All Bacteria Function.....	11
4 Results	15
4.1 Time Taken for Execution of Sequential and Parallelised Applications.....	15
4.2 Profiling Reports	15
4.3 Speed Up Graph	17
5 Reflection	18
5.1 Issues Encountered	18
6 Appendix	20
6.1 Original Code	20
6.2 High Performance and Parallel Code.....	25

6.3	Detailed Software Architecture.....	30
6.4	Bacteria Class – Part 1 Software Architecture	31
6.5	Bacteria Class – Part 2 Software Architecture	32
6.6	Compare Bacteria Software Architecture	33

List of Figures

Figure 1: Basic View of Software Architecture	1
Figure 2: Read Input File Software Architecture	2
Figure 3: Compare All Bacteria Software Architecture	2
Figure 4: AcMNPV Animo Acid File Content Example	3
Figure 5: Get Data from *.faa File	3
Figure 6: Calculate Variables in Bacteria Class – Part 1	4
Figure 7: Other Values Calculated in Bacteria Class	4
Figure 8: Bacteria Class Ending.....	5
Figure 9: Detail View of Compare Bacteria	6
Figure 10: Total CPU (ms, %) of Functions in Bioinformatics – Genome Similarity Using Frequency Vectors Application	6
Figure 11: Original Code for Compare All Bacteria Function	11
Figure 12: Parallelised Code for Compare All Bacteria Function	11
Figure 13: Original Loading Bacteria.....	12
Figure 14: Parallel Loading Bacteria Four Threads Max	12
Figure 15: Parallel Loading Bacteria Eight Threads Max	12
Figure 16: Sequential Application Memory and CPU Usage	13
Figure 17: Parallelised Application Memory and CPU Usage	13
Figure 18: Parallelised Nested For-Loop in Order.....	14
Figure 19: Correlation in Order.....	14
Figure 20: Correlation not in Order.....	14
Figure 21: Sequential CPU Profiler	16

Figure 22: Parallelised CPU Profiler.....	16
Figure 23: Speed Up Graph of Sequential vs Parallelised Over Threads	17
Figure 24: Restructured Nested For-Loop Attempt.....	18

List of Tables

Table 1: Sequential and Different Thread Count for Parallelised Application	
---	--

Run-Time	15
----------------	----