LOUISIANA STATE UNIVERSITY - SHREVEPORT

INDEPENDENT STUDY

Independent Study in Cluster Computing and Virtualization

Authors:

Jay JAIN
Blaise LEONARDS
Joshua SCHOOLS

Supervisor:
Dr. Alfred MCKINNEY

A document submitted in fulfillment of the requirements for CSC 495 (Independent Study)

in the

Department of Computer Science

November 12, 2017

"For over a decade prophets have voiced the contention that the organization of a single computer has reached its limits and that truly significant advances can be made only by interconnection of a multiplicity of computers. "

Gene Amdahl

LOUISIANA STATE UNIVERSITY - SHREVEPORT

Abstract

Dr. Alfred McKinney
Department of Computer Science

Independent Study

Independent Study in Cluster Computing and Virtualization

by
Jay JAIN
Blaise LEONARDS
Joshua SCHOOLS

Understanding the fundamentals of parallel computing in the age of big data is integral to a comprehensive computer science education. Over the course of the Fall 2017 semester, major mathematical and programming concepts were identified and implemented on a self-built Raspberry Pi cluster. The cluster consisted of a master node (Raspberry Pi 3) and four slave nodes (Raspberry Pi Zeros). Using the `Python` message passing interface (`mpi4py`), the value of pi was calculated using the Monte Carlo method at different iterations. The latency, throughput, and efficiency of the parallel processes were compared. Second, the `Docker` virtualization interface was used to automate parallel processes in big data processing using `Apache Hadoop` and image analysis using the `scikit-learn` package in `Python`.

Acknowledgements

The authors would like to thank Dr. McKinney for giving us the opportunity to participate in this independent study and his enduring kindness. We would also like to thank Mr. Phillip C.S.R. Kilgore for his words of encouragement and guidance through the project. Lastly, we would like to thank Dr. Cvek and Dr. Trutschl for allowing us to house our Raspberry Pi cluster in their laboratory.

Contents

List of Abbreviations

MPI	**M**essage **P**assing **I**nterface
RPi	**R**aspberry **Pi**

Chapter 1

Background

1.1 Introduction

The idea for this independent study came from multiple conversations by the authors regarding the lack of a parallel or cluster computing course at the university. Subsequently, we have realized the importance of large scale processing in the age of big data. Understanding and analyzing big sets of data is a necessity in an age where almost all electronic devices (large and small) are connected to a network and feeding in trillions of data points per second.

We decided to utilize a set of Raspberry Pi computers as a small-scale example of a cluster computing platform. Initially, we thought that utilizing commodity hardware such as servers for our cluster would be the way to go. Unfortunately, we do not have access to such hardware, but after further research, we concluded that the Raspberry Pi would be sufficient to demonstrate concepts in parallel and cluster computing.

1.2 Raspberry Pi

The Raspberry Pi is a low-cost, single-board computer that is capable of running light-weight GNU/Linux operating systems such as Raspbian. The Raspberry Pi utilizes the system on a chip (SoC) architecture in order to integrate components of the computer such as the microprocessor, graphics processing unit, and WiFi device. The Raspberry Pi is fairly easy to configure as there is a large, open-source community willing to help with setup issues. Additionally, the Raspberry Pi Foundation has put in a lot of effort to provide detailed and up-to-date documentation.

1.2.1 ARM Architecture

The Raspberry Pi central processing unit utilizes the ARM Architecture developed by the ARM Holdings. The ARM architecture paradigm is extremely relevant as there have been over 100 billion devices produced (as of 2017) that utilize the ARM instructuction set architecture. Most handheld devices including iPads and gaming consoles utilize ARM cores.

ARM, also known as Advanced RISC (Reduced Instruction Set Computing) Machine requires less transistors than x86 processors (usually found in personal computers) which make it an attractive candidate for embedded systems seeking to lower costs and energy consumption on devices. The ARM architecture supported only a core-bit width of 32, but the newest version of ARM, ARMv8 now supports both 32 and 64 bits as of 2011. Note, the Raspberry Pi utilizes ARMv7. ARMv7 adheres to the load/store architecture which separates instructions into memory access and ALU (Arithmetic Logic Unit) operations. Memory access is simply the process of transfering data from the memory to registers. ALU operations consist of the actual operations on the loaded memory.

1.2.2 Raspberry Pi 3

The Raspberry Pi 3 Model B consists of a quad-core 64-bit CPU with one gigabyte of RAM in the system on a chip configuration. Additionally, the RPi 3 is equipped with wireless LAN and Bluetooth capabilities. It is important to note that while the WiFi interface is speedy enough for basic internet usage, it does cause considerable bottlenecks when utilizing a parallel processing interface. In fact, the 802.11n WiFi speeds were clocked around 45 Mbits/second, while the USB Gigabit LAN clocked around 320 Mbits/second. Configuring USB Gigabit LAN with a Raspberry Pi cluster does take a considerable amount of time to configure as MAC addresses and USB devices have to be specifically assigned in order to ensure a functional RPi cluster.

1.2.3 Raspberry Pi Zero

1.3 Message Passing Interface (MPI)

If you are familiar with LaTeX, then you should explore the directory structure of the template and then proceed to place your own information into the *THESIS INFORMATION* block of the `main.tex` file. You can then modify the rest of this file to your unique specifications based on your degree/university. Section **??** on page **??** will help you do this. Make sure you also read section **??** about thesis conventions to get the most out of this template.

If you are new to LaTeX it is recommended that you carry on reading through the rest of the information in this document.

Before you begin using this template you should ensure that its style complies with the thesis style guidelines imposed by your institution. In most cases this template style and layout will be suitable. If it is not, it may only require a small change to bring the template in line with your institution's recommendations. These modifications will need to be done on the `MastersDoctoralThesis.cls` file.

1.3.1 mpi4py

1.4 Docker

1.4.1 References

The `biblatex` package is used to format the bibliography and inserts references such as this one (**Reference1**). The options used in the `main.tex` file mean that the in-text citations of references are formatted with the author(s) listed with the date of the publication. Multiple references are separated by semicolons (e.g. (**Reference2**; **Reference1**)) and references with more than three authors only show the first author with *et al.* indicating there are more authors (e.g. (**Reference3**)). This is done automatically for you. To see how you use references, have a look at the `Chapter1.tex` source file. Many reference managers allow you to simply drag the reference into the document as you type.

Scientific references should come *before* the punctuation mark if there is one (such as a comma or period). The same goes for footnotes[1]. You can change this but the

[1] Such as this footnote, here down at the bottom of the page.

most important thing is to keep the convention consistent throughout the thesis. Footnotes themselves should be full, descriptive sentences (beginning with a capital letter and ending with a full stop). The APA6 states: “Footnote numbers should be superscripted, [...], following any punctuation mark except a dash.” The Chicago manual of style states: “A note number should be placed at the end of a sentence or clause. The number follows any punctuation mark except the dash, which it precedes. It follows a closing parenthesis.”

The bibliography is typeset with references listed in alphabetical order by the first author’s last name. This is similar to the APA referencing style. To see how LaTeX typesets the bibliography, have a look at the very end of this document (or just click on the reference number links in in-text citations).

A Note on bibtex

The bibtex backend used in the template by default does not correctly handle unicode character encoding (i.e. "international" characters). You may see a warning about this in the compilation log and, if your references contain unicode characters, they may not show up correctly or at all. The solution to this is to use the biber backend instead of the outdated bibtex backend. This is done by finding this in `main.tex`: *`backend=bibtex`* and changing it to *`backend=biber`*. You will then need to delete all auxiliary BibTeX files and navigate to the template directory in your terminal (command prompt). Once there, simply type `biber main` and biber will compile your bibliography. You can then compile `main.tex` as normal and your bibliography will be updated. An alternative is to set up your LaTeX editor to compile with biber instead of bibtex, see here for how to do this for various editors.

1.4.2 Tables

Tables are an important way of displaying your results, below is an example table which was generated with this code:

```
\begin{table}
\caption{The effects of treatments X and Y on the four groups studied.}
\label{tab:treatments}
\centering
\begin{tabular}{l l l}
\toprule
\tabhead{Groups} & \tabhead{Treatment X} & \tabhead{Treatment Y} \\
\midrule
1 & 0.2 & 0.8\\
2 & 0.17 & 0.7\\
3 & 0.24 & 0.75\\
4 & 0.68 & 0.3\\
\bottomrule\\
\end{tabular}
\end{table}
```

You can reference tables with `\ref{<label>}` where the label is defined within the table environment. See `Chapter1.tex` for an example of the label and citation (e.g. Table 1.1).

TABLE 1.1: The effects of treatments X and Y on the four groups studied.

Groups	Treatment X	Treatment Y
1	0.2	0.8
2	0.17	0.7
3	0.24	0.75
4	0.68	0.3

1.4.3 Figures

There will hopefully be many figures in your thesis (that should be placed in the *Figures* folder). The way to insert figures into your thesis is to use a code template like this:

```
\begin{figure}
\centering
\includegraphics{Figures/Electron}
\decoRule
\caption[An Electron]{An electron (artist's impression).}
\label{fig:Electron}
\end{figure}
```

Also look in the source file. Putting this code into the source file produces the picture of the electron that you can see in the figure below.

FIGURE 1.1: An electron (artist's impression).

Sometimes figures don't always appear where you write them in the source. The placement depends on how much space there is on the page for the figure. Sometimes there is not enough room to fit a figure directly where it should go (in relation to the text) and so LaTeX puts it at the top of the next page. Positioning figures is the job of LaTeX and so you should only worry about making them look good!

Figures usually should have captions just in case you need to refer to them (such as in Figure 1.1). The `\caption` command contains two parts, the first part, inside the square brackets is the title that will appear in the *List of Figures*, and so should be short. The second part in the curly brackets should contain the longer and more descriptive caption text.

The `\decoRule` command is optional and simply puts an aesthetic horizontal line below the image. If you do this for one image, do it for all of them.

LaTeX is capable of using images in pdf, jpg and png format.

1.4.4 Typesetting mathematics

If your thesis is going to contain heavy mathematical content, be sure that LaTeX will make it look beautiful, even though it won't be able to solve the equations for you.

The "Not So Short Introduction to LaTeX" (available on CTAN) should tell you everything you need to know for most cases of typesetting mathematics. If you need more information, a much more thorough mathematical guide is available from the AMS called, "A Short Math Guide to LaTeX" and can be downloaded from: `ftp://ftp.ams.org/pub/tex/doc/amsmath/short-math-guide.pdf`

There are many different LaTeX symbols to remember, luckily you can find the most common symbols in The Comprehensive LaTeX Symbol List.

You can write an equation, which is automatically given an equation number by LaTeX like this:

```
\begin{equation}
E = mc^{2}
\label{eqn:Einstein}
\end{equation}
```

This will produce Einstein's famous energy-matter equivalence equation:

$$E = mc^2 \tag{1.1}$$

All equations you write (which are not in the middle of paragraph text) are automatically given equation numbers by LaTeX. If you don't want a particular equation numbered, use the unnumbered form:

```
\[ a^{2}=4 \]
```

1.5 Sectioning and Subsectioning

You should break your thesis up into nice, bite-sized sections and subsections. LaTeX automatically builds a table of Contents by looking at all the `\chapter{}`, `\section{}` and `\subsection{}` commands you write in the source.

The Table of Contents should only list the sections to three (3) levels. A `chapter{}` is level zero (0). A `\section{}` is level one (1) and so a `\subsection{}` is level two (2). In your thesis it is likely that you will even use a `subsubsection{}`, which is level three (3). The depth to which the Table of Contents is formatted is set within `MastersDoctoralThesis.cls`. If you need this changed, you can do it in `main.tex`.

1.6 In Closing

You have reached the end of this mini-guide. You can now rename or overwrite this pdf file and begin writing your own `Chapter1.tex` and the rest of your thesis. The easy work of setting up the structure and framework has been taken care of for you. It's now your job to fill it out!

Good luck and have lots of fun!

Guide written by —
Sunil Patel: www.sunilpatel.co.uk
Vel: LaTeXTemplates.com

Appendix A

Frequently Asked Questions

A.1 How do I change the colors of links?

The color of links can be changed to your liking using:
`\hypersetup{urlcolor=red}`, or
`\hypersetup{citecolor=green}`, or
`\hypersetup{allcolor=blue}`.
If you want to completely hide the links, you can use:
`\hypersetup{allcolors=.}`, or even better:
`\hypersetup{hidelinks}`.
If you want to have obvious links in the PDF but not the printed text, use:
`\hypersetup{colorlinks=false}`.