Specification report for project 'Raspberry Pi Computing Cluster'

Author: Ben Hague (201146260)

Project Supervisor: Dr James Walsh

Project Assessor: Dr Mohammad Hasan

Department of Electrical Engineering and Electronics

12 October 2018

Abstract

Cluster Computing is used widely in industry for computing large quantities of data in what would be an unrealistic timescale for an individual normal computer. During the last 5 years a massive decrease in price has led to a massive increase in availability of small low power single board computers. Combining the technologies used to process large data with clusters and cheaper single board computers there is a potential to substantially lower the cost of processing large data. This report outlies the specification for a project investigating how a cluster could be constructed and investigate the benefits and drawbacks of using lower power hardware to conduct calculations on large quantities of data.

Declaration

I confirm that I have read and understood the University's definitions of plagiarism and collusion from the Code of Practice on Assessment. I confirm that I have neither committed plagiarism in the completion of this work nor have I colluded with any other party in the preparation and production of this work. The work presented here is my own and in my own words except where I have clearly indicated and acknowledged that I have quoted or used figures from published or unpublished sources (including the web). I understand the consequences of engaging in plagiarism and collusion as described in the Code of Practice on Assessment (Appendix L).

Contents

Introduction

Cluster computing is used on a day to day bases to compute large amounts of data in far less time than possible by normal methods. It is used to split tasks into sets of smaller tasks that can be executed concurrently. This report outlines a rough timescale for each part of the project to produce a small computing cluster, the overall goals of the project and how the project will be validated.

Project Description

The aim of the project is to produce a cluster of Raspberry Pi computers capable of executing plasma code in a suitable period. A key objective of the project are to investigate the economic effectiveness of using arrays of cheaper single board computers above using a more expensive workstations or even traditional clusters. Over the project it will be important to ensure that software and hardware are correctly configured and working effectively.

Project Specification

The project has been divided into 4 separate categories, Hardware, Software, Benchmarking

and Paperwork. Each of these categories is a distinct deliverable which must be done to achieve success in the project. Within each of the Deliverables there is a small number of work packets which relate to the goals which must be completed for the section to succeed.

The outcome of this Project is to create a 6 node 24 Core Cluster out of Raspberry Pi 3 units, the cluster should also be able to run Plasma Simulation code from the Plasma Theory and Simulation Group (PTSG). The cluster should also be able to demonstrate the speed difference between a single node and a full cluster and show how this compares to using a traditional workstation.

Specification and aims for each deliverable:

- Hardware - The outcome of this deliverable is to create a 6 node 24 Core Cluster out of Raspberry Pi 3 units.
    - Construct Hardware Network
    - Design Casing
    - Construct Casing for Cluster

- Software – The outcome of this deliverable is to use an MPI to allow the cluster to work together and execute code, the cluster should also be able to run Plasma Simulation code from the Plasma Theory and Simulation Group (PTSG) [1]
  - Build Linux Software for network
  - Set up and install MPI software
  - Compile Benchmarking code
- Benchmarking – The outcome of this deliverable is to Test, Benchmark and improve the cluster built. The cluster should also be able to demonstrate the speed difference between a single node and a full cluster and show how this compares to using a traditional workstation.
  - Construct Benchmarking Tests
  - Conduct Cluster Benchmarking
  - Conduct Single Node Benchmarking
  - Conduct Workstation Benchmark
- Paperwork – The outcome of this deliverable is to ensure that all the correct permissions and orders have been put forward at the correct time to allow the other parts of the project to be conducted in good time.
  - Project Forms (Appendix 3,4,5)
  - Specification Report
  - Order Form
  - Order Casing
  - Poster
  - Report

Methodology

The cluster is going to be based on the Beowulf style cluster computer. This means that one node is dedicated as the master node, the others are all designated as slave nodes. The master node uses a Message Passing Interface (MPI) to transfer data and commands to the slave nodes [2]. This makes one of the key tasks to install and set up a working MPI instance.

Following this, the next important set up will be to compile the code written by PTSG [1]. This has been done before for raspberry pi [3]. It is important to ensure that the codes are compiled compatible with the MPI and able to effectively distribute the load needed and provide a higher level of processing power.

Project Plan

The Project has been broken down into discrete work packets which appear on the Gantt chart shown in Appendix 1. Each packet of work has been scheduled with an appropriate deadline making it easy to identify the appropriate workload to be doing at any given time.

Work Packets and Deadlines

- Identify software dependencies and requirements by 31st October 2018
- Constructed hardware network by the 30th November 2018
- Build Linux software for network by the 30th November 2018
- Set up a working MPI with 6 nodes by 17th December 2018
- Compile Plasma Software for benchmarking by 31st January 2019
- Begin benchmarking the cluster in February 2019
- Calculate baseline benchmark from Single Raspberry Pi by 28th February 2019
- Calculate baseline benchmark from PC by 15th March 2019
- Construct casing for Raspberry Pi cluster by 1st April 2019

During this period as work is done a log will be kept of how work is going, any key discoveries, obstacles and solutions. Documentation for each section of the project will be made to produce a user manual as to how to use the cluster at the end of operation.

The project has been planned and split up in a work breakdown structure (WBS), a Gantt chart and network diagram is shown in Appendix 1 and 2 respectively.

Project Rationale and Industrial Relevance

This project aims to show how we could reduce the power consumption and therefore relative running costs associated with using cluster computers. Regardless of power consumption the project also shows that a lower cost cluster computing set up, would allow learning of how-to code for large numbers of nodes without using expensive and highly demanded supercomputer time. Problems which scale well for parallel are known as embarrassingly parallel examples of these are processes such as Monte Carlo Simulations or Simulations where the next calculation does not depend on the result of the previous one. Using this cluster to explore the suitability of small computing clusters when calculating plasma simulations will be useful those people who are conducting research in the plasma field.

Literature Review

During the research period of this project I have investigated basic principles and different types of cluster setups how they are used and what the advantages are [4]. I also specifically researched instances where Raspberry Pi clusters have been made for different uses. I specifically noted a conference that happening in Cambridge, USA where they discussed the use of a cluster for detecting anomalies in power grids. [5] The algorithms used for this application happened to be extremely scalable and led them to the conclusion that using an 8 node cluster matched the performance of a workstation for a significantly lower cost whilst using under half the power.

Conclusion

This project aims to construct a cluster of single board low power computers. The idea behind doing this is to show how cluster computers work and understand how to use them effectively. The end goal is to run plasma codes with higher efficiency then when a PC or traditional cluster is used. Each of the key deliverables has defined requirement for success and each of the work packets has been analysed and given a suitable time period and deadline. This is shown in the Network diagram in Appendix 2.

References

[1] Plasma Theory and Simulation Group, Electrical Computer Engineering Department, Michigan State University, "Plasma Theory and Simulation Group," [Online]. Available: https://ptsg.egr.msu.edu/.

[2] D. Nath, "Running an MPI Cluster within a LAN," MPI Tutorial, 2015. [Online]. Available: http://mpitutorial.com/tutorials/. [Accessed 08 10 2018].

[3] Particle In Cell Consulting LLC, "Fun with Raspberry Pi – plasma simulation code performance," Particle In Cell Consulting LLC, 16 05 2016. [Online]. Available: https://www.particleincell.com/2016/raspberry-pi-performance/. [Accessed 08 10 2018].

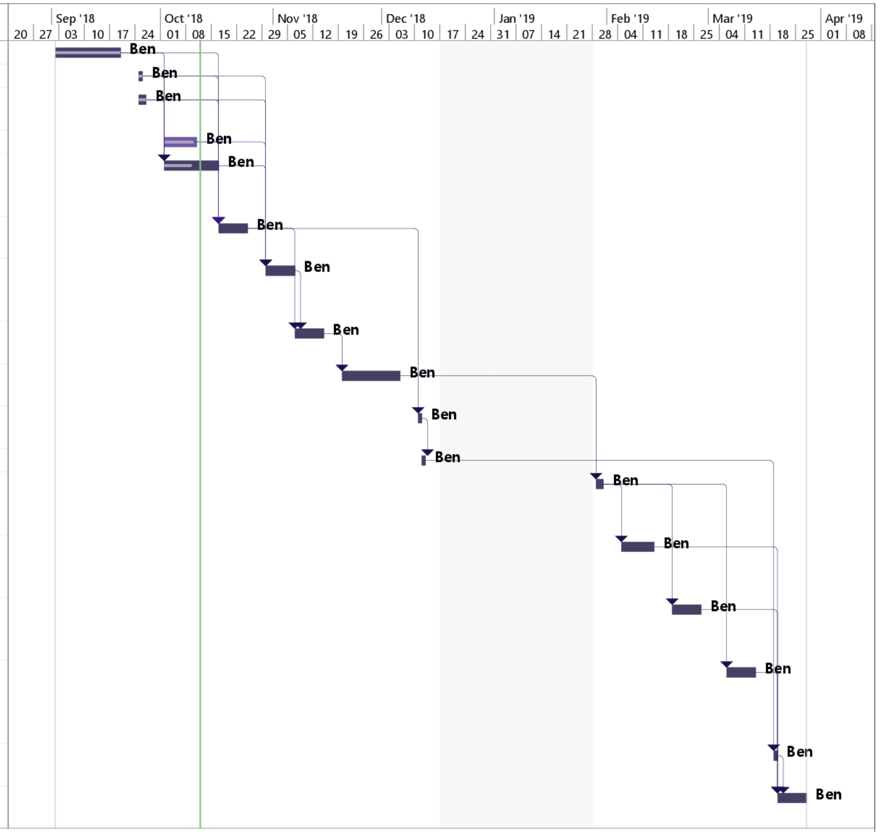
[4] The Camber Group, "Clustering: A basic 101 tutorial," IBM Developer Works, 03 04 2002. [Online]. Available: https://www.ibm.com/developerworks/aix/tutorials/clustering/clustering.html. [Accessed 08 10 2018].

[5] K. Candelario, C. Booth, A. S. Leger and S. J. Matthews, "Investigating a Raspberry Pi cluster for detecting anomalies in the smart grid," in *Investigating a Raspberry Pi cluster for detecting anomalies in the smart grid*, Cambridge, MA, USA, 2017.

Appendix

Appendix 1

ID		Task Mode	Task Name	Duration	Start	Finish	Predecessors	Resource Names
1			Research	6 days	Sun 02/09/18	Wed 19/09/1		Ben
2			Risk Assessment	1 day	Tue 25/09/18	Tue 25/09/18		Ben
3			Ethical Requirements	2 days	Tue 25/09/18	Wed 26/09/18		Ben
4			Requirements Form	4 days	Tue 02/10/18	Wed 10/10/1		Ben
7			Identify software dependencies and requirements	5 days	Tue 02/10/18	Tue 16/10/18	1,2,3	Ben
11			Construct Hardware Network	3 days	Wed 17/10/18	Wed 24/10/18	1,2,3,4	Ben
8			Build Linux software for network	3 days	Tue 30/10/18	Tue 06/11/18	7,2,3,4	Ben
9			Set up working MPI install	3 days	Wed 07/11/18	Wed 14/11/18	11,8	Ben
10			Compile code for Benchmarking	6 days	Tue 20/11/18	Wed 05/12/18	9	Ben
12			Design Case	1 day	Tue 11/12/18	Tue 11/12/18	11	Ben
5			Order Casing	1 day	Wed 12/12/1	Wed 12/12/1	12	Ben
14			Construct Benchmarking Tests	2 days	Tue 29/01/19	Wed 30/01/19	10	Ben
15			Conduct first round of Cluster Benchmarking	4 days	Tue 05/02/19	Wed 13/02/19	14	Ben
16			Carry out Benchmark on Single Node	3 days	Tue 19/02/19	Tue 26/02/19	14	Ben
17			Carry out Benchmark on PC for real life comparison	3 days	Wed 06/03/19	Wed 13/03/19	14	Ben
13			Construct Case	1 day	Tue 19/03/19	Tue 19/03/19	5	Ben
6			Design and Print Poster	3 days	Wed 20/03/19	Wed 27/03/19	13,15,16,17	Ben

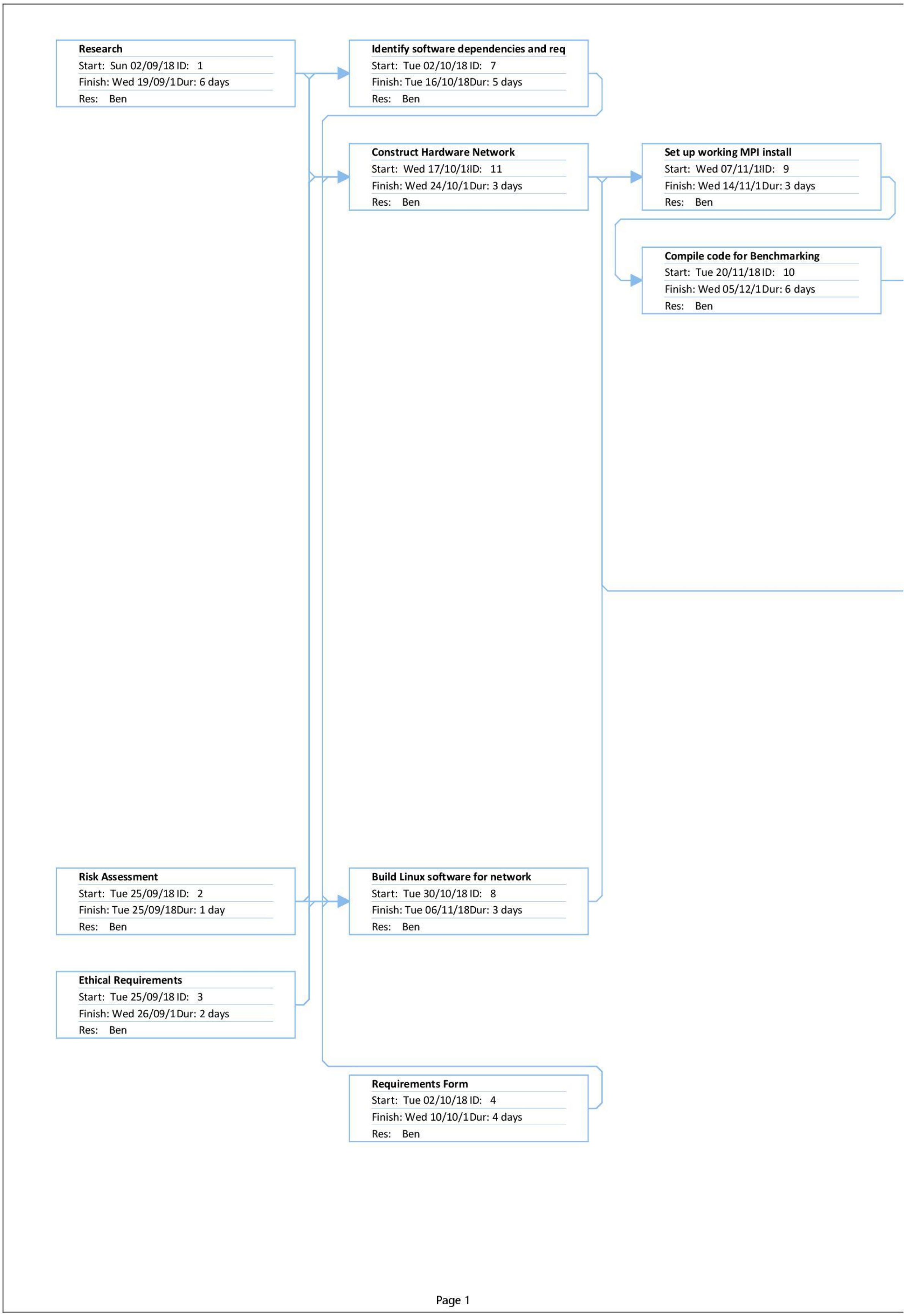


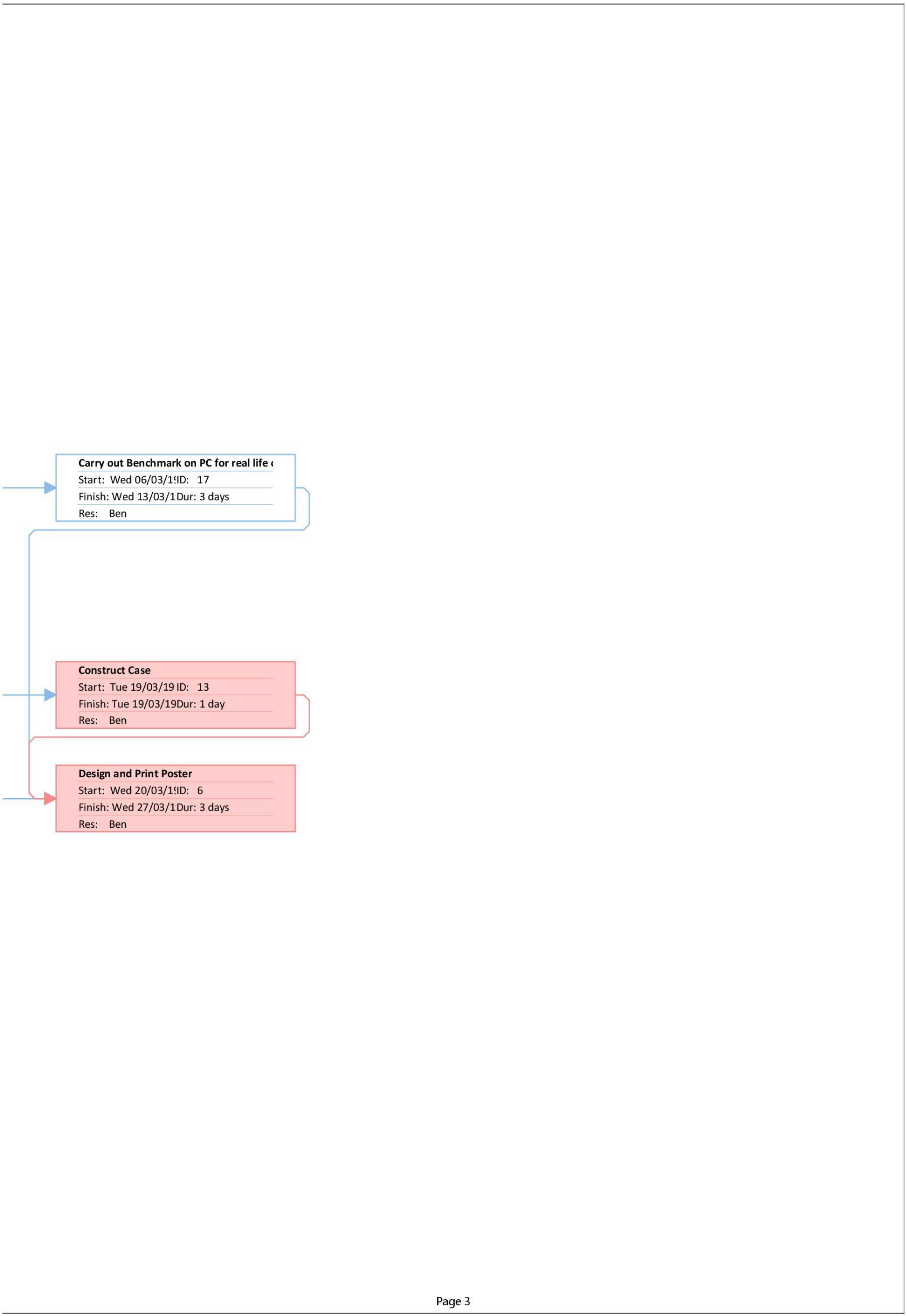
Project: Project Plan.mpp
Date: Fri 12/10/18

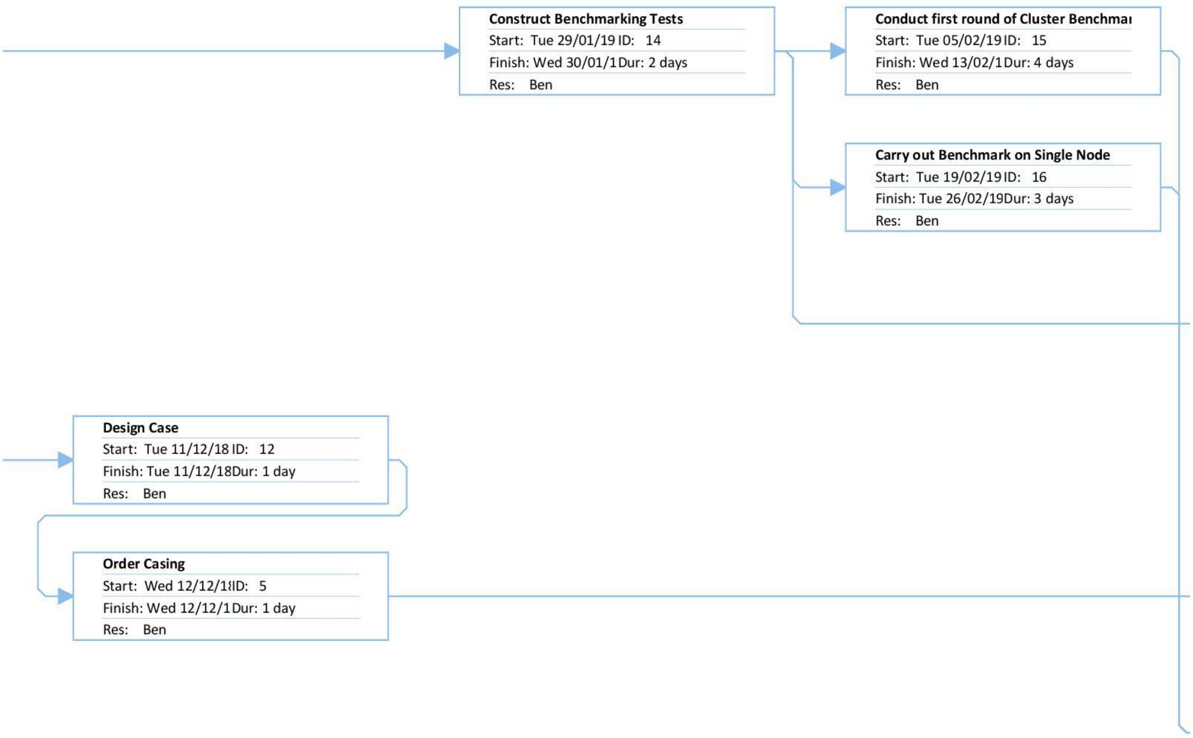
Task	Project Summary	Manual Task	Start-only	Deadline
Split	Inactive Task	Duration-only	Finish-only	Progress
Milestone	Inactive Milestone	Manual Summary Rollup	External Tasks	Manual Progress
Summary	Inactive Summary	Manual Summary	External Milestone	

Page 1

Appendix 2


Research
Start: Sun 02/09/18 ID: 1
Finish: Wed 19/09/1 Dur: 6 days
Res: Ben
Identify software dependencies and req
Start: Tue 02/10/18 ID: 7
Finish: Tue 16/10/18 Dur: 5 days
Res: Ben
Construct Hardware Network
Start: Wed 17/10/1 ID: 11
Finish: Wed 24/10/1 Dur: 3 days
Res: Ben
Set up working MPI install
Start: Wed 07/11/1 ID: 9
Finish: Wed 14/11/1 Dur: 3 days
Res: Ben
Compile code for Benchmarking
Start: Tue 20/11/18 ID: 10
Finish: Wed 05/12/1 Dur: 6 days
Res: Ben
Risk Assessment
Start: Tue 25/09/18 ID: 2
Finish: Tue 25/09/18 Dur: 1 day
Res: Ben
Build Linux software for network
Start: Tue 30/10/18 ID: 8
Finish: Tue 06/11/18 Dur: 3 days
Res: Ben
Ethical Requirements
Start: Tue 25/09/18 ID: 3
Finish: Wed 26/09/1 Dur: 2 days
Res: Ben
Requirements Form
Start: Tue 02/10/18 ID: 4
Finish: Wed 10/10/1 Dur: 4 days
Res: Ben
Page 1


Carry out Benchmark on PC for real life c
Start: Wed 06/03/1! ID: 17
Finish: Wed 13/03/1 Dur: 3 days
Res: Ben
Construct Case
Start: Tue 19/03/19 ID: 13
Finish: Tue 19/03/19 Dur: 1 day
Res: Ben
Design and Print Poster
Start: Wed 20/03/1! ID: 6
Finish: Wed 27/03/1 Dur: 3 days
Res: Ben
Page 3


Construct Benchmarking Tests
Start: Tue 29/01/19 ID: 14
Finish: Wed 30/01/1 Dur: 2 days
Res: Ben
Conduct first round of Cluster Benchmar
Start: Tue 05/02/19 ID: 15
Finish: Wed 13/02/1 Dur: 4 days
Res: Ben
Carry out Benchmark on Single Node
Start: Tue 19/02/19 ID: 16
Finish: Tue 26/02/19 Dur: 3 days
Res: Ben
Design Case
Start: Tue 11/12/18 ID: 12
Finish: Tue 11/12/18 Dur: 1 day
Res: Ben
Order Casing
Start: Wed 12/12/1 ID: 5
Finish: Wed 12/12/1 Dur: 1 day
Res: Ben

Appendix 3

UNIVERSITY OF LIVERPOOL

DEPARTMENT OF ELECTRICAL ENGINEERING AND ELECTRONICS

Ethical Approval Questionnaire 2017-2018

Final Year BEng (ELEC340) and Year 3 MEng (ELEC440)

Student Name: Benjamin Hague ___________________ Module: ~~ELEC340 /~~ ELEC440 (delete one)

Supervisor: Dr James Walsh ____________________ Student ID No: ___________ 201146260

Project Title: Raspberry Pie computing Cluster ____

Formal ethical approval must be obtained for all research projects '*involving research on human subjects or human tissues or databases of personal information to be carried out by University staff or students on University premises, or at any location, where there is no other acceptable provision for ethical consideration*'. Final year projects (ELEC340) and year 3 MEng projects (ELEC440) involving human participation must be undertaken in a way that safeguards the dignity, rights, health, safety, and privacy of those involved.

It should be noted that this policy covers **all** research methodologies including such activities as informal interviews, accessing personal files in an archive, or on-line data gathering. The requirement to obtain ethical review applies with equal force to projects undertaken by undergraduate students. For these projects, it is the responsibility of the supervisor to ensure that the ethical issues of the research are fully assessed and that formal ethical approval is obtained before the project commences.

Does your project involve any human participants (including situations where you are a participant as well as the investigator)?	~~YES~~	NO
Does your project involve any human tissues (including your own)?	~~YES~~	NO
Does your project involve any databases of personal information (including your own personal information)?	~~YES~~	NO
Does your project involve experiments using animals?	~~YES~~	NO

Delete either YES or NO on each line above

If any of the answers given above are YES then you, along with your project supervisor, must investigate the requirement to apply for ethical approval. Details of how to apply for ethical approval can be found at www.liv.ac.uk/intranet/research-support-office/research-ethics/ (for human participation) and at https://www.liv.ac.uk/research-integrity/biomedical-research/ (for use of animals).

Student Signature: ________________________ Date: 28-9-2018__________

Supervisor's Signature: ___ ______________________ Date: _28-9__________

Appendix 4

UNIVERSITY OF LIVERPOOL

RISK ASSESSMENT FORM

School/Department: EEE	**Building: EEE**
Task: Building Raspberry Pi Computer Cluster	
Persons who can be adversely affected by the activity: Only individual doing the work	

Section 1: Is there potential for one or more of the issues below to lead to injury/ill health (tick relevant boxes)

People and animals/Behaviour hazards

Allergies		Too few people		Horseplay		Repetitive action	☑	Farm animals	
Disabilities		Too many people		Violence/aggression		Standing for long periods		Small animals	
Poor training		Non-employees		Stress		Fatigue		Physical size, strength, shape	
Poor supervision		Illness/disease		Pregnancy/expectant mothers		Awkward body postures		Potential for human error	
Lack of experience		Lack of insurance		Static body postures		Lack of or poor communication		Taking short cuts	
Children		Rushing		Lack of mental ability		Language difficulties		Vulnerable adult group	

What controls measures are in place or need to be introduced to address the issues identified?

Identified hazards	CURRENT CONTROLS	RISK SCORE	ADDITIONAL CONTROLS REQUIRED (To include responsibilities and timescales)	RESIDUAL RISK SCORE
Repetitive action caused strains	Take Regular Breaks Ensure Good Posture	2	NA	2

Eye Strain	**Take Regular breaks**	**1**	**NA**	**1**

Section 2: Common Workplace hazards. Is there potential for one or more of the issues below to lead to injury/ill health (tick relevant boxes)

Fall from height		Poor lighting		Portable tools		Fire hazards		Chemicals		Asbestos	
Falling objects		Poor heating or ventilation		Powered/moving machinery		Vehicles		Biological agents		Explosives	
Slips, trips, falls		Poor space design		Lifting equipment		Radiation sources		Waste materials		Genetic modification work	
Manual handling		Poor welfare facilities		Pressure vessels		Lasers		Nanotechnology		Magnetic devices	
Display screen equipment		Electrical equipment	☑	Noise or vibration		Confined spaces		Gases		Extraction systems	
Temperature extremes		Sharps		Drones		Cryogenics		Legionella		Robotics	
Home working		Poor signage		Overseas work		Overnight experiments		Unusual events		Community visits	
Late/lone working		Lack of/poor selection of PPE		Night work		Long hours		Weather extremes		Diving	

Section 3: Additional hazards: are there further hazards NOT IDENTIFIED ABOVE that need to be considered and what controls are in place or needed? (list below)

Additional hazards	CURRENT CONTROLS	RISK SCORE	ADDITIONAL CONTROLS REQUIRED (To include responsibilities and timescales)	RESIDUAL RISK SCORE

What controls measures are in place or need to be introduced to address the issues identified?

Identified hazards	CURRENT CONTROLS	RISK SCORE	ADDITIONAL CONTROLS REQUIRED (To include responsibilities and timescales)	RESIDUAL RISK SCORE
Electric Shock	Only low voltages are to be used Avoid coming into direct contact with powered circuits	1	NA	1

Section 4: Emergency arrangements (List any additional controls that are required to deal with the potential emergency situation)

Emergency situation	Additional control required

Risk assessor (signature)............................. ....Date....... 28/09/18................ Authorised by
(signature)................. ....Date...............

Appendix 5

Student Name	**Ben Hague**	**Student ID Number**	**201146260**	**Contact Email** (In case of problems with orders etc)	**sgbhague@liv.ac.uk**
Project Title	**Raspberry Pie computing cluster**				
Supervisor	**Dr J Walsh**				
Order ID No.	**(Assigned by 4th floor Technicians on receipt of form)**				

This form is to be completed by the student in co-operation with their supervisor. On completion the form **must** be signed by the supervisor in the space provided on the rear on this form, and then returned to the Technicians. Forms must be submitted even for those students with nil requirements.

Type of project

Is the project completely software based?	~~YES~~/NO	Delete as appropriate
Is the project completely Hardware based?	~~YES~~/NO	Delete as appropriate
Is the project mixed software/Hardware based?	YES/~~NO~~	Delete as appropriate

Small parts/PCB manufacture etc.

You must NOT approach any of the departmental workshops directly. All such requests should be made through the 4th floor Technicians from whom further information can be obtained. Note – Considerable delays may be experienced regarding the manufacture of small parts &/or PCB's for student projects dependent upon the workshop work loading at the time of request submission.

Location

Please state any special requirement(s) that will affect the location of the project within the laboratory area. Software projects will NOT be allocated a specific location until shortly before the bench inspections. Details of the arrangements for the bench inspections will be issued during session 2017/18 semester 2

Additional equipment, other than standard laboratory equipment to be supplied by supervisor

Equipment / Components	Available from

Initial Start-up form 2018-2019

Supervisor Signature Date

No order will be processed without the supervisor's signature - Please allow for delivery delays both within the University and external. Especially if ordering non-standard components. These delays may be up to **2 weeks** or even longer if non Standard parts ordered.

Components							**Technician use only**		
	Supplier	**Order Code**	**Description**	**Unit Price**	**Qty**	**Total**	**O**	**R**	**C**
1	RS Online	137-3331	Raspberry Pi 3 Model B+	28.39	4	113.56			
2	RS Online	124-9640	16 GB SD Card	8.52	6	51.12			
3	RS Online	136-3019	8 Port Gigabit Ethernet Switch	20.92	1	20.92			
4	RS Online	411-368	Ethernet Patch Cable	1.92	7	13.44			
5	RS Online	135-4176	Plug In Power Supply	6.00	6	36.00			
6									
7									
8									
9									
10									
11									
12									

Note - Please give as much information as possible when ordering components. Orders for Amazon/Ebay or similar will be rejected

Comments

Page 2 of 2