CS 596: Introduction to Parallel Computing
Homework Comments 1:
Using the Student Cluster/Unix Ops/Timing.

Mary Thomas

Department of Computer Science
Computational Science Research Center (CSRC)
San Diego State University (SDSU)

Due: 09/09/14
Posted: 09/18/14
Updated: 09/18/14
Table of Contents

1 Homework #1 Overview: Running Using the student cluster
   - Homework #1a: Accessing the student cluster
   - Homework #1b: Homework directories
   - Homework #1c: Unix operations
   - Homework #1d: Pacheco example code
   - Homework #1e: Histogram program
   - Homework #1f: Timing the Histogram program

2 What to turn in

3 Comments
   - General Comments
   - Timing Histogram/Data Count Max
Homework #1 Overview

- Due: 09/09/14

This homework involves demonstrating that you can log onto the student cluster. Tasks:
  - obtain a user account and logon
  - create the correct homework directory structure
  - perform simple unix operations to get you familiar with the system
  - install the pacheco demo codes into your directories
  - compile, run, test, time a histogram program
Description: obtain a user account and logon using ssh

- Once you are confirmed in the class, we will create an account using your last name (or using first name initials + last name in the case of duplicate last names).
- This will usually be done after the second day of class.
- What to turn in: some image or evidence that you logged on: such as a listing of your directory on tuckoo
Homework #1: Getting cluster information:

- `[gidget:~] mthomas% nslookup tuckoo.sdsu.edu
  Server: 10.0.1.1
  Address: 10.0.1.1#53
  Non-authoritative answer:
  Name: tuckoo.sdsu.edu
  Address: 130.191.127.136

- The cluster is on the internal SDSU campus network with no external login allowed. You can access the cluster from any on campus machine, including the ROHAN Academic Computing system, rohan.sdsu.edu.

- You create the account using your WebPortal account. See:
  - Home Page: http://www-rohan.sdsu.edu/
  - Create Rohan Account:
    http://www-rohan.sdsu.edu/raccts.shtml
Remote SSH login

- Launch SSH terminal on your computer
- SSH onto rohan:
  
  ```
  %ssh rohanUserName@rohan.sdsu.edu
  ```

- SSH onto tuckoo:
  
  ```
  %ssh tuckooUserName@tuckoo.sdsu.edu
  ```

- locate the class homework source code directory:
  
  ```
  /cs596
  ```
Once you have account information, you will create a homework directory where I will look for all assignment material. You are free to create other directories for development and testing (e.g. dev), but these directories must only contain specific material for the assignment.

- log onto student cluster: tuckoo.sdsu.edu (130.191.127.136)
- create a homework directory call hw
- create a sub directory called hw1
- use unix command `chmod` to set the directory accesses so only you and the instructor can read/access the codes

What to turn in: evidence that you completed this: an image, listing, etc.
Homework #1c: Test these Unix operations:

- `cat /etc/motd`
  - Note 1: try these from your home directory
- `whoami, date, uname -a`
- `cd ~, pwd`
- `ls`, optional arguments `[-al, -R]`
- `mkdir`
- `chmod`, test arguments such as `[-R]`

Create, compile and run a serial "Hello USER" program in C or Fortran (where USER is your username).

Compiler commands: use the specialized parallel library compiler commands

- C code: `mpicc -o myprogram myprogram.c`
- C code: `mpif90 -o myprogram myprogram.f90`

where is the command installed?

% locate mpicc

What to turn in: evidence that you completed this: images, text file which contains the output captured, session output.
Homework #1d: Install the Pacheco demo codes

Description:

- tar files are located in /cs596/pacheco_examples
- you may want to read the Unix man pages or Web pages to learn about the commands for tar and gzip
- you will want to locate the histogram code in the IPP codebase (ch2)
- What to turn in: evidence that you completed this: directory listing.
run the code for different variables:

usage: ./histogram

< bin_count >< min_meas >< max_meas >< data_count >

use the following test cases (8 combinations):

- bin_count = [5, 20]
- min_meas = 1
- max_meas = [500, 3000]
- data_count = [50, 500]

What to turn in: evidence that you completed this:
screen image, or text copy of output.
Homework #1 Overview: Running Using the student cluster

Homework #1e: Histogram program

```
[mthomas@tuckoo pacheco]$ cd intro-par-pgmng-pacheco/
[mthomas@tuckoo intro-par-pgmng-pacheco]$ ls
```
```
total 876
```
```
dirwx----- 3 mthomas mthomas 4896 Mar 1 16:19 .
dirwx----- 4 mthomas mthomas 4896 Mar 4 11:22 ...
dirwx----- 8 mthomas mthomas 4896 Feb 14 2014 ipp-source
```
```
-rw------- 1 mthomas mthomas 880640 Oct 16 2012 ipp-source.tar
[mthomas@tuckoo intro-par-pgmng-pacheco]$ cd ipp-source
```
```
total 52
```
```
dirwx----- 8 mthomas mthomas 4896 Feb 14 2014 .
dirwx----- 3 mthomas mthomas 4896 Mar 1 16:19 ..
dirwx----- 2 mthomas mthomas 4896 Jan 19 2011 ch2
dirwx----- 2 mthomas mthomas 4896 Apr 27 13:57 ch3
dirwx----- 2 mthomas mthomas 4896 Apr 23 13:48 ch4
dirwx----- 3 mthomas mthomas 4896 Nov 1 2012 ch5
dirwx----- 2 mthomas mthomas 4896 Jan 15 2012 ch6
```
```
rw------- 1 mthomas mthomas 14515 May 26 2011 INDEX
```
```
dirwx----- 7 mthomas mthomas 4896 Feb 14 2014 ipp-source
```
```
rw------- 1 mthomas mthomas 1694 Jan 7 2011 README
[mthomas@tuckoo ipp-source]$ cd ch2
[mthomas@tuckoo ch2]$ ls
```
```
total 28
```
```
dirwx----- 2 mthomas mthomas 4896 Jan 19 2011 .
dirwx----- 6 mthomas mthomas 4896 Feb 14 2014 ..
rw------- 1 mthomas mthomas 8838 Jan 19 2011 histogram.c
[mthomas@tuckoo ch2]$ mpicc -o histogram histogram.c
[mthomas@tuckoo ch2]$ ./histogram 10 1 1500 198
1.000-150.900: XXXXXXX
150.900-300.800: XXXXXXXX
300.800-450.700: XXXXXXXX
450.700-600.600: XXXXXXXX
600.600-750.500: XXXXXXXX
750.500-900.400: XXXXXXXX
900.400-1050.300: XXXXXXXXXX
1050.300-1200.200: XXXXXXXX
1200.200-1350.100: XXXXXXXXXX
1350.100-1500.000: XXXXXXXXXX
[mthomas@tuckoo ch2]$ 
```
Homework #1f: Timing the Histogram code

- Time how long the code takes to run as a function of Problem Size:
  \[ \text{ProbSize} = \text{data\_count} \]
  
  Wallclock Time: \( T_{\text{wall}} \)

- use the following test cases:
  - \( \text{bin\_count} = 10 \)
  - \( \text{min\_meas} = 1 \)
  - \( \text{max\_meas} = 50000 \)
  - Vary \( \text{data\_count} = 10^n \), where \( n = 0, 1, 2, N_{\text{max}} \)
  - What is \( N_{\text{max}} \)? Explain the limit.

- Modify how the code prints out the results: you don’t need to
  "plot" the histogram

- You only need to printout the bins, ranges, and the count (not all
  the X’s).

- Plot your test results using excel, Matlab, by hand: \( T_{\text{wall}} \) vs \( \text{ProbSize} \)
Suggestions for timing:

- Use unix `date` function, but not very sensitive
- Use C/Fortran internal timer
- You can find example timer code in `/cs596/tools/code_timer.c`
Homework #1f: Timing - What to turn in

A table of your test data and a plot of the results. Include labels.

<table>
<thead>
<tr>
<th>ProbSize</th>
<th>$T_{wall}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td>10</td>
<td>0.2</td>
</tr>
<tr>
<td>100</td>
<td>0.5</td>
</tr>
<tr>
<td>1000</td>
<td>0.6</td>
</tr>
<tr>
<td>10000</td>
<td>0.9</td>
</tr>
</tbody>
</table>
Put homework into a directory:
/596/ < your_username > /hw/hw1
include the source code(s), compiled binaries
see sections above for what to include
Write a simple report (this can be TEXT, Word, PDF Doc).
See each section for what to turn in.
Turn in hard copy at start of class.

Once the submission timeline has closed
DO NOT CHANGE THE FILE TIMESTAMPS!
HW directory listing example

```
[mthomas@tuckoo hw_dir_ex3 ls -R

 total 12
 drwx------ 3 mthomas mthomas 4096 Aug 28 12:23 .
 drwx------ 2 mthomas mthomas 4096 Aug 28 12:22 ..
 drwx------ 5 mthomas mthomas 4096 Aug 28 12:22 username

./username:
 total 20
 drwx------ 5 mthomas mthomas 4096 Aug 28 12:23 .
 drwx------ 3 mthomas mthomas 4096 Aug 28 12:23 ..
 drwx------ 2 mthomas mthomas 4096 Aug 28 12:23 dev
 drwx------ 5 mthomas mthomas 4096 Aug 28 12:22 hw
 drwx------ 2 mthomas mthomas 4096 Aug 28 12:23 misc

./username/dev:
 total 8
 drwx------ 2 mthomas mthomas 4096 Aug 28 12:22 .
 drwx------ 5 mthomas mthomas 4096 Aug 28 12:22 ..

./username/hw:
 total 20
 drwx------ 5 mthomas mthomas 4096 Aug 28 12:23 .
 drwx------ 5 mthomas mthomas 4096 Aug 28 12:23 ..
 drwx------ 2 mthomas mthomas 4096 Aug 28 12:23 hw1
 drwx------ 2 mthomas mthomas 4096 Aug 28 12:23 hw2
 drwx------ 2 mthomas mthomas 4096 Aug 28 12:23 hw3

./username/hw/hw1:
 total 8
 drwx------ 5 mthomas mthomas 4096 Aug 28 12:23 ..
-rw-rw-r-- 1 mthomas mthomas 0 Aug 28 12:24 data1.dat
-rw-rw-r-- 1 mthomas mthomas 0 Aug 28 12:24 file1.c
-rw-rw-r-- 1 mthomas mthomas 0 Aug 28 12:24 file2.c

./username/hw/hw2:
 total 8
 drwx------ 2 mthomas mthomas 4096 Aug 28 12:23 .
 drwx------ 5 mthomas mthomas 4096 Aug 28 12:23 ..

./username/hw/hw3:
 total 8
 drwx------ 2 mthomas mthomas 4096 Aug 28 12:23 .
 drwx------ 5 mthomas mthomas 4096 Aug 28 12:23 ..

./username/misc:
 total 8
 drwx------ 2 mthomas mthomas 4096 Aug 28 12:23 .
 drwx------ 5 mthomas mthomas 4096 Aug 28 12:23 ..
[mthomas@tuckoo hw_dir_ex3 ls]
```
Your graded homework has a grading *rubric* sheet:
- It explains how the HW was graded and weighted
- It will include comments.

Copy scripts are being used – preserving creation data, recursive copy.

Most of you did fine on 1a-1d.
- Some had extra codes/directories in the hw dir.
- You can run any test cases you wish, but if the HW specifies certain test cases, be sure to run what them and turn them in.
- include source code.

Histogram timing: tricky
- Not all of you did the right cases
- Not all of you found the maximum value for *data_count*

Reduce paper:
- Please use two sided where possible
- Please use single line space for code printouts
- Use two or 3 columns if that works for code (with separation lines)
Maximum number of data count items due to precision.

Increase *data_count* by using:

- In most cases, program dies at $10^8$, but sometimes it dies at $10^9$
- Use *double* instead of *float* gets $10^9$
- Use *long long data_count* gets $10^9$
- Anything larger than $10^{10}$ causes seg-fault.
- T. Smith: removed "=" from\n  \( if data\_count \geq bin\_max \)