Moving along - due to 16 March 2018 23:59

B.K. Muïte and Joonas Puura

Chapter 1

Introduction

In this practical homework, the focus is on five areas:

- Broad knowledge of available scientific computing infrastructures
- Using the Rocket and Vedur clusters in Tartu
- Distinguishing between different types of parallel computers
- Submitting a job to a parallel computer that uses a queue job management system
- Comparing performance of BLAS libraries

Submission and grading of this homework

In order to get points for this homework you need to submit the following:

- Get an account on Tartu HPC cluster
- Answer questions with yellow background.

You should submit this homework as a LaTeX writeup. To do so, get an account at https://sharelatex.mt.ut.ee and share your writeup with us.
Chapter 2
Scientific computing infrastructures

You have access to several scientific computing infrastructures. As a member of the University of Tartu, for some of your academic work you should consider using:


Many of the methods introduced here will also apply on other remote scientific computing infrastructures. In Europe, resources are offered through PRACE ([http://www.prace-ri.eu/](http://www.prace-ri.eu/)). Other notable scientific computing infrastructures are

- XSEDE [https://www.xsede.org/](https://www.xsede.org/)

Some of the above scientific infrastructure providers also have separate agreements for commercial use, and some of the methods you will learn about are also applicable for running applications on computational and web hosting resource providers such as

- Alibaba [https://www.alibabacloud.com/](https://www.alibabacloud.com/)
- Amazon web services [https://www.aws.com](https://www.aws.com)
- Baidu cloud [https://cloud.baidu.com/](https://cloud.baidu.com/)
- Digital Ocean [https://www.digitalocean.com/](https://www.digitalocean.com/)
- Google [https://cloud.google.com/](https://cloud.google.com/)
- Rackspace [https://www.rackspace.com/](https://www.rackspace.com/)
- Tendo [https://tendo.lt/](https://tendo.lt/)
Many of the above infrastructure providers try to have documentation and online tutorials, which may be useful to you with some adaptation to where you are running the program.

**Homework question 1**
Find a cloud service provider not listed above, and not listed by anyone else in the course mattermost group and list it in the course mattermost group. Add a link to user documentation (in any language) that may be helpful if provided - if not indicate no user documentation. Also in your submission, indicate which cloud service provider you listed.

**ANSWER**

**COMMENT**

**GRADE**
Chapter 3

Getting an account

Get an account on the University of Tartu HPC cluster by sending an email to request one to support@hpc.ut.ee [https://www.hpc.ut.ee/en_US/web/guest/access](https://www.hpc.ut.ee/en_US/web/guest/access)

You can find answer to many of the questions by making use of search engine or for example making use of HPC center’s website.

3.1 Logging in

Homework question 2

a) Why is it not a good idea to run intensive jobs on the head node?
   ANSWER
   COMMENT
   GRADE

b) What is a batch scheduling system and why would you want to have one on a shared computing resource?
   ANSWER
   COMMENT
   GRADE

c) How might job priorities differ on the supercomputers, Rocket/Vedur (capacity systems) and Taihulight/K computer/Titan/Earth Simulator 3 (capability systems)?
   ANSWER
   COMMENT
   GRADE

d) What is backfill in job scheduling?
   ANSWER
   COMMENT
   GRADE

Homework question 3

What CPU is on Rocket (Manufacturer, model and base clock speed)? You can get some information by logging in to the cluster

```
ssh your_username@rocket.hpc.ut.ee
```

and then typing:

```
cat /proc/cpuinfo
```

once logged into Rocket. On some machines the login node CPU is different than the compute node CPU, in which case you can start an interactive job to get processor information (with
Homework question 4
What CPU is on Vedur?

Homework question 5
Log in to the rocket cluster:
```bash
ssh your_username@rocket.hpc.ut.ee
```
check the partitions at the cluster:
```bash
scontrol show partitions
```
What are the names of the partitions?

Homework question 6
Create a script (you can do this in a text editor on your own machine and then transfer the file, or using text editors on Rocket such as vi, nano or emacs), which prints out environment variables and runs hostname command:
```bash
#!/bin/sh

echo "Job array"
printenv SLURM_JOBID
printenv SLURM_ARRAY_JOB_ID
printenv SLURM_ARRAY_TASK_ID
hostname
```
Submit the script
```bash
sbatch --array=1-3 --time=00:05:00 script
```
What were the task ID’s? On which machine did the job run? In computations, sometimes a node can have a failure, while other nodes are healthy. Why might this make it useful to know task ID’s on which a job is run?

ANSWER
COMMENT
GRADE
Chapter 4

Parallel computers and accelerators

The main aim of this section is to differentiate between different kinds of parallel computers. Please cite any references you use in answering these questions.

**Homework question 7**
What is a shared memory supercomputer?
ANSWER
COMMENT
GRADE

**Homework question 8**
What is a distributed memory supercomputer?
ANSWER
COMMENT
GRADE

**Homework question 9**
Why might shared memory programming be easier than distributed memory programming?
ANSWER
COMMENT
GRADE

**Homework question 10**
Why might a distributed memory programming style result in more efficient code than a shared memory programming style?
ANSWER
COMMENT
GRADE

**Homework question 11**

a) What is a computational accelerator and why might you want to use one?
   ANSWER
   COMMENT
   GRADE

b) What are the main differences between CUDA, OpenMP 4.5, OpenCL and OpenACC for programming accelerators?
   ANSWER
   COMMENT
   GRADE
Homework question 12

a) What is a heterogeneous parallel computer and what is the most powerful heterogeneous parallel computer on the Top500 list?

ANSWER

COMMENT

GRADE

b) What is a homogeneous parallel computer and what is the most powerful homogeneous parallel computer on the Top500 list?

ANSWER

COMMENT

GRADE
Using supercomputers and other remote resources typically requires some skills that are usually acquired through practice, rather than in formal courses. Thus far, you have logged in to a remote computer and transferred files to and from a remote computer. In this section, we will introduce some more Linux commands. You will need to use ssh to log in to the Rocket cluster. A description of how to use ssh from Putty on Windows is available at https://hpc.ut.ee/en_US/ssh. If you use a Linux or Mac computer, open up a terminal and use ssh from the terminal.

Homework question 13
What do the following commands do when logged on to Rocket cluster using ssh.

- module avail
  ANSWER
  COMMENT

- module load
  ANSWER
  COMMENT

- module purge
  ANSWER
  COMMENT

Homework question 14
Why modules be useful for managing loaded software packages on a shared high performance computing infrastructure?

ANSWER
COMMENT
GRADE

Homework question 15
What is SLURM and what role does it have?

ANSWER
COMMENT
GRADE
Homework question 16
Gathering information from SLURM. Log into the Rocket cluster:

```
ssh your_username@rocket.hpc.ut.ee
   then type:
   sacct -a
```
What information do you get?
ANSWER
COMMENT
GRADE

Homework question 17
Gathering information from SLURM. Log into the Rocket cluster:

```
ssh username@rocket.hpc.ut.ee
   then type:
   squeue
```
What information do you get?
ANSWER
COMMENT
GRADE
Chapter 6
BLAS and LAPACK

**Homework question 18**

a) What is BLAS?

b) What is LAPACK?

c) In what programming language were BLAS and LAPACK initially written in?

d) Why might you pay for a BLAS and LAPACK implementation even though some are freely available?

e) How are BLAS functions used in machine learning, linpack or some other application area that you find interesting?

**Homework question 19**

There are several BLAS and LAPACK libraries available. The libraries performances may differ on different machines. To minimize the time taken to solve heavy calculation tasks, an efficient library should be used. Several well known libraries are:

- Reference BLAS
  and LAPACK available at
  [http://www.netlib.org/lapack/index.html](http://www.netlib.org/lapack/index.html)
- GOTO BLAS available at
  [https://www.tacc.utexas.edu/research-development/tacc-software/gotoblas2](https://www.tacc.utexas.edu/research-development/tacc-software/gotoblas2)
- OpenBLAS available at
  [http://www.openblas.net/](http://www.openblas.net/)
- ATLAS BLAS available at
- BLIS available at
  [https://github.com/flame/blis](https://github.com/flame/blis)
- Intel Math Kernel Library available at
- ULM blas library, available at
  [http://apfel.mathematik.uni-ulm.de/~lehn/ulmBLAS/](http://apfel.mathematik.uni-ulm.de/~lehn/ulmBLAS/)

Compare the performance of these libraries on Rocket or Vedur for double precision dense matrix
multiplication (DGEMM). You should consider matrix sizes of

\[ 8^2, 32^2, 33^2, 55^2, 63^2, 65^2, \]
\[ 100^2, 128^2, 156^2, 200^2, 255^2, \]
\[ 256^2, 300^2, 511^2, 512^2, 513^2, \]
\[ 768^2, 1000^2, 1024^2, 1536^2, 2048^2. \]

Your results should be presented in tabular and graphical form. Please also tabulate and plot the difference between the naive algorithm and the library result as a function of matrix size. Source codes and example submission scripts should also be included. Example code to install and run on Rocket is included. You are encouraged to modify the provided codes, or use other wrapper codes (e.g. written in C/C++) to call the libraries. You are also encouraged to try other compilers, for example Intel, Open64 and if using C/C++ CLANG.
Chapter 7

Feedback

**Homework question 20**
In order to improve this course for the future we ask you to give feedback on how long completing those tasks took. It would be great if you could give time estimates separately for each chapter.

ANSWER
COMMENT
GRADE
Bibliography

[1] BLAS webpage http://www.netlib.org/blas/ Checked on 4 March 2018


