Using R on the LSHTM High Performance Computing (HPC)

Peninah Murage
Peninah.Murage@lshtm.ac.uk
(PHES PHP)
Objectives

• Brief overview of the LSHTM HPC
  • What is the HPC?
• Getting started
  • LSHTM wiki page
  • Joining
  • Login: WINSCP & PuTTY
• Example using R
• Other issues
• Contacts
**Brief overview of the HPC**

- Also know as ‘Cluster’, ‘super computers’ - used interchangeably
- Combined computing power to deliver higher performance
- Performance and processing speed exceeds a typical desktop computer
- Used for computationally intensive tasks
  - Need large memory
  - Large datasets
  - Repetitive tasks
  - Not very useful for small tasks!
- Supports a variety of software (Stata, R, Python)

**Accessible memory per user = 8GB**

**Accessible memory per user = 200 GB**
**Getting started – for Windows only**

LSHTM HPC wiki page
http://wiki.lshtm.ac.uk/hpc/index.php5/Main_Page

**To join:**
Send an email to the ITS Helpdesk (servicedesk@lshtm.ac.uk) with name, username, department, brief description of proposed work on cluster, software requirements

**To log in:**
A new user account on the HPC + a new home directory
Use your usual LSHTM username and password
Copy files to and from HPC using WinSCP
Login in to HPC – Windows example

Your local drive

Your HPC drive

hpclogin.lshtm.ac.uk
Getting started

To submit a job:

- PuTTY software provides remote terminal access to the HPC
- Check LSHTM wiki page on how to configure PuTTY
- Once again use your usual LSHTM username and password
Typical HPC system layout

- Login Nodes
- Compute Nodes
- Disk
- SSH Terminal
- Batch System
- Upload / Download Data
Brief example

• Preparing a task/job
  • Transfer files to your new HPC drive
  • Check that the job works ...run part of the job interactively or on own PC

• Brief description of my task
  save R scripts on HPC drive

• Submitting a “simple” job
  • batch/ shared system
  • use a ‘.txt file’ – specify required memory
  • running the script on PuTTY
  • example of script –
    running a job = \textit{qsub myjob.txt}
    every job is issued a jobID
    deleting a job = \textit{qdeljobID}
Brief example using R

• Email notifications
  • Job started
  • Job completed

• Submit one job with multiple tasks
  E.g. process 5 tasks using 5 different data files

Run job on one textfile ‘.txt’

```bash
#!/bin/bash
#$ -N merge_hpc
#$ -M Peninah.Murage@lshtm.ac.uk -m be
#$ -q short.q
#$ -l mem_free=90G,h_vmem=100G
#$ -t 1-5

R CMD BATCH merge_hpc${SGE_TASK_ID}.R merge_hpcout${SGE_TASK_ID}.out.R
```

-t 1:5 specifies the number of sequential tasks
The job is submitted in 5 tasks and will create 5 R output files

• Parallel processing?
Other issues

- Queuing
  show all your on jobs, `qstat`
  other jobs on queue `qstat -u '*'`
  details of specified job `qstat -j jobid`

- Installing packages
  Login to ‘PuTTY’ and run R interactively

- Loading different versions of R
  Presently not possible at LSHTM!
  See University of Sheffield example here

- Service desk and LSHTM wiki page contacts
  Steven Whitbread (Datacentre and Infrastructure Manager)