What is an R package?

A standardized way to

- 1. Collect together a group of related R functions
- 2. Insist on some level of documentation [help files, vignettes, demos, reference manual]
- 3. Carry out some level of testing
- 4. Present a front-end to the user where internal functions, that the user shouldn't have to call, are invisible to the user
- 5. Compile and link source code in C, C++ or Fortran

What is a computer cluster?

What is cloud computing? How does it compare with cluster computing? Concretely, how does AWS compare to Flux?

Flux architecture (Google: flux umich)

- 126 compute nodes with 20 cores and 96 GB RAM
- 124 compute nodes with 16 cores and 64 GB RAM.
 492 compute nodes with 12 cores and 48 GB RAM
- Large memory: ten compute nodes with either 32 or 40 cores and 1 TB RAM.

The main practical difference to working on a cluster is that jobs need to be submitted to a queue.

- In practice, this means that you write a PBS (Portable Batch System) script which requests the required resources and runs the job when those resources become available.
- An example will be given as homework.

Queuing happens at two levels.

- To submit a job, you must have access to an **allocation**. The Statistics department currently has an allocation of 150 cores, to which you should all have access.
- If you submit a job requiring 50 cores to the Stats allocation, and 120 cores are currently in use, your job must wait.
- Jobs are also queued at a cluster-wide level, since the combined size of all allocations exceeds the number of cores in Flux. Flux buys cores to target a reasonable quality of service.

In order to use Flux, you need

- 1. A Flux login account
- 2. An MToken
- 3. An allocation.

- A Flux login account is needed in order to login to the Flux cluster and to prepare and run jobs.
- You should all have one of these from the orientation.
- To login to Flux, ssh to flux-login.engin.umich.edu. You'll be asked for your MToken code (see below) and your Kerberos password.
- You can connect to Flux from any wired campus network or using MWireless. Otherwise, you'll need to use a VPN to connect, or ssh to login.itd.umich.edu first, and then ssh to flux-login from there.

- An **MToken** (or a Software Token) is required to authenticate to the cluster. When logging in, you will need to give both the passcode it displays at time of login as well as your UMICH password in order to authenticate.
- You should have an MToken from the orientation.

http://www.mais.umich.edu/mtoken/

Allocation

In order to submit jobs to the Flux compute nodes, you need a Flux allocation against which this use is charged. The Statistics department maintains a departmental allocation named stats_flux, which you may use for this purpose. Your PBS script preamble should contain the following three lines:

- #PBS -A stats_flux
 #PBS -l qos=flux
- **#**PBS -q flux

HPC 101: High Performance Computing Workshop

http://arc-ts.umich.edu/hpc101/

Prerequisites: HPC 100 or equivalent. This course assumes familiarity with the Linux command line.

or Google: HPC 101 umich

The website has slides for the workshop that provide a good introduction to Flux.

Why is multi-processor computing of increasing importance to Statistics (and elsewhere)?

At a smaller scale, similar ideas are emerging in neuromorphic chip design: http://www.sciencemagazinedigital.org/ sciencemagazine/10_october_2014?pg=47# pg46 What are the implications of cluster architecture for developing statistical theory and methodology?