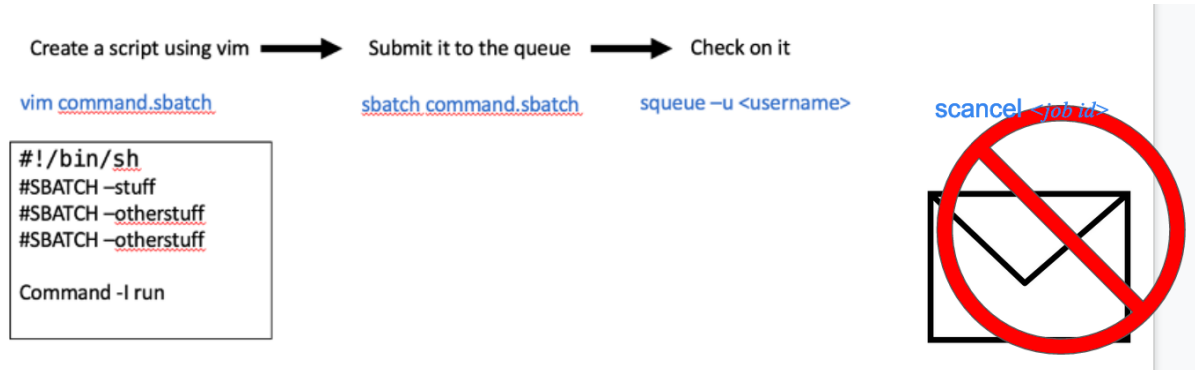# Short Read Sequencing Analysis Workshop

Day 3 – Compute Cluster Usage and Data Transfer

Overview

Slurm Batch Script Directives

`#!/bin/sh` This is called a shebang and must be the first line in a shell script. It tells the computer you are writing in bash

`#SBATCH -p <queue>` Specify job queue to run the job. Queue names are specific to machine you are on! Fiji has "short" as the default. AWS has "compute".

`#SBATCH --time=00:00:00` hrs:minutes:seconds

`#SBATCH --mem=<memory>`

`#SBATCH --ntasks=<number>` specify how many cores are needed for the job

`#SBATCH --job-name=<jobname>` give the job a name

`#SBATCH --mail-type=ALL`

`#SBATCH --mail-user=<you@email.com>`

`#SBATCH --error=<path/file.err>` specify name and location of error file

`#SBATCH --output=<path/file.out>` specify name and location of log file

Slurm Command Line Utilities

`$ sbatch <jobscript.sbatch>` submit a job to the queue

`$ squeue` view the status of all jobs on cluster

`$ squeue -u <username>` view status of jobs submitted by *<username>*

`$ squeue -j <job id>` view the status of a specific job with job ID

`$ scancel <job id>` delete/cancel a specific job with job ID

Tips

- Make multiple directories before you start: one for for scripts, one for output, one for e_and_o files.
- Name your sbatch scripts with <scriptname>.sbatch. Then you will know it is a sbatch script.
- Make a template.sbatch. Then copy it each time you need to make a sbatch script.

  - Default suggestions in your template
    - `ntasks should be 1 unless the program tells you otherwise`
      - `Sometimes ntasks are called` cores/processors
    - `Time should be 1 hour by default`
    - `Memory should be 1gb by default`
    - `I make my default error and out files`
      `/scratch/Users/<username>/e_and_o/<jobname>.%j.err`
      `/scratch/Users/<username>/e_and_o/<jobname>.%.%j.out`
      `.%j stands for the job number`
- Error and out directories must exist before you submit the job!
- On this computer the queue name is compute. On fiji the default queue is short. To find out the queues that exist on your supercomputer type sinfo.
- When you are trying a new program:
  1. Guess at your SBATCH resources
  2. Run one infile (the biggest input file you have)
  3. Check what you really used
  4. Fix your SBATCH resources
  5. Run the rest of your files
- How much resources did I use
  1. While the job is running:
     - sstat --format=AveCPU,AvePages,AveRSS,AveVMSize,JobID -j 2050904.batch #where 2050904 is the jobid
     - scontrol show jobid -dd 2050904 #where 2050904 is the jobid
  2. Or for completed jobs
     - sacct --format="CPUTime,MaxRSS,AveCPU,AvePages,AveRSS,AveVMSize,JobID,MaxVMSize" -j 2050904.batch #where 2050904 is the jobid

- The short hand for this document is that <this> means file "this" in without the <>
- Don't copy anything in RED
- Because of hidden characters, don't copy and paste from this document.