



# Cluster Computing

Dan Barker  
Research Computing Admin



# Presentation Overview

Nomenclature

Why Use a Cluster? - GWAS Example

Batch Computing Clusters

Web Portal

Live Demonstration

# Nomenclature

Cluster Computing =

High Performance Computing (HPC) =

Super Computing

Cluster = Supercomputer



# Why use a Cluster? - GWAS Example

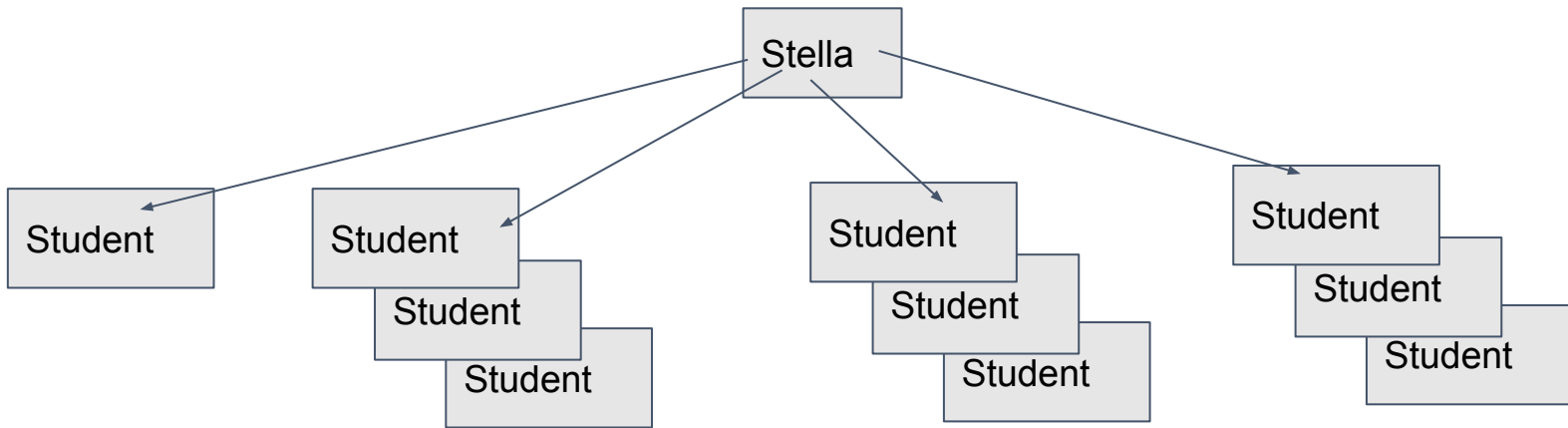
Genetic Data Samples		Samples	Disease(0/1)
Variant1	A/A A/T T/T	ID1	0
Variant2	C/C C/C G/G	ID2	1
.	.	.	.
.	.	.	.
.	.	.	.
Variant1,000,000	A/T T/T A/A	ID50k	1

Regression analysis of the disease vector against a variant takes 5 seconds

5 seconds/variant \* 1,000,000 variants = 5,000,000s = ~58 days

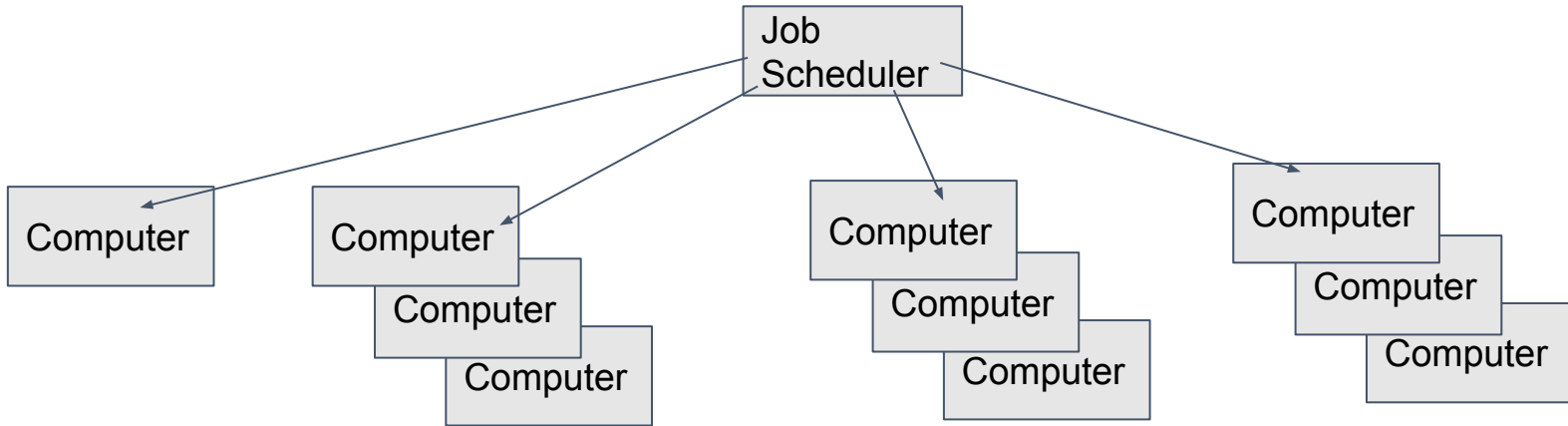
# Faster GWAS

Lab director Stella decides there is a faster way to do this. Stella divides the data set into 10 sets of 100k variants and gives one to everyone in the lab to run on their laptops. Now we get results in 5.8 days.

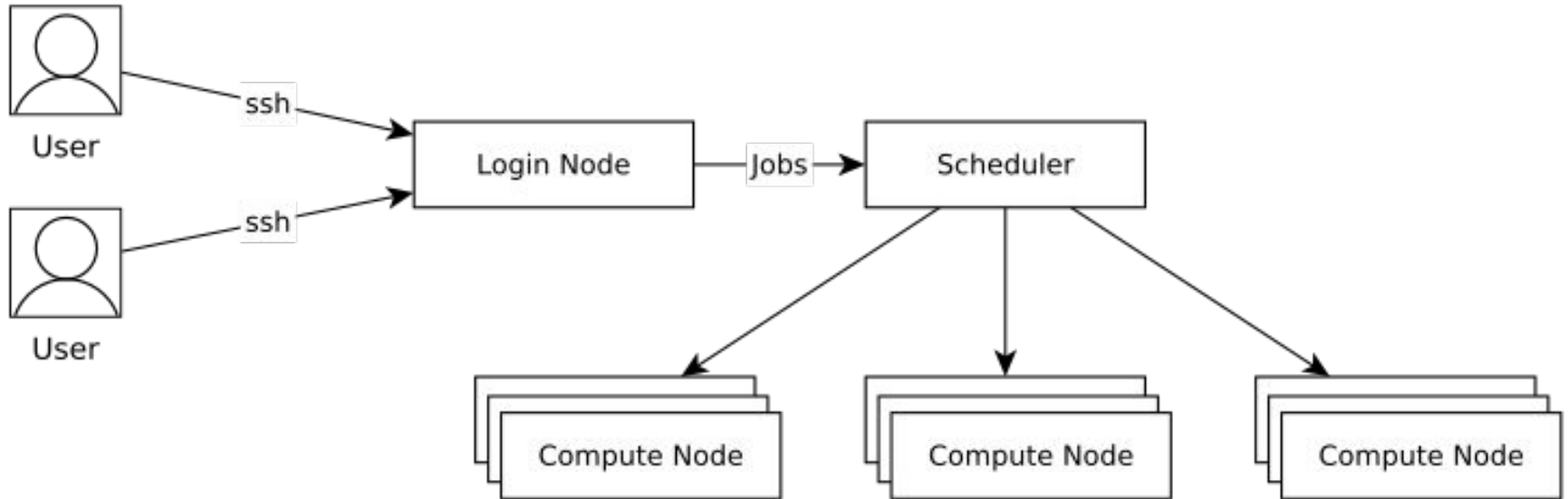


# Cluster

Stella acts as a cluster job scheduler.  
Laptops act like cluster compute nodes.



# Cluster Computer in a Schematic



# Cluster Computer in a Picture





# R in Batch Mode

R script that can execute top to bottom without error

```
[danbarke@idran biostat625]$ cat script.R  
print("Entering loop number 1")  
Sys.sleep(5)
```

R CMD BATCH --no-save --no-restore script.R <script.Rout>

R output saved to script.Rout by default. Can change file name with optional argument



# R in Batch Mode Output

```
[danbarke@idran biostat625]$ R CMD BATCH --no-save --no-restore script.R
```

```
[danbarke@idran biostat625]$ cat script.Rout
```

```
~
```

```
> print("Entering loop number 1")
```

```
[1] "Entering loop number 1"
```

```
> Sys.sleep(5)
```

```
~
```



# Batch Job Files

```
#!/bin/bash  
#SBATCH --job-name=hello_world  
#SBATCH --time=10:00  
#SBATCH --mail-type=end,fail  
#SBATCH --mem=1g  
#SBATCH --cpus-per-task=1
```

```
R CMD BATCH --no-save --no-restore script.R <script.Rout>
```



# Demonstration on the Biostatistics Cluster

<https://biostat-login.sph.umich.edu/>

armis2.arc-ts.umich.edu



# Download examples

<https://github.com/danbarke/biostat625>



# Getting Help

Cluster Computing

Dan Barker ([danbarke@umich.edu](mailto:danbarke@umich.edu))

R Programming

Mike Kleinsasser ([mkleinsa@umich.edu](mailto:mkleinsa@umich.edu))

4017 SPH II