

Maintaining a publicly used analysis infrastructure



Keeping everything running when everyone's an analyst

Madelaine Gogol
Computational Biology Core
Stowers Institute
Kansas City, Missouri

Where we were, where we're going

In the past...	In the future...
Analysts did analysis	Many people do analysis
A small group of people shared computational resources	More people sharing, more diverse needs
Communication about and tracking load on servers/clusters is fairly straightforward	more complex – more groups, more people involved
Users pretty comfortable with Linux	Training and mentoring is needed

BRACE YOURSELVES



**THE BIOLOGISTS ARE
COMING**

My institute's (current) approach to sharing computational resources

- Most servers / clusters are currently open
- A few are dedicated to particular groups
- Some clusters require training to get access
- We made “server usage guidelines”
 - Use < half of available memory and cores on a server
- Encourage use of htop
- Informal “Policing” via email

```
1 [|||||] 97.4%
2 [|||||] 89.0%
3 [|||||] 194.1%
4 [|||||] 89.5%
5 [|||||] 91.5%
6 [|||||] 89.0%
7 [|||||] 194.0%
8 [|||||] 90.2%
Mem [|||||] 4130/8192
Swp [|||||] 0/1024MB

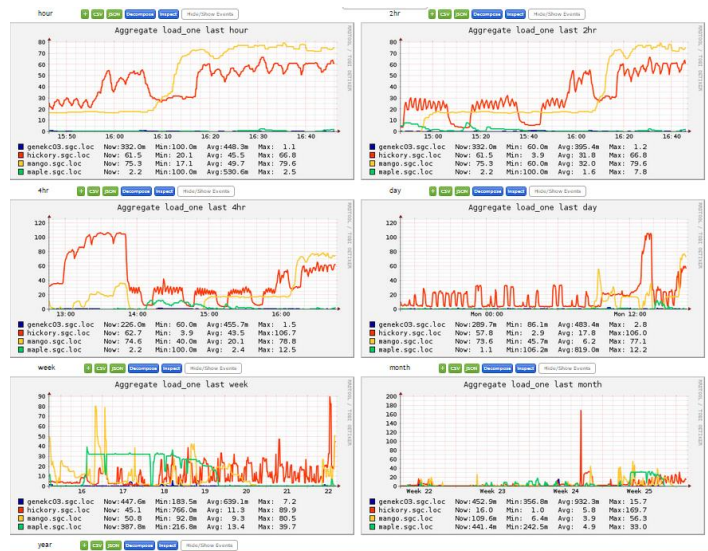
Tasks: 246 total, 2 running
Load average: 8.96 4.85 3.01
Uptime: 10 days, 22:11:00

PID USER   PRI  NI  VIRT  RES  SHR  S  CPU% MEM%  TIME+  Command
33899 miHenry 53   0 5700M 817M  0 C 739 10.0 0:11.29 /usr/bin/java -server -XX:+HeapDumpOnOutOfMemoryError -Xms512m -Xmx512m -XX:NewSize=128m -XX:MaxNewSize=128m ->
23429 miHenry 53   0 5657M 556M  0 S 11.0 6.8 0:03.55 /usr/bin/java -server -XX:+HeapDumpOnOutOfMemoryError -Xms512m -Xmx512m -XX:NewSize=128m -XX:MaxNewSize=128m ->
26551 miHenry 97   0 1289M 249M  0 S  0.0 3.0 0:25.66 /Applications/Google Chrome.app/Contents/Versions/36.0.1985.143/Google Chrome Helper.app/Contents/MacOS/Google
  111 root    79   0 3716M 101M  0 S  0.0 1.2 1:06.87 /System/Library/Frameworks/ApplicationServices.framework/Frameworks/CoreGraphics.framework/Resources/WindowSer
53437 miHenry 33   0 1057M 166M  0 S  0.0 2.0 0:00.55 /Applications/Google Chrome.app/Contents/Versions/36.0.1985.143/Google Chrome Helper.app/Contents/MacOS/Google
```

Challenges we've had

- Encouraging people to be mindful, use htop
- Resource hog programs, novice users
- When do we add computational resources?

nagios / ganglia



Analyst grumbling

March 2014 Archives by thread

- Messages sorted by: [\[subject \]](#) [\[author \]](#) [\[date \]](#)
- [More info on this list...](#)

Starting: *Mon Mar 3 11:02:08 CST 2014*

Ending: *Fri Mar 28 15:46:36 CDT 2014*

Messages: 36

- [\[linux-discuss\] hickory](#) *Patison, Ariel*
 - [\[linux-discuss\] hickory](#) *McGee, Jenny*
 - [\[linux-discuss\] hickory](#) *Meier, Samuel*
 - [\[linux-discuss\] hickory](#) *McGee, Jenny*
- [\[linux-discuss\] Hickory](#) *Patison, Ariel*
 - [\[linux-discuss\] Hickory](#) *Dietz, Dustin*
 - [\[linux-discuss\] Hickory](#) *Cook, Malcolm*
 - [\[linux-discuss\] Hickory](#) *Dietz, Dustin*
 - [\[linux-discuss\] Hickory](#) *Cook, Malcolm*
 - [\[linux-discuss\] Hickory](#) *Dietz, Dustin*
 - [\[linux-discuss\] Hickory](#) *Cook, Malcolm*
 - [\[linux-discuss\] Hickory](#) *Dietz, Dustin*
 - [\[linux-discuss\] Hickory](#) *Cook, Malcolm*
 - [\[linux-discuss\] Hickory](#) *Dietz, Dustin*
- [\[linux-discuss\] Color Exhibit](#) *Gogol, Macéline*
- [\[linux-power-users\] desktop reboot](#) *Miller, Danny*
 - [\[linux-power-users\] desktop reboot](#) *Seidel, Christopher*
 - [\[linux-power-users\] desktop reboot](#) *Patison, Ariel*
- [\[linux-discuss\] cirros installation](#) *Seidel, Christopher*
 - [\[linux-discuss\] cirros installation](#) *Gogol, Macéline*
 - [\[linux-discuss\] cirros installation](#) *Seidel, Christopher*

Other Analysis Resources

- Genomes, annotations, alignment indexes
- Databases (e.g. GO, BLAST)
- Software Installation, Updates
- Public Sequencing Data
- Galaxy
- Rstudio Server

How about you?

- Balancing user access vs. user responsibility
- Breakdown of “private” versus “public” resources
- Challenges faced and solutions
- Tracking load, deciding when to add resources
- Managing genomes/software/data

