GMOD in the Cloud
http://gmod.org/wiki/Cloud

Scott Cain[1] and the GMOD Consortium

[1]Ontario Institute for Cancer Research, Toronto, Ontario, Canada, M5G 0A3

With decreasing costs of sequencing technology there is an increasing need for computing resources and tools to work with the data. An alternative to building and maintaining a large computing infrastructure in-house is to use existing networked computing systems ("cloud computing") to make use of preconfigured and extensible servers. Here we present several efforts implementing GMOD software tools using Amazon Web Services (AWS). First, a GBrowse2 server has been prepared with options for importing available data from Amazon storage (EBS) for several common organisms and well as tools for creating additional rendering servers as needed to improve the responsiveness of the website. Next, a community annotation server is available with Chado, GBrowse2, JBrowse and Tripal to facilitate taking computational annotations (for instance from MAKER) and making them available to interested users in the community to review and improve upon. Additionally there are two data analysis cloud images: one for sequence annotation based on GMOD tools called CloVR, and a cloud implementation of Galaxy called Cloudman.

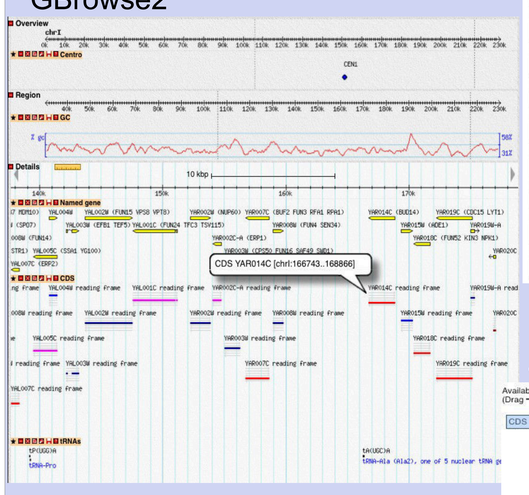
GMOD in the Cloud

Installed software:

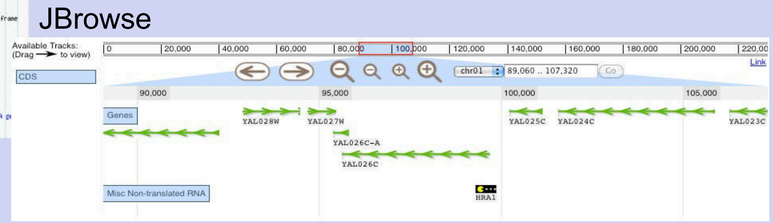
Tripal	Drupal-based web frontend
Chado	Generic organism DB schema
GBrowse	Venerable genome browser
JBrowse	Fast, AJAX genome browser
Sample data	Saccharomyces cerevisiae

Tripal

GBrowse2



JBrowse



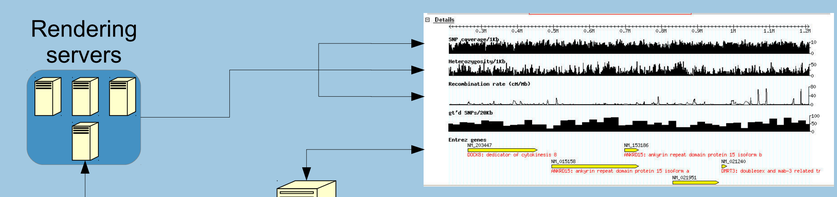
Why use cloud.gmod.org?

- As a community annotation server, where you don't have to set up the server.
- As an example implementation to guide your own efforts.
- As a test bed for software you are developing.

For more information, including the AMI ID of the current release, see http://gmod.org/wiki/Cloud.
For a demo, see http://cloud.gmod.org/.

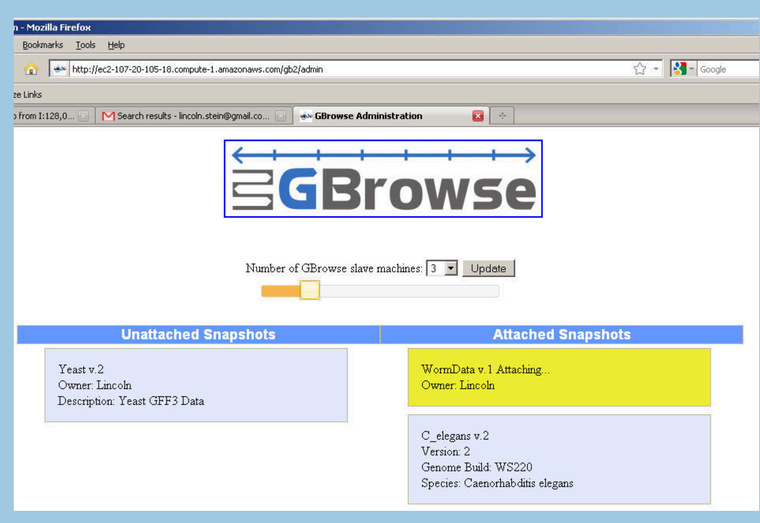
GBrowse2

GBrowse2 is a highly configurable and flexible web-based genome browser used by hundreds of organizations



To speed up GBrowse, tracks can be rendered by any number of rendering slaves, spreading out the work.

The admin interface lets you add and remove slave rendering nodes as needed, as well as mount standard data sets.



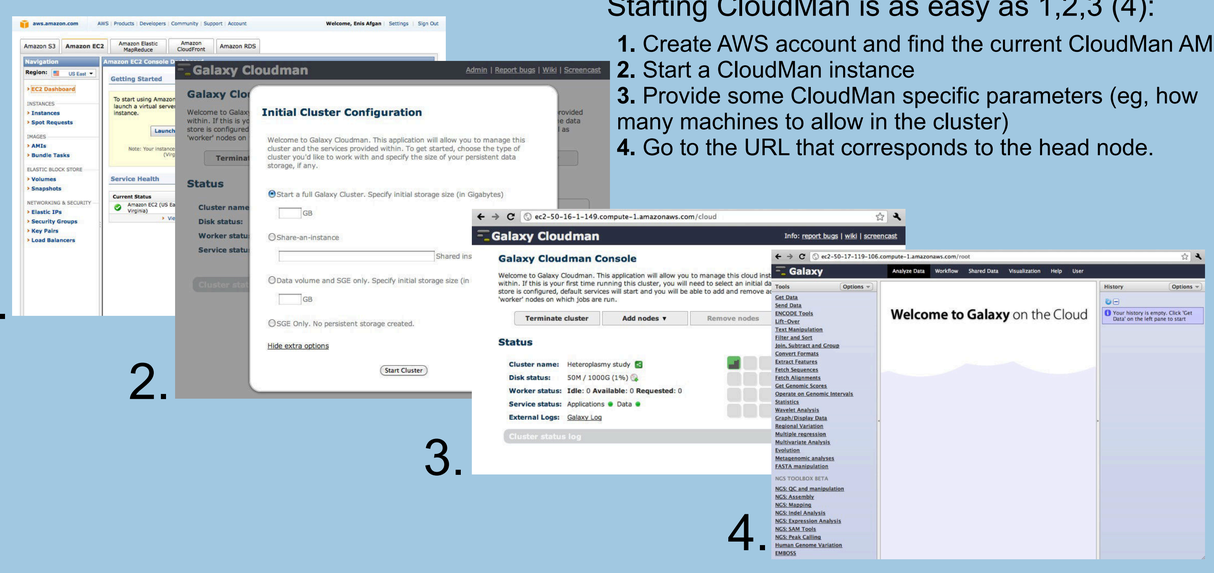
The GBrowse2 AMI is not publicly available yet; please watch the GBrowse mailing list and http://gmod.org/wiki/Cloud for announcements.

Galaxy CloudMan

Use Galaxy's CloudMan to get the power of Galaxy without the usage or data restrictions.

Starting CloudMan is as easy as 1,2,3 (4):

1. Create AWS account and find the current CloudMan AMI
2. Start a CloudMan instance
3. Provide some CloudMan specific parameters (eg, how many machines to allow in the cluster)
4. Go to the URL that corresponds to the head node.



Powerful computation on the cheap!

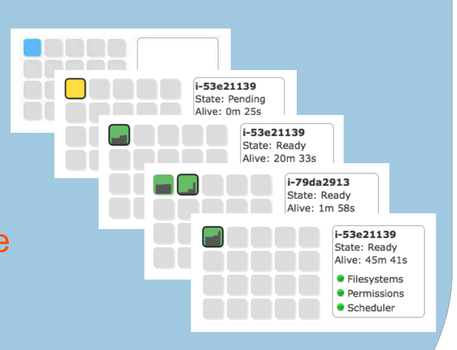
By allowing CloudMan to dynamically change the number of nodes in the cluster, you get both good performance and good value, as when you need more nodes to do computations, they'll be added, and when you are done, they'll be shutdown, so you only pay when you need the power. Also, if you aren't using CloudMan for a while, you can turn it off altogether.

Fixed cluster size

5 nodes	Computation time: 9 hrs Computation cost: $20
20 nodes	Computation time: 6 hrs Computation cost: $50

Dynamic cluster size

1 to 16 nodes	Computation time: 6 hrs Computation cost: $20

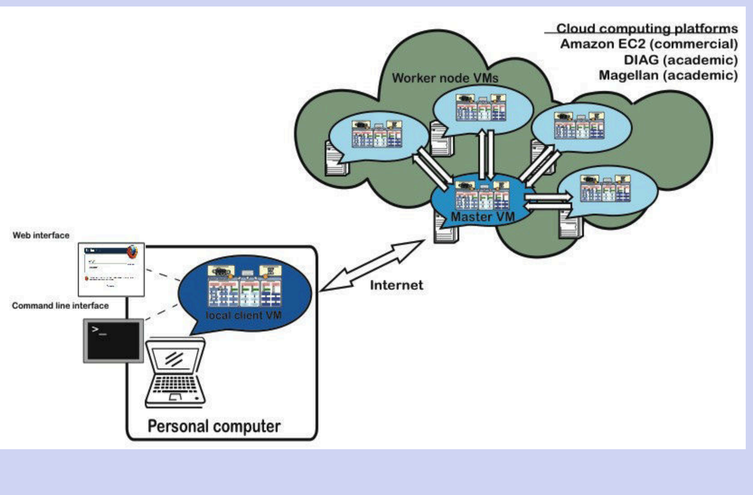


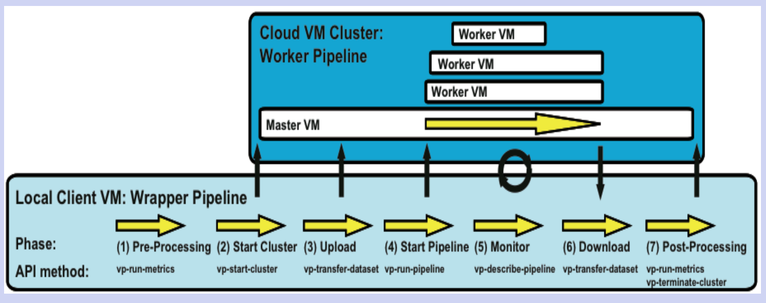
For more information, see http://usecloudman.org/

CloVR

Cloud Virtual Resource is an automated sequence analysis pipeline powered by GMOD tools Workflow and Ergatis.

To use CloVR, download the virtual machine that runs locally on your desktop and interacts with your nodes in the cloud. Note that in addition to AWS, CloVR supports other cloud infrastructures like DIAG and Magellan.





The local virtual machine monitors and directs the master node in the analysis cluster in the cloud, and the master node in the cloud directs computational work in the cluster.

For more information, please see http://clovr.org/

GMOD is supported by a specific cooperative agreement from the USDA Agricultural Research Service, and by NIH grants co-funded from the National Human Genome Research Institute and the National Institute of General Medical Sciences. GBrowse development is also funded by the Ontario Ministry of Research and Innovation.

The Galaxy project is funded by NIH, NSF, Penn State, Emory, and the Pennsylvania Department of Public Health.

The CloVR project is being funded through the Microbial Genome Sequencing program (USDA-CSREES-AFRI-001967) of the National Science Foundation (NSF) and the Recovery Act Limited Competition for NIH Grants: Research and Research Infrastructure "Grand Opportunities" (RC2) (RFA-OD-09-004) of the National Human Genome Research Institute (NHGRI/NIH).