Reference genome based sequence variation (SNP and INDEL) detection

Computational Resources for the Workshop

Robert Bukowski (bukowski@cornell.edu)

5/25/2011

In order to meet the anticipated high demand for computational resources during our joint CBSU/3CPG workshops, we teamed up with **iPlant Collaborative** (http://www.iplantcollaborative.org/about) and accepted their kind offer to make their **iPlant Atmosphere** cloud resources available to workshop participants.

While iPlant Atmosphere cloud is expected to be the primary computational resource for the workshop, the exercises may also be completed on the workstations of the CBSU/3CPG BioHPC Lab (http://cbsu.tc.cornell.edu/lab/lab.aspx). If you prefer to follow this route, contact us at cbsu@cornell.edu.

iPlant Atmosphere cloud

The complete description of iPlant Atmosphere cloud services can be found at https://pods.iplantcollaborative.org/wiki/display/atmosphere/iPlant+Atmosphere.

From the point of view of the workshop, it is sufficient to view Atmosphere as a collection of powerful Linux (Centos 5) workstations (also referred to as 'Virtual Machines', 'VMs', or 'instances') which you can launch on demand, then connect to remotely, do the work and/or transfer files, and terminate when you no longer need them. Each such workstation is equipped with 4 CPU cores and 12 GB of memory. Each workshop user account comes with 50 GB of personal storage on a so-called EBS (Elastic Block Storage) volume. The EBS volume can be attached to any Atmosphere instance, and data written to the volume is preserved after the instance is terminated.

Below we describe steps you have to take to use the iPlant Atmosphere cloud during the workshop.

Apply for iPlant Atmosphere account

Navigate to http://www.iplantcollaborative.org/forms/preview-atmosphere-cbsu and fill out the form as completely as possible. Once your application is processed (which may take up to 24 hours), you will receive a welcome e-mail message. Use the link provided in that e-mail to set your Atmosphere password. At this point your Atmosphere account will be ready.

Install prerequisite software on your PC

To access Atmosphere cloud, you will need the following software installed on your PC or Mac:

Firefox

Needed to start and terminate Atmosphere instances and to attach/detach EBS volumes. If you do not have Firefox yet, install the latest version (http://www.mozilla.com/en-US/firefox/new/).

<u>VNC Enterprise Edition Viewer (Stand-alone Viewer)</u>

This is the preferred client program for connecting to Atmosphere instances. It provides full access to the instance's graphical desktop. Go to http://www.realvnc.com/cgi-bin/download.cgi. Click "Proceed to downloads" (you do not have to fill out the form before clicking). Scroll down the list until you see „stand-alone viewers". Download the "Executable" appropriate for your operating system, saving it anywhere on your hard drive, for example, on the Desktop (before download starts, you will need to "check" that you understand the user agreement). Once the executable file is on your hard drive, you can launch it simply by double-clicking on its icon. No special installation is required. If you prefer, you can download and install the full version of the product (rather than just a stand-alone executable), but it is not necessary for the workshop.

<u>SFTP client for file transfers (optional)</u>

Any data transfer between your PC and the workshop machine is not necessary to complete the workshop. However, you may wish to transfer some result files to your PC to analyze them later, open them using your favorite spreadsheet program, etc. In such a case, you will need an SFTP client installed on your PC. For Windows PCs, we recommend **Filezilla** (http://filezilla-project.org/) or **WinScp** (http://winscp.net/eng/index.php). When connecting to your VM from an SFTP client, use your Atmosphere user ID and password, and the VM's IP address instead of the machine name. See the 3CPG Lab document http://cbsu.tc.cornell.edu/lab/doc/Remote_access.pdf for more info about options for remote access to a Linux machine.

<u>SSH client for accessing VM (optional)</u>

While the VNC client is expected to be the primary way to access Atmosphere VMs, it is also possible to log in by opening an ssh terminal window using an SSH client, such as **Putty** (http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html). The ssh window is just a text terminal – you can run all workshop commands in it, but it will be more difficult to run any graphical software, such as IGV viewer.

Launch your Atmosphere instance (aka Virtual Machine, VM)

If you do not yet have your Atmosphere account, apply for it at http://www.iplantcollaborative.org/forms/preview-atmosphere-cbsu. Please note: it may take some time (up to 24 hours) before your application is processed.

Start the Firefox browser and navigate to https://atmo-beta.iplantcollaborative.org/login/. Log in with your Atmosphere user name and password. Note: the web page will not work correctly with browsers other than Firefox! If the browser displays a message that additional plugins are needed, click a button to install those plugins and re-load the page.

Once logged in, click on the "CBSU Workshop Image" icon. You should see a message box notifying you that the launch of your VM is started and will take about 15 minutes to complete. Click "OK" and be

patient. While you wait, you can click the "Dashboard" tab and then on "Running instance index" to see the status of your VM (in the beginning – it will be "pending", when the VM is ready – it will change to "running"; you may need to refresh the page to see this change).

Once your VM is ready to use, its status on the "Dashboard" will change to "running". Also, you will receive an e-mail message containing the IP address of your VM. Save this message – you will need the IP address to access your VM.

Click on the ">>" icon in the top-left corner of the page and then expand the "Cloud Instances" tab. You will see an entry starting with "CBSU Workshop …". Click on this entry. On the right, you will see a screen with all kinds of information about your VM, including its IP address (listed as "Public DNS"). Click on the tab "Attach volume" on top of this screen and then click on the volume ID that shows up. After a few seconds, your personal EBS volume will be attached to your VM. In the non-cloud world, this is equivalent to plugging in, say, your USB drive into a computer.

The process of launching your cloud VM is now complete. If you want, you may log out of the web page and close Firefox.
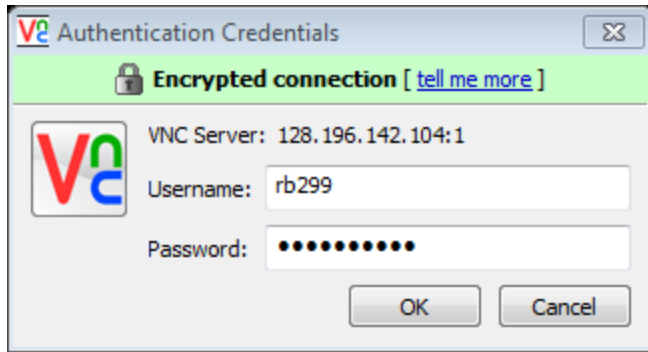
Use your Atmosphere VM

We assume that you have already launched your Atmosphere VM and attached your personal EBS volume to it via the web interface. In this Section, we explain how to connect to your VM from your PC, how to configure your attached EBS volume for use, how to work in your VM, and how to terminate it when no longer needed.

Connect

Launch your VNC viewer on your PC by double-clicking on its icon. You will see the following window:
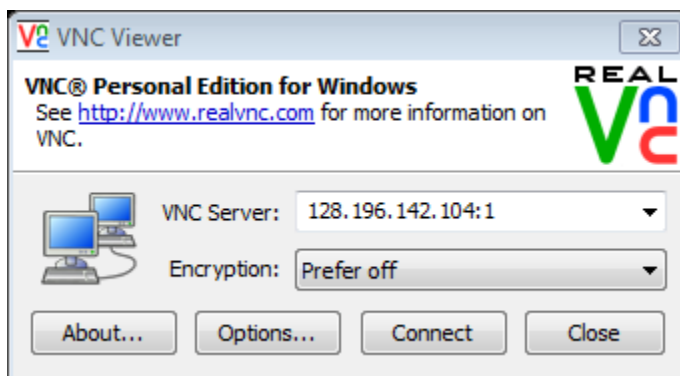
In the VNC Server box, enter the IP address of your VM (in the example above, it is 128.196.142.106 – yours will be different!) followed by ":1" and click "Connect". If prompted, accept the VNC server's signature. You will see the dialog

Supply your Atmosphere user name (<u>instead of rb299</u>) and password, then click "OK". You will see a screen similar to this:
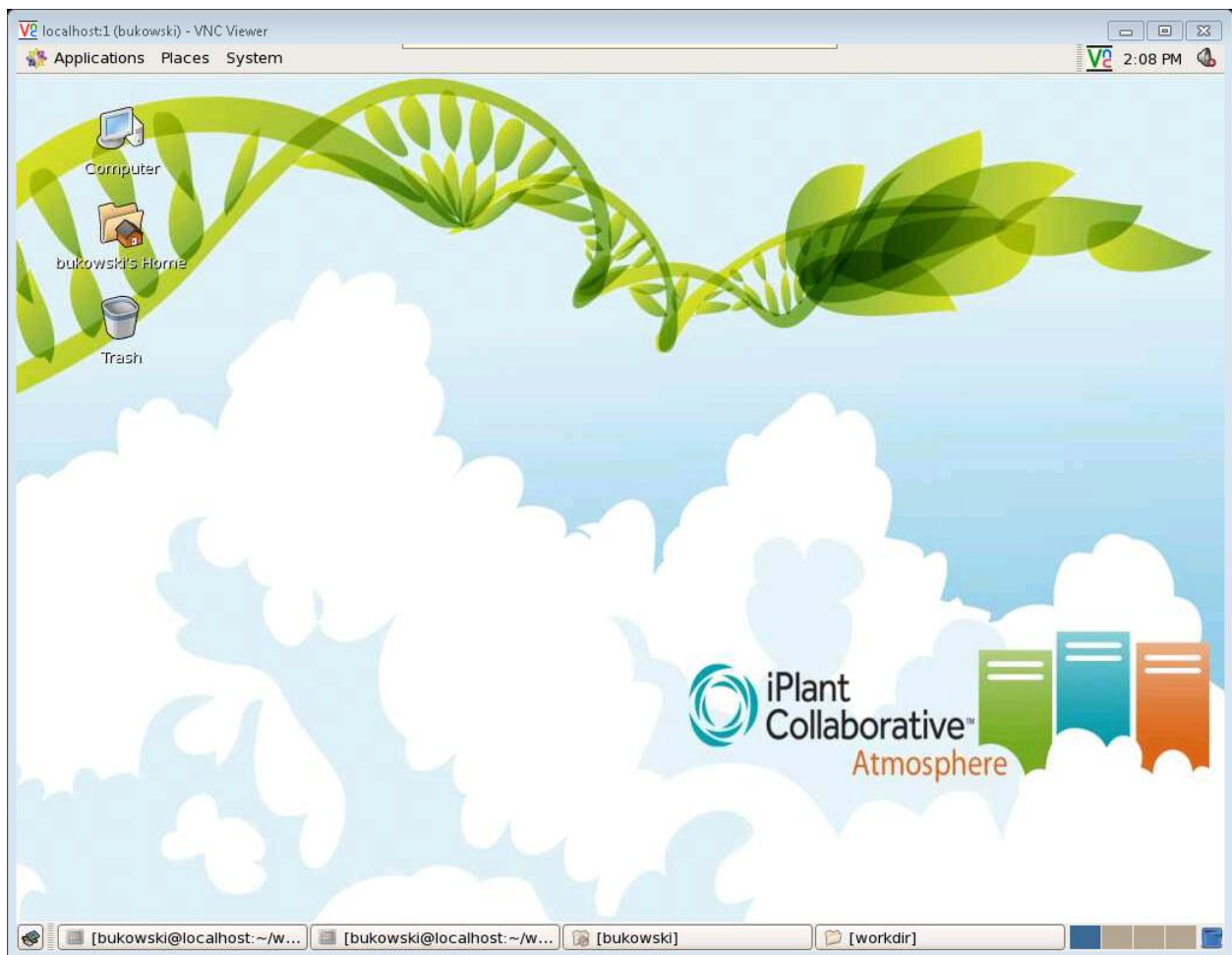
This is the desktop of your VM. You can use it as if you were sitting in front of the machine's console. Right-click anywhere on the desktop and select "Open terminal" to launch an interactive terminal window. You can open multiple such terminals, if you wish.

Attach your personal EBS volume for the first time

If you are using your EBS volume for the first time, it needs to be formatted – a simple process explained in this section. If your personal EBS volume has already been formatted (e.g., if you are connecting for the second time), you can skip to the next section.

In the terminal window, type the following commands (hitting "Enter" after each command):

```
cd

mkdir workdir

sudo /sbin/mkfs /dev/sdb
```

The last command will ask for your password (provide your Atmosphere password here). When you see a question 'proceed anyway?', answer 'y' (and hit "Enter"). When the last command completes (which may take 15 seconds or so), run the following two commands replacing **rb299** with your own Atmosphere user ID (if asked, provide your Atmosphere password):

```
sudo mount /dev/sdb  /home/rb299/workdir

sudo chown rb299 workdir
```

The volume is now mounted **under /home/rb299/workdir** (with rb299 replaced with your own Atmosphere ID) and ready to use.

Attach your previously formatted personal EBS volume

Use this section if your EBS volume is already formatted (typically, formatting step is performed upon the first log-in to an Atmosphere instance).

If the **workdir** directory is not yet present, create it by typing the following commands in the terminal window (you may skip these two commands if **workdir** already exists in your home directory):

```
cd

mkdir workdir
```

Then type the following, replacing **rb299** with your own Atmosphere user ID . If asked, provide your Atmosphere password:

```
sudo mount /dev/sdb  /home/rb299/workdir
```

The volume is now mounted under **/home/rb299/workdir** (with rb299 replaced with your own Atmosphere ID) and ready to use.

Working in your Atmosphere VM

The VM behaves a lot like a regular Linux (CentOS 5, to be specific) desktop. The workshop exercises are performed mostly by typing commands in the terminal window (you can open multiple terminal windows by right-clicking anywhere within the desktop background and selecting "Open terminal"). Here are several hints to remember:

- If you launched a command (e.g., a program) from a terminal window in the background (i.e., with an '&' at the end of the command), do not close that window. Otherwise, your program will be prematurely terminated. You can minimize the window to de-clutter the desktop.
- During the exercise, you can <u>disconnect</u> from the VM using the "Disconnect" icon available (among other icons) on top of your VNC viewer. Disconnecting from the VM will <u>not</u> terminate your session or any of your running programs. You can return to your VM any time by launching the VNC client again and reconnecting. Once you reconnect, you should see your VM the way you left it. Your programs will still be running, unless they finished in the meantime while you were gone. Your EBS volume will be still attached and mounted as well (check directory **/home/<your_ID>/workdir** – you should see all your files there) –do not attempt to mount it again!
- Do <u>NOT</u> log out of the VM using the desktop's **System->Log Out** mechanism. It will make it hard to log in again. Use the "Disconnect" icon (see above) instead.
- Occasionally you may experience network problems (your VM is physically located far away – in Arizona) resulting in a loss of connection to your VM. This does <u>not</u> kill your session or any of your running programs. By the time you attempt to reconnect to your VM, chances are the network issues will be resolved.
- Almost all usable disk space available to your VM is on your personal EBS volume, i.e., in **/home/<your_ID>/workdir**. Thus, all your work should be performed there. Also, remember that after your VM is terminated, all your data will be lost except those stored on your personal EBS volume. The specific workshop instructions frequently prompt you to make sure you operate in **/home/<your_ID>/workdir**.
- The workshop assignments do not require any file transfer between your PC and the VM. However, if you would like to examine some output files using your own computer, you can fetch these files using any SFTP client. Please consult the 3CPG website for advice on SFTP clients. Use your VM's IP address (the same one used in the VNC client) with your Atmosphere user ID and password to connect.

Terminate your Atmosphere VM

When your work is finished, you need to terminate your Atmosphere VM. This involves un-mounting your EBS volume (in the terminal window of the VM), then detaching the volume from the instance and terminating the instance (both done using the web interface). <u>Attention: un-mounting and detaching the EBS volume prior to instance termination is essential for the integrity of your data stored there !!!</u>

To un-mount your EBS volume, type the following in the terminal window of your VM:

```
cd

sudo umount /dev/sdb
```

Provide your Atmosphere password, if prompted. You may now disconnect from your VM (using, for example, the "Disconnect" icon on top of the VNC client).

Open Firefox browser on your local PC, navigate to web interface to https://atmo-beta.iplantcollaborative.org/login/ and log in using your Atmosphere user name and password. Click on the ">>" icon (top-left corner of the page) and then click on "Cloud instances" tab and the entry corresponding to your VM. On the screen to the right, click on "Detach volume" button on top bar. After this operation completes, click on "Terminate" button.

Your VM is now terminated. Its IP address you used to access it is no longer valid and you cannot re-connect to it.

However, your EBS volume is preserved with all the data you deposited on it during your work. If needed, you can launch another VM (as described in the beginning) at a later time and attach your EBS volume to it. This way, you can complete your workshop assignments piece by piece, possibly over several days, launching VMs only when you have time to work on them, and terminating VMs when you do not need them.