

Creating Storage Volumes For Large Files

Transferring large data sets every time an instance is launched can be time-consuming. Storage volumes can be created and mounted to instances much like a removable drive. They also serve as a convenient way to archive files and backup data off-site.

1. Log in to your AWS account. The AWS console will display.
2. In the left sidebar, click “Volumes” under the heading Elastic Block Store
3. Click “Create Volume”. A window will appear allowing you to select the size of the volume
4. Click “Create”. You now have a persistent storage volume with AWS that you can attach to any running instance. Upload speeds often become a problem when transferring large files. By storing large files in this volume, you can work on projects using any connection.
5. Your volume should appear. Select it and click “Attach Volume”. A window will allow you to attach the volume to a running instance. Note the Device menu, this will be important to remember when mounting the volume.

(FIRST TIME ONLY!) Format the drive using the following command (this will delete any data if repeated):

```
mkfs -t ext2 <Device from step 5>
```

When using an AWS instance, the /mnt directory is your working directory for large files. Mount the volume to an existing directory within your instance (for example a directory named “backup”):

```
mount <Device> /mnt/backup
```

Files can now be transferred to your volume for use. Any files copied to the folder the volume is attached to will be copied to that volume. Remember that every time an instance is terminated, all data is lost. This is a convenient way to manage an ongoing project without leaving an instance running. Transferring files to the volume can be easily scripted, or if you are in a hurry you can copy the entire workspace (the /mnt directory) to a folder on the volume.

Resources

SEQanswers Wiki and Forum

The wiki describes the function of many useful tools and links to the files. There are also great how-to articles which describe some basic workflows and provide reviews for various software packages. The forum is a great resource for advice, help, and scripts from some very talented members.

BioPython Tutorial and Cookbook

Python is a very useful language for parsing sequence data. BioPython has many scripts useful for bioinformatics as built-in functions allowing for easy sequence conversion, generating histograms and statistics, etc. It also has an installation guide.

Python Documentation - docs.python.org/index.html

Contains tutorials and a database of every aspect of Python and sample usages. A bit technical for a quick-reference, but very complete.

LinuxCommand - linuxcommand.org/writing_shell_scripts.php

Beginners guide to shell scripts. How to create workflows and pipelines that will automate otherwise time-consuming tasks and using powerful tools to process large quantities of data quickly.

Software Carpentry – software-carpentry.org

Beginners guides to various programming languages and shell scripts.

Stack Overflow – stackoverflow.com

A question and answer site for any programming language. Posting a question will usually result in a quick reply or link. Very useful for cleaning up scripts and problem solving. Many beginner questions have already been asked and answered with useful snippets of code or scripts.

SeqHack – seqhack.blogspot.com

Scripts related to my bioinformatic projects will be archived at this website.