



# CYVERSE<sup>®</sup>

Transforming Science Through Data-Driven Discovery

## CyVerse Visual Interactive Computing Environment

**Jason Williams**

Cold Spring Harbor Laboratory, DNA Learning Center;

CyVerse EOT Lead

[williams@cshl.edu](mailto:williams@cshl.edu)

 [@JasonWilliamsNY](https://twitter.com/JasonWilliamsNY)

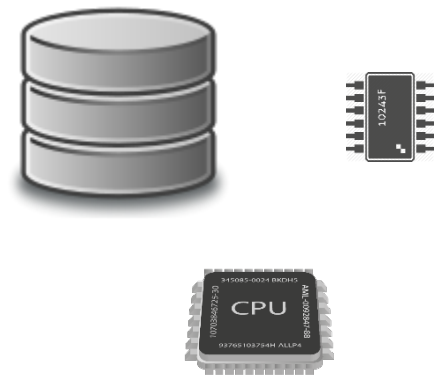


CyVerse is supported by the National Science Foundation under Grant Nos. DBI-0735191, DBI-1265383, and DBI-1743442

# Why use VICE?

# How (most) applications in the Discovery Environment work

```
[root@localhost ~]# ping -q fa.wikipedia.org
PING text.pmtpa.wikimedia.org (208.80.152.2) 56(84) bytes of data.
^C
--- text.pmtpa.wikimedia.org ping statistics ---
 1 packets transmitted, 1 received, 0% packet loss, time 0ms
 rtt min/avg/max/mdev = 540.528/540.528/540.528/0.000 ms
[root@localhost ~]# pwd
/root
[root@localhost ~]# cd /var
[root@localhost var]# ls -la
total 72
drwxr-xr-x. 18 root root 4096 Jul 30 22:43 .
drwxr-xr-x. 23 root root 4096 Sep 14 20:42 ..
drwxr-xr-x.  2 root root 4096 May 14 00:15 account
drwxr-xr-x. 11 root root 4096 Jul 31 22:26 cache
drwxr-xr-x.  3 root root 4096 May 18 16:03 db
drwxr-xr-x.  3 root root 4096 May 18 16:03 empty
drwxr-xr-x.  2 root root 4096 May 18 16:03 games
drwxrwx--T.  2 root gdm  4096 Jun  2 18:39 gdm
drwxr-xr-x. 38 root root 4096 May 18 16:03 lib
drwxr-xr-x.  2 root root 4096 May 18 16:03 local
lrwxrwxrwx.  1 root root   11 May 14 00:12 lock -> ../run/lock
drwxr-xr-x. 14 root root 4096 Sep 14 20:42 log
lrwxrwxrwx.  1 root root   10 Jul 30 22:43 mail -> spool/mail
drwxr-xr-x.  2 root root 4096 May 18 16:03 nis
drwxr-xr-x.  2 root root 4096 May 18 16:03 opt
drwxr-xr-x.  2 root root 4096 May 18 16:03 preserve
drwxr-xr-x.  2 root root 4096 Jul  1 22:11 report
lrwxrwxrwx.  1 root root   6 May 14 00:12 run -> ../run
drwxr-xr-x. 14 root root 4096 May 18 16:03 spool
drwxrwxrwt.  4 root root 4096 Sep 12 23:50 tmp
drwxr-xr-x.  2 root root 4096 May 18 16:03 yp
[root@localhost var]# yum search wiki
Loaded plugins: langpacks, presto, refresh-packagekit, remove-with-leaves
rpmfusion-free-updates                               | 2.7 kB    00:00
rpmfusion-free-updates/primary_db                    | 206 kB   00:04
rpmfusion-nonfree-updates                            | 2.7 kB    00:00
updates/metalink                                     | 5.9 kB    00:00
updates                                                | 4.7 kB    00:00
updates/primary_db                                   73% [=====] | 62 kB/s | 2.6 MB    00:15 ETA
```



# How (most) applications in the Discovery Environment work

```
[root@localhost ~]# ping -q fa.wikipedia.org
PING text.pmtpa.wikimedia.org (208.80.152.2) 56(84) bytes of data.
^C
--- text.pmtpa.wikimedia.org ping statistics ---
 1 packets transmitted, 1 received, 0% packet loss, time 0ms
 rtt min/avg/max/mdev = 540.528/540.528/540.528/0.000 ms
[root@localhost ~]# pwd
/root
[root@localhost ~]# cd /var
[root@localhost var]# ls -la
total 72
drwxr-xr-x. 18 root root 4096 Jul 30 22:43 .
drwxr-xr-x. 23 root root 4096 Sep 14 20:42 ..
drwxr-xr-x.  2 root root 4096 May 14 00:15 account
drwxr-xr-x. 11 root root 4096 Jul 31 22:26 cache
drwxr-xr-x.  3 root root 4096 May 18 16:03 db
drwxr-xr-x.  3 root root 4096 May 18 16:03 empty
drwxr-xr-x.  2 root root 4096 May 18 16:03 games
drwxrwx--T.  2 root gdm  4096 Jun  2 18:39 gdm
drwxr-xr-x. 38 root root 4096 May 18 16:03 lib
drwxr-xr-x.  2 root root 4096 May 18 16:03 local
lrwxrwxrwx.  1 root root   11 May 14 00:12 lock -> ../run/lock
drwxr-xr-x. 14 root root 4096 Sep 14 20:42 log
lrwxrwxrwx.  1 root root   10 Jul 30 22:43 mail -> spool/mail
drwxr-xr-x.  2 root root 4096 May 18 16:03 nis
drwxr-xr-x.  2 root root 4096 May 18 16:03 opt
drwxr-xr-x.  2 root root 4096 May 18 16:03 preserve
drwxr-xr-x.  2 root root 4096 Jul  1 22:11 report
lrwxrwxrwx.  1 root root   6 May 14 00:12 run -> ../run
drwxr-xr-x. 14 root root 4096 May 18 16:03 spool
drwxrwxrwt.  4 root root 4096 Sep 12 23:50 tmp
drwxr-xr-x.  2 root root 4096 May 18 16:03 yp
[root@localhost var]# yum search wiki
Loaded plugins: langpacks, presto, refresh-packagekit, remove-with-leaves
rpmfusion-free-updates                               | 2.7 kB    00:00
rpmfusion-free-updates/primary_db                    | 206 kB   00:04
rpmfusion-nonfree-updates                            | 2.7 kB    00:00
updates/metalink                                     | 5.9 kB    00:00
updates                                                | 4.7 kB    00:00
updates/primary_db                                   73% [=====] | 62 kB/s | 2.6 MB    00:15 ETA
```



# How (most) applications in the Discovery Environment work

[root@localhost ~]# ping -q fa.wikipedia.org  
PING text.pmtpa.wikimedia.org (208.80.152.2) 56(84) bytes of data.  
64 bytes from 208.80.152.2: icmp\_seq=1 ttl=64 time=0.042 ms  
64 bytes from 208.80.152.2: icmp\_seq=2 ttl=64 time=0.042 ms  
64 bytes from 208.80.152.2: icmp\_seq=3 ttl=64 time=0.042 ms  
64 bytes from 208.80.152.2: icmp\_seq=4 ttl=64 time=0.042 ms  
64 bytes from 208.80.152.2: icmp\_seq=5 ttl=64 time=0.042 ms  
64 bytes from 208.80.152.2: icmp\_seq=6 ttl=64 time=0.042 ms  
64 bytes from 208.80.152.2: icmp\_seq=7 ttl=64 time=0.042 ms  
64 bytes from 208.80.152.2: icmp\_seq=8 ttl=64 time=0.042 ms  
64 bytes from 208.80.152.2: icmp\_seq=9 ttl=64 time=0.042 ms  
64 bytes from 208.80.152.2: icmp\_seq=10 ttl=64 time=0.042 ms  
---  
ping statistics:  
 0 packets transmitted, 0 received, 0% success rate  
 0 bytes transmitted, 0 bytes received, 0% success rate

DISCOVERY ENVIRONMENT Muscle-3.8.31 Apps

← Back Documentation Details

### Muscle-3.8.31

This is a Perl wrapper that sanitizes sequence names before passing them to: MUSCLE v3.8.31 by Robert C. Edgar This software is donated to the public domain. Please cite: Edgar, R.C. Nucleic Acids Res 32(5), 1792-97.

- 1 Analysis Info
- 2 Parameters
- 3 Advanced Settings (optional)
- 4 Review and Launch

← Back Next →

Step 1: Analysis Info

Analysis Name\*  
Muscle-3.8.31\_analysis1  
Enter analysis name

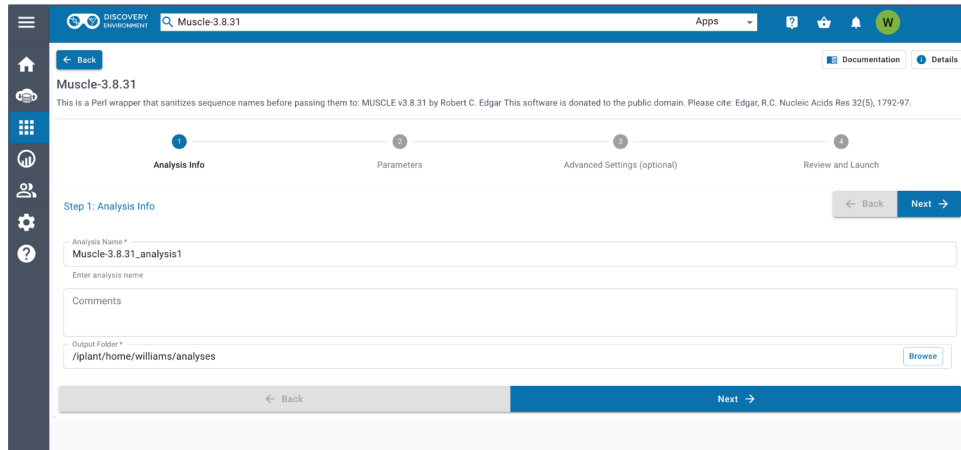
Comments

Output Folder\*  
/iplant/home/williams/analyses  
Browse

← Back Next →

updates | 4.7 kB | 00:00  
updates/primary\_db 73% [=====] | 62 kB/s | 2.6 MB | 00:15 ETA

# How (most) applications in the Discovery Environment work

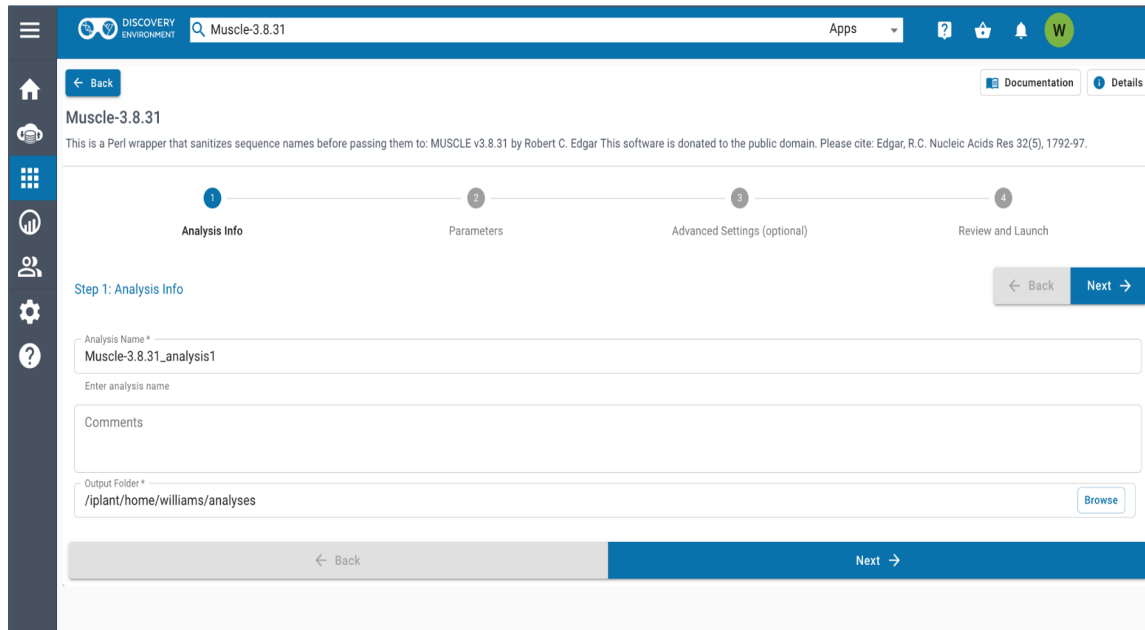


(non-interactive)

# What if you want to...

- See and adjust intermediate steps in an analysis
- Use an application that has a graphical user interface (GUI)
- Make data visualization easy for you and your collaborators\students

# What VICE application enable

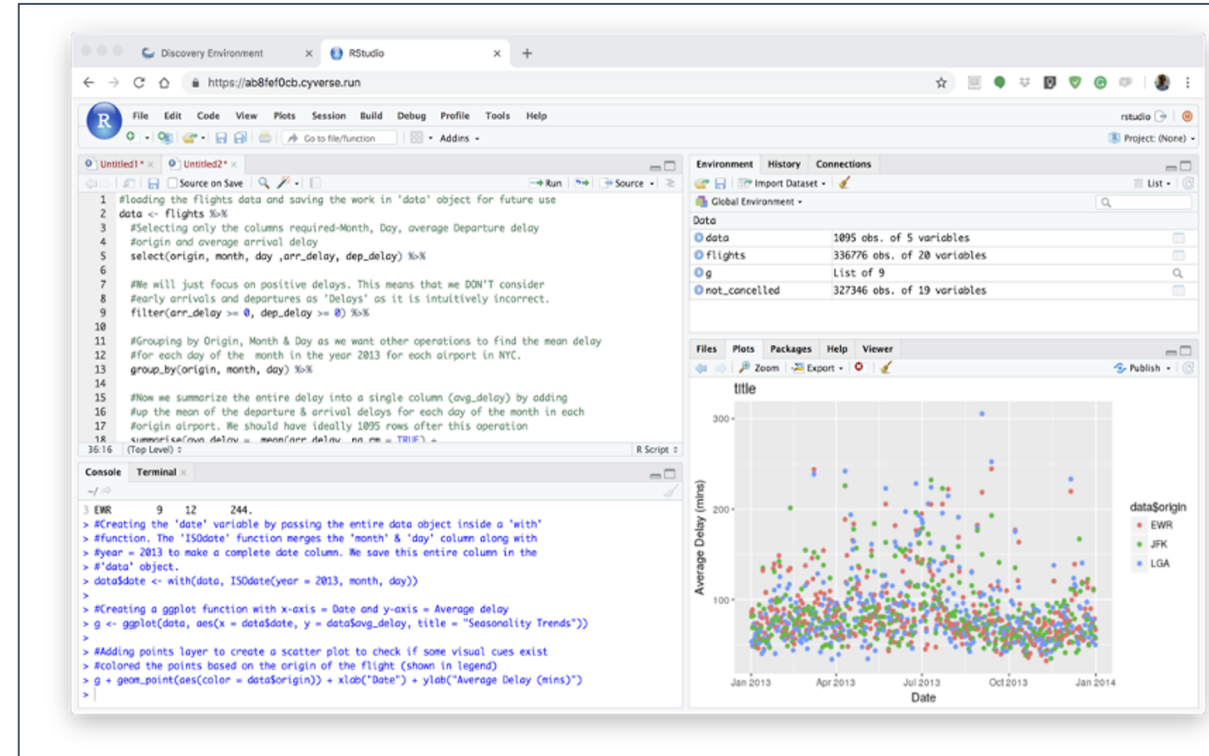


(Interactive)



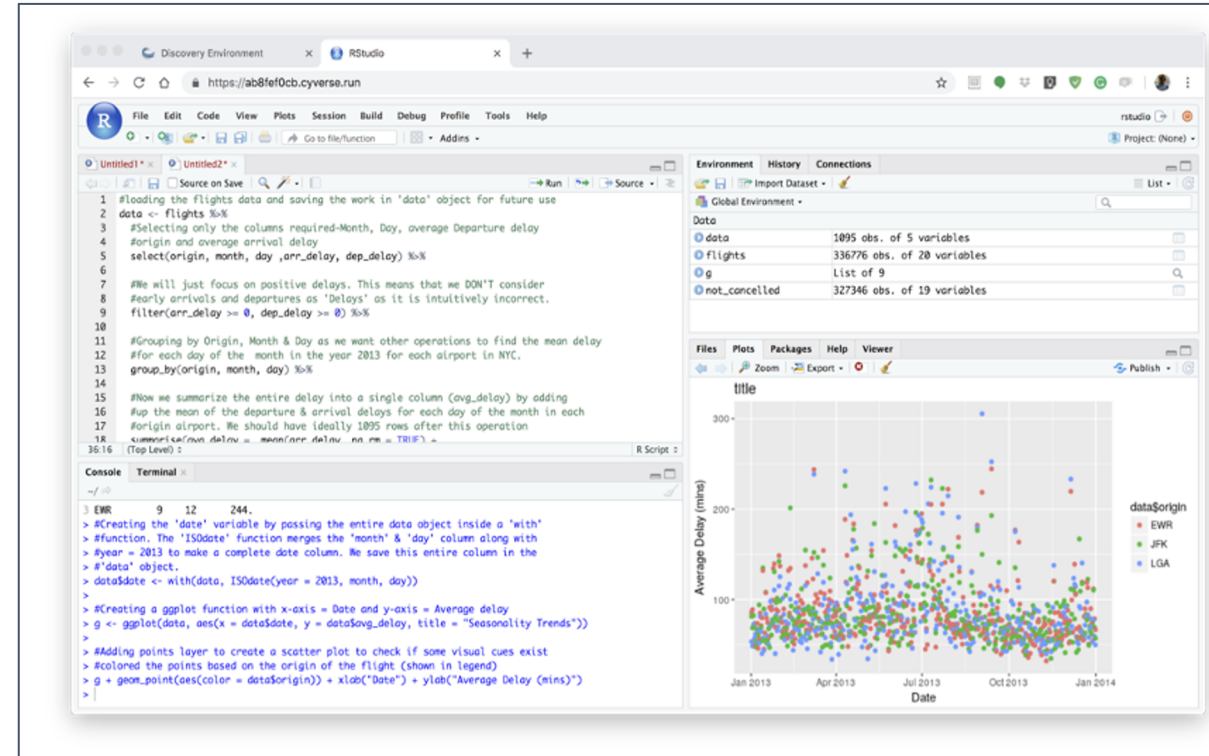
# Visual Interactive Computing Environment (VICE)

- Interactively manipulate analyses and visualize data



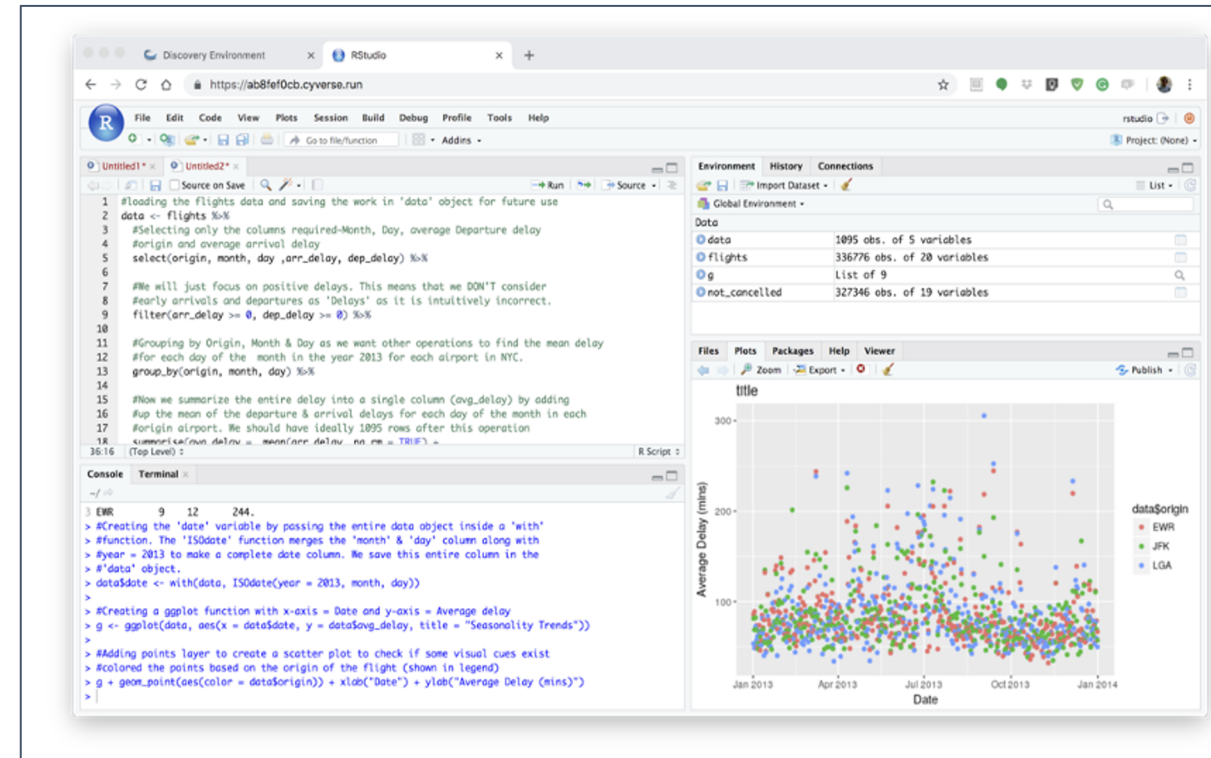
# Visual Interactive Computing Environment (VICE)

- Interactively manipulate analyses and visualize data
- Run popular applications such as Jupyter, RStudio, RShiny, Linux shell and more



# Visual Interactive Computing Environment (VICE)

- Interactively manipulate analyses and visualize data
- Run popular applications such as Jupyter, RStudio, RShiny, Linux shell and more
- Allows users to launch web applications packaged into the DE (Docker)

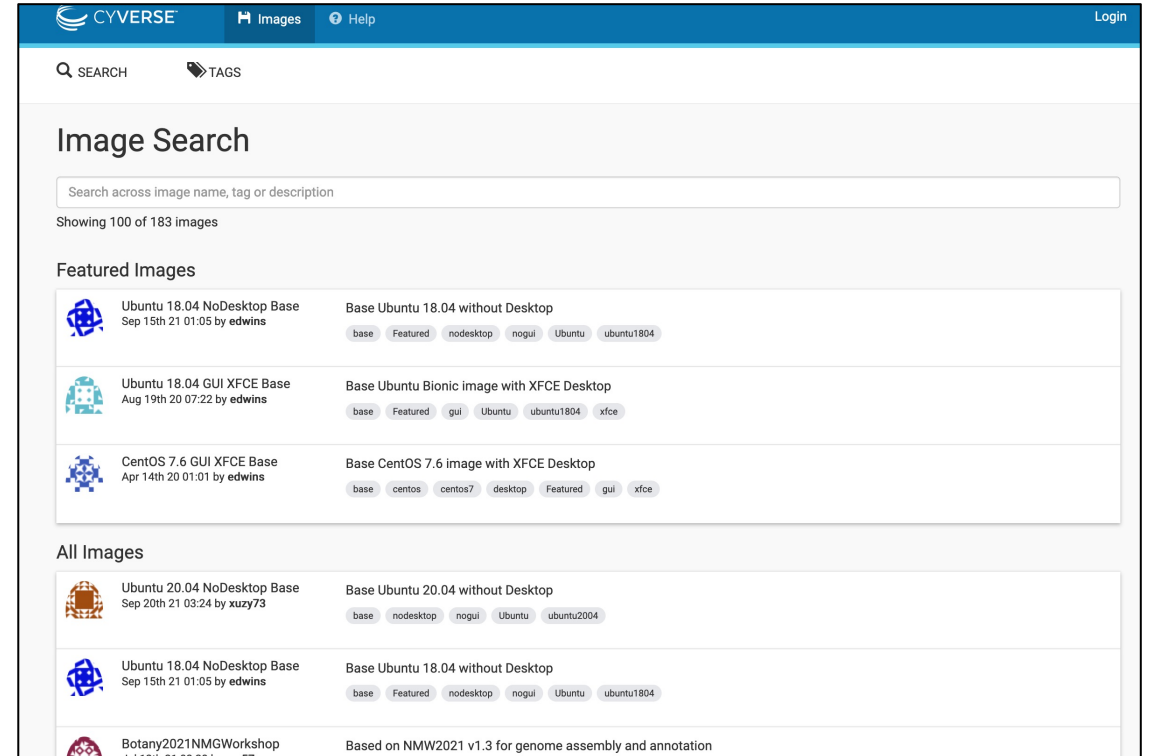


# VICE advantages

- No restriction on data limits
- Secure environment
- User can select/create custom computing environments
- Longer run times
- Highly scalable
- Share both apps and analyses

# Focusing the Atmosphere use case

- Users who need to work with a GUI
- Using one or a few apps (again often involving visualization)
- Atmosphere still available for development purposes; JetStream for general purpose use



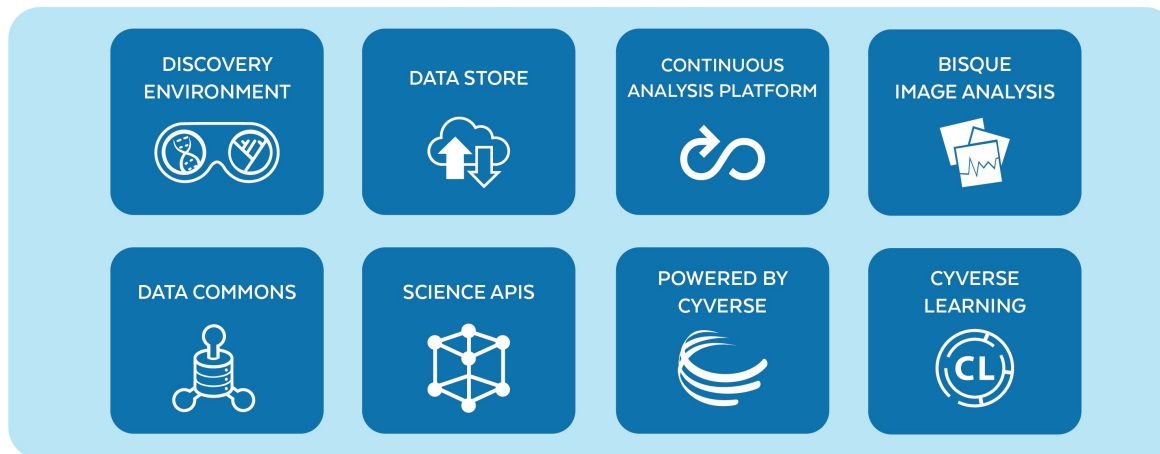
# VICE reminders

- Available on the US ([cyverse.org](http://cyverse.org)) DE
- Authorization required to use ([user.cyverse.org](http://user.cyverse.org))
- Runtimes are for 48 hours by default
- Data transfer and mounting

# VICE reminders

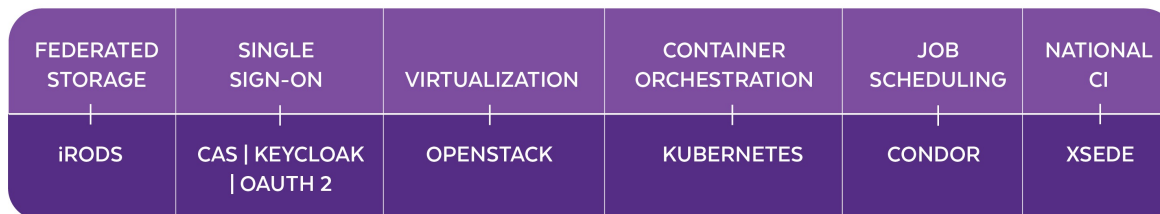
Ease of Use ↑  
Flexibility ↓

## PRODUCTS



← Ready to use Platforms

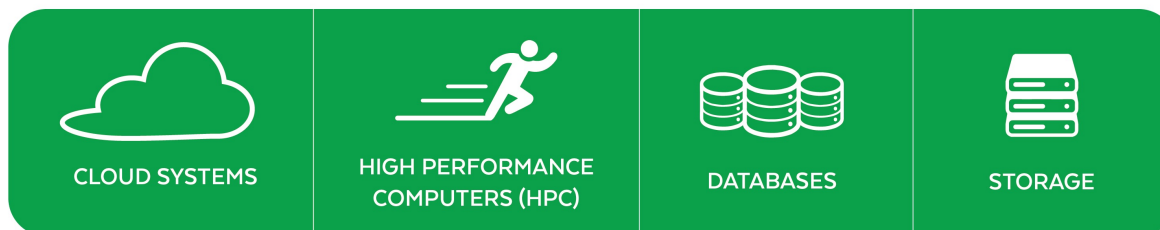
## SERVICES



← Extensible Services

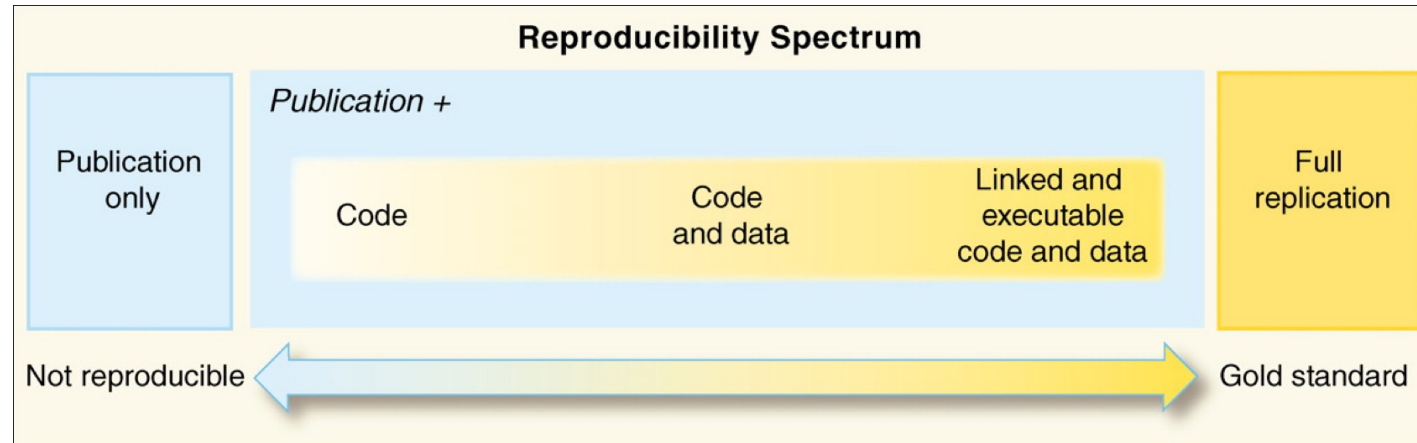
← Established CI Components

## HARDWARE RESOURCES



← Foundational Capabilities

# VICE reminders

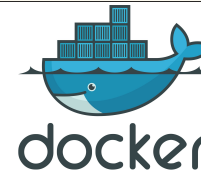


Peng, 2011 Science DOI: 10.1126/science.1213847

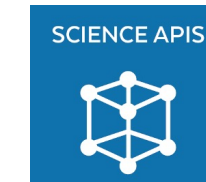
3rd party technologies



iRODS



CyVerse CI Platforms



External Repositories and Compute



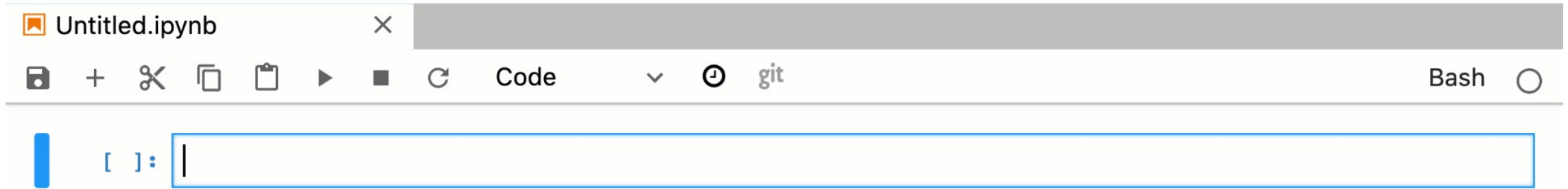


# VICE example use case

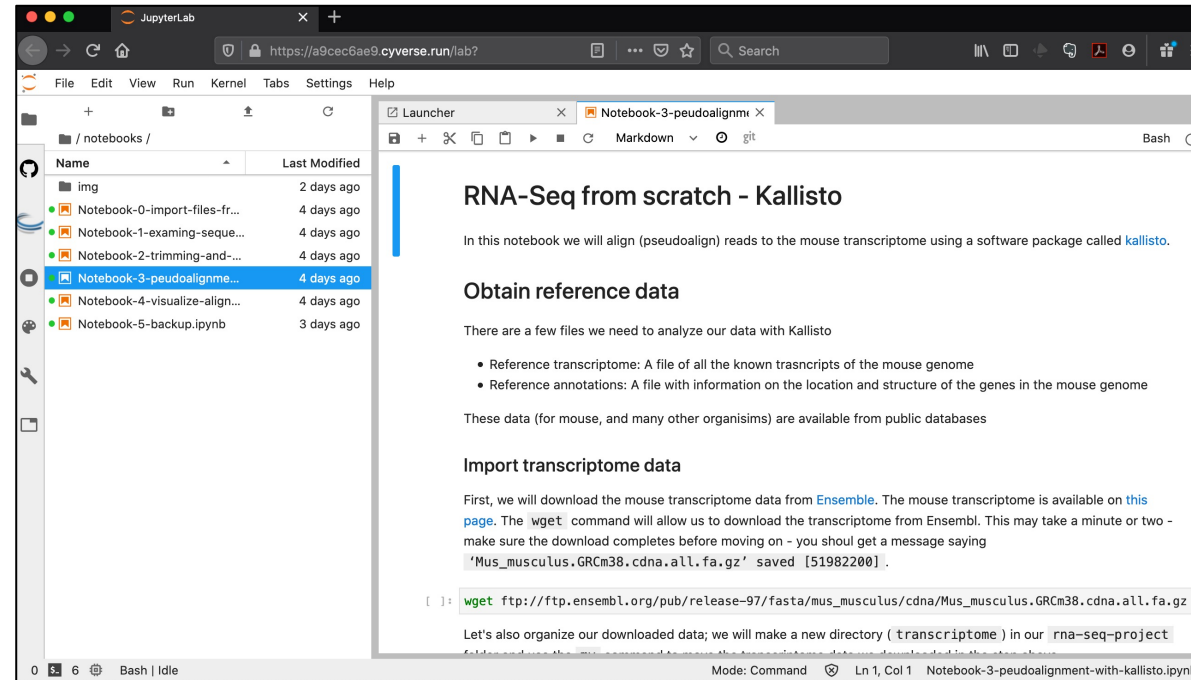


Project Jupyter exists to develop open-source software, open-standards, and services for interactive computing across dozens of programming languages.

# VICE example use case



# VICE example use case: RNA-Seq w/Jupyter



0 - Import from SRA



1 - Sequence QC



2 - Sequencing Trimming



3 - Read Pseudoalignment



4 - Read Visualization



5 - Data Backup



# VICE example use case: RNA-Seq w/Jupyter

## Import transcriptome data

First, we will download the mouse transcriptome data from [Ensemble](#). The mouse transcriptome is available on [this page](#). The `wget` command will allow us to download the transcriptome from Ensembl. This may take a minute or two - make sure the download completes before moving on - you should get a message saying

```
'Mus_musculus.GRCm38.cdna.all.fa.gz' saved [51982200].
```

```
[ ]: wget ftp://ftp.ensembl.org/pub/release-97/fasta/mus_musculus/cdna/Mus_musculus.GRCm38.cdna.all.fa.gz
```

Let's also organize our downloaded data; we will make a new directory ( `transcriptome` ) in our `rna-seq-project` folder and use the `mv` command to move the transcriptome data we downloaded in the step above.

```
[ ]: mkdir -p /home/gea_user/rna-seq-project/transcriptome && mv Mus_musculus.GRCm38.cdna.all.fa.gz /home/
```

## Index transcriptome

We will now use Kallisto's indexing function to prepare the transcriptome for analysis. The "Index" is a lookup table for the transcriptome that allows it to be more easily searched by Kallisto. First let's organize our files by creating a new directory to hold our kallisto work.

```
[ ]: mkdir -p /home/gea_user/rna-seq-project/kallisto
```

**Code is there when you need it**

# VICE example use case: RNA-Seq w/Jupyter

2. Complete the code below to trim your reads. Make sure that you change `INPUT` and `OUTPUT` to reflect your file(s). If you are working with more than one file, see the bonus exercise.

- `INPUT` - should be the name of your file
- `OUTPUT` - should be changed so that the prefix of the file (e.g. SRR5017XXX) is followed by `_trimmed.fastq.gz`

```
[ ]: trimmomatic SE INPUT OUTPUT -threads 8 SLIDINGWINDOW:4:20 MINLEN:75
```

3. Move your trimmed files to our previously created output folder (no need to change this code)

```
[ ]: mv *_trimmed.fastq.gz /home/gea_user/rna-seq-project/trimmed-reads
```

4. Complete the cells below to change directories into the trimmed-reads folder and run `fastqc`

```
[ ]: cd
```

```
[ ]: fastqc
```

Using the Jupyter Labs files browser you can navigate to the `trimmed-reads` folder in `rna-seq-project` to view your fastqc results

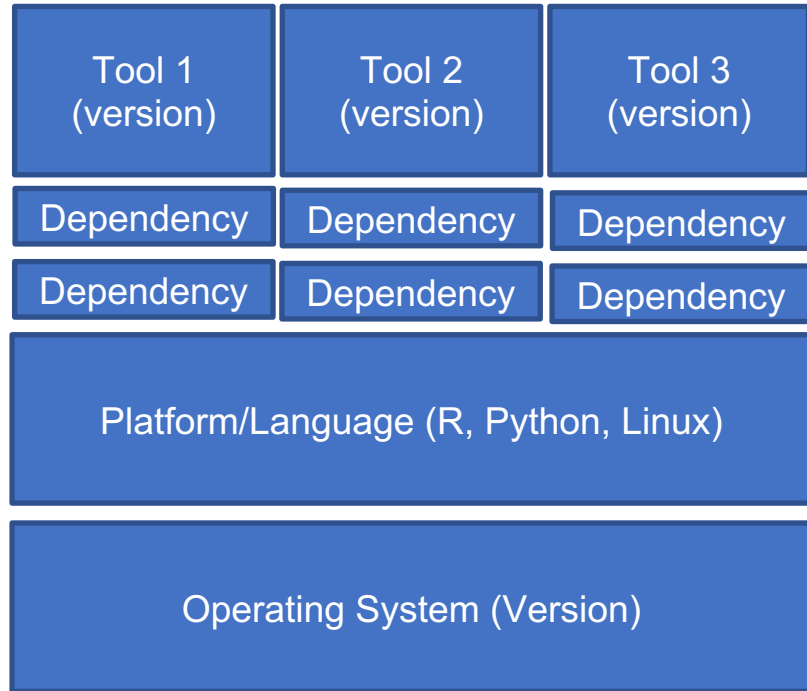
**Teaching code/command line with faded examples**

# Advanced – Customizing VICE

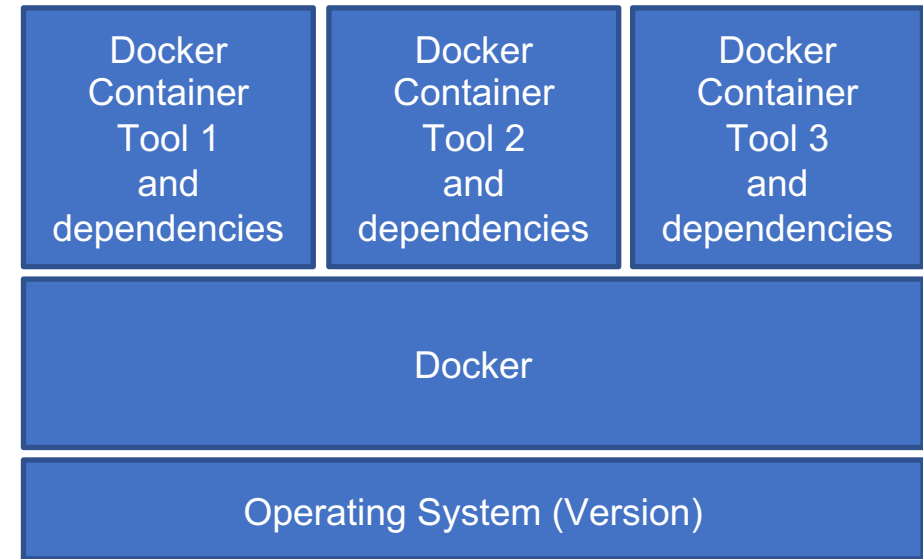


# Advanced – Customizing VICE

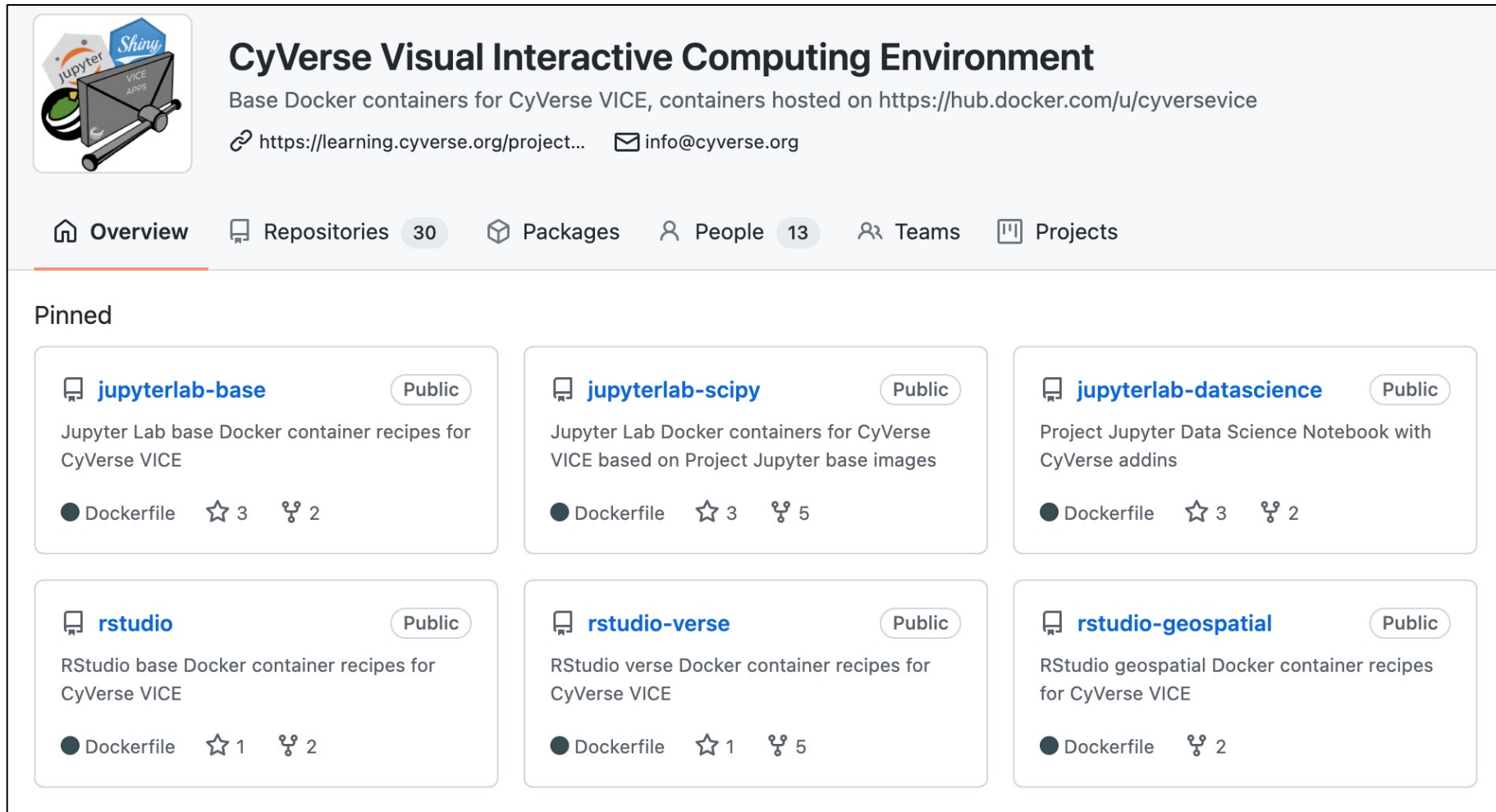
## Tools and dependencies separate



## Tools and dependencies bundled



# Advanced – Customizing VICE



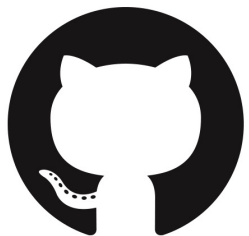
The screenshot shows a GitHub profile for 'CyVerse Visual Interactive Computing Environment'. The profile includes a repository count of 30, 13 people, and various navigation tabs like Overview, Repositories, Packages, People, Teams, and Projects. The 'Pinned' section displays six Docker container recipes for JupyterLab and RStudio, each with a 'Public' label, a 'Dockerfile' icon, and star/fork counts.

**CyVerse Visual Interactive Computing Environment**  
Base Docker containers for CyVerse VICE, containers hosted on <https://hub.docker.com/u/cyversevice>  
<https://learning.cyverse.org/project...> [info@cyverse.org](mailto:info@cyverse.org)

Overview Repositories **30** Packages People **13** Teams Projects

Pinned

- jupyterlab-base** (Public)  
Jupyter Lab base Docker container recipes for CyVerse VICE  
● Dockerfile ☆ 3 🍴 2
- jupyterlab-scipy** (Public)  
Jupyter Lab Docker containers for CyVerse VICE based on Project Jupyter base images  
● Dockerfile ☆ 3 🍴 5
- jupyterlab-datascience** (Public)  
Project Jupyter Data Science Notebook with CyVerse addins  
● Dockerfile ☆ 3 🍴 2
- rstudio** (Public)  
RStudio base Docker container recipes for CyVerse VICE  
● Dockerfile ☆ 1 🍴 2
- rstudio-verse** (Public)  
RStudio verse Docker container recipes for CyVerse VICE  
● Dockerfile ☆ 1 🍴 5
- rstudio-geospatial** (Public)  
RStudio geospatial Docker container recipes for CyVerse VICE  
● Dockerfile 🍴 2



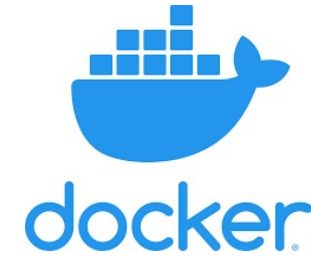


# Advanced – Customizing VICE

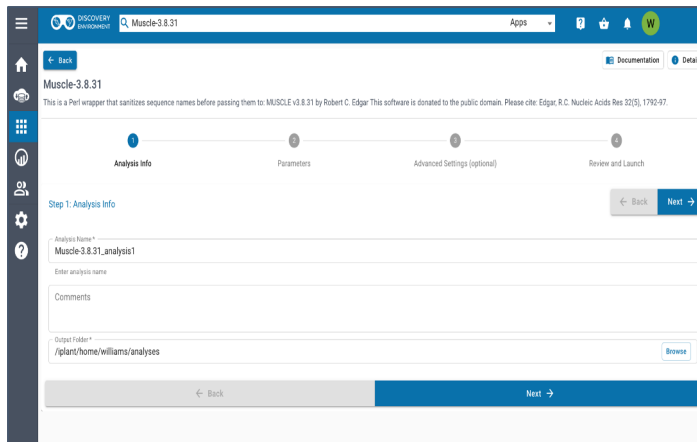
## 1. Write a Dockerfile

```
# Install bioconda packages
RUN /opt/conda/bin/conda install -y -q fastqc=0.11.7=5
RUN /opt/conda/bin/conda install -y -q trimmomatic=0.38=0
RUN /opt/conda/bin/conda install -y -q sra-tools=2.9.6=hf484d3e_0
RUN /opt/conda/bin/conda install -y -q parallel-fastq-dump=0.6.5=py_0
RUN /opt/conda/bin/conda install -y -q kallisto=0.46.0=hb6a4e58_0
RUN /opt/conda/bin/conda install -y -q fastp=0.20.0=hdbcaa40_0
#
```

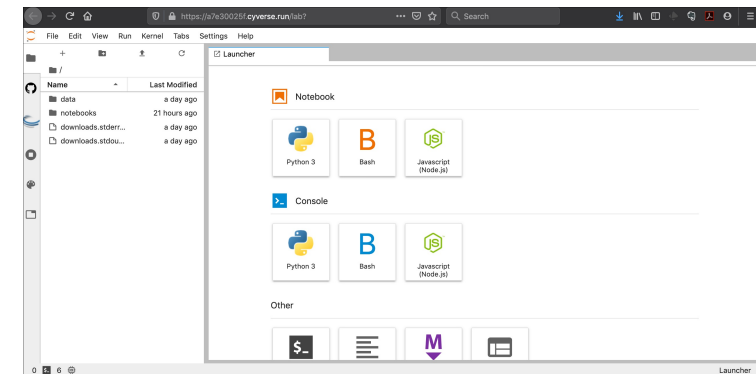
## 2. Push container to CyVerse



## 3. Create an interface and get button



## 4. Use your tools



# Learning.cyverse.org


🏠 [CyVerse Learning Center Home](#)  
latest

Search docs

**WELCOME TO CYVERSE**

- Getting Started
- Quick Start Guides
- Platform Guides
- Frequently Asked Questions
- Glossary
- Tool and App Integration
- Tutorials
- Power Users
- Workshops
- Webinars
- Contributing to the Learning Center

Docs » Getting Started [Edit on GitHub](#)



[🏠 Learning Center Home](#)

## Welcome to the CyVerse Learning Center

The CyVerse Learning Center is a release of our learning materials in the popular “Read the Docs” formatting.

## Getting Started

**Account Creation**

[Create a CyVerse account](#)

### What to do first

- [Upload your data](#)
- [Launch an analysis](#)
- Watch a [Getting Started Webinar](#)